

Unidade de Resposta Audível com Interação por Voz: Análise de Requisitos de um Estudo de Caso

Valéria Farinazzo Martins

Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie
Rua da Consolação, 930 – CEP 01302-907 – São Paulo – SP – Brasil

valeria.farinazzo@mackenzie.br

Abstract. *Hardware and communication media evolution brought out human-computer interfaces which use a telephone keyboard or the user voice, instead of the computer keyboard and mouse. By using the voice recognition and synthesis, it is possible to achieve a more natural communication way between user and the machine. This work aims to define the main characteristics and technologies used in these types of system, including the Interactive Voice Response (IVR), which implements the interface between the user and the system available information. It also aims to show, in a case study, the user and system requirements, as well as the implementation issues.*

Resumo. *A evolução do hardware e dos meios de comunicação trouxe interfaces homem-computador que usam um teclado de telefone ou a voz do usuário, em vez do teclado e do mouse do computador. Ao utilizar o reconhecimento de voz e síntese de voz, é possível obter uma forma de comunicação mais natural entre o utilizador e a máquina. Este trabalho tem como objetivo definir as principais características e tecnologias utilizadas nestes tipos de sistema, incluindo a Unidade de Resposta Audível (URA), que implementa a interface entre o usuário e as informações disponíveis do sistema. Também se pretende mostrar, em um estudo de caso, os requisitos do usuário e do sistema, bem como as questões de implementação.*

1. Introdução

Com o desenvolvimento cada vez mais promissor dos meios de comunicação, desde a década de 90, uma nova categoria de interfaces começou a fazer parte do cotidiano, em que o usuário interage com o sistema através da fala ou de teclas de tom pelo telefone, e o sistema responde através de síntese de voz e/ou de falas pré-gravadas; estes sistemas nasceram da ideia de autoatendimento via telefone. Neste caso, o usuário, mesmo inexperiente, deve conseguir navegar pelo sistema e obter as informações necessárias. Ele deve ser capaz de interagir com um sistema que “fala”, “escuta” e “entende”.

O estado da arte em reconhecimento/síntese de voz já permite que sistemas automáticos sejam desenvolvidos para trabalhar em condições reais [San-Segundo et al 2005]. Empresas como a Nuance (2013), através do sistema Dragon (2013), a IBM

ViaVoice e a Philips Speech – compradas há alguns anos pela Nuance e a iSpeech (2013) têm investido no desenvolvimento de sistemas de voz para domínios restritos. Assim, sistemas com estas tecnologias já são comercialmente viáveis em uma grande variedade de aplicações de *Call Center*. Estas tecnologias estão, aos poucos, substituindo o menu *touch-tone* (navegação por menu) que nunca conquistou os usuários mais exigentes.

Desenvolver aplicações com reconhecimento e síntese de voz é particularmente diferente de desenvolver, por exemplo, aplicações Web. Há requisitos principalmente não funcionais que não são considerados em aplicações que utilizam mouse e teclado, tais como a naturalidade da comunicação por voz e a prestimosidade da interface. As peculiaridades dos sistemas com interface de voz fazem surgir questões ligadas a critérios de usabilidade de interface. O estudo destes critérios fez surgir o presente artigo.

O artigo está organizado da seguinte forma: no item 2 são apresentados os fundamentos conceituais Reconhecimento de Voz, Síntese de Voz, Unidade de Resposta Audível e aplicações. No item 3 é descrito o estudo de caso real. Já no item 4 são apresentados os requisitos para este tipo de aplicação. O item 5 traz alguns aspectos de implementação. Finalmente, no item 6, são apresentadas as conclusões sobre o trabalho.

2. Fundamentação Teórica

2.1. Reconhecimento de Voz

Reconhecimento de voz pode ser visto como um processo pelo qual o computador pode converter um sinal de fala acústico em um texto. A partir daí, ele é capaz de executar ações sobre a interpretação do pedido [Cohen, Giangola e Balogh 2004] [Damasceno, Pereira e Brega 2005] [Martins, Brasileiro e Fernandes 2012] [Martins et al 2013].

O processo de converter voz em texto envolve os seguintes passos [Dantas 2000], [Lamel, Minker e Paroubek 2000], [San Segundo et al 2005], [Cohen, Giangola e Balogh 2004], [Mctear 2002] [Deng e Huang, 2004] [Martins, Brasileiro e Fernandes 2012]: a) O sistema captura a palavra dita pelo usuário como um sinal acústico analógico; b) O sistema faz a conversão deste sinal analógico em componentes digitais; c) O sinal digital é fracionado em sons distintos ou segmentos; cada segmento corresponde a um som específico, tais como consoantes e vogais; d) O software de reconhecimento de voz faz a classificação do som, determinando combinações possíveis entre os segmentos de som e as representações fonéticas; e) A aplicação de reconhecimento de voz busca a palavra ou frase que mais combina com o que disse o usuário.

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]. Sobre o vocabulário utilizado, quanto maior seu tamanho, maiores serão a complexidade, as exigências e a precisão do sistema. O tamanho do vocabulário depende da aplicação a ser desenvolvida [Martins, Brasileiro e Fernandes 2012].

2.2. Síntese de Voz

Síntese de voz, do inglês Text-to-Speech (TTS), é o processo que converte texto em voz. O sintetizador recebe um texto na forma digital e o transforma em ondas sonoras, isto é, faz uma vocalização do texto. Um programa de síntese de voz é útil nas situações em que o usuário não pode desviar a atenção para ler algo ou não tem acesso ao texto escrito, seja porque a informação está distante, porque deseja realmente ouvir o texto ou porque o usuário sofre de alguma deficiência visual [Guilhoto e Rosa 2001].

Embora a tecnologia TTS ainda não reproduza com fidelidade a qualidade da voz humana gravada - vale a pena destacar que, até o momento, os sintetizadores de voz não conseguem representar entonação - ela tem melhorado muito nos últimos anos. Tipicamente, o uso da voz humana pré-gravada está atrelado aos *prompts* e ao envio de mensagens para os usuários. No entanto, determinadas aplicações, como leitores de e-mail e notícias, têm informações muito dinâmicas. 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. Unidade de Resposta Audível

A utilização de URA está se tornando uma maneira popular de automação e otimização do atendimento telefônico, além de ser um canal de comunicação entre empresas. Assim como na Internet, há um esquema eficiente de relacionamento com clientes, abrindo possibilidades de diferenciação, onde a criatividade e a flexibilidade tecnológica passam a ser fundamentais para a área empresarial [Dillman et al 2009].

A URA conecta usuários utilizando um telefone com as informações que eles necessitam. Pode ser definida como um dispositivo em um sistema telefônico capaz de reproduzir as mensagens ao usuário, e também suportar as entradas deste e disparar requisições de ações para o sistema. É uma solução que permite interação com o usuário através de tecnologias como Reconhecimento de Voz Natural e de tons DTMF (Dual-Tone Multi-Frequency), Text-to-Speech e autenticação de voz [Salvador e Serrano 2003].

Os sistemas que utilizam entrada, apenas via teclado do telefone, necessitam de um menu sonoro, denominado *touch tone*; muitas vezes estes sistemas são cansativos para o usuário, pois é necessário escutar muitos menus (e em muitos níveis) para atingir a opção desejada. Embora existam técnicas, como o corte do menu, que visam melhorar a interação, estes sistemas geralmente não atingem a satisfação do usuário e começam a ser substituídos por entradas por voz. Apesar das URAs terem começado com interação por teclas do telefone, é por meio do reconhecimento de voz que se tem as maiores vantagens, pois a fala é inerente ao ser humano e sua comunicação é natural e simples. Estes sistemas, mais diretos, permitem que o usuário atinja mais rapidamente seus objetivos.

A Figura 1 mostra como é realizada a interação entre o usuário e o sistema para acesso a informações desejadas.

2.4. Aplicações Comerciais

As aplicações que utilizam URA variam bastante em grau de complexidade, visando à facilidade e agilidade de obtenção de informações tanto para o usuário quanto para o atendente humano que a ligação eventualmente possa ser direcionada. Podem ser utilizadas como parte fundamental de um sistema ou como um módulo dentro de um

sistema de telefonia mais complexo, como um sistema de celulares pré-pagos [Salvador e Serrano 2003]. São exemplos de sistemas que utilizam URA [Estabel, Moro e Santarosa, 2006] [Martins, Brasileiro e Fernandes 2012]:

- Discador Automático por Voz/Agenda - Permite que se possa realizar chamadas telefônicas a números previamente cadastrados, consultar os números telefônicos, enviar um e-mail utilizando comandos de voz, através de um telefone fixo ou celular.
- Envio de Mensagens Através do Correio Eletrônico - Através do telefone, pode-se enviar uma mensagem de correio eletrônico a um contato previamente cadastrado. É gerado um arquivo de voz que é indexado quando o e-mail é enviado.
- Notificações de Erros ou Condições Críticas no Sistema - Qualquer sistema de telefonia pode utilizar uma URA para notificar seus clientes de erros ocorridos no sistema ou condições críticas alcançadas pelos usuários, como, por exemplo, falta de créditos para executar uma chamada desde um celular pré-pago.

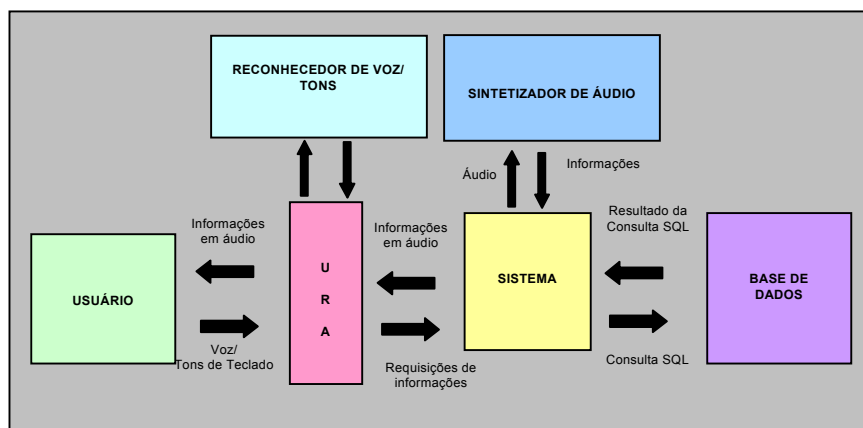


Figura 1. Arquitetura Básica de uma URA

3. Definição de um Estudo de Caso

O estudo de caso apresentado é um sistema de atendimento e transferência computadorizada de chamadas. Trata-se de um filtro que pode ser colocado quando se deseja diminuir o número de chamadas direcionadas diretamente para a telefonista. Trata-se de um caso real que foi implementado para um cliente. Os requisitos, assim como os protótipos gerados de maneira iterativa foram validados junto ao cliente e usuários reais do sistema.

O sistema deve ser capaz de reconhecer, através de voz, o nome da pessoa com quem o usuário deseja falar, ou o nome do departamento ou ainda o número do ramal. O usuário também pode, através de teclas, digitar o número do ramal, se preferir uma entrada por DTMF.

Para os casos de reconhecimento de voz, é utilizado um grau de confiabilidade para saber qual foi a taxa de reconhecimento da fala do usuário. Caso este grau de confiabilidade (configurável no sistema) seja baixo, a URA deve emitir a mensagem “Desculpe-nos, poderia repetir mais um vez?” a fim de que o sistema possa tentar reconhecer a entrada com um grau de confiabilidade maior. Se o sistema não compreender a entrada do usuário, depois de um número de tentativas de obter a

resposta, também configurável, o sistema transfere o usuário diretamente para a telefonista, tendo a URA emitido a mensagem “Desculpe-nos, estamos transferindo sua chamada para a telefonista”. Se, contudo, o grau de confiabilidade é médio, a URA deve emitir a mensagem “Você deseja falar com” composta pelo nome do usuário ou nome do departamento ou número do ramal, a fim de confirmar a entrada. Pode ainda o sistema reconhecer a entrada do usuário com grau de confiabilidade alto, em que não é necessário confirmar a entrada do usuário.

Uma vez que o sistema tenha a entrada correta do usuário, por voz ou DTMF, o sistema deve realizar uma consulta a uma base de dados que contém o ramal desejado. Outra informação vinda do resultado da consulta é se o sistema está autorizado a transferir a chamada diretamente para o ramal ou direcionar para a telefonista. No primeiro caso, a URA emite a mensagem de “Estamos transferindo sua chamada para” composta com o nome da pessoa ou do departamento ou o número do ramal. Assim, o sistema conecta o áudio do usuário com o do ramal desejado e desconecta o áudio da URA que fica liberada para um novo atendimento de chamada. Se o sistema tiver que transferir diretamente para a telefonista, a URA dará a mensagem “Estamos transferindo sua chamada para a telefonista” e o sistema conecta o áudio do usuário com a da telefonista e desconecta a URA.

4. Definição dos Usuários e Requisitos do Sistema

Requisitos são as necessidades do sistema e/ou do usuário; dizem respeito ao que o sistema deve ou não fazer e como fazer a fim de serem amigáveis para o usuário.

Um sistema deve ter seus requisitos muito bem especificados para que os erros desta fase não sejam transmitidos a fases subsequentes de projeto e implementação, desapontando os usuários, aborrecendo desenvolvedores ou mesmo podendo inviabilizar o sistema [Pressman 2004].

Seguindo uma sequência, é possível identificar os requisitos necessários desta fase:

- Definir os usuários com relação a experiências, atitudes, necessidades, capacidades humanas e motoras e desempenho;
- Qual é o ambiente em que a aplicação será usada e em qual contexto;
- Especificar os requisitos comuns a essas aplicações, os funcionais, os requisitos específicos da aplicação e os referentes à interface.

4.1. Definição dos Usuários, Tarefas e Ambientes

A definição dos usuários, tarefas e ambiente é essencial para a especificação de técnicas apropriadas de IHC que visem à criação de sistemas utilizáveis.

A eficácia da interface é um fator crucial. Questões vinculadas à utilização da representação da informação devem permitir uma nova compreensão dos dados na aplicação. Outro fator importante refere-se à usabilidade da interface. Ela deve ser avaliada de acordo com critérios como taxa de erros, tempo de execução de uma tarefa, carga de trabalho e avaliações subjetivas dos usuários. Finalmente, considerar o valor da interface no ambiente de trabalho e/ou físico em que o usuário, de fato, executa a tarefa em questão.

Os usuários. Para este tipo de sistema, não há como saber quem serão seus efetivos usuários, pois serão todas as pessoas que desejarem se comunicar com um determinado

departamento na empresa ou com alguém especificamente. É necessário que seja desenvolvido um sistema que não tenha necessidade de treinamento prévio, pois os usuários querem ser transferidos rapidamente e de maneira eficaz. Para um sistema de atendimento e transferência computadorizada de chamadas, é possível tratar usuários experientes de forma diferenciada dos principiantes. Assim, usuários experientes podem cortar a mensagem vinda da URA através de uma fala. O usuário experiente pode ainda digitar o número do ramal. Esta ação dá agilidade ao processo de transferência da chamada.

A tarefa. Na especificação dos requisitos de uma aplicação com URA, é importante que sejam definidas as metas da aplicação, as tarefas que serão necessárias para se alcançar essas metas, e como as tarefas serão realizadas no ambiente. Neste caso, o usuário pode interagir com o sistema através de voz ou de tons DTMF.

O ambiente. Em aplicações guiadas por voz, deve-se ter em mente que o usuário pode estar num ambiente físico silencioso ou com muita interferência sonora, pode estar utilizando um celular ou um telefone fixo. Assim, deve haver ajustes que controlam o volume com que a voz chega ao equipamento de reconhecimento de voz, uma tolerância a ruído e certo grau de liberdade na gramática para não tomar todas as palavras que chegam através do telefone [Salvador e Serrano 2003].

4.2. Requisitos Funcionais Comuns a Aplicações de Interface de Voz

Os requisitos comuns a aplicações de URA são os requisitos funcionais que devem ser considerados quando são projetadas praticamente todas as aplicações deste tipo, entre os quais é possível citar [Salvador e Serrano 2003]:

- Veracidade - a precisão com que a interface responde as interações do usuário, fazendo-o se sentir como se estivesse conversando com uma pessoa.
- Capacidade de Reconfiguração - é a habilidade que o sistema possui de ser modificado facilmente. Para uma URA, a facilidade de se trocar as mensagens dadas ao usuário deve ser considerada como ponto crucial, assim como configurar o grau de confiabilidade de compreensão das entradas do usuário.
- Multiusuário - deve permitir que vários usuários possam usar o sistema ao mesmo tempo. A URA deve permitir tantos usuários quanto o número de canais de entrada do PABX em que está inserida.
- Interatividade - é a capacidade do usuário poder se expressar e o sistema responder às entradas do usuário, seja através de tons DTMF ou voz.

Além disto, segundo [Mitchell 2007], [Rocha e Baranauskas 2003] e [Nielsen 2003], para se obter uma boa interface, é necessário que certos requisitos sejam atendidos:

- Diversidade - A interface deve suportar todas as classes de usuários, identificar cada usuário e adaptar-se a ele. Algumas estratégias são lançadas neste caso: desvio da chamada para atendentes ao perceber certa dificuldade do usuário; ajudas e exemplificações de uso colocadas no final de frases. Já para usuários experientes, deve haver certa diferenciação para não aborrecê-los e gerar agilidade em suas ações, como, por exemplo: ser possível cortar uma frase sem escutá-la inteira.

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

Para o estudo de caso em questão, foi determinado que ele deveria atender, ao máximo, os requisitos comuns e de interface. Assim:

- **Multiusuário:** determinado pelo número de canais do PABX disponíveis que chegam a URA.
- **Capacidade de reconfiguração:** é atingida através da fácil alteração de mensagens ao usuário, grau de confiabilidade de compreensão do usuário, grau de liberdade da gramática, nível de ruído permitido, etc.
- **Diversidade:** sistema deve guiar usuários principiantes e experientes para que ambos atinjam seus objetivos. Algumas estratégias serão explanadas na seção 5.
- **Flexibilidade:** Para este tipo de interface é possível realizar entrada de dados através de tons DTMF ou voz; ainda através de voz, é possível dizer o nome da pessoa, ou do departamento ou o número do ramal.
- **Prestimosidade:** Neste caso, se o usuário não consegue se comunicar de forma eficaz com o sistema, então este deve transferi-lo para a telefonista antes que o usuário se aborreça.

- Eficiência, Satisfação e Passividade: o sistema deve reconhecer para onde a chamada deve ser transferida rapidamente para não frustrar o usuário. Se muitas chamadas são transferidas para a telefonista porque o sistema não reconheceu a entrada do usuário, então o sistema deve ser reajustado. O sistema também deve guardar em arquivo todas as chamadas bem sucedidas e mal sucedidas a fim de que seja possível analisar o desempenho do sistema.
- Naturalidade: a própria comunicação por voz permite que a interface seja mais natural.

4.3. Algumas Recomendações sobre Interface de Voz

O projeto de interfaces de voz é bastante crítico em sistemas que trabalham como *Front end* com o usuário. Mesmo algo simples como o *touch-tone* pode ser bem ou mal projetado. Com voz, um projeto de interface deve colocar o usuário em um diálogo agradável, e não ambíguo, projetado para reconhecimento de alta precisão. Dantas (2000) descreve algumas orientações sobre a seleção de bons serviços ativados por voz:

- Um padrão de voz e um estilo próprios devem ser fixados, para permitir “personalidade” para o serviço, seja pelos apelos criados ou como são gravados.
- Múltiplas opções de navegação devem ser oferecidas. É crucial entender se tanto os usuários experientes quanto os novatos se sentirão cómodos com as instruções passo a passo e as opções de linguagem natural disponibilizadas, ou se é conveniente tratá-los diferentemente.
- Os serviços devem ser amigáveis; recomenda-se o uso de frases educadas, tais como: “sinto muito, eu não entendi” ao invés de frases técnicas como “este dado não é válido”.
- O sistema deve ser constituído de forma a induzir o usuário a solicitar naturalmente informações específicas ao invés de questões abertas. Por exemplo, deve perguntar ao usuário “Com quem você deseja falar?” ao invés de “Em que posso ajudá-lo?”.

5. Aspectos de Implementação

Os principais aspectos de implementação referentes à interface do estudo de caso são destacados a seguir.

- Controle de entradas mal sucedidas do usuário: É possível configurar a quantidade de vezes que se deseja tentar obter a correta entrada do usuário. Se, nestas tentativas, o sistema não obtiver a compreensão da fala do usuário, a chamada é transferida para a telefonista.
- Corte do menu: A fim de não frustrar e irritar usuários mais experientes que não queiram ouvir a mensagem completa vinda da URA, é possível realizar a entrada por voz; o sistema, então, corta a reprodução de áudio e tenta fazer o reconhecimento desta entrada do usuário.
- Ajuste da Porcentagem de Reconhecimento: É possível ajustar o nível em que o sistema reconheceu determinada entrada do usuário. Compreensão baixa: o sistema informa que não compreendeu corretamente a entrada e solicita que o

usuário tente novamente. Compreensão média: é feita a confirmação do usuário, repetindo a entrada que supostamente foi dada. Compreensão alta: o sistema executa a ação, podendo mencionar qual será a ação a ser executada.

- Ajuste de Interferência Sonora: Configuração do nível de ruído permitido na chamada. Isto é importante quando o usuário está num ambiente físico com barulho.

6. Conclusões

Foi verificado que sistemas que utilizam Unidade de Resposta Audível como interface devem ter seus requisitos muito bem definidos, e, além disto, devem ser levados em conta, os requisitos de interface e também as metas de interface, já que este tipo de sistema dificulta o processo de treinamento de classes de usuários. O sistema deve ser bastante natural ao usuário, como se a interação fosse entre duas pessoas.

Outro cuidado que se deve ter é o de não frustrar nenhuma classe destes potenciais usuários, proporcionando facilidades de uso àqueles usuários experientes.

O sistema deve ser bastante rápido em suas interações com o usuário, além de ser configurado de forma que se tenha um reconhecimento de voz bastante eficaz.

A evolução destas classes de sistemas está intrinsecamente relacionada ao estudo e ao avanço das áreas de reconhecimento de voz, autenticação e síntese de voz.

O reconhecimento de voz já começa a ser o método mais cômodo para a entrada de dados, edição de textos e computação por conversação, uma vez que a fala é o meio mais fácil e natural para diversas aplicações. Atualmente, o reconhecimento de voz permite que usuários recuperem informações de empresas e realizem transações até mais rapidamente que através da web, mediante uma linha telefônica, se o sistema for bem projetado.

Referências

- Cohen, M. H., Giangola, J. P., Balogh, J. (2004). Voice User Interface Design, Addison Wesley, ISBN 0-321-18576-5, 368 páginas.
- 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.
- Dantas, E. B. (2000). Telemarketing, A Chamada para o Futuro. Editora Atlas, 4ª edição, São Paulo.
- Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Social Science Research, 38(1), 1-18.
- Estabel, L.B.; Moro, E. L.; Santarosa, L. C. (2006). 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.
- Ghilhoto, P. J. S., Rosa, S.P.C.S. (2001). Reconhecimento de Voz. Trabalho de Síntese para a disciplina de Sistemas Multimídia, Licenciatura em Engenharia Informática, Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Ispeech. Disponível em: <http://www.ispeech.org/>. Acesso em: 31/01/2013.

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, In: 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>

Martins, V. F.; Santos, A. G.; Rodrigues, F. A.; Okumura, M. H.; Sakoda, T. J.; Guimarães, M. P. (2013). Análise, projeto e implementação de uma aplicação utilizando interface de voz com o usuário. In: Computer on the beach, 2013, Florianópolis. Anais do Computer on the beach 2013.

Mctear, M. F. (2002). Spoken Dialogue Technology: Enabling the Conversational User Interface, In: ACM Computing Surveys, Vol. 34, No. 1, pp. 90–169.

Mitchell, P.P. (2007). A step-by-step guide to usability testing, Lincoln, NE: iUniverse.

Nielsen, J. (2003). Usability Engineering. Academic Press, Cambridge, MA.

Nuance. Disponível em: <http://www.nuance.com/for-individuals/by-product/dragon-for-pc/index.htm>. Acesso em 31/01/2013.

Pressman, R.S. (2004). Software Engineering. 7th ed., Addison-Wesley.

Rocha, H. V., Baranauskas, M. C. C. (2003). Design e avaliação de interfaces humano-computador, NIED, Instituto de Computação, Unicamp, Campinas.

Salvador, V. F. M., Serrano, D. (2003). Especificação de requisitos aplicados a sistemas que utilizam URA e um estudo de caso. In: Cadernos do Centro Universitário São Camilo, v.9, n.4, ISSN 0104-5865.

San-Segundo, R.; Montero, J. M.; Macías-Guarasa, J.; Ferreiros, J.; Pardo, J. M. (2005). Knowledge-Combining Methodology for Dialogue Design in Spoken Language Systems, In: International Journal of Speech Technology 8, 45-66, Springer Science + Business Media.