

## Análise, projeto e implementação de uma aplicação utilizando interface de voz com o usuário

Valéria Farinazzo Martins<sup>1</sup>, Almir Gonçalves dos Santos<sup>1</sup>, Felipe Antunes Rodrigues<sup>1</sup>, Marcos Haruo Okumura<sup>1</sup>, Thomas Jun Sakoda<sup>1</sup>, Marcelo de Paiva Guimarães<sup>2</sup>

<sup>1</sup>Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie  
Rua da Consolação, 930 Cep 01302-907 - Consolação - São Paulo - SP – Brasil

<sup>2</sup>Universidade Federal de São Paulo/membro do Programa de mestrado da Faculdade  
Campo Limpo Paulista, São Paulo, Brasil

valeria.farinazzo@mackenzie.br, almirsantos@hotmail.com,  
felipearodrigues@gmail.com, marcosokumura@gmail.com,  
thomas.sakoda@gmail.com, marcelodepaiva@gmail.com

**Abstract.** Recent technological advances have allowed using voice recognition and synthesis to develop a wide range of applications with interfaces based on voice. However many of them are developed without using a development methodology, generating problems related with the project concept. This work presents phases of development to be followed during the development of an application that is based on voice recognition and synthesis. It was developed an application to validate the described phases.

**Resumo.** Avanços tecnológicos recentes proporcionaram que o reconhecimento de voz e a síntese de voz fossem utilizados em aplicações com interfaces baseadas em comandos de voz. Contudo, muitas delas ainda são desenvolvidas sem o uso de uma metodologia para seu desenvolvimento, caindo em problemas na concepção de seus projetos. Este trabalho apresenta as fases de desenvolvimento, seguindo uma metodologia, de uma aplicação utilizando as tecnologias de reconhecimento e síntese de voz, de acordo com suas peculiaridades. Uma aplicação foi desenvolvida a fim de validar estas fases descritas neste trabalho.

### 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 as de Realidade Virtual, Realidade Aumentada, baseadas em interfaces gestuais e em voz – surge como uma forma de tentar tornar estas aplicações mais naturais e mais fáceis de serem usadas, a fim de gerar maior satisfação aos usuários não especialistas [Oliveira Neto, Salvador e Kawamoto 2010].

Para Cohen, Giangola e Balogh (2004), no início do crescimento e uso das tecnologias de fala, na década de 90, o maior obstáculo a ser superado pelas tecnologias

era o ceticismo sobre as suas capacidades de reconhecimento e síntese, que, por várias décadas, havia sido prometida e por várias vezes os resultados foram desapontadores. Após vários anos, com a evolução da tecnologia de reconhecimento de voz e o aumento do poder computacional, este anseio tornou-se realidade. Com a robustez da tecnologia, tornou-se possível fornecer interações muito mais eficazes em diversas situações.

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 captura 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].

A fala é o meio de comunicação mais importante entre os homens, e é de extrema importância para a socialização entre os indivíduos, pois também é meio mais rápido de troca de informações entre as pessoas. Assim, a tecnologia de reconhecimento de voz visa justamente permitir que as pessoas interajam com as máquinas e assim possam tornar a interação o mais natural possível; porém, esta tecnologia possui algumas particularidades e necessidades especiais, como alta compreensão; isto é, deverá entender os diferentes sotaques e dicções próprios do usuário e alta *performance* no processamento das mensagens recebidas.

Este trabalho mostra, com detalhes, o desenvolvimento de uma aplicação VUI, desde a fase de análise de requisitos, até a avaliação. O trabalho se baseia em uma metodologia concebida a partir de Cohen et al (2004), Martins, Brasiliano e Fernandes (2012) e em Lamel, Minker e Paroubek (2000). Além disso, discute as peculiaridades de tais aplicações. Estas técnicas buscam diminuir falhas no desenvolvimento de aplicações de reconhecimento de voz e consequentemente falhas que possam ocorrer na utilização do sistema.

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 voos e reserva de hotéis, auxílio à lista telefônica, guia de restaurantes, bares e filmes. Nesse trabalho é apresentada como estudo de caso uma aplicação para acesso dos alunos ao sistema de notas e faltas de uma instituição de curso superior.

O artigo está organizado como segue: na seção 2 é apresentada a tecnologia VUI e seu funcionamento; na seção 3, por intermédio do estudo de caso, são explicadas e exemplificadas as fases que compõe o processo de desenvolvimento de aplicações VUI, focadas em suas peculiaridades. Finalmente, na seção 4 são apresentadas as conclusões deste trabalho.

## 2. Voice User Interface

Conceitua-se VUI como a interação do usuário com a máquina através de comando de voz ou fala para a execução de processos. Segundo Cohen, Giangola e Balogh (2004), ela é um tipo de interface na qual um usuário interage enquanto está se comunicando com uma aplicação de linguagem falada. Os elementos de VUI incluem *prompts*, gramáticas, e lógica do diálogo. Os *prompts*, ou sistemas de mensagens, são todas as falas gravadas ou sintetizadas apresentadas ao usuário durante o diálogo. As gramáticas definem o que o usuário pode dizer em resposta a cada *prompt*. O sistema é limitado às

palavras, sentenças e frases que fazem parte da gramática [Lai 2000], [Shneiderman 2000], [Cohen, Giangola and Balogh 2004], [Chamberlain et al 2006], [Bosch, Oostdjik and Ruitter 2004].

As VUI assemelha-se à interface de sistemas convencional, contudo, segundo Cohen, Giangola e Balogh (2004), suas características básicas para entrada e saída de dados são:

- Input de dados: realizado através da linguagem de voz (linguagem falada);
- Output de dados: comunicação audível.

O sistema de diálogo pode ser direto ou com iniciativa mista. No primeiro a interação do usuário é através de perguntas que o próprio sistema faz e espera uma resposta; ou através de informações organizadas as quais são apresentadas ao usuário, e a partir delas será feita a escolha dentre as opções disponíveis. Já o sistema de iniciativa mista tem como propósito de fazer com que o usuário trabalhe com o sistema de um modo mais maleável. O sistema captura informações relevantes para continuar o diálogo, as quais não são passadas no processo de resposta do usuário.

Algumas aplicações suportam ambas as iniciativas. Elas devem tratar as ocorrências e variações do diálogo como a descoberta, recuperação e a prevenção de erros e comandos universais.

A principal peça que compõe uma Interface de Voz é o 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**

O processo de reconhecimento de voz se dá da seguinte maneira:

- 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;
- 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;

- 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;

Segundo Guilhoto e Rosa (2001) a aplicação do processamento de voz é dividida em quatro tipos:

- Comando por voz: o sistema recebe um trecho de fala, que é transformado em texto, identificando então a ação que o sistema irá realizar. Neste caso, o processo é simples, pois o sistema tem o conhecimento de todos os comandos que o usuário pode realizar;
- Fala natural ou contínua: O sistema recebe um conjunto de palavras que tenham um sentido semântico (frase), que podem ser uma ou mais frases. Esse conjunto de palavras é transformado em frase(s), processado e transformado em texto. Esse tipo de tecnologia é geralmente voltada para a criação de documentos, escrita de emails, entre outras;
- Síntese da voz: é o processo no qual o sintetizador recebe o texto na forma digital e o transforma em ondas sonoras. Os programas de síntese de voz são mais úteis em ocasiões em que o usuário não tem acesso ao texto escrito ou não pode desviar a atenção para ler algo, seja porque o usuário tem alguma deficiência visual ou o texto está distante;
- Autenticação de voz: a autenticação de voz é utilizada para identificar alguém, pois a voz é única para cada indivíduo. Os sistemas que usam essa tecnologia podem ser úteis para permitir o acesso de uma pessoa a uma determinada função.

O estado da arte em tecnologia de voz já permite que sistemas automáticos sejam desenvolvidos para trabalhar em condições reais (SAN-SEGUNDO et al, 2005). Empresas como a Nuance, através do sistema Dragon (2012), a IBM ViaVoice e a Philips Speech – compradas há alguns anos pela Nuance – a Microsoft, através do sistema Speech (Microsoft, 2012) e a iSpeech (2012) têm investido no desenvolvimento de sistemas de voz para domínios restritos.

### **3. Sistema de Terminal Informativo Acadêmico (TIA) por Voz**

Atualmente, dentro da universidade foco do estudo, existem muitos terminais de acesso ao TIA (Terminal Informativo Acadêmico), onde os alunos matriculados conseguem visualizar informações sobre suas notas, faltas, situação financeira, entre outros dados referentes à vida acadêmica. O acesso a este terminal também pode ser feito externamente ao campus da universidade, bastando, ao aluno, utilizar um computador com acesso à internet e acessar a página correspondente.

Esta aplicação apresenta a ideia de acesso ao terminal do aluno, sem a necessidade de utilização dos terminais existentes e da internet. O objetivo é permitir que o TIA seja acessado efêtuada através de um telefone, por comandos e síntese de voz.

O processo utilizado para o desenvolvimento da aplicação teve como base Cohen et al (2004), Martins, Brasileiro e Fernandes (2012) e em Lamel, Minker e Paroubek (2000). Ele é composto pelas seguintes fases: Definição dos requisitos, Projeto de Alto Nível, Projeto Detalhado, Desenvolvimento, Testes e *Tuning*.

### 3.1. Definição dos Requisitos

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, foi composta por:

- **Definição dos usuários finais:** para esse sistema, o usuário poderá ser qualquer pessoa regularmente matriculada na Universidade e que possua acesso ao TIA. Além disso, esse usuário deverá possuir um telefone fixo ou móvel para poder fazer a comunicação ao portal. Não está prevista a utilização do sistema por pessoas com necessidades especiais (auditiva e afonia).
- **Definição do Ambiente:** os usuários poderão acessar o sistema de qualquer lugar que tenha acesso a um telefone: no trabalho, na escola ou na sua casa. Porém, é necessário que este ambiente possa pouca interferência sonora.
- **Definição das tarefas:** O sistema autenticará o aluno devidamente matriculado e fornecerá informações de notas ou faltas referentes à matéria que o usuário necessita consultar. O sistema deverá fornecer um *feedback* ao usuário toda vez que o mesmo solicitar uma informação referente à nota ou à falta de uma certa matéria. O aluno consulta as notas e faltas após realizar a autenticação – por isso, no diagrama existe a dependência entre “Consultar notas e Faltas” e “Autenticar usuário”.

De acordo com Sommerville (2011), requisitos não funcionais são aqueles que descrevem apenas atributos do sistema ou do ambiente. Conforme Salvador, Oliveira Neto e Kawamoto (2008), os requisitos não funcionais podem ser utilizados em qualquer aplicação, e estão presentes em aspectos de *software* e *hardware*. Assim, para os requisitos não funcionais para esta aplicação, definiu-se:

- Metáforas: utilização de palavras que são intuitivas ao usuário;
- Minimização de memória: fazer com que o usuário seja induzido a escolher alguma opção que o sistema fornece, minimizando seu esforço;
- Mecanismos de ajuda e prevenção de erros: o sistema deverá prover ajuda ao usuário sempre que houver a necessidade. Em um sistema de voz, o sistema deverá interagir com opções se o usuário não tiver a iniciativa ou criar as estratégias de diálogos, fornecendo opções de funções ao usuário, prevenindo erros ao mesmo tempo;
- Veracidade: o sistema deverá mostrar o que o usuário necessita dentro das suas expectativas;
- Capacidade de confirmação do sistema: capacidade que o sistema de voz deve ter de confirmar se aquilo que foi solicitado pelo usuário é aquilo que o usuário deseja;

- Naturalidade: modo de o sistema agir de forma mais natural com o usuário;
- *Feedback*: em um sistema de voz, trata-se de um requerimento importante, pois dará um retorno ao usuário quanto ao seu objetivo. Esse retorno deverá ser compreensível para o usuário. A falta de *feedback* pode dar a sensação ao usuário de que o sistema está inoperante.

### 3.2. Projeto de Alto Nível

A seguir serão descritos as Tecnologias de Entrada e Saída utilizado no TIA com VUI e as tecnologias empregadas:

- **Tecnologia de Entrada**: Para este projeto utilizou-se um equipamento telefônico convencional com discagem por tom;
- **Tecnologia de Saída**: Da mesma forma que a tecnologia de entrada utiliza um telefone comum, utilizou-se o mesmo dispositivo para a saída das mensagens audíveis e as informações contidas no sistema;
- **Tecnologia de Suporte e Computacional**: Para o desenvolvimento da aplicação foi utilizada uma estação de trabalho *Desktop* convencional com acesso a internet e com sistema multimídia capaz de gravar e executar comandos de voz. O acesso à Internet foi necessário, pois usou-se a ferramenta que utiliza a VXML (*Voice Extensible Markup Language*) (2012). Depois de codificada toda a aplicação, os arquivos desenvolvidos foram armazenados em um provedor de hospedagem com suporte a arquivos deste formato. Esse mesmo provedor também teve como responsabilidade a disponibilidade de um número fixo associado ao endereço da TIA;

### 3.3. Projeto Detalhado

Nesta etapa, foram planejadas todas as ações da aplicação. A Figura 2 descreve o fluxo principal do sistema, que compreende as funcionalidades do início ao fim do TIA, mostrando seus processos e interações envolvidas.

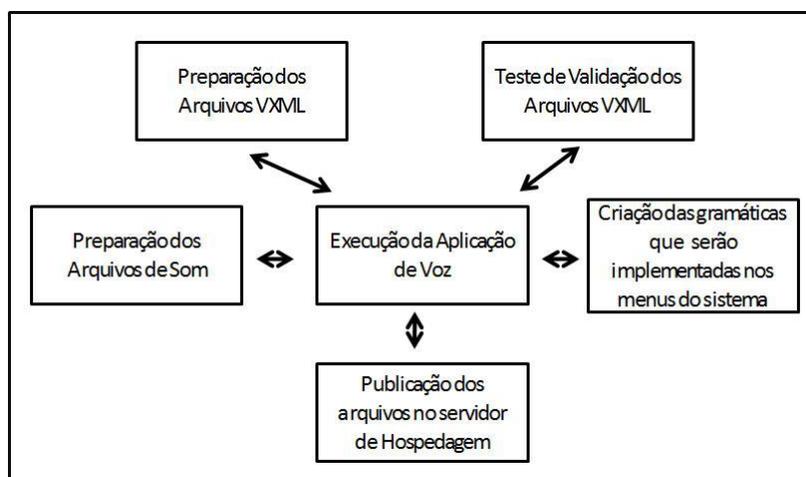
- 1 - Usuário, utilizando um telefone, disca o número do atendimento automatizado TIA;
- 2 - O sistema emite a mensagem de “boas-vindas”;
- 3 - O sistema solicita ao aluno que digite ou fale sua identificação acadêmica (TIA);
- 4 - O sistema solicita que o aluno informe a senha (a mesma utilizada no TIA via web);
- 5 - O sistema autentica a identificação e a senha;
- 6 - Após a identificação o sistema pergunta ao usuário se ele deseja saber informações sobre Notas ou Faltas;
- 7 - O usuário poderá escolher entre qualquer uma das duas opções;
- 8 - Caso o aluno opte por Notas
  - a. O sistema solicita ao aluno que informe o nome completo da matéria;
  - b. O sistema informa todas as notas que estão cadastradas para a matéria selecionada;
- 9 - Caso o aluno opte por Faltas
  - a. O sistema solicita ao aluno que informe o nome completo da matéria;
  - b. O sistema informa todas as notas que estão cadastradas para a matéria selecionada;
- 10 - O sistema pergunta ao aluno se ele deseja sair ou voltar ao menu de opções;

**Figura 2: Fluxo principal do sistema**

### 3.4. Desenvolvimento

Nesta fase, foi realizada a implementação do sistema. Para o desenvolvimento da aplicação, utilizando a linguagem VXML, foi utilizada a ferramenta VBuilder versão 2.0. Os arquivos desenvolvidos foram hospedados em um provedor.

As principais atividades de desenvolvimento desta da TIA são mostradas na Figura 3. Além de atividades relacionadas a codificação da aplicação, esse sistema demandou de atividades específicas, como a definição de uma gramática e a preparação de sons.



**Figura 3 – Atividades de Implementação da Aplicação**

### 3.5. Testes e *Tuning*

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

Depois de completar a fase de testes, ocorreu a instalação do TAI no ambiente do usuário. Esse procedimento foi realizado em duas etapas: na primeira fase instalou-se o piloto, em que o sistema foi disponibilizado para um número limitado de usuários. Os dados foram coletados e analisados a fim de se possibilitar melhorias. Foram realizadas três iterações de coleta e análise de dados. Após esta fase, o sistema foi finalmente disponibilizado 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.

Já os testes de usabilidade ocorreram com 16 usuários finais; neste teste foram abordadas questões relacionadas a: visibilidade do status do sistema, facilidade de utilização do sistema, reconhecimento ao invés de relembração, help e documentação, consistência e padrões. Estas medidas foram retiradas da avaliação heurística proposta por Nielsen (1990) que, embora tenham sido planejadas para aplicações web, podem ser reinterpretadas para aplicações VUI. Estas medidas foram então abordadas em um questionário pós-teste, utilizando a escala Likert, respondido pelos usuários ao término do uso da aplicação.

Os resultados da avaliação de usabilidade são apontados a seguir:

- Visibilidade do status do sistema: 31,25% dos usuários mostraram-se muito satisfeito; 43,75% satisfeito e 25% indiferentes.
- Facilidade de utilização do sistema: 18,75% das pessoas mostraram-se muito satisfeitas; 37,5% satisfeitas. 18,75% indiferentes; e 25% insatisfeitas. Isso relata que não houve uma homogeneidade em relação à facilidade de uso. O uso de uma interface não visual pode ser o motivo deste sentimento do usuário; não há retenção das informações que se perdem, à medida que são faladas.
- Reconhecimento ao invés de relembração: 12,5% dos usuários mostraram-se muito satisfeitos; 18,75% satisfeitos; 31,25% indiferentes; 25% insatisfeitos e 12,5% muito insatisfeitos. A razão para esta heterogeneidade deve-se também à falta de retenção das informações quando são somente faladas. Muitas vezes o usuário necessita reativar o sistema para ter novamente a informação.
- Help e documentação: 43,75% dos usuários mostraram-se muito satisfeitos, assim como 43,75% se mostraram satisfeitos e apenas 12,5% mostraram-se indiferentes. O sistema sempre apresentava, ao usuário, mensagens de como acionar o *help*, quando necessário.
- Consistência e padrões: 62,5% dos usuários se mostraram muito satisfeitos, enquanto 37,5% satisfeitos.

#### 4. Conclusões

O contexto atual de desenvolvimento e pesquisa de interfaces do usuário 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 nosso cotidiano. Este trabalho reafirma a importância deste tipo de interfaces ao propor uma aplicação baseada em voz a fim de tornar a interação com o usuário cada vez mais natural e intuitiva. Ao longo deste trabalho foram discutidas e apresentadas as etapas do desenvolvimento e as tecnologias utilizadas na criação da aplicação TIA, servindo de guia para a construção de novas aplicação com reconhecimento de voz.

Vale ressaltar que, durante a implementação do sistema, notou-se a necessidade de uma maior atenção a ser dada para um sistema que é altamente susceptível a falhas, devido ao reconhecimento dos comandos de voz, e isso deve ser priorizado em todo seu ciclo de vida, uma vez que, esse ponto é o fator mais crítico e decisivo para sua concretização. Por isto, nota-se a importância da atendente como parte da solução para sistemas que sejam o canal de comunicação com o cliente, pois o sistema deve ser capaz de atender o usuário além da pré-definições realizadas pelo sistema de voz (comandos e gramática).

Os pontos que foram pior avaliados pelos usuários trata-se de “Facilidade de utilização do sistema” e “Reconhecimento ao invés de relembração”. Estes dois quesitos merecem maior atenção, pois se referem ao fato da voz permitir apenas informações não visuais, ou seja, elas desaparecem tão logo sejam ditas, fazendo com que o usuário necessite de maior atenção para fixação da informação. Porém, de uma maneira geral, o sistema foi bem avaliado pelo público-alvo da pesquisa.

#### Referências

- BOSCH, L., OOSTDIJK, N., RUITER, J. P. (2004), Turn-taking in social talk dialogues: temporal, formal, and functional aspects, *SPECOM-2004*, pp 454-461.
- CHAMBERLAIN, J. ELLIOTT, G., KLEHR, M., BAUDE, J. (2006) *Speech User Interface Guide, RedPaper IBM*, <http://www.redbooks.ibm.com/redpapers/pdfs/redp4106.pdf>
- COHEN, M. H., GIANGOLA, J. P., BALOGH, J. (2004) *Voice User Interface Design*, Addison Wesley, ISBN 0-321-18576-5, 368 pp.
- DAMASCENO, E. F.; PEREIRA, T. V.; BREGA, J. R. F. (2005) Implementação de Serviços de Voz em Ambientes Virtuais. In *INFOCOMP Journal of Computer Science*, v.4, n3, p.67-73.
- DENG, L., HUANG, X. (2004). Challenges in adopting speech recognition, *Commun. ACM* 47, 1, pp. 69-75.
- GUILHOTO, P. J. dos S.; ROSA, S. P. C. de S. Reconhecimento de Voz: Sistemas Multimídia. Universidade de Coimbra, Coimbra - Portugal. 2001/2002.
- ISPEECH. Disponível em: <http://www.ispeech.org/>. Acesso em: 31/01/2013.
- LAI, J. (2000) Conversational Interfaces, *Communications of the ACM*, Vol. 43, No 9, pp 24 – 27.

- LAMEL, L.; MINKER, W.; PAROUBEK, P. (2000) “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.
- MARTINS, V.F; BRASILIANO, A.; FERNANDES, L.F. (2012) Interface do Usuário Baseada em Voz Como Ferramenta para Promover o Ensino/Aprendizagem de Língua Estrangeira, *REAVI - Revista Eletrônica do Alto Vale do Itajaí*; 1ª edição, 2012, disponível em: <http://www.revistas.udesc.br/index.php/reavi/issue/view/251>
- MICROSOFT, Disponível em: <http://www.microsoft.com/en-us/download/details.aspx?id=10121>. Acesso em 31/01/2013.
- NIELSEN, J. (2001). Ten usability heuristics. [Online]. Available: <http://www.useit.com/papers/heuristic/>
- NUANCE, Disponível em: <http://www.nuance.com/for-individuals/by-product/dragon-for-pc/index.htm>. Acesso em 31/01/2013.
- OLIVEIRA NETO, J. S. de; SALVADOR, V. F. M.; KAWAMOTO, A. L. (2010) “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.
- SALVADOR, V. F. M.; NETO, J. S. O.; KAWAMOTO, A. L. (2008) Requirement Engineering Contributions to Voice User Interface. In First International Conference on Advances in Computer-Human Interaction, ACHI, Saint Luce Martinique. IEEE Computer Society.
- SAN-SEGUNDO, R. et al. (2005). Knowledge-combining methodology for dialogue design in spoken language systems. In: International Journal of Speech Technology n. 8, p. 45-66, Springer Science + Business Media.
- SHNEIDERMAN, B. (2000) The Limits of Speech Recognition, *Communications of the ACM*, Vol. 43, No 9, pp 24 – 27.
- SOMMERVILLE, I (2011). Engenharia De Software. 9ª Edição. Editora: Pearson Education – Br.
- VoiceXML Development Guide. Disponível em <http://www.vxml.org/>. Último acesso em: 08/11/2012.