Avaliação de aspectos de sincronização de Livros Falados Digitais

Carlos Duarte*, Luís Carriço*, Hugo Simões, Teresa Chambel*, Nuno Guimarães*

* Departamento de Informática
LaSIGE

Faculdade de Ciências da Universidade de Lisboa
Campo Grande, Edifício C5, 1749-016 Lisboa, Portugal

{cad, lmc, hsimoes, tc, nmg}@di.fc.ul.pt

Resumo

Os livros falados, para além de poderem atingir audiências como os invisuais, possibilitam a "leitura" em situações onde a visão está momentaneamente indisponível. Neste artigo, apresenta-se uma plataforma de desenvolvimento de livros falados. Os livros produzidos permitem interacção multimodal, de forma a tentar alargar ao máximo o espectro de utilizadores. Para além disso, apresentam-se os primeiros passos dados no sentido de enriquecer a apresentação do livro, imergindo o leitor numa experiência multimédia. São também apresentados os resultados de testes de usabilidade, efectuados para validar as opções de implementação tomadas.

Abstract

Talking books, besides helping the blind and print-disable population access to books, also allow “reading” in situations where the vision in temporarily unavailable. This paper presents a digital talking books production framework. The digital talking books broaden the scope of possible users by allowing for multimodal interaction. The paper also presents the first approaches to book enrichment, in an attempt to surround the reader in a multimedia environment. The results of usability tests conducted for validation of the current framework implementation are also presented.

1. Introdução

A utilização conjunta da palavra escrita e falada abre novas perspectivas de exploração dos livros. A construção de um livro em formato digital, que permita acompanhar a leitura do texto, com a reprodução duma gravação digital da narração do conteúdo, traz vários benefícios aos leitores. Em relação às cassetes áudio com a gravação analógica dos livros, o formato digital não traz apenas uma melhoria da qualidade de reprodução. Outras possibilidades de interacção com os livros, que dantes seriam extremamente difíceis ou até impossíveis, tornam-se agora acessíveis. Por exemplo, procurar uma palavra ou frase no livro, é agora equivalente a fazer uma procura num documento de texto digital, enquanto antes existia a necessidade de fazer a procura directamente na cassete áudio (eventualmente tendo de escutar a totalidade da obra). Também a criação de apontamentos pessoais é agora possível, através da gravação da voz do leitor, ficando o apontamento sincronizado com o ponto da leitura em que foi feito, permitindo assim a sua reprodução na altura correcta.

Um dos sectores da população que poderá tirar mais vantagens deste formato é o dos invisuais e pessoas com dificuldades de visão. No entanto, outros segmentos da população também podem beneficiar duma plataforma com estas características. A capacidade de interacção multimodal do livro, alarga as possibilidades de leitura, a situações em que o leitor se encontra a realizar tarefas que ocupam a visão, como por exemplo, durante a condução de viaturas, ou em actividades de vigilância.

Esta plataforma digital de reprodução de livros permite também pensar em alargar a experiência de leitura tradicional, e transformá-la numa imersão num ambiente multimédia. Surge desta forma, a possibilidade de composição da apresentação do conteúdo, aproveitando recursos que se encontrem disponíveis, e que possam enriquecer o livro, criando uma nova forma de “contar histórias”. Esta evolução passa pela introdução de novos elementos durante a “leitura”, duma forma coerente com o conteúdo do livro. Exemplos possíveis de apontar, passam pela introdução de uma música de fundo, de sons ambientes relacionados com o local da acção, de imagens e

vídeos que possam complementar a informação disponibilizada pelo texto original, etc.

Para poder efectuar este enriquecimento é necessário (1) um repositório de "media" a ser integrada nos livros, e (2) uma forma de classificar o conteúdo dos livros para poder integrar a "media" coerentemente na história.

Neste artigo apresenta-se uma plataforma de construção de livros digitais [Duarte, et al. 2003; Carriço, et al. 2003], dando particular relevância aos novos desenvolvimentos efectuados ao nível da apresentação, e aos testes de usabilidade entretanto conduzidos para validar algumas das opções tomadas.

Os livros falados digitais são construídos a partir de uma cópia digital do texto, e de uma gravação da narração do mesmo, ambas fornecidas pela Biblioteca Nacional, um dos parceiros do projecto. Através de um processo automático de alinhamento do texto e do áudio é determinado, para cada palavra do texto, o seu instante na gravação. É assim possível a construção, de uma forma automática, de uma apresentação sincronizada. O texto é processado de forma a ficar em formato XML. Informação adicional, necessária para indexação e sincronização, é acrescentada. Todo o processo decorre de uma forma automática, até se obter uma versão do livro com o texto e o áudio sincronizados. O leitor poderá ler o livro num monitor enquanto ouve uma narração sincronizada, ou poderá optar por usar apenas uma das modalidades. Para interagir com o livro poderá usar comandos de voz, teclado e rato, duma forma coordenada ou independente. Actualmente encontra-se em construção o repositório de "media", e em desenvolvimento uma forma de auxiliar a classificação do conteúdo do livro, com vista a atingir um processo de realizar essa classificação também automaticamente.

Na próxima secção apresenta-se um resumo do trabalho realizado nesta área, focando livros falados digitais, interfaces não visuais e métodos de análise de conteúdos. De seguida descreve-se a plataforma de construção automática dos livros falados digitais, com particular atenção para as tecnologias adoptadas. Na secção seguinte descreve-se a criação das interfaces para os livros falados digitais. Os resultados dos testes de usabilidade já conduzidos são apresentados de seguida. Finalmente apresentam-se as conclusões e o trabalho futuro a desenvolver.

2. Trabalho Relacionado

2.1. Livros Falados Digitais

Os Livros Falados Digitais (LFDs) foram pensados como um meio de facilitar o acesso da comunidade de invisuais e pessoas com deficiências visuais aos livros. Algumas organizações, em cooperação com membros dessas comunidades, desenvolveram normas respeitantes aos LFDs. Na Europa, o Daisy Consortium, com a colaboração da European Blind Union [European Blind Union 1996], desenvolveu uma norma. Nos Estados Unidos, um trabalho semelhante foi realizado pelo National Information Standards Organization (NISO), em colaboração com a The National Library Service for the Blind and Physically Handicapped. Como resultado do trabalho destas e doutras organizações, várias normas foram desenvolvidas e evoluíram independentemente. No entanto, o Daisy Consortium e a NISO decidiram cooperar, tendo daí resultado o desenvolvimento da mais importante especificação de LFDs, a norma ANSI/NISO z39.86 [ANSI/NISO 2002].

O desenvolvimento destas normas possibilita a categorização dos LFDs de acordo com as funcionalidades que disponibilizam, o que também reflecte a complexidade

inerente ao LFD. Assim, um LFD pode pertencer a uma das seguintes categorias [Daisy Consortium 2002]: áudio completo e apenas o título; áudio completo e controlo de navegação; áudio completo, controlo de navegação e texto parcial; áudio e texto completos; texto completo e áudio parcial; texto completo, sem áudio.

A plataforma de criação de livros falados apresentada neste artigo, permite o desenvolvimento da classe mais complexa de LFDs, com áudio e texto completos, não deixando no entanto de ser possível, gerar todos os LFDs com as características acima descritas.

De acordo com a lista de características e funcionalidades publicada pela NISO [NISO 1999], um LFD deve garantir capacidades de navegação básicas (avançar um caracter, palavra, linha, frase, parágrafo ou página de cada vez, e navegar para segmentos específicos do LFD), avanço e recuo rápidos, narração a velocidades variáveis, navegação através do índice ou de um Ficheiro de Controlo de Navegação (que deve permitir ao utilizador obter facilmente uma visão geral do conteúdo do livro), leitura de anotações, acesso a referências cruzadas, marcações, procura e outras capacidades.

No entanto, não são apresentadas na norma, quaisquer soluções específicas de implementação. Essas soluções deverão levar em consideração, não só os aspectos relacionados com as especificações propostas, mas também a natureza não visual do ambiente de execução. Conseguir transmitir ao utilizador as diferenças entre anotações, referências cruzadas, navegação estrutural, e sincronização, sem ambiguidade, e mantendo uma interface coerente, levanta problemas de usabilidade [Morley 1998].

2.2. Interfaces de Voz

O trabalho desenvolvido na área das interfaces não visuais pode indicar algumas formas de abordar as questões enfrentadas.

"Browsers" de voz são dispositivos que exibem pelo menos uma das seguintes características: (1) conseguem apresentar páginas web num formato áudio; (2) conseguem reconhecer fala e utilizá-la para controlar a navegação. A interface dos "browsers" de voz partilha com a interface dos LFDs alguns problemas comuns:

- O formato áudio é um meio temporal. Uma página apresentada visualmente pode exibir imagens, tabelas e texto simultaneamente, num formato espacial, que é rapidamente processado pelo sistema de percepção humano. Em contraste, texto falado, só pode apresentar uma palavra de cada vez.
- A emissão de comandos por voz, e o processamento de áudio, são actividades que utilizam as memórias de trabalho e de curto prazo, entrando em conflito com tarefas de planeamento e de resolução de problemas. A informação visual é processada por sistemas cognitivos distintos [Oviatt, et al. 2000].
- Os inevitáveis erros de reconhecimento.

A investigação na área dos sistemas multimodais, já tornou claro que as entradas de voz apresentam vantagens em determinadas situações [Oviatt, et al. 2000]. Alguns estudos [Van Buskirk e LaLomia 1995; Christian, et al. 2000] indicam que "as melhores tarefas para utilizar entradas de voz são aquelas em que o utilizador deve emitir comandos breves usando um vocabulário pequeno".

As características da interacção com um LFD são então vantajosas para a adopção de uma interface por voz, dado que um número relativamente pequeno de comandos pode ser utilizado para implementar as funcionalidades desejadas. No entanto,

algumas limitações podem surgir, se, por exemplo, para seguir uma ligação do índice, o leitor tiver de dizer o título do capítulo.

A investigação sobre a eficiência da voz como um modo de entrada ainda não é conclusiva [Haller, et al. 1984; Martin 1989; Visick, et al. 1984], apesar de indicar um aumento do tempo necessário para completar tarefas [Van Buskirk e LaLomia 1995; Christian, et al. 2000]. Algumas das recomendações feitas para a construção de "browsers" de voz, podem ser adoptadas para o projecto de LFDs:

- As ligações devem ser texto facilmente pronunciável.
- As ligações devem ser curtas (poucas palavras).
- Deve evitar-se ligações com sons semelhantes.
- Deve desenvolver-se alternativas às ligações numeradas, dado que estas causam sobrecarga cognitiva.

Formas de transmitir ao utilizador a estrutura dos documentos e auxiliar a navegação, em ambientes não visuais, também já foram alvo de investigação. Em [Goose e Moller 1999] é proposto o uso de áudio 3D. Para transmitir o conteúdo dos documentos, como a presença de ligações e cabeçalhos, foram já estudadas várias técnicas, donde se pode destacar o emprego de ícones auditivos [Gaver 1993; Blattner, et al. 1990] e de combinações de colunas de som e efeitos sonoros [James 1997]. Os resultados obtidos com estas metodologias não são, no entanto, conclusivos, sendo por isso justificada a construção de uma bancada de produção que permita testar diferentes configurações, de forma a explorar a utilização destas e de outras técnicas.

2.3. Análise de Conteúdo

Várias abordagens podem ser empregues no processo de classificação do conteúdo de textos, variando em complexidade, e variando nas aproximações, mesmo dentro da área do processamento de linguagem natural, desde aproximações racionalistas [Chomsky 1986] até aproximações estatísticas [Manning e Schütze 2001]. Um sistema de processamento de linguagem natural deve ser capaz de clarificar o sentido das palavras, a sua categoria, a estrutura sintáctica e o contexto semântico. Recentemente tem sido dado maior ênfase à descoberta de soluções práticas, que possam ser aplicadas a texto não previamente processado, ou seja, texto como existe no "mundo real". Estes objectivos tendem a favorecer aproximações estatísticas ao processamento de linguagem natural, porque são melhores para aprendizagem automática e para clarificação de significados [Manning e Schütze 2001].

No tocante à língua Portuguesa deve referir-se que há muito pouco trabalho desenvolvido. Por estes motivos, as aproximações à análise de conteúdos apresentadas neste artigo são baseadas na vertente estatística, e ainda de baixa complexidade.

3. Produção de Livros Falados Digitais

Os LFDs gerados pela plataforma automática são construídos a partir de cópias digitais do texto e da narração da obra, ambas fornecidas pela Biblioteca Nacional. O primeiro passo do processo de criação do LFD é determinar os alinhamentos entre as palavras no texto e na narração. A metodologia de determinação automática dos alinhamentos encontra-se descrita em [Serralheiro, et al. 2002]. O resultado deste processo é um ficheiro com o tempo em que cada palavra é proferida na narração. Para além das palavras, são também identificados no ficheiro os silêncios, isto é, as alturas em que há pausas na leitura. As palavras que constam do ficheiro de

alinhamentos nem sempre são idênticas às encontradas no ficheiro com o texto. Por exemplo, quando no ficheiro de texto se encontra “101”, no ficheiro de alinhamentos surge a versão literal do número, ou seja, as três palavras “cento e um”, cada uma com o correspondente tempo de início. Outra situação que pode causar discrepância entre os dois ficheiros prende-se com a existência de erros de reconhecimento quando o ficheiro de texto é gerado, por exemplo, a partir duma digitalização do livro original.

Este processo de alinhamento automático das palavras é da responsabilidade de outro dos parceiros do projecto. Do ponto de vista do trabalho apresentado no artigo identificam-se três entradas para o gerador de LFDs: (1) uma cópia digital do texto, (2) uma narração do texto, também em formato digital, e (3) o ficheiro de alinhamento entre as palavras da cópia e da narração.

A partir deste momento, o processo segue de uma forma automática até se concluir a produção de um LFD pronto para “leitura”. O processo encontra-se dividido em três fases: fase 0 – pré-processamento; fase 1 – produção do LFD; fase 2 – construção da apresentação. Na fase 0 processam-se os ficheiros de entrada, de forma a obter o mesmo conteúdo, mas num formato normalizado. Na fase 1 procede-se à reestruturação de todos os elementos que digam respeito ao conteúdo do LFD, e à sua informação estrutural e de navegação. Na fase 2, utilizando uma especificação de apresentação, constrói-se a versão interactiva do LFD. Na figura 1 pode ver-se um resumo de todo o processo.

A divisão do processo de produção em três fases foi motivada pela necessidade de criar LFDs, duma forma automática, para audiências com características e capacidades muito diferentes. Assim, a especificação de apresentação permite criar várias interfaces para o mesmo livro, sendo inclusivamente possível alterar os modos de entrada e de saída utilizados. O próprio processo de enriquecimento do livro sai facilitado. Deste modo, a separação entre conteúdo e apresentação é total, factor importante para a produção automática de LFDs que possam ser adaptados a cada um dos seus leitores.

Para a implementação do processo de criação automática de LFDs foi utilizada a plataforma Apache Cocoon [Langham e Ziegler 2002]. Esta plataforma utiliza XML e XSLT como tecnologias base, numa arquitectura baseada em “pipelines”, facilitando dessa forma a reutilização de componentes.

De seguida descrevem-se duma forma mais detalhada as fases 0 e 1 de produção de LFDs, sendo a fase de produção da interface do livro focada na próxima secção.

3.1. Pré-processamento

A fase de pré-processamento tem com objectivo restabelecer a informação sintáctica do livro, traduzindo e normalizando os diferentes formatos em que se encontram as entradas. Esta necessidade justifica-se pelo desconhecimento da estrutura usada para representar o conteúdo do livro na cópia digital fornecida. Para além disso é necessário efectuar uma reposição do alinhamento, devido às diferenças de conteúdo quando comparados a cópia do texto e o ficheiro de alinhamentos. O resultado do processo de alinhamento é um conjunto de tempos para cada palavra dita na narração, o que se pode denominar “alinhamento de som”. Aquilo que é necessário para efectuar uma apresentação sincronizada entre o texto e a narração, é uma lista de tempos para cada palavra constante do texto, que pode ser denominado por “alinhamento de texto”. É por isso necessário proceder a uma reposição do alinhamento resultante do processo de extracção de tempos. Assim, para além da situação anteriormente referida, causada pela presença de números em formato

numérico, que surgem no ficheiro de alinhamentos sob a forma literal, outras situações produzem inconsistências: o uso de abreviaturas no texto, que depois surgem expandidas no ficheiro de alinhamento; os sinais de pontuação, que não estão presentes no ficheiro de alinhamento; e erros de reconhecimento quando o texto é obtido a partir de um processo de digitalização. Em [Duarte, et al. 2003] encontra-se uma descrição dos métodos utilizados para resolver estes problemas.

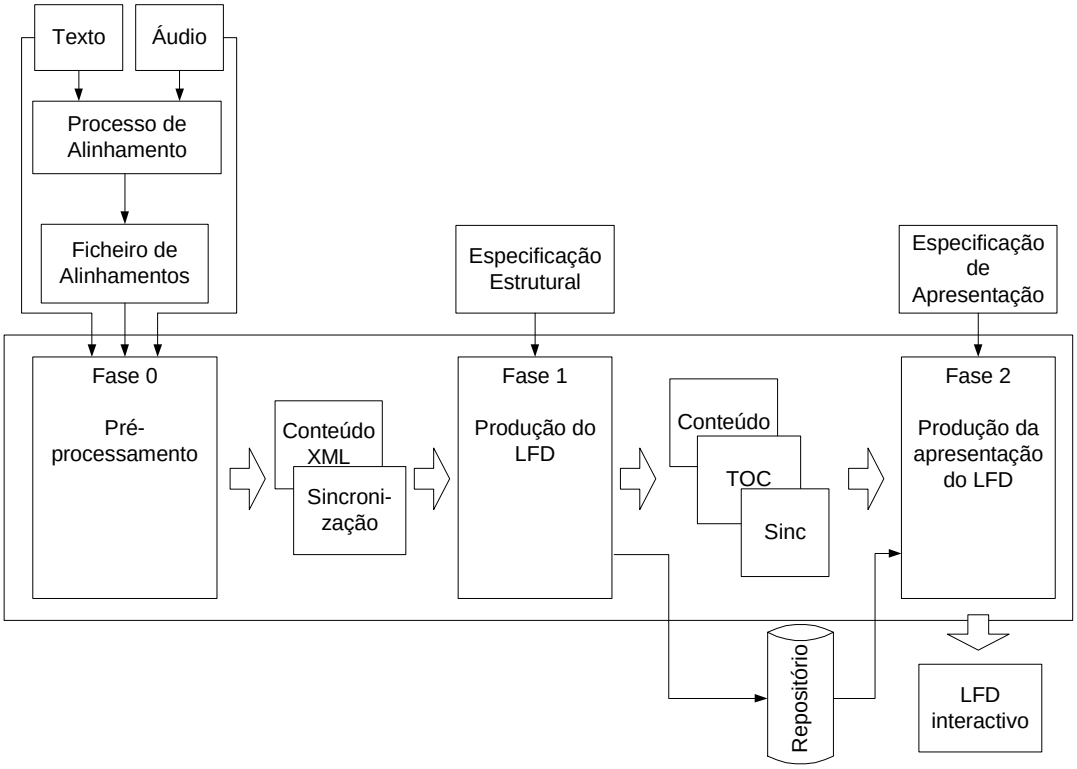

Figura 1- Processo de produção de LFDs

A fase 0 produz então um ficheiro normalizado com o conteúdo do livro, um ficheiro com a informação de sincronização para cada palavra existente no texto do livro, e, opcionalmente, um ficheiro com metadados no formato Dublin Core, caso estes tenham sido disponibilizados pelo fornecedor do conteúdo do livro.

3.2. Produção do LFD

A fase de produção do LFD parte destes dados, e recorrendo a uma especificação estrutural, gera os aspectos referentes ao conteúdo do LFD, independentemente da apresentação, que será alvo de atenção na próxima fase. A especificação estrutural cobre os aspectos relativos à estrutura e navegação do LFD, à identificação de unidades de sincronização e ao processo de análise de conteúdo. Informação relevante é fornecida para a criação de índices e tabelas de conteúdos que podem ser utilizados para navegação no livro. Outro dos aspectos relevantes processados nesta fase é referente às capacidades de sincronização. Apesar de a norma ANSI/NISO para LFDs prever apenas a utilização de uma unidade de sincronização (a palavra), esta plataforma de produção de LFDs tem a capacidade de gerar livros com mais do que uma unidade de sincronização. A vantagem de ter várias unidades de sincronização é aumentar as possibilidades de interacção. As unidades de sincronização podem ser usadas com diferentes fins. Por exemplo, durante a apresentação do livro, a frase que está a ser proferida muda de cor. Para conseguir esse efeito é necessária uma

sincronização à frase. No mesmo livro, quando o utilizador executa uma procura por uma palavra, o texto passa a ser lido a partir da primeira palavra encontrada, sendo para isso necessário uma sincronização à palavra. Desta forma se pode observar que diferentes funcionalidades ou formas de interacção podem necessitar de diferentes unidades de sincronização.

Assim, a fase 1 gera um ficheiro XML com o conteúdo do livro, organizado hierarquicamente segundo os seguintes elementos: livro, volume, capítulo, secção, subsecção, parágrafo, frase, oração e palavra, eventualmente limitado pela granularidade de sincronização; um ficheiro XML com a árvore de navegação de conteúdo (semelhante a uma tabela de conteúdos); e um ficheiro XML com informação respeitante às várias unidades de sincronização solicitadas.

Também relevante na fase 1 é o processamento realizado para analisar o conteúdo do texto, e a possível extracção de conteúdo para reutilização em apresentações de outros livros. Actualmente encontram-se em desenvolvimento para a plataforma de construção de LFDs as primeiras aproximações à determinação de palavras chave para classificação de conteúdo. Sendo este o trabalho inicial nesta área, as aproximações são ainda de baixa complexidade.

Através de uma análise estatística do conteúdo do livro, realiza-se uma filtragem de forma a obter uma lista de palavras relevantes. A partir da lista de palavras são determinadas quais as palavras chave do texto. O conjunto de palavras chave pode ser determinado por mais do que um método. Uma hipótese é escolher as *n* primeiras classificadas da lista. Outra hipótese é definir um limiar, e todas as palavras com um número de ocorrências acima do limiar farão parte do conjunto de palavras chave. Esse limiar deverá ser definido em função do número total de palavras do texto, pois a importância de uma palavra que ocorre dez vezes num texto de mil palavras, não é a mesma se o texto tiver cem mil palavras.

Futuramente, pretende-se evoluir esta aproximação determinando a raiz das palavras. Desta forma, palavras que tenham raiz comum serão contadas como uma só, sendo escolhida como palavra chave a raiz desse conjunto de palavras.

4. A Interface do LFD

Na fase de construção da apresentação, são abordados os temas respeitantes à interacção com o LFD. Partindo do conteúdo gerado na fase anterior, podem ser construídas formas diferentes de apresentar o mesmo livro, e também escolher diferentes modos de interacção.

Os modos de interacção podem variar entre teclado, rato e comandos de voz para as entradas, ou uma combinação deles. Para as saídas, o áudio ou a apresentação visual podem ser empregues, individualmente ou de forma coordenada. Para além dos modos de interacção empregues na interface, a especificação da apresentação pode ainda determinar outros pormenores, tais como a gramática empregue para reconhecer os comandos de voz dos utilizadores, o nível de detalhe de apresentação da tabela de conteúdos (apresentar só os capítulos ou incluir também secções e subsecções), a forma como é apresentada a sincronização entre a narração e o texto escrito, a visualização das anotações feitas pelo utilizador, e a forma de informar o utilizador da presença de uma anotação quando esta é alcançada durante a narração do texto. Também o nível de enriquecimento do livro com elementos do repositório de “media” é detalhado na especificação da apresentação. Este enriquecimento é guiado pelas palavras chaves determinadas na fase anterior. Quando um tema dominante do livro é

determinado, existe a opção de adicionar à apresentação os conteúdos do repositório de elementos que estejam classificados com esse tema.

Na figura 2 mostra-se uma apresentação construída para o livro "O Senhor Ventura" de Miguel Torga. Esta apresentação, a ser visualizada no browser Internet Explorer 6, foi construída recorrendo à linguagem HTML+TIME [Microsoft Corporation 2002], a implementação da Microsoft da linguagem SMIL [Bulterman 2001]. Esta linguagem fornece as capacidades de sincronização necessárias para uma apresentação coordenada do texto e da narração. Outras apresentações foram criadas recorrendo à linguagem HTIMEL [Chambel, et al. 2001] para sincronização. Para reconhecimento de voz foi utilizada a plataforma Microsoft Speech.

Figura 2 – Uma apresentação do livro "O Senhor Ventura".

Nesta apresentação o texto é exibido visualmente enquanto é narrado. A sincronização é apresentada mudando a cor da frase que está a ser proferida. O utilizador pode interagir com o livro usando teclado, rato e comandos de voz, aproveitando assim a natureza multimodal da interacção. Na sua maioria, os comandos podem ser emitidos em qualquer dos modos. A tabela de conteúdos apresenta os diferentes capítulos do livro e pode ser usada para navegação, quer seleccionando o capítulo com o rato, quer emitindo vocalmente o comando "ir para" seguido do número do capítulo. Do lado direito do ecrã, o utilizador pode ver as anotações que já foram criadas. Nesta apresentação são exibidas as primeiras palavras de cada anotação, mas a especificação de apresentação pode produzir resultados diferentes. Pode ser exibida a data e hora de criação da apresentação, ou também pode ser exibida a anotação na sua totalidade. Ao seleccionar a anotação com o rato, o seu conteúdo é mostrado, e a narração do livro, bem como o texto exibido, avança para o ponto de leitura em que a anotação foi criada. O mesmo resultado pode ser alcançado

dizendo “ler nota” seguido do número da anotação. Um sinal sonoro é reproduzido sempre que a narração chega a um ponto do livro onde foi criada uma anotação. A introdução de anotações pode ser efectuada seleccionando o botão “adicionar” ou dizendo o comando “adicionar”. Actualmente a aplicação suporta anotações de texto e imagens, e futuramente virá a ser adicionado o suporte para anotações áudio e vídeo. O utilizador pode também parar e reiniciar a apresentação a qualquer momento, com os comandos “pause” e “play”.

Para este livro, uma das palavras chaves determinadas pela análise estatística da fase anterior é “mar”. Assim, a apresentação pode ser enriquecida através da adição de conteúdos presentes no repositório de elementos que estejam catalogados como relativos a “mar”, nomeadamente a inclusão de segmentos áudio com o som do mar.

5. Testes de Usabilidade

Para avaliar a interface dos LFDs construídos conduziram-se, até ao momento, duas sessões de avaliação. O principal objectivo foi avaliar aspectos de usabilidade da interface, e o grau de satisfação dos utilizadores. Uma característica particular avaliada prende-se com a forma como é transmitido ao utilizador a sincronização entre o texto apresentado no ecrã e o texto narrado. Para isso foram utilizados duas interfaces para o mesmo livro. A primeira interface é a apresentada na secção 4 (figura 2). Nesta interface a sincronização é transmitida ao utilizador através da mudança de cor das palavras que estão a ser narradas. Mais concretamente é alterada a cor de todas as palavras entre dois silêncios. A sincronização mudando a cor palavra a palavra revelou-se impraticável, excepto em exemplos de dimensão reduzida, pelos recursos computacionais necessários para visualizar os livros construídos dessa forma. A segunda interface (figura 3) apresenta a sincronização através de um indicador da linha que se encontra a ser narrada. Nesta interface não existe alteração da cor das palavras. Todos os outros componentes são idênticos nas duas interfaces.

Os testes foram realizados com dezasseis pessoas, tendo cada uma das interfaces sido testada por oito utilizadores. Nenhuma das pessoas sofria de problemas visuais ou auditivos. A interface foi explicada a cada um dos utilizadores, sendo-lhes disponibilizado de seguida o tempo que necessitassem para se familiarizar com ela. Quando assinalavam estar preparados, era-lhes apresentada uma lista de 12 tarefas. Estas tarefas exigiam navegação no livro para responder a questões sobre o conteúdo e para criação de anotações. Os utilizadores eram livres de escolher o modo como emitiam os comandos. No final das tarefas os utilizadores responderam a um questionário.

Em termos de facilidade de utilização foi considerado pelos utilizadores (tabela 1) que seguir ligações, quer da tabela de conteúdos, quer das anotações, era mais fácil com o rato do que com comandos de voz. No entanto, essa diferença esbate-se quando se considera as acções de criação de anotações e de controlo da reprodução.

Quanto à utilidade das funcionalidades disponibilizadas, a maioria dos utilizadores considerou a existência de ligações da tabela de conteúdos para o texto, e o controlo da reprodução indispensáveis. As anotações foram consideradas muito úteis. A possibilidade de interagir multimodalmente foi considerada pouco útil por 19% dos utilizadores, muito útil por 56% e indispensável por 25%.

A nível de satisfação, enquanto que todos os utilizadores se mostram satisfeitos com a interacção por rato e teclado, houve dois utilizadores que se revelaram pouco satisfeitos com os comandos de voz.

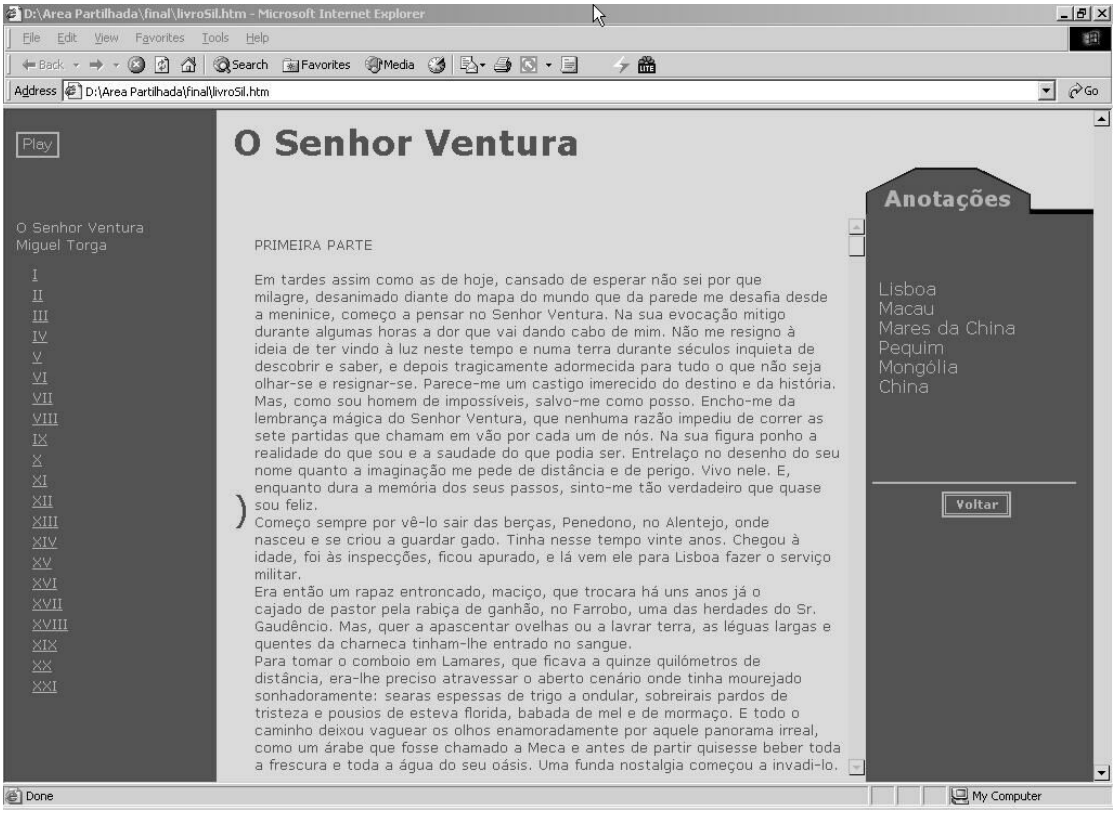

Figura 3 – Segunda interface para o livro “O Senhor Ventura”

Os comentários dos utilizadores e a observação dos testes permitiram identificar alguns problemas de usabilidade das interfaces testadas. Uma das questões mais vezes levantada refere-se à ausência de identificação do capítulo que está a ser narrado. Dada a falta de confiança no reconhecedor de voz, esta questão causa insegurança aos utilizadores quando seguem ligações da tabela de conteúdos através de comandos de voz. Problema semelhante ocorre quando seguem ligações das anotações. Uma outra questão, também relevante, é a não existência de números nas anotações, o que forçava os utilizadores a terem de contá-las para usar comandos de voz, quando desejavam seguir a ligação respectiva. Outros problemas mencionados prendem-se com a inexistência de um mecanismo de procura (no entanto o disponibilizado pelo “browser” pode ser utilizado), e com a sensibilidade e ineficácia do reconhecedor de voz.

Tabela 1 – Facilidade de Utilização dos modos de entrada

	Ligações da Tabela de Conteúdos		Ligações das Anotações		Criação de Anotações		Controlo da reprodução	
	Rato	Voz	Rato	Voz	Rato	Voz	Rato	Voz
Não Usado	-	13%	6%	25%	13%	31%	6%	6%
Muito Difícil	-	6%	-	6%	-	6%	6%	-
Difícil	-	38%	6%	31%	6%	-	-	13%
Fácil	19%	38%	31%	31%	50%	31%	38%	31%
Muito Fácil	81%	6%	56%	6%	31%	31%	50%	50%

No que diz respeito à avaliação das apresentações da sincronização, os resultados não podiam ser mais esclarecedores. Enquanto que todos os utilizadores da primeira interface detectaram falhas de sincronização, e referiram o desconforto causado por essas falhas, nenhum utilizador da segunda interface referiu qualquer falha de sincronização. No entanto, um destes utilizadores comentou que seria preferível uma forma de sincronização que acompanhasse a palavra narrada.

6. Conclusões e Trabalho Futuro

Neste artigo apresentou-se uma plataforma de desenvolvimento de livros falados digitais. Esta plataforma, a partir de cópias do texto e de uma narração em formato digital, permite construir apresentações do livro que combinam texto e áudio duma forma sincronizada. O "leitor" pode interagir multimodalmente com o livro, usando o teclado, rato e comandos de voz. A construção do livro é automática, recorrendo à plataforma Apache Cocoon para controlar o processo. O conteúdo do livro, bem como informação de navegação e sincronização são representados usando XML. Para obter apresentações sincronizadas já se utilizaram duas linguagens, HTML+TIME e HTIMEL.

Durante o processo de produção, o conteúdo do livro é separado da sua apresentação. Desta forma é possível reutilizar recursos para construir diferentes apresentações para o mesmo livro. Esta separação também permite efectuar o enriquecimento do livro de uma forma mais expedita.

Um livro enriquecido contém conteúdos acrescentados pela plataforma, no sentido de melhorar o entretenimento ou a produtividade do leitor. Para enriquecer o livro é necessário um repositório de elementos que possam ser utilizados com esse fim, e um processo de classificação de conteúdos dos livros, para determinar quais os elementos escolhidos. Neste artigo descreveu-se o trabalho realizado até agora, no sentido de conseguir automatizar o processo de classificação de conteúdos. Uma aproximação inicial, pouco complexa, foi detalhada. Outras aproximações, mais complexas, terão de ser desenvolvidas para possibilitar uma análise semântica dos livros. Um próximo desenvolvimento passa pela reunião de palavras com a mesma raiz.

As capacidades disponibilizadas por esta plataforma, sugerem o desenvolvimento futuro de livros falados capazes de se adaptarem em tempo real ao seu leitor. Para isso, será necessário iniciar o desenvolvimento de modelos de utilizador, identificando quais as características destes que podem ser usadas para adaptar o livro. Funcionalidades que permitam ao utilizador indicar o seu grau de satisfação acerca de determinada característica da interface, ou que se apercebam desse grau através da observação das acções do utilizador serão preponderantes para permitir implementar capacidades de adaptação do livro.

Os resultados dos testes de usabilidade conduzidos até ao momento indicam a importância de uma sincronização, entre o texto escrito e o narrado, sem falhas. Quando são perceptíveis pelo utilizador, as falhas de sincronização provocam desconforto e dificuldades de manutenção de concentração nos momentos em que ocorrem. Uma das interfaces testadas, utilizando unidades de sincronização com menor detalhe (linha em vez de palavras) conseguiu eliminar as falhas de sincronização e os problemas a elas associados.

Outro dos problemas de usabilidade identificados diz respeito à falta de informação acerca do capítulo do livro que está a ser lido, o que, conjuntamente com a falta de confiança no reconhecedor de voz, fez com que vários utilizadores preferissem utilizar o rato para seguir as ligações da tabela de conteúdos.

De notar que, apesar do desempenho por vezes pouco eficaz do reconhecedor de voz, apenas 2 utilizadores se mostraram pouco satisfeitos com os comandos de voz, e que 81% dos utilizadores consideram que a possibilidade de interagir multimodalmente é muito útil ou indispensável.

Futuros testes irão permitir avaliar soluções para os problemas de usabilidade identificados nestas interfaces. Testes efectuados usando interfaces que disponibilizem apenas comandos de voz como modo de entrada, irão permitir descobrir novos problemas de usabilidade, bem como identificar os comandos de voz que os utilizadores preferem, quando não se encontram restringidos a uma gramática definida pela aplicação.

7. Referências

ANSI/NISO, "Specifications for the Digital Talking Book (ANSI/NISO Z39.86-2002)", 2002, http://www.niso.org/standards/resources/Z39-86-2002.html

Blattner, M. et al, "Earcons and icons: Their Structure and Common Design Principles", in Ephraim P. Glinert (Ed.), Visual Programming Environments: Applications and Issues, IEEE Computer Society Press, Los Alamitos, CA, 1990, 582-606.

Bulterman, "Smil 2.0: Overview, concepts, and structure", IEEE Multimedia, 8, 4, (2001), 82-88.

Carriço, L. et al, "Spoken Books: Multimodal Interaction and Information Repurposing", Proceedings of the 10th International Conference on Human-Computer Interaction, Creta, Grécia, 2003.

Chambel, T. et al, "Hypervideo on the Web: Models and Techniques for Video Integration", International Journal of Computers & Applications, 23, 2, (2001), 90-98.

Christian, K. et al, "A Comparison of Voice Controlled and Mouse Controlled Web Browsing", Proceedings of ASSETS 2000, ACM Conference on Assistive Technologies, Arlington, VA, 2000, 72-79.

Chomsky, N, Knowledge of Language: Its Nature, Origin and Use, Prager, New York, 1986.

Daisy Consortium, "Daisy Structure Guidelines", 2002, http://www.daisy.org/publications/guidelines/sg-daisy3/structguide.htm

Duarte et al, "Producing DTB from audio tapes", Proceedings of the Fifth International Conference on Enterprise Information Systems ICEIS2003 (Volume 3), Angers, França, 2003, 582-585.

European Blind Union, Reaching Forward to the 21st century: User Requirements for the Next Generation of Talking Books, RNIB, London, 1996.

Gaver, W., "Synthesizing Auditory Icons", Proceedings of the ACM INTERCHI 1993, Amsterdão, Holanda, 1993, 228-235.

Goose, S. e C. Moller, "A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure", Proceedings of the seventh ACM international conference on Multimedia (Part 1), Orlando, Florida, 1999, 363-371.

Haller, R. et al, "Comparison of Input Devices for Correction of Errors in Office Systems", Proceedings of INTERACT 84, Londres, Reino Unido, 1984, 177-182.

James, F., "Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext", Proceeding of the International Conference on Auditory Display ICAD97, Palo Alto, California, 1997, 97-103.

Langham, M. e C. Ziegler, Cocoon: Building XML Applications, New Riders, 2002.
Manning, C. e H. Schütze, Foundations of Statistical Natural Language Processing, MIT Press, Cambridge, Massachusetts, 2001.
Martin, G., “The Utility of Speech Input in User-Computer Interfaces”, International Journal of Man-Machine Studies, 30, 4, (1989), 355-375.
Microsoft Corporation, HTML+TIME 2.0 Reference, 2002, http://msdn.microsoft.com/workshop/author/behaviors/reference/time2_entry.asp
Morley, S., “Digital Talking Books on a PC: A Usability Evaluation of the Prototype DAISY Playback Software”, Proceedings of ASSETS 1998, ACM Conference on Assistive Technologies, Marina del Rey, California, 1998, 157-164.
NISO, Document Navigation Features List, 1999, http://www.loc.gov/nls/z3986/background/navigation.htm
Oviatt, S. et al, “Designing the user interface for multimodal speech and gesture applications: State-of-the-art systems and research directions”, Human Computer Interaction, 15, 4, (2000), 263-322.
Serralheiro, A. et al, “Spoken Book Alignment Using WFSTS”, Proceedings of HLT2002, Human Language Technology Conference, San Diego, California, 2002.
Van Buskirk, R. e M. LaLomia, “A Comparison of Speech and Mouse/Keyboard GUI Navigation”, Proceedings of ACM CHI 1995 (Volume 2), 1995, 96.
Visick, D. et al, “The Use of Simple Speech Recognizers in Industrial Applications”, Proceedings of INTERACT 84, Londres, Reino Unido, 1984.