

Controle de geração de histórias Interativas através de comandos de voz

Luiz José Schirmer Silva

Cesar Tadeu Pozzer

Departamento de Informática
Pontifícia Universidade Católica do Rio de Janeiro

Departamento de Computação Aplicada
Universidade Federal de Santa Maria

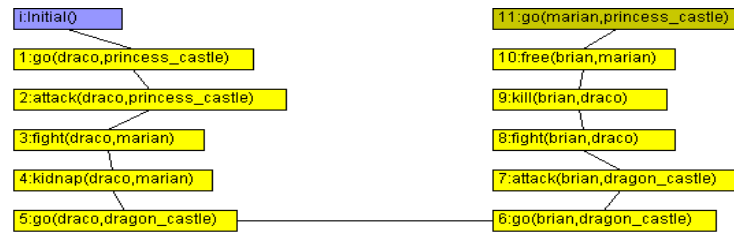


Figura 1: diagrama representando uma história gerada por requisições do usuário

Resumo

Muitas pesquisas em *Storytelling* concentram-se na criação de técnicas para permitir a fácil interação dos usuários com a história sendo contada. A interação com o usuário pode se dar através da interpretação de comandos de voz, porém a maioria destes sistemas faz uso de interfaces de usuário pouco atrativas ou faz uso de comandos de voz específicos. Este trabalho apresenta o desenvolvimento de um modelo para interação baseado em processamento de linguagem natural para o sistema *Logtell*, isto é, um modelo de interação com o usuário o qual utiliza mecanismos de interpretação da fala para interação com a narrativa. O sistema proposto transforma frases ditas pelos usuários em texto o qual é analisado posteriormente juntamente com um módulo gerenciador de histórias. Sendo assim, isso permite aos usuários interagir com os personagens através de fala influenciando no desfecho de um enredo apresentado.

Palavras-chave: *Storytelling*; Linguagem Natural; Comandos de Voz.

Contato dos autores:

lschirmer@inf.puc-rio.br
pozzer@inf.ufsm.br

1. Introdução

Com o advento dos sistemas de televisão digital, muitas características encontradas atualmente em jogos estão sendo empregadas em ambientes onde o espectador não é considerado apenas um agente passivo, mas este mesmo é capaz de interagir com o conteúdo apresentado. Segundo Forman et al. [2000], existe uma grande tendência de buscar a convergência entre os principais tipos de entretenimento digital, sejam eles TV, *games* ou cinema, onde é possível que diferentes agentes interajam e gerem experiências colaborativas.

O uso de histórias interativas tem sido abordado tanto em aplicações para jogos digitais como também em protótipos para televisão digital. Todas elas buscam transformar o usuário em um agente ativo em determinada história apresentada.

No que se diz respeito à narração de histórias, *Storytelling* é um modelo de entretenimento digital que faz uso de técnicas e ferramentas que permitem criar, guiar e visualizar histórias interativas em um meio computacional. *Storytelling* pode ser considerado como um sistema que engloba variadas técnicas de conhecimento, como inteligência artificial e computação gráfica, para o processamento e elaboração de uma história [Mateas, 1997]. Agentes autônomos que possuem características como personalidade própria, personagens virtuais dentre outras características comuns a games derivam de pesquisas relacionadas à *Storytelling* [Spierling et al., 2002]. Neste contexto, o usuário deixa de ser um espectador passivo, onde ele tem a possibilidade de interagir com uma história a fim de guiar ações dos agentes inseridos nela e gerando diferentes versões de uma mesma história. Novos jogos e sistemas para televisão digital têm incorporado técnicas que possibilitem que o jogador ou telespectador interaja com o sistema de uma forma aparentemente mais natural. Por exemplo, a possibilidade do usuário poder utilizar a fala para influenciar eventos definidos em uma narrativa se mostra muito mais vantajosa do que digitar textos, usar o mouse ou usar botões do controle remoto.

Baseando-se nestas premissas, este trabalho apresenta o desenvolvimento de um sistema que utiliza interpretação de fala aplicada ao *Storytelling*. Neste sistema o usuário age como um elemento ativo no contexto da história onde cada entrada fornecida é inicialmente transformada em texto para posteriormente ser processada em sentenças de lógica de primeira ordem. Elas serão usadas como regras na definição do fluxo da história sendo gerada. Estas

regras são usadas como base para entradas de um módulo responsável pela análise de requisições, qual foi incorporado ao sistema de *Storytelling Logtell* [CIARLINI et al. 2005].

Este artigo está organizado da seguinte maneira. Na próxima seção apresentam-se os trabalhos relacionados ao tema. As seções 3 e 4 apresentam o processo de desenvolvimento do sistema proposto e a avaliação de resultados. A conclusão e trabalhos futuros encontram-se na seção 5.

2. Trabalhos Relacionados

Para a geração de enredos em *Storytelling*, o sistema deve agir sobre algum tipo de dado para representar a história, onde estes dados são um conjunto de regras, cenários e eventos. Um dos grandes desafios para a criação dinâmica de enredos é tornar a trama coerente. Devido a isso, grande parte dos sistemas desenvolvidos depende fortemente de um modelo de história criado pelo autor. O usuário apenas decide qual o rumo que ela deve tomar de acordo com os eventos acontecidos até então. Sendo assim dada uma história genérica, o usuário pode interferir nas ações do enredo, enviando comandos a um módulo responsável por interpretá-los. Com relação à geração de contos, Ciarlini [1999] propôs um gerenciador interativo de histórias baseado em um modelo formal para a especificação de eventos na narrativa através de operações lógicas. Este sistema possui um módulo em Prolog que tem o papel de gerar histórias interativas, a partir de um determinado contexto, de acordo com um conjunto de regras e inferências lógicas, respeitando as condições necessárias para que um evento possa ocorrer. Através deste sistema é permitido ao usuário interagir com os personagens virtuais da história, dando-lhe ordens para influenciar no desfecho de um enredo que lhe é apresentado. Dada uma história, o sistema fica encarregado de gerar o “ponto inicial” e ainda gerenciar as requisições feitas por usuários. Sendo assim em um enredo onde se tem heróis e vilões bem definidos, um herói só poderia interagir com algum vilão caso este já tivesse realizado ações que desencadeassem uma reação do herói, como sequestrar outro personagem ou atacar alguém. Desta forma ao se ter as entradas do usuário processadas estas podem ser enviadas ao gerenciador que irá garantir que as ordens ou sugestões dadas possam ser executadas dado o estado corrente da narrativa.

No que diz respeito à interatividade, existem vários trabalhos relacionados ao modo de interação com narrativas de histórias, onde a maioria dos projetos encontrados faz uso de interfaces gráficas (GUI) tradicionais [Ciarlini et al., 2005][Grasbon and Braun 2001]. Nestes modelos são apresentadas caixas de diálogo contendo botões e outros atributos que podem comprometer a experiência proporcionada ao usuário. Apesar disso existem sistemas que buscam construir um modelo mais atrativo para o usuário final. Lima et

al. [2012] desenvolveram um protótipo onde o usuário é capaz de interagir com uma história através de técnicas de realidade aumentada. O sistema permite aos usuários interagir livremente com os personagens virtuais, através de objetos desenhados em folhas de papel. Apesar disso, este modelo apresenta algumas limitações onde acaba sendo incapaz de reconhecer alguns dos desenhos do usuário. Além disso, existem ainda técnicas para interação as quais fazem uso da fala, desta forma, o sistema não se limita a comandos específicos, nem a interfaces que poderiam até comprometer a linguagem visual da história gerada. Lima et al. [2012] propuseram um mecanismo que define como os usuários irão interagir com um sistema de *Storytelling* utilizando linguagem natural. O sistema apresentado faz uso de um *parser* probabilístico para o processamento dos comandos advindos dos usuários, cujo objetivo é transformá-los em sentenças lógicas capazes de serem usadas em algum interpretador de comandos.

A principal diferença entre o sistema apresentado neste trabalho é o modo em que são analisadas e transcrevidas as frases ditas pelo usuário, bem como o tratamento dado à sinônimos e outras estruturas de linguagem. Não é necessário que o usuário possua um conhecimento prévio de comandos específicos, pois a aplicação encarrega-se de interpretar frases e transcrevê-las em um padrão lógico usado pelo sistema.

3. Processo de desenvolvimento

Para o desenvolvimento deste trabalho foram utilizados conceitos relacionados ao processamento de linguagem natural (PLN) para interpretar os comandos enviados pelos usuários ao sistema. O sistema faz uso da captura da fala do usuário através de microfones ou do sensor Kinect [MICROSOFT, 2013]. Além disso, o protótipo foi concebido de forma a operar como uma arquitetura cliente-servidor. O cliente apenas fica encarregado de transcrever as entradas em áudio para texto, que posteriormente serão enviadas ao servidor. Isto faz com que o cliente possa ser implementado em diferentes arquiteturas, pois a interação como o módulo gerador de histórias não é realizada de forma direta, ficando a cargo dele apenas a captura de áudio. Desta forma justifica-se transformar as entradas do usuário apenas em texto, pois as requisições serão de fato analisadas apenas pelo servidor da aplicação.

O servidor é responsável por transcrever os comandos do usuário, previamente transformados em textos, em sentenças lógicas. Estes sentenças são usadas por um gerenciador de interação a fim de garantir que os estados da história gerada são válidos. É neste ponto em que uma história é de fato processada. O Servidor através deste gerenciador é responsável por processar uma história definida pelo autor do sistema e apresentar o estado inicial do enredo e suas possíveis alterações. O gerenciador de interação é baseado no

trabalho proposto por Ciarlini et al. [2005] onde é especificado um modelo formal para a lógica de eventos e comportamento dos personagens na história. Este módulo trata requisições do usuário através de uma lógica em Prolog a fim de validá-las, verificando a coerência da história gerada até o momento. O servidor ainda é responsável por apresentar o resultado aos usuários, que aqui, para testes preliminares, é apresentado através de um diagrama contendo os eventos correntes da história. Além disso, o sistema foi implementado de modo que mais de um cliente possa interagir com o servidor ao mesmo tempo. A Figura 2 apresenta um diagrama da arquitetura proposta.

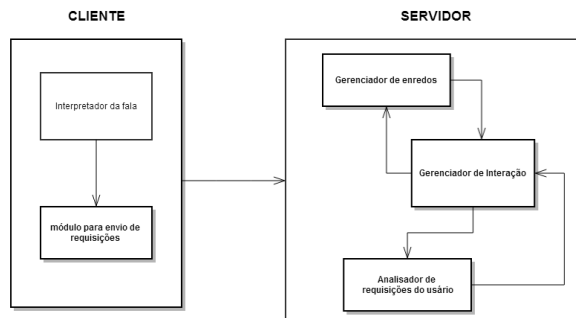


Figura 2: Arquitetura do sistema proposto.

Dadas as diferentes opções e definições de eventos para um determinado enredo, o utilizador pode interagir com a ferramenta a fim de obter uma grande variedade de versões de uma história, respeitando as regras para manter o enredo coerente.

3.1 Captura de áudio

Para interpretar as frases ditas pelo usuário, foi implementado um módulo em C# onde são identificadas as palavras-chave dado o contexto da história definida pelo sistema. Para tanto foi utilizada a API Speech da Microsoft [MICROSOFT, 2013], que é compatível também com o kit de desenvolvimento para o Kinect. Para a utilização deste módulo foi criado um "dicionário", contendo palavras que podem ser consideradas pelo sistema. Este dicionário contém aproximadamente 100 palavras sensíveis ao contexto da história a ser gerada além de possuir outras, como pronomes e verbos, e inclui ainda uma pequena lista de sinônimos. Para a identificação de voz, foi preciso criar e carregar uma gramática que indica o método usado para o reconhecimento. A API Speech possui uma classe denominada grammar que permite carregar uma gramática livre de contexto especificada no formato Microsoft Speech API (SAPI) XML. As palavras definidas no dicionário alimentam esta gramática utilizada pelo programa no reconhecimento de termos relevantes, onde são consideradas juntamente com o padrão de língua utilizado, o qual neste caso é o inglês. Desta forma, é possível transformar frases ditas pelo usuário em conjuntos de frases em formato de texto quando os comandos identificados são comparados com as palavras contidas no dicionário. Cada frase dita pelo usuário é transformada em texto e para que seja

confirmado o seu envio ao servidor é necessário falar o comando "continue". As frases processadas são enviadas através de uma implementação via sockets.

3.2 O módulo gerenciador de interação

No sistema proposto é preciso realizar uma análise das sentenças enviadas pelo usuário. Foi implementado um módulo na linguagem Java, o qual funciona na forma de um servidor que recebe os comandos vindos do cliente, onde estes são analisados a fim de serem validados pelo gerenciador de interação. Os textos enviados ao servidor foram analisados sintaticamente com a construção de uma árvore abstrata e com a análise da dependência entre os termos das sentenças enviadas. Em uma etapa posterior, uma análise semântica para extrair o significado das frases também é realizada. Para a análise sintática foi utilizado o Stanford Parser [STANFORD, 2012], sendo este um parser probabilístico com a finalidade de analisar a relação de dependência entre os termos das frases e produzir também uma árvore da estrutura de cada frase. Por exemplo, a sentença "the knight must attack the dragon", poderia ser representada pela árvore sintática gerada e pelas relações de dependência encontradas conforme a Figura 3.

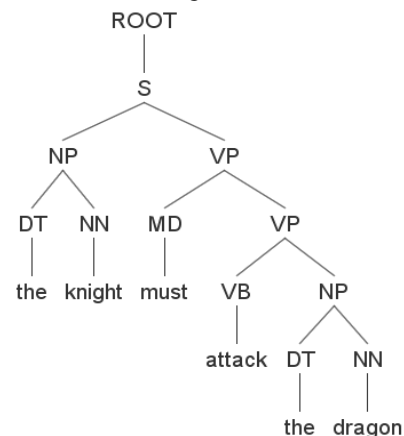


Figura 3: Árvore sintática da expressão "the knight must attack the dragon" gerada com o parser.

Nota-se que pela análise da frase enviada pode-se extrair relações binárias identificando quem realiza ou sofre alguma ação. Desta forma a sentença enviada pode ser facilmente transcrita para uma lógica capaz de ser interpretada. Assim é possível extrair, por exemplo, a seguinte sentença em lógica de primeira ordem: *Attack(Knight,Dragon)*. Posteriormente com as entradas processadas, as sentenças obtidas são enviadas ao gerenciador de interação, o qual é responsável por analisá-las. Dada o contexto da história, cada ação considera pré e pós-requisitos para ser executada. Sendo assim esta etapa do processo executa a interação de fato com a história. Caso alguma das requisições do usuário possa ser atendida, ela é adicionada a um diagrama contendo a história corrente e então apresentada ao usuário. Caso contrário, é simplesmente descartada. Em resumo a história é transcrita em um modelo formal baseado em regras explícitas, as quais o

sistema permite que sejam modificadas pelo usuário a fim de gerar diferentes versões da mesma história.

4. Avaliação e Resultados

A história criada como modelo de teste neste sistema é baseada em contos de fadas. Existem um cavaleiro herói, um dragão vilão e a vítima do dragão a ser resgatada. A interação com a história é permitida a partir do momento em que o vilão captura sua vítima, onde o usuário pode dar “dicas” ao herói do que é necessário fazer como, por exemplo, investigar o local do sequestro ou ir até o castelo do vilão.

O sistema foi testado em diferentes computadores com microfones comuns e também com o Kinect. Apesar das limitações decorrentes dos ruídos no ambiente, o identificador de comandos de voz se mostrou eficaz. Aproximadamente 80% das requisições, de um conjunto de 30 frases diferentes submetidas ao sistema, puderam ser atendidas, porém vale salientar a dificuldade a qual a biblioteca Speech possui em diferenciar palavras com fonemas parecidos. Em alguns casos, o sistema não conseguiu compreender o comando dito pelo usuário, o que pode atrapalhar a execução da aplicação. Outra limitação do sistema diz respeito às palavras que podem ser identificadas. Apesar de um dicionário contendo aproximadamente 100 palavras ter sido criado, ainda há palavras às quais não puderam ser identificadas.

Diferentemente do módulo responsável pela interpretação da fala, os resultados obtidos da interação com o gerenciador se mostrou eficiente, sendo que 9 em cada 10 requisições puderam ser processadas pelo sistema. Para fins de teste, são apresentados, ao usuário, os eventos da história corrente na forma de um diagrama conforme a Figura 1.

5. Conclusão e Trabalhos Futuros

O processo para transcrever a fala do usuário em uma lógica capaz de o gerenciador de interação utilizar representa a maior contribuição deste trabalho. Esta é uma tarefa complexa onde ainda existe certa dificuldade em processar sinônimos e algumas estruturas frasais. Apesar disso, dada uma análise qualitativa do sistema a maioria das requisições processadas foram atendidas. O sistema proposto visa eliminar caixas de diálogo e a constante interação dos usuários através de comandos de mouse ou teclado. Com a utilização de comandos de voz para a geração de enredos, o modo como a pessoa utiliza o sistema torna-se mais natural, pois em uma abordagem mais “tradicional”, o espaço em tela necessário para apresentar uma GUI poderia competir com o espaço necessário para renderizar as cenas da história, atrapalhando a experiência deste tipo de sistema, onde devido ao espaço limitado, certos detalhes da história apresentada poderiam passar despercebidos. Desta forma este sistema pode transformar a geração de

histórias interativas em uma experiência mais agradável do que os métodos tradicionais. Futuramente pretende-se realizar novos testes com o sistema a fim de aprimorar o mecanismo de interpretação da fala. Isto permitirá que um número maior de palavras possa ser adicionado ao dicionário e um maior número de sentenças possa ser processado.

Referências

- CIARLINI, A., 1999. *Geração Automática de enredos*. Tese de Doutorado em Informática, Pontifícia Universidade Católica do Rio de Janeiro.
- CIARLINI, A., POZZER, C. T., FURTADO, A. L. AND FEIJÓ B., 2005. *A logic-based tool for interactive generation and dramatization of stories*. Proceedings of the International Conference on Advances in Computer Entertainment Technology, p. 133-140.
- FORMAN, P., JOHN, R. W. S., 2000. The Future of digital entertainment/creating convergence. *Scientific American*, v. 283, p. 50-56.
- GRASBON, D., BRAUN, N., 2001. A morphological approach to interactive storytelling. Proceedings of Cast01, Living in Mixed Realities, Sankt Augustin, Germany, p. 337-340.
- LIMA, E. S., FEIJÓ, B., CIARLINI, A. E. M., POZZER, C. T., FURTADO, A. L. AND SILVA, F. A. G., 2012. *A Multi-User Natural Language Interface for Interactive Storytelling in TV and Cinema*. Proceedings of the XI Brazilian Symposium on Computer Games and Digital Entertainment (SBGames 2012), Brasília, Brazil, p. 154 - 161.
- LIMA, E. S., FEIJÓ, B., BARBOSA, S. D. J., FURTADO, A. L., CIARLINI, A. E. M., AND POZZER, C. T., 2011. *Draw Your Own Story: Paper and Pencil Interactive Storytelling*. Proceedings of the 10th International Conference on Entertainment Computing (ICEC 2011), Vancouver, Canadá, Outubro 2011, p. 1-12.
- MATEAS, M., 1997. *An oz-centric review of interactive drama and believable agents*. Technical report, School of Computer Science, Carnegie Mellon University.
- SPIERLING, U., BRAUN, N., IURGEL, I. AND GRABSON, D., 2002. *Setting the Scene: playing digital director in interactive Storytelling and creation*. *Computer & Graphics*, v. 26, p. 31-44.
- MICROSOFT, 2013. Microsoft Speech API 5.4. Disponível em: [www.msdn.microsoft.com/enus/library/ee125663\(v=vs.85\).aspx](http://www.msdn.microsoft.com/enus/library/ee125663(v=vs.85).aspx) [Acesso em 16 de Novembro 2013].
- MICROSOFT, 2013. Microsoft Kinect SDK. Disponível em: <http://www.microsoft.com/enus/kinectforwindows/> [Acesso em 16 de Novembro 2013].
- STANFORD, 2012. The Stanford NLP (Natural Language Processing) Group. Disponível em: <http://nlp.stanford.edu/downloads/lex-parser.shtml> [Acesso em 2 de Dezembro de 2013].