

Avaliação de Interfaces Multimodais para Livros Falados Digitais com foco Não Visual

Carlos Duarte Teresa Chambel Hugo Simões Luís Carriço
LaSIGE / Faculdade de Ciências da Universidade de Lisboa
{cad,tc,hsimoes,lmc}@di.fc.ul.pt

Eduardo Santos, Guilherme Francisco, Sérgio Neves, Ana Catarina Rua, João Robalo, Tiago Fernandes
Faculdade de Ciências da Universidade de Lisboa
{i32097,i32169,i32536,i31586,i32059,i29059}@alunos.di.fc.ul.pt

Sumário

Este artigo apresenta um estudo de avaliação de interfaces multimodais para Livros Falados Digitais, tendo em conta diversos tipos de utilização, inspirados quer por diferentes características dos utilizadores quer pelas situações de uso. São apresentados: uma análise de trabalhos relevantes na área e requisitos a ter em conta; o Rich Book Player, um dispositivo de reprodução de livros falados digitais, multimodal e adaptativo; e um estudo de usabilidade das suas funcionalidades e capacidades de interacção, com ênfase em aspectos não visuais.

Palavras-chave

Livros Falados Digitais, Interfaces Multimodais, Interfaces Não Visuais, Acessibilidade, Usabilidade, Avaliação

1. INTRODUÇÃO

Com o desenvolvimento da sociedade da informação, os pressupostos sobre o utilizador típico têm vindo a alterar-se. Este já não é, necessariamente, uma pessoa que utiliza a tecnologia essencialmente no ambiente de trabalho e sem necessidades especiais de interacção. A convergência das tecnologias da informação, telecomunicações e electrónica de consumo, introduziu alterações radicais no mercado, colocando uma maior ênfase na componente residencial e na mobilidade, e consequentemente no tipo e contexto de uso das aplicações interactivas. Em face desta mudança de paradigma, o público alvo é bem mais diverso e heterogéneo, quer em termos de conhecimentos e envolvente cultural, quer de capacidades e contextos de utilização, e os dispositivos de acesso mais diversificados em termos de plataformas tecnológicas. Neste contexto, tornam-se cada vez mais relevantes os aspectos de acessibilidade e usabilidade para todos, em qualquer lugar e em qualquer momento [Stephanidis2000].

Os livros falados, tradicionalmente gravados em meios analógicos, disponibilizam versões de livros impressos, tornando-os acessíveis a pessoas ou contextos de utilização diversos do tradicional. As pessoas com deficiências visuais têm sido algumas das que mais têm beneficiado com esta tecnologia, apesar das limitações de usabilidade que estes livros apresentam. Em versão digital, estes livros podem disponibilizar um leque mais rico de capacidades, colmatando problemas de usabilidade encontra-

dos nas versões analógicas, incluindo melhores formas de navegação e a possibilidade de armazenar vários livros em dispositivos de menores dimensões. Com a possibilidade de integrar diversos meios, como o áudio, o texto do livro e imagens, os livros falados digitais podem ser mais versáteis e facilitar a disponibilização de múltiplas modalidades de interacção, o que pode ser atractivo para leitores com dificuldades perceptuais permanentes ou temporárias, como é o caso dos cegos, ou pessoas com outras deficiências visuais, motoras, ou outro tipo de dificuldade em lidar com versões impressas dos livros [Moreno2000]. As capacidades multimédia e multimodais destes livros podem ser exploradas para enriquecimento da experiência de leitura destas pessoas, mas também de pessoas sem necessidades especiais, ou em contextos situacionais que imponham temporariamente restrições em termos cognitivos, perceptuais e de atenção [Gazzaniga1998].

O objectivo do presente estudo consistiu na avaliação de interfaces multimodais de livros falados digitais, tendo em conta diversos tipos de utilização, inspirados quer por diferentes características dos utilizadores quer pelas situações de uso. Para tal, foram analisados trabalhos relevantes na área e requisitos a ter em conta. Foi ainda feito um estudo de usabilidade do Rich Book Player [Duarte2006], um dispositivo de reprodução de livros falados digitais, multimodal e adaptativo, desenvolvido

no contexto do projecto de investigação RiCoBA¹ – Rich Content Books for All.

Este trabalho resultou de uma colaboração entre elementos da equipa do RiCoBA, sendo um deles docente da disciplina de Interfaces Pessoa Máquina do 2º. Ano da Licenciatura em Eng. Informática da Faculdade de Ciências da Universidade de Lisboa, e dois grupos de três estudantes dessa disciplina. Cada um destes dois grupos incluiu um estudante invisual. A sua experiência como utilizadores de tecnologias com foco na acessibilidade, em particular com interfaces não visuais, aliada à experiência dos restantes colegas em interfaces visuais, e num contexto de aprendizagem sobre a concepção, desenvolvimento e avaliação de interfaces, contribuiu para um cenário mais rico de trabalho. Este estudo surgiu por um lado como um desafio, por outro como uma oportunidade, permitindo enquadrar estes alunos num contexto de aprendizagem tradicionalmente mais focado para as dimensões visuais da interacção.

Depois desta introdução, a secção 2 introduz os conceitos e apresenta os requisitos e o estado da arte que definem o contexto deste trabalho. A secção 3 descreve a aplicação de reprodução de livros falados digitais Rich Book Player, evidenciando os princípios subjacentes e as principais funcionalidades. A secção 4 apresenta a configuração dos testes de avaliação efectuados, sendo os seus resultados apresentados e discutidos na secção 5. O artigo termina com a apresentação das principais conclusões e identificação de direcções para desenvolvimentos futuros.

2. CONTEXTO

Esta secção introduz os conceitos, requisitos e estado da arte nos aspectos que definem o contexto em que o trabalho se insere.

2.1 Multimodalidade e Acessibilidade

As interfaces multimodais permitem que o utilizador tire partido do seu modo ou modos mais naturais de comunicação, em cada tarefa ou situação, numa interacção com o computador. Estas interfaces podem ser usadas em dispositivos de grandes ou pequenas dimensões [Cohen2004]. Outra das suas vantagens é a desambiguação mútua, podendo a informação de uma das fontes ajudar a resolver ambiguidades na informação, reduzindo erros [Oviatt1999]. Por exemplo, num livro falado, a visualização de uma palavra escrita pode ajudar a desambiguar o sentido de palavras homófonas.

Para tirar partido das capacidades de processamento dos utilizadores (incluindo a atenção, a memória de trabalho e a tomada de decisões), os designers de interfaces multimodais deverão conceber os seus sistemas de forma a maximizar as habilidades cognitivas e físicas dos seus utilizadores. Algumas directivas que apontam nesse sentido incluem: 1) evitar a apresentação desnecessária de informação em duas modalidades diferentes, em situações em que o utilizador tem que prestar atenção simul-

tânea às duas fontes para compreender a informação, evitando sobrecarga cognitiva [Kalyuga1999]; 2) Maximizar as vantagens de cada modalidade, para reduzir sobrecarga da memória nalgumas tarefas ou situações [Wickens1992]. Exemplos desta combinação incluem: o uso combinado de apresentação visual combinado com interacção manual de informação espacial; e o uso combinado de apresentação áudio combinada com interacção por voz.

As abordagens tradicionais à acessibilidade, consideradas de natureza reactiva, têm visado a criação de configurações alternativas de hardware e software que permitam aos utilizadores com necessidades especiais aceder a aplicações originalmente desenvolvidas para utilizações tradicionais. As duas abordagens tecnológicas para a acessibilidade são: a adaptação ao nível do produto, onde cada aplicação é tratada em separado e criada uma versão de acessibilidade alternativa; e a adaptação ao nível do ambiente, onde a intervenção é feita de forma a disponibilizar software e hardware que permitam que este, e consequentemente as aplicações nele executadas, sejam acessíveis por meios alternativos.

As Interfaces Unificadas, Unified User Interfaces [Stephanidis2000], disponibilizam um conjunto de metodologias, técnicas e ferramentas que permitem a adopção de um paradigma mais proactivo para lidar com os requisitos das pessoas e situações em que há necessidades especiais. O conceito base consiste na criação de artefactos com interfaces únicas que encapsulam elementos alternativos de interacção mais adequados a cada pessoa ou situação. Um exemplo de uma abordagem reactiva na concepção de livros falados seria deixar a leitura do livro a cargo de um leitor de ecrã, como os que são usados pelas pessoas com dificuldades visuais no acesso à maioria das aplicações. De acordo com uma abordagem proactiva, a aplicação que suporta os livros falados disponibiliza versões lidas por locutores humanos, ou tem a capacidade de sintetização de voz, concebidos e integrados na própria aplicação. Esta abordagem tem-se mostrado viável mesmo em interfaces de uso geral e grande escala, como é o caso do web browser AVANTI [Stephanidis1997] que, de acordo com os autores, tem o potencial de apresentar melhores resultados que as interfaces dedicadas, desenvolvidas explicitamente para utilizadores com necessidades especiais, quer em termos de acessibilidade e qualidade da interacção, quer em termos de eficácia e eficiência económica a longo prazo.

2.2 Livros Falados Digitais

Os Livros Falados Digitais (LFD) surgem como uma evolução dos livros falados analógicos, disponibilizando meios de interacção inspirados na forma como interagimos com livros impressos. Os livros falados, tradicionalmente gravados em meios analógicos, como as cassetes áudio, disponibilizam versões de livros impressos falados ou lidos por pessoas. Quando comparados com os livros impressos, os livros falados apresentam algumas limitações: sendo gravados sequencialmente, são limitados no acesso aleatório; Não é possível interagir com os

¹ hcm.di.fc.ul.pt/ricoba/

livros, marcando páginas ou efectuando anotações; Não é possível aceder à forma escrita das palavras, uma queixa frequente dos utilizadores; Apenas disponibilizam uma versão dos livros; As formas de lidar com notas de rodapé, lidas quando referenciadas ou fora da sequência de leitura do livro, não agradam a todos nem em todas as situações de leitura. Os livros falados digitais permitem maior flexibilidade na leitura, criar marcas de página, evidenciar porções de texto e fazer anotações. O texto do livro pode ser apresentado de forma sincronizada com a sua leitura em áudio, permitindo a localização e visualização da forma escrita das palavras lidas. Também é mais fácil localizar, saltar ou ler as notas de rodapé quando for mais oportuno.

As pessoas com dificuldades visuais têm sido das que mais têm beneficiado com o uso de livros falados, para quem os livros impressos são tipicamente inacessíveis, com a excepção das versões em Braille, nem sempre disponíveis e limitadas em capacidades. No entanto, estes livros têm outros adeptos, e nos países onde são mais divulgados, é mais frequente encontrar pessoas que ouvem livros falados por exemplo no carro, enquanto conduzem, ou em casa, por vezes enquanto desempenham outras tarefas. Alguns dos problemas reportados quer em livros falados quer em livros Braille para material estruturado [Aldrich1988, Arons1997, Petrie1996] incluem: a necessidade de ter mais de uma cassete áudio ou mais de um volume com o livro Braille, para cada livro impresso; é frequentemente difícil e moroso navegar e localizar informação, especialmente em várias cassetes ou volumes; é difícil fazer uma leitura rápida de livros baseados em cassetes áudio.

O consórcio Daisy [Daisy] é uma entidade responsável pela publicação de normas e especificações decorrentes da investigação no domínio dos livros falados digitais dos últimos anos. A National Information Standards Organization (NISO), em cooperação com o consórcio Daisy, publicou em 2002 a norma actual de LFD, a ANSI/NISO z39.86-2002 [NISO2002], também conhecida como DAISY 3.0. A norma foca a estrutura e conteúdo necessários à produção de livros falados, mas não inclui especificações para os dispositivos e aplicações de leitura destes livros. O Playback Device Features List [NISO1999], um documento auxiliar criado durante o desenvolvimento da norma, descreve as principais características que estes deverão ter, incluindo por exemplo: a não dependência de um dispositivo visual, tabelas de conteúdo navegáveis, acesso a níveis mais finos de detalhe, facilidade de aceder directamente a um destino, capacidade de adicionar informação como destaques e anotações, leitura de anotações, estabelecimento de marcações, avanço e recuo rápidos, disponibilização de fala humana e electrónica. Este documento não tem um carácter normativo e não apresenta soluções específicas de concretização. Consequentemente, os leitores de livros desenvolvidos de acordo com as normas de LFD adoptam diferentes soluções para a apresentação e interacção com os livros falados, muitas delas sofrendo de proble-

mas de usabilidade significativos, condicionando muitas vezes a acessibilidade das audiências a que se destinam.

2.3 Dispositivos de Reprodução de Livros Falados Digitais

Como sugerido pela especificação ANSI/NISO z39.86-2002, um dispositivo de reprodução, ou leitor, de um LFD é multimodal, apresentando o texto e o áudio narrado de forma sincronizada. São identificadas várias situações em que a interacção multimodal pode ser benéfica para o utilizador. Por exemplo, o reconhecimento de fala pode ser usado como modalidade de entrada, e a presença de marcações de página podem ser feitas de forma visual ou auditiva. No entanto, a maioria dos leitores de livros não fazem uso de múltiplas modalidades, nem vão muito para além da narração sincronizada com o texto.

O DAISY Playback Software é descrito como o primeiro dispositivo de reprodução de livros falados digitais num PC [Morley1998a], apresentando melhorias sobre o dispositivo dedicado PlexTalk Player [Plextor] também para livros DAISY. O leitor para PC foi concebido para utilizadores avançados que requerem acesso rápido e eficiente a livros estruturados. Foi concebido para PC, como aplicação autónoma que não faz uso de leitores de ecrã adicionais; é dirigido a utilizadores cegos, com deficiências de visão, ou visuais, sendo orientado para interacção por teclado e rato, com monitores capazes de apresentar letras de diversas dimensões; e permite pesquisas de texto nos títulos. Os comandos foram concebidos para serem consistentes e fáceis de usar, através do teclado normal, do teclado numérico, ou através do rato. A avaliação efectuada identificou falta de informação estrutural em interacções não visuais, devido à falta de indicações sonoras de informação que visualmente se distingue por tipo de letra diferente. A navegação entre secções também foi considerada confusa, em especial nas secções iniciais ou finais em navegação circular. No entanto, a maioria dos comandos foram considerados fáceis de usar e úteis. Para resolver os problemas encontrados, os autores sugerem a inclusão de sons não falados, e a adição de novos comandos.

Em [Duarte2005] é feita uma avaliação baseada em heurísticas de oito leitores de livros falados digitais, desenvolvidos de acordo com as normas DAISY e ANSI/NISO, incluindo os leitores AMIS 1.3, Book Wizard Reader, EaseReader, eClipseReader, gh Player 2.0, TAB Player, TPB Reader e Victor Reader Soft. Foram avaliadas as capacidades de navegação, as possibilidades de personalização oferecidas, a implementação de marcas de livros e de anotações, e a utilização de diferentes modalidades. Muitos dos leitores analisados não seguem várias das recomendações da norma ANSI/NISO. A maioria destas limitações não afecta fortemente interacções visuais, mas a interacção não visual é fortemente penalizada. De notar ainda que a maioria destes leitores não seguem as recomendações sobre a combinação criteriosa de modalidades referida anteriormente, resultando por vezes em experiências de leitura excessivamente

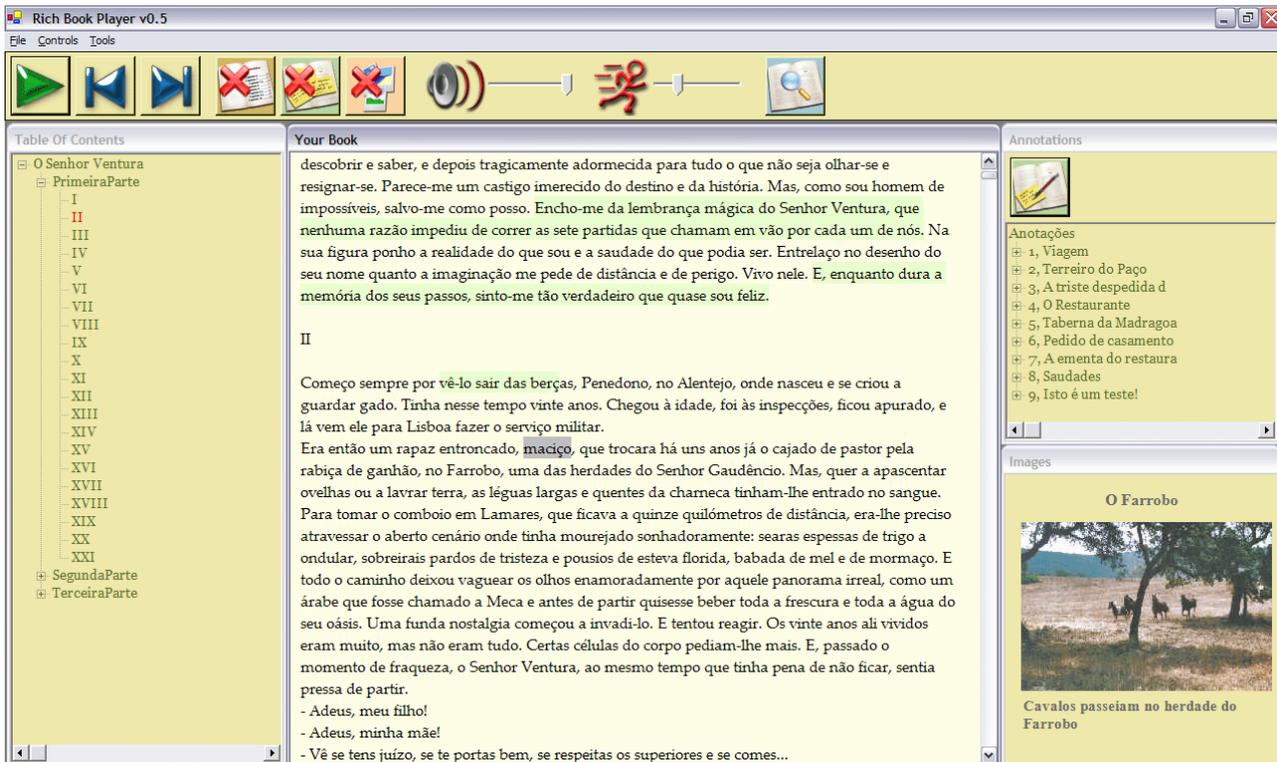


Figura 1 – O Rich Book Player, com todos os componentes visíveis.

complexas, com problemas de usabilidade e acessibilidade.

2.4 Outras interfaces não visuais

O estudo e desenvolvimento de interfaces não visuais tem sido efectuado noutras áreas relacionadas. Os sistemas hipermedia para invisuais [Petrie1996, Morley1998b] abordam aspectos semelhantes aos dos livros falados, com uma maior ênfase na definição, percepção e navegação de ligações, e operações de localização no hiperespaço. Um sistema que suporta interacção visual e não visual para a web é descrito em [Ntoa2005].

Um aspecto significativo na apresentação de livros é a apresentação de informação complexa de forma sumariada, como é o caso da informação contida numa tabela ou num mapa. Em [Kildal2006] é descrita uma interface não visual para tabelas, baseada na sonorização dos seus valores, por linha ou coluna, em diferentes níveis de detalhe. Em [Zhao2004] é apresentada a sonorização de informação geo-referenciada.

Em [Crispien1996] é apresentado um ambiente auditivo espacial que combina a apresentação espacial de informação com interfaces baseadas em gestos e fala. O utilizador é rodeado por um anel virtual contendo objectos de interacção auditivos, num diálogo de navegação multimédia. Os áudiojogos² constituem um exemplo de uma área de aplicação onde se exploram estes tipos de navegação, sendo o efeito de Doppler uns dos usados para percepção espacial.

² www.audiogames.net/

3. RICH BOOK PLAYER

Os livros falados combinam a apresentação visual e a narração de conteúdos literários de forma a melhorar quer a usabilidade, quer a acessibilidade a esses conteúdos. A combinação da representação áudio com o texto dos livros, que aparentemente seria apenas direccionada para utilizadores sem problemas de visão, permite oferecer funcionalidades, como a pesquisa, impraticáveis em livros com suporte apenas áudio, e que trazem benefícios acrescidos a utilizadores com baixa visão e cegos.

3.1 Concepção

Com o objectivo de tornar o alcance destes livros ainda mais abrangente, desenvolveu-se o conceito de Livro Falado Enriquecido [Carriço2003], que combina, para além do texto e do áudio, outros componentes que se encontrem disponíveis, e que podem incluir, por exemplo, imagens, vídeos, música, ou outros sons. Para além disso, os livros falados deverão oferecer aos seus utilizadores a possibilidade de criar marcações e anotar o conteúdo.

De modo a cumprir estes e outros requisitos enunciados anteriormente, desenvolveu-se uma aplicação, o Rich Book Player (Figura 1), que permite a reprodução de Livros Falados Digitais Enriquecidos. Esta aplicação foi desenvolvida com base na FAME (Framework for Adaptive Multimodal Environments) [Duarte2006]. A concepção da aplicação baseada nesta framework permitiu desenvolver, de raiz, uma aplicação que considera os diferentes aspectos de interacção e apresentação, aplicáveis à diversidade de potenciais utilizadores e situações de uso, indo assim ao encontro do conceito de interface

unificada [Stephanidis2000]. Desta forma, a aplicação será capaz de funcionar tanto em ambientes que disponham apenas de dispositivos que permitam entradas e saídas visuais, como em ambientes que disponham apenas de dispositivos áudio, como em ambientes onde seja possível combinar os dois tipos de dispositivos. É assim possível, com a mesma aplicação, oferecer as mesmas funcionalidades a uma diversidade de utilizadores que inclui os utilizadores sem deficiências, os utilizadores com baixa visão ou cegos, e os utilizadores com deficiências motoras mas que possam recorrer a comandos vocais.

Também ao nível dos conteúdos, a aplicação possibilita dois modos de operação, de modo a cobrir o máximo possível de situações de utilização. O modo de operação preferencial recorre a gravações áudio dos livros. O modo alternativo, a ser empregue em situações em que, por exemplo, não haja possibilidade de armazenar as gravações áudio, utiliza síntese de voz para transmitir o conteúdo.

Tanto para o reconhecimento, como para a síntese de voz empregou-se o Microsoft Speech 5.1. De notar que, quer para um, quer para outro, só se encontravam disponíveis motores de reconhecimento em Inglês, tendo sido por isso necessário adaptar a fonética dos comandos que eram fornecidos a ambos os motores, de modo a permitir o funcionamento da aplicação em Português. Por exemplo, para que o comando “início” pudesse ser reconhecido pelo motor de reconhecimento em Inglês teve de ser escrito como “nyssi”. Este factor poderá ter condicionado os resultados dos testes à frente apresentados, visto que se verificou uma evolução positiva na taxa de reconhecimento do reconhecedor com o avançar da utilização. De forma a minimizar este efeito, foi dada especial atenção à selecção dos comandos disponíveis para utilizar. A selecção dos comandos e do retorno a dar ao utilizador são aspectos fundamentais na concepção de interfaces baseadas em áudio.

3.2 Funcionalidades

Actualmente, o Rich Book Player disponibiliza as seguintes funcionalidades aos seus utilizadores:

Navegação – no Rich Book Player, a navegação pode ser realizada de diferentes formas:

- Através da selecção livre de um ponto de leitura no texto, sendo a narração continuada a partir desse ponto;
- Através da selecção do capítulo ou secção desejada na tabela de conteúdos;
- Através de comandos que permitem avançar (recuar) para a próxima (anterior) frase ou parágrafo ou secção ou capítulo;
- Através de comandos que permitem ir directamente para o início ou fim do livro;
- Através de comandos que permitem ir para a anotação seguinte ou anterior.

Marcações e anotações – o Rich Book Player permite aos seus utilizadores criarem marcações em qualquer parte do texto, bem como seleccionarem excertos do texto e associarem-lhes anotações. O utilizador pode posteriormente consultar as várias anotações criadas, bem como utilizá-las para navegar no texto. De cada vez que o utilizador fecha a aplicação, é criada, automaticamente, uma marcação que permite recomeçar a leitura desse ponto quando o utilizador regressar à aplicação. O conteúdo do livro que tenha sido anotado é apresentado com uma cor de fundo diferente para assinalar essa situação.

Apresentação de outros conteúdos multimédia – o Rich Book Player permite a apresentação de conteúdos multimédia associados a partes do livro. Esses conteúdos podem incluir imagens, músicas ou outros sons. O comportamento por omissão é avisar o utilizador sobre a presença desses conteúdos, deixando a este a possibilidade de os ver ou ignorar. Este comportamento da aplicação evolui de forma a reflectir as preferências do utilizador quanto à exibição destes conteúdos.

Pesquisa – é possível ao utilizador efectuar pesquisas no Rich Book Player. O utilizador pode navegar nos resultados da pesquisa, sendo possível avançar ou recuar pelos vários resultados retornados pela pesquisa. O texto pesquisado é apresentado com uma cor de fundo diferente, de forma a realçá-lo, bem como a frase em que está inserido, que é apresentada com outra cor de fundo, de forma a contribuir para uma melhor visualização e mais rápida identificação do resultado da pesquisa.

Controlo da narração – é permitido ao utilizador controlar a narração áudio do texto, através de comandos básicos como parar ou iniciar a leitura. É ainda possível controlar o volume da narração, bem como a sua velocidade.

Sincronização – o Rich Book Player apresenta um mecanismo de sincronização visual que permite acompanhar no ecrã a palavra que está a ser narrada. Essa palavra é destacada alterando a sua cor de fundo. O utilizador, se pretender, pode seleccionar como objecto de destaque a frase ou o parágrafo a que a palavra pertence. A unidade de sincronização também é adaptada pela aplicação aquando do aumento da velocidade de narração, visto que a velocidades mais elevadas, acompanhar visualmente a palavra que está destacada começa a exigir maiores esforços cognitivos aos utilizadores.

Caminhos alternativos de leitura – é possível definir caminhos alternativos à narração principal. Actualmente, a única possibilidade é definir como texto a ser lido o conteúdo que foi alvo de anotações. No futuro, outros mecanismos de suporte a esta funcionalidade serão incluídos, possibilitando ao utilizador marcar texto com diferentes categorias (análogo ao procedimento de sublinhar texto com diferentes cores) e definir caminhos de leituras a partir dessas marcações.

Disposição automática dos componentes visuais – o Rich Book Player apresenta visualmente quatro componentes: o conteúdo principal do livro, a tabela de conteúdos, as anotações, e um espaço para apresentação de outros con-

teúdos como imagens ou vídeos. Estes componentes podem surgir ou desaparecer, quer por ordem do utilizador, quer por decisão da aplicação. O utilizador pode alterar a distribuição espacial dos componentes, e o Rich Book Player automaticamente adapta a posição dos outros componentes, movendo-os e alterando o seu tamanho, de acordo com dois requisitos: ocupar o espaço que tem disponível, e minimizar os movimentos do componente que apresenta o conteúdo principal do livro. Este último requisito tem como objectivo limitar ao máximo a possível distração causada por movimentos deste componente, que será o seguido com mais atenção pelo leitor, visto que alterações do tamanho ou da posição do componente levam a uma alteração da distribuição do texto no ecrã, o que pode levar a uma desorientação temporária do leitor que se vê obrigado a encontrar novamente o texto que estava a ler.

Mecanismos de percepção – o Rich Book Player consegue transmitir aos seus utilizadores a posição em que se encontram dentro do livro de várias formas. Visualmente, essa informação é transmitida: 1) mudando na tabela de conteúdos, a cor correspondente à entrada do capítulo que está a ser lido; e 2) através da posição da barra de deslocamento vertical na janela do conteúdo principal do livro, de forma a ter sempre visível o texto a ser lido. Através de áudio, de cada vez que se entra dentro de um novo capítulo, a sua identificação é anunciada por síntese de voz. Para além disso, outra informação de percepção é transmitida pelo Rich Book Player, relativamente à presença de texto anotado ou de texto com outros conteúdos multimédia anexados. O utilizador é informado através de avisos visuais sob a forma de ícones a piscar, e de uma voz sintetizada que indica o tipo de conteúdo ou a existência de uma anotação. O texto que tenha sido anotado também é apresentado com uma cor de fundo diferente.

Modos de interacção – o Rich Book Player permite aos seus utilizadores optarem por utilizar o teclado e o rato, ou comandos de voz, como modalidades de entrada. As modalidades de saída incluem voz, gravada e sintetizada, texto, imagens, e sons.

4. CONFIGURAÇÃO DOS TESTES DE AVALIAÇÃO

A avaliação preliminar do Rich Book Player decorreu no laboratório de usabilidade do LaSIGE, Laboratório de Sistemas Informáticos de Grande Escala da Faculdade de Ciências da Universidade de Lisboa. Durante a realização dos testes, procedeu-se à captura dos comandos de voz utilizados no mesmo computador onde foi executada a aplicação. Outro computador capturou as saídas áudio e vídeo que eram apresentadas ao utilizador. Com uma câmara de vídeo capturou-se a actividade do utilizador durante o teste.

Durante os testes, foi utilizado o livro "O Senhor Ventura" da autoria de Miguel Torga, gravado previamente por uma locutora profissional.

Os testes foram efectuados em duas fases. Os resultados da primeira fase permitiram identificar aspectos a melhorar e conceber novos comandos para colmatar insuficiências detectadas, em especial na interacção não visual. Alguns foram introduzidos na aplicação antes da segunda fase de testes, outros foram testados utilizando a técnica do feiticeiro de Oz.

No início de cada sessão de testes, era apresentado ao utilizador o contexto de uso da aplicação, sendo-lhe depois permitido um período de familiarização com esta de cerca de dez minutos. De seguida, eram propostas ao utilizador um conjunto de tarefas que implicavam exercitar diversas funcionalidades da aplicação, desde as capacidades de navegação e pesquisa, até à criação de anotações. Depois de terminadas as tarefas, era pedido ao utilizador para preencher um questionário que cobria vários aspectos, como a utilidade das funcionalidades disponibilizadas, a satisfação com o desempenho da aplicação, e a facilidade de utilização.

Desta forma, foi possível durante uma sessão de testes conseguir três tipos de medidas através de três métodos de avaliação empírica [Jordan1998]. Os questionários permitiram obter medidas com origem nos utilizadores, a experiência controlada permitiu realizar observações, e os registos de utilização permitiram obter medidas de uso através das diferentes capturas empregues.

De referir ainda que, numa segunda fase de experiências, se empregou a técnica do feiticeiro de Oz para avaliar a introdução de novas funcionalidades, mais direccionadas para a interacção não visual, como a possibilidade de controlar o nível de retorno áudio oferecido pela aplicação, e a introdução de um comando de voz para pedir à aplicação para dizer em que ponto do livro (capítulo, secção e parágrafo) a leitura se encontra.

No total, realizaram-se sessões de testes com vinte utilizadores, alunos da Faculdade de Ciências da Universidade de Lisboa, sendo doze dos utilizadores do sexo masculino e oito do sexo feminino. Nenhum dos participantes no teste tinha deficiências visuais. Para poder avaliar as diferentes funcionalidades e possibilidades de utilização da aplicação, os testes foram executados em três cenários de uso distintos: um cenário em que o utilizador dispunha da possibilidade de utilizar todos os modos de entrada e saída visuais e áudio; outro em que o utilizador tinha disponíveis todos os modos de saída, mas permitindo apenas entradas por comandos de voz; e, por fim, um cenário em que a interacção visual não era utilizada quer para entradas quer para saídas. Desta forma foi possível avaliar o desempenho da aplicação nos diferentes cenários, e também observar quais as modalidades preferidas dos utilizadores quando dispõe de alternativas.

5. RESULTADOS PRELIMINARES

Os testes decorreram em duas fases. Uma primeira fase permitiu aos estudantes que organizavam os testes de avaliação familiarizarem-se com os procedimentos utilizados em experiências controladas, e realizar uma análise inicial aos resultados, de modo a fundamentar sugestões

de melhorias e alterações à aplicação. Uma segunda fase, realizada recentemente, permitiu aos mesmos estudantes avaliar as suas propostas, tendo sido empregue a técnica do feiticeiro de Oz, nos casos em que não foi possível introduzir em tempo útil na aplicação as alterações propostas. As duas fases foram separadas por um intervalo de duas semanas.

5.1 Primeira fase de testes

A primeira fase de testes incidiu principalmente na avaliação da facilidade de utilização e utilidade de algumas das funcionalidades oferecidas pelo Rich Book Player: navegação a partir da tabela de conteúdos, procura de palavras, criação de anotações, e controlo sobre a velocidade de leitura.

A navegação usando a tabela de conteúdos foi considerada normal ou fácil por 55% dos utilizadores que reportaram tê-la utilizado, recorrendo ao uso do rato. 18% consideraram a navegação na tabela de conteúdos difícil, e 27% não utilizaram a tabela de conteúdos para navegar. Quando são utilizados comandos de voz, os valores são semelhantes, registando-se, no entanto, um decréscimo na percentagem de utilizadores que consideram a navegação normal ou fácil, 46%, e um acréscimo nos que consideram a navegação difícil, 27%. A mesma percentagem de utilizadores reporta não ter utilizado comandos de voz para navegar na tabela de conteúdos. 73% dos utilizadores consideram muito útil ou indispensável esta funcionalidade, enquanto que os restantes 27% consideram-na pouco útil ou inútil.

A procura de palavras foi considerada de fácil utilização por 100% dos utilizadores que empregaram o rato e o teclado para a executar. Sendo que a procura baseada exclusivamente em comandos de voz não se encontra desenvolvida, devido à impossibilidade de dispor de um reconhecedor de voz com gramática livre, não é possível estabelecer uma comparação entre as duas modalidades para esta funcionalidade.

A procura de palavras foi considerada muito útil ou indispensável pela totalidade dos utilizadores.

A possibilidade de criação de anotações foi também avaliada. Quando empregando o rato e o teclado, 46% dos utilizadores consideraram-na uma tarefa de fácil execução, 27% de difícil execução e os restantes 27% não a utilizaram. Quanto recorreram a comandos de voz, apenas 36,5% consideraram a tarefa normal ou fácil, a mesma percentagem de utilizadores que consideraram a tarefa difícil, enquanto que 27% não utilizou comandos de voz para efectuar esta tarefa.

A criação de anotações foi considerada muito útil por 64% dos utilizadores, e pouco útil pelos restantes 36%.

No que diz respeito à possibilidade de controlar a velocidade de narração, esta foi considerada de fácil utilização por 100% dos entrevistados, quando se recorre ao rato, e por 75% dos entrevistados quando se empregam comandos de voz. A totalidade dos entrevistados considerou esta funcionalidade como muito útil ou indispensável.

Ainda nesta fase, avaliou-se a utilidade dos mecanismos de percepção sonoros que alertam para a presença de anotações e imagens. Estes mecanismos consistiam numa voz sintetizada que emitia as palavras “anotação” ou “imagem” conforme a situação. 56% dos utilizadores consideraram os alertas para a presença das anotações muito úteis ou indispensáveis, tendo esse valor subido para 89% quando o alerta dizia respeito a imagens. Esta diferença pode ser explicada pelo facto de as anotações disporem de um mecanismo de percepção que as imagens não possuem, nomeadamente, a alteração da cor de fundo do texto que se encontra anotado. Dessa forma, os utilizadores normovisuais podem identificar prontamente qual o texto que está anotado, já se encontrando conscientes dessa situação antes da narração chegar a esse texto. Esta situação não se verifica para as imagens, sendo por isso o alerta mais valioso para o utilizador.

Foi ainda pedido aos utilizadores para classificarem o Rich Book Player numa escala de um a cinco, tendo sido a classificação média no final desta primeira fase de testes de 3,8.

5.2 Segunda fase de testes

Através de algumas entrevistas abertas com utilizadores que participaram nas sessões controladas, e através da análise das capturas de ecrã e das gravações vídeo dessas sessões, os estudantes participantes no estudo elaboraram um conjunto de funcionalidades que poderiam ser acrescentadas à aplicação, e prepararam novas sessões de testes, de forma a avaliar as suas propostas. Como as maiores limitações identificadas foram ao nível do retorno em situações de interacção não visual, as propostas também foram maioritariamente dirigidas nesse sentido.

Para a realização desta fase de testes, recorreu-se à técnica do feiticeiro de Oz, de forma a poder simular funcionalidades ainda não desenvolvidas.

Uma das propostas passou pelo aumento da quantidade de informação retornada pela aplicação após a execução de comandos de navegação. Em condições normais, o retorno indica o capítulo em que a leitura se encontra, de cada vez que há uma mudança de capítulo. Nas condições do teste, a informação passava a incluir indicações sobre o parágrafo, e a ser disponibilizada de cada vez que se navegava para um parágrafo diferente. Os resultados mostraram que a percentagem de utilizadores que consideram a situação de retorno normal agradável (86%) é superior à percentagem que considera a situação de retorno completo agradável (71%).

Um conjunto de outros comandos foi testado, de forma a perceber qual a sua utilidade. Um desses comandos permitia ao utilizador saber a sua localização no livro (capítulo e parágrafo) a qualquer altura. Segundo 86% dos entrevistados, este comando é muito útil, sendo pouco útil para apenas 14%. É de notar que este comando será particularmente útil em situações de interacção não visual, visto que nas situações em que há retorno visual, 83% dos utilizadores consideraram normal ou fácil perceber a sua localização no livro.

Outro comando que, segundo 57% dos utilizadores seria útil, é um comando para pedir à aplicação que sintetize uma lista das anotações criadas até ao momento.

Foram ainda testados comandos que permitem controlar outros aspectos da narração. Um comando de voz para retirar o som da narração foi considerado útil por apenas 29% dos utilizadores. Isto pode dever-se à existência de outros comandos que permitem controlar o volume, e ao facto de as pessoas preferirem ouvir a narração. Um comando para elevar a velocidade de narração para o valor máximo instantaneamente também só foi considerado útil por parte de 29% dos utilizadores. Isto pode justificar-se pelo facto de a narração na sua velocidade máxima ser de muito difícil compreensão, e pela existência de variados comandos de navegação, que dispensam dessa forma a necessidade de avançar na narração à velocidade mais rápida, que é necessária em meios que não suportam o acesso aleatório aos seus conteúdos.

De referir ainda que se manteve a tendência, já evidenciada durante as primeiras sessões de teste, de considerar a utilização da rato e teclado como mais fácil do que os comandos de voz.

No fim dos testes foi novamente pedido aos utilizadores que classificassem a aplicação na mesma escala de um a cinco, sendo que desta feita, com a introdução, ainda que simulada, de novas funcionalidades, a classificação média subiu para 3,9.

Numa perspectiva de análise aos próprios testes, será interessante notar a evolução que se verificou nos estudantes que organizaram e realizaram os testes. Esse factor foi mais notório durante a realização dos testes com a técnica do feiticeiro de Oz, em que a preparação para a realização dos testes e a coordenação entre os vários intervenientes evoluiu à medida que se foram desenvolvendo.

6. CONCLUSÕES

Este artigo apresenta os resultados de um estudo de avaliação de usabilidade de interfaces multimodais para Livros Falados Digitais. Os Livros Falados Digitais combinam a apresentação visual do seu conteúdo com a sua narração. A sua população alvo principal é composta por utilizadores com deficiências visuais, apesar de que, com desenvolvimentos que enriquecem os conteúdos e a sua apresentação, estes livros tenham diversificado os seus utilizadores. As características intrínsecas destes livros, tornam-nos particularmente aliciantes para exploração de uma plataforma de apresentação multimodal, que ofereça capacidades de interacção áudio e visuais, quer ao nível das entradas, quer ao nível das saídas. Assim, este artigo começou por apresentar um conjunto de recomendações para o desenvolvimento de dispositivos de reprodução de Livros Falados Digitais, baseadas no conceito de interface unificada. De seguida apresentou-se o protótipo de leitor utilizado nos testes de avaliação.

O leitor desenvolvido explora o uso de múltiplos meios e modalidades, disponibilizando funcionalidades e acessibilidade melhoradas. Através de capacidades adaptativas,

permite ainda dar resposta a diversos contextos de uso, ambientes de leitura e características dos utilizadores. Este leitor resolve grande parte das limitações encontradas nos leitores actualmente disponíveis, respeitando as recomendações apresentadas, e permitindo melhorar a experiência de leitura de utilizadores com diferentes tipos de capacidade e em diversos contexto de utilização.

Os resultados dos testes de avaliação preliminares realizados permitiram aferir a utilidade das funcionalidades disponibilizadas, bem como perceber dificuldades de utilização experimentadas. Neste aspecto, as maiores dificuldades prenderam-se com problemas com o desempenho do reconhecedor de voz utilizado, por este ser um motor de reconhecimento de Inglês, o que obrigou a uma adaptação da fonéticas das palavras a reconhecer. Apesar de tudo, com o treino resultante da utilização, o desempenho do reconhecedor foi melhorando ao longo dos testes. Foi ainda possível realizar uma série de experiências recorrendo à técnica do feiticeiro de Oz, que permitiram avaliar funcionalidades ainda não desenvolvidas no protótipo, principalmente relacionadas com a interacção não visual, como qual o nível de retorno áudio sobre a posição de leitura no livro, tendo sido concluído que a introdução de uma funcionalidade que permita transmitir ao utilizador a sua localização no livro seria muito útil.

Futuramente, a aplicação será ainda melhorada com os resultados destes testes de avaliação preliminares, estando planeada a realização de mais testes, direccionados para a avaliação dos mecanismos de adaptação incluídos no Rich Book Player.

7. AGRADECIMENTOS

O trabalho apresentado neste artigo é parcialmente baseado em investigação financiada pela Fundação para a Ciência e Tecnologia através do projecto POSC/EIA/61042/2004.

8. REFERÊNCIAS

- [Aldrich1988] Aldrich, F.K. e Parkin, A. Tape Recorded Textbooks for the Blind: A Survey of Producers and Users. *The British Journal of Visual Impairment*, 1(1), 3-6, 1988.
- [Arons1997] Arons, B. SpeechSkimmer: A System for Interactively Skimming Recorded Speech. *ACM Transactions on Computer-Human Interaction*, Vol 4, No 1, 3-38, March 1997.
- [Carriço2003] Carriço, L., Guimarães, N., Duarte, C., Chambel, T., and Simões, H. Spoken Books: Multimodal interaction and information repurposing. In Proceedings of HCI'2003, International Conference on Human-Computer Interaction, Crete, Greece, June 2003, 680-684.
- [Cohen2004] Cohen, P., e McGee, D. Tangible multimodal interfaces for safety-critical applications. *Communications of the ACM*, 47(1), 41-46, 2004.
- [Crispien1996] Crispien, K., e Fellbaum, K. A 3D-Auditory Environment for Hierarchical Navigation in Non-Visual Interaction. In Proceedings of

- ICAD'96, International Conference on Auditory Display, Palo Alto, California, November 4-6, 1996.
- [Daisy] Daisy Consortium. Welcome to the DAISY Consortium Web site. Retrieved June 12, 2006, from <http://www.daisy.org>
- [Duarte2006] Duarte, C., e Carriço, L. A Conceptual Framework for Developing Adaptive Multimodal Applications. In Proceedings of the 11th ACM International Conference on Intelligent User Interfaces (IUI 2006), pp. 132-139, Sydney, Australia, 2006
- [Duarte2005] Duarte, C., e Carriço, L. Users and Usage Driven Adaptation of Digital Talking Books. In Proceedings of the 11th International Conference on Human-Computer Interaction (HCI 2005), Las Vegas, Nevada, USA, 2005.
- [Gazzaniga1998] Gazzaniga, M. S., Ivry, R. B., e Mangun, G. R. Cognitive Neuroscience – the Biology of the Mind. W. W. Norton & Company, 1998.
- [Jordan1998] Jordan, P. W. Methods for Usability Evaluation. In An Introduction to Usability (pp. 51-80). London - Bristol: Taylor & Francis.
- [Kalyuga1999] Kalyuga, S, Chandler, P, e Sweller, J. Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351-371, 1999.
- [Kildal2006] Kildal, J., e Brewster, S.A. Non-Visual Overviews of Complex Data Sets. In Proceedings of ACM CHI' 2006, Montréal, Québec, Canada, April 22-27, 2006.
- [Moreno2000] Moreno, R., e Mayer, R. E. A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, 97, 117-125, 2000.
- [Morley1998a] Morley, S. Digital Talking Books on a PC: A Usability Evaluation of the Prototype DAISY Playback Software. In Proceedings of ASSETS'98, the Third Annual ACM Conference on Assistive Technologies, Los Angeles, CA, 1998.
- [Morley1998b] Morley, S., Petrie, H., O'Neill A., and McNally, P. Auditory Navigation in Hyperspace: Design and Evaluation of a Non-Visual Hypermedia System for Blind Users. In Proceedings of ASSETS'98, the Third Annual ACM Conference on Assistive Technologies, Los Angeles, CA, 1998.
- [NISO2002] NISO. Specifications for the Digital Talking Book. Retrieved June 12, 2006, from <http://www.niso.org/standards/resources/Z39-86-2002.html>
- [NISO1999] NISO. Digital Talking Book Player Features List. Retrieved June 12, 2006, from <http://www.loc.gov/nls/z3986/background/features.htm>
- [Ntoa2005] Ntoa, S., e Stephanidis, C. ARGO: A System for Accessible Navigation in the World Wide Web. *ERCIM News*, 61, April 2005.
- [Oviatt1999] Oviatt, S. Mutual disambiguation of recognition errors in a multimodal architecture. In Proceedings of Conference on Human Factors in Computing Systems, 576-583, Pittsburgh, Pennsylvania, USA, May 1999.
- [Petrie1996] Petrie, H., Morley, S., McNally, P. O'Neill, A-M. e Majoe, D. Initial Design and Evaluation of an Interface to Hypermedia Systems for Blind Users. In Proceedings of Eighth ACM Conference on Hypertext, (Hypertext'96), ACM, 48-56, 1996.
- [Plextor] Plextor Co. Ltd, Japan. <http://www.plextor.com/about.htm#JAPAN>
- [Stephanidis2000] Stephanidis, C. Universal Access Through Unified User Interfaces. In Proceedings of 15th Technology and Persons with Disabilities Conference, Los Angeles, CA, USA, March 20-25, 2000.
- [Stephanidis1997] Stephanidis, C., Paramythis, A., Karagiannidis, C., and Savidis, A. Supporting Interface Adaptation: the AVANTI Web Browser. In Proceedings of the 3rd ERCIM Workshop on User Interfaces for All, 1997.
- [Wickens1992] Wickens, C. Engineering Psychology and Human Performance. Harper Collins, 1992.
- [Zhao2004] Zhao, H., Plaisant, C., Shneiderman, B., and Duraiswami, R. Sonification of Geo-Referenced Data for Auditory Information Seeking: Design Principle and Pilot Study. In Proc. International Conference on Auditory Display (ICAD), 2004.