

Integração do Mecanismo de Síntese de Fala MaryTTS ao Avatar Educação

Integration of Synthesis Speech Mary TTS Mechanism to Avatar Education

Fabio Alexandre de Holanda
Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
famhs@ecom.poli.br

Danilo de Souza Barbosa
Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
dbs3@ecom.poli.br

Alexandre Magno Andrade Maciel
Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
amam@ecom.poli.br

Resumo *O Avatar Educação consiste num assistente virtual que utiliza personagens animados com voz sintética com o objetivo de melhorar a comunicação e a interação no Ambiente Virtual de Aprendizagem. Apesar da sua considerável contribuição na área educacional, alguns problemas na qualidade da voz sintética têm sido relatados pelos os usuários que utilizam o ambiente. Este artigo possui como objetivo a integração do mecanismo de síntese de fala MaryTTS ao Avatar Educação. Para isso foi realizado a comunicação dos módulos cliente e servidor do MaryTTS ao módulo de síntese do FIVE e uma adaptação do middleware Avatar Vocálico. Esta integração possibilitou a utilização dos serviços do mecanismo MaryTTS com as suas respectivas técnicas, através do seu serviço web e, assim gerar novos áudios com uma melhor qualidade na sua síntese de fala.*

Palavras-Chave: *Síntese de fala, MaryTTS, FIVE e Avatar Educação*

Abstract *The Avatar Education is a virtual assistant that uses animated characters with synthetic voice with the objective of improve communication and interaction in Virtual Learning Environment. Despite their considerable contribution in education, some problems in the quality of synthetic speech have been reported by users who use the environment. This paper aims to integrate the speech synthesis engine to MaryTTS Avatar Education. To this was accomplished the communication of MaryTTS the client and server modules to the FIVE synthesis module and an adaptation of middleware Avatar Vocálico. This integration possible the use of the mechanism MaryTTS services with their respective techniques, through its web service and thus generate new audio with higher quality in your speech synthesis.*

Keywords: *Speech synthesis, MaryTTS, FIVE Avatar Education*

1 Introdução

A disponibilização da aprendizagem pela internet, relacionado com o interesse crescente da sociedade para realizar a sua qualificação por meio de recursos computacionais, mudaram a visão do ensino à distância, e a maneira de como deliberar o processo de ensino e aprendizagem [1]. Através dessa evolução novas pesquisas multidisciplinares, envolvendo ciência e educação tem contribuído significativamente no planejamento, desenvolvimento e uso de recursos tecnológicos, a fim de oferecer novas perspectivas de interação e comunicação dentro do contexto educativos [2].

O Ambiente Virtual de Aprendizagem (AVA) é uma ferramenta que realiza uma interação entre os usuários, além de auxiliar com recursos de comunicação assíncrona e síncrona. Entre esses recursos estão as listas de discussão, fóruns, chat, videoconferência e ferramenta para gerenciamento de atividades e avaliação que são os mais utilizados. [3] Apesar disso, esse ambiente causa uma sensação de elevada recebimento de informação para os estudantes, que geralmente confundem a grande quantidade de conteúdo disponível por essa plataforma de ensino [4].

Esta situação é considerada muito comum em cursos na modalidade de ensino à distância, pelo alto grau de interação dos alunos, por meio do ambiente e uma grande fluxo de informação diário. Nesse contexto, a carga excessiva de informações apresentadas de forma não estruturada, pode causar efeitos adversos, tais como falta de compromisso nas ações do curso, ou em casos mais graves de evasão dos alunos [5].

Em Maciel *et al.* [6] foram construídas, a partir de uma base de dados própria contendo cerca de quatro horas de áudio, um conjunto de vozes masculina e feminina, tomando como base a ferramenta de síntese de fala HTS (*HMM-based Speech Synthesis Toolkit*) [7] que foi integrado ao FIVE. Apesar dos bons resultados obtidos a solução apresentou problemas de portabilidade que foram solucionados em Sousa *et al.* [8], utilizando a plataforma MaryTTS (*Modular Architecture for Research in sYnthesis*) [9]. Já em Maciel *et al.* [4] foi construída uma arquitetura de desenvolvimento portátil por meio de um *middleware* chamado Avatar Vocálico responsável pela geração das sequências de animação e por realizar a comunicação com o Framework FIVE.

2 Revisão da Literatura

Esta revisão da literatura apresenta os conceitos essenciais para o entendimento do problema de pesquisa e para a integração do mecanismo de síntese de fala MaryTTS ao

Avatar Educação. Dentre os conceitos apresentados, analisamos a síntese de fala e as técnicas de síntese baseada em seleção de unidade concatenativa e baseada em HMM, e as ferramentas Avatar Educação, o framework FIVE e o sistema de síntese de fala MaryTTS.

2.1 Síntese de Fala

Síntese de fala é o processo de geração de uma réplica sintética de um sinal de fala, como forma de transmitir mensagem de uma máquina para uma pessoa [10]. Duas características são essenciais na classificação dos sistemas de síntese de fala: a qualidade do sinal de fala e o tamanho do vocabulário. Na qualidade do sinal existe as suas características de naturalidade e inteligibilidade. Já o tamanho do vocabulário existe as amostras de áudio pré-gravadas possuem qualidade real de sinal, contudo possuem vocabulário limitado, já os sistemas de *Text-To-Speech* (TTS) são capazes de gerar fala com boa qualidade com um vocabulário ilimitado [11].

O sistema TTS é o processo de conversão de um determinado texto, para discurso sintético. Sendo composto formado por duas fases: a primeira consiste no Processamento de Linguagem Natural (PLN), onde a mensagem de entrada é transcrita em uma representação de nível fonético e a segunda consiste na Geração da Fala em que as formas de onda de voz são geradas e a saída acústica é produzida. Essas duas fases são respectivamente chamadas de síntese de alto nível e síntese de baixo-nível [12]. Uma versão simplificada do processo é apresentada na Figura 1.



Figura 1: Descrição do processo de síntese de fala [12]

De acordo com Tokuda *et al.* [7], há diferentes abordagens ou técnicas que podem ser utilizadas para a geração de formas de onda do discurso de uma locução. As abordagens mais comuns são: a síntese de voz para a seleção e concatenação de unidades e, a síntese de voz com base em parâmetros calculados pelo Hidden Markov Models (HMM).

2.1.1 Síntese baseada em seleção de unidades

A seleção da unidade ainda é a abordagem dominante para a síntese de voz de alta qualidade, no entanto sintetizadores baseados em HMMs foram muito apreciados recentemente por causa da possibilidade de obtenção de discurso de boa qualidade a partir de um pequeno banco de dados, e devido ao fato de que as características de voz podem ser modificadas [13].

De acordo com Taylor [10], historicamente, sistemas de síntese de seleção da unidade é considerada uma extensão natural do sistema Concatenativo, em que são utilizadas pequenas unidades, pré-gravadas de discurso, como fonemas, difonemas e trifonemas para construir o enunciado. No caso em que não apenas um, mas centenas de realizações de cada unidade de discurso fonética estão presentes em um inventário, um processo de seleção da unidade deve ocorrer a fim de criar a seqüência da sua unidade sintética [14].

2.1.2 Síntese baseada em HMN

A síntese baseado em HMM tem sido um tema importante na pesquisa de síntese de voz e utilizado em todo mundo por ambas as organizações acadêmicas e comerciais. Ele consiste em duas fases principais, a fase de treinamento e a fase de síntese. Na fase de treinamento, deve ser decidido que modelos devem ser treinados, utilizando os recursos que são extraídos e colocado no vetor com suas características [11].

O algoritmo de Baum-Welch é usado com os vetores de características para a produção de modelos para cada allophones. Um modelo geralmente consiste de três estados que representam o início, no meio e no final do fonemas. A fase de síntese consiste em dois passos: em primeiro lugar, os vectores de características para uma dada seqüência de allophones têm que ser estimados. Em segundo lugar, um filtro é implementado para transformar esses vectores de características em sinais de áudio [15].

2.2 Ferramentas

As ferramentas utilizadas para integração do sistema Mary ao Avatar Educação foram o plugin Avatar Educação, o framework FIVE e o sistema de síntese de fala MaryTTS.

2.2.1 Avatar Educação

O Avatar Educação consiste num assistente virtual que utiliza personagens animados com voz sintética com o objetivo de melhorar a comunicação e a interação no Ambiente Virtual de Aprendizagem (AVA) Moodle [4]. O seu funcionamento ocorre por um serviço Web onde é enviado solicitações ao Avatar Vocálico [16], middleware responsável pela comunicação com o FIVE [17], que por sua vez tem a função de gerar voz sintetizada. A Figura 2 mostra a visão geral da arquitetura do Avatar Educação.

O processo de execução do avatar educação iniciou com uma consulta ao banco de dados do Moodle, no qual realiza uma comparação no seu log de consultas. Caso a

comparação tenha resultado inalterado desde a última consulta, então nada é feito em relação a solicitação realizada via web service para o Avatar Vocálico, ou seja, essa síntese já foi realizado, não sendo assim necessário refazer as operações de geração de voz, precisando somente buscar os arquivos no repositório do Avatar Vocálico [4].

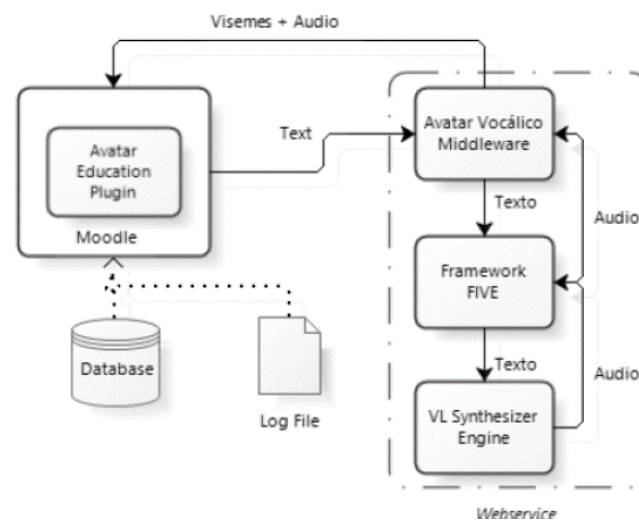


Figura 2: Arquitetura do Avatar Educação

Caso o resultado dessa comparação tenha dado diferente dos registros do log, é enviado uma mensagem de texto para o Avatar Vocálico, que é responsável pela geração da animação dos visemas, além de enviar solicitação de síntese e fala para o FIVE. Por sua vez o FIVE utiliza o motor de síntese VL Synthesizer para geração do áudio que é devolvido ao Avatar Vocálico. Os arquivos de áudio e visemas são disponibilizado no repositório do Avatar Vocálico, no qual são chamados pelo plugin, para assim realiza a geração da animação.

2.2.2 Framework FIVE

O FIVE (*Framework for an Integrated Voice Environment*) é um *framework* construído com o objetivo de auxiliar o desenvolvimento integrado de aplicações com interfaces de voz. Sendo possível gerar motores de síntese de voz, de reconhecimento de fala e de verificação de locutor, por meio de uma API (*Application Programming Interface*) própria e instanciar estes motores em múltiplas aplicações [17].

A construção do FIVE foi realizado em consideração os requisitos propostos pela comunidade acadêmica na área de processamento de voz. Pelo qual foi observado nesta pesquisa a necessidade da construção de uma ferramenta que auxilie a aplicação com interface baseado em voz, com uma rápida curva de aprendizagem, capacidade de extensão entre as técnicas de extração de características

e classificação de padrões e a portabilidade entre ambientes computacionais [5].

A arquitetura do framework FIVE é composta por cinco módulos, no qual os módulos básicos são representados por três módulos de processamento de fala: motor de reconhecimento automático de fala (ASR) responsável pela extração de informação de uma mensagem, contida em um sinal de fala, motor texto-fala (TTS) responsável pela criação de uma réplica sintetizada a partir de uma mensagem textual e o motor de verificação automático de fala (ASV) responsável pela verificação do locutor com o propósito de restringir o acesso à informação.

O FIVE além dos módulos básicos possui dois módulos intermediários: o módulo interface de programação (API) que consiste numa comunicação entre os módulos básico que é integrado ao arcabouço de classes do CORE que realiza a execução central do *framework* e ao GUI (*Graphical User Interface*) que possui uma interface gráfica que auxilia o desenvolvimento de projetos necessários para a camada de aplicação; e o módulo APP que consiste numa possível aplicação dessa arquitetura em outra plataforma como no ambiente virtual de aprendizagem [17].

2.2.3 Sistemas MaryTTS

O sistema MaryTTS é uma plataforma open-source, multilingual Text-to-Speech Synthesis escrito em Java, originalmente desenvolvido como um projeto colaborativo de DFKIs Idioma Lab Tecnologia e pelo Instituto de Fonética da Universidade de Saarland, e agora mantido pelo grupo de processamento Multimodal Speech no Cluster de Excelência MMCI e DFKI [18]. MARY significa arquitetura modular para a pesquisa sobre a síntese de voz e essa arquitetura modular é capaz de suportar vários idiomas [19, 20], treinados através de várias técnicas [21, 22].

A arquitetura do MaryTTS é formada pela estruturação do processamento de linguagem natural e pela geração de voz sintética. A arquitetura do processamento de linguagem natural é composta por quatro partes [23]:

1. O pré-processamento ou normalização do texto que realiza a expansão de abreviatura e numeral;
2. O processamento de linguagem natural que realiza a análise linguística;
3. O cálculo dos parâmetros acústicos que realiza a entonação de voz;
4. O sintetizador de voz que cria um arquivo de som a partir de uma sequência de fonemas.

Já a arquitetura utilizada para a geração da voz é composta pelo serviço realizado entre o cliente e o servidor. O seu funcionamento é formada por um servidor principal que faz pedidos para o servidor executa alguma tarefa e recebe o resultado do processamento do cliente. Um servidor pode suportar várias línguas e recebe pedidos dos clientes em uma porta especificada pelo seu servidor [23].

3 Integração

O processo de integração do mecanismo de síntese de fala Mary ao Avatar educação foi proposto para realizar a geração de novas vozes sintética com as técnicas baseado em Seleção de Unidade Concatenativa e baseado em HMM. Sendo essa integração proposta em três fases: instalação do sistema Mary ao servidor WEB, modificação do módulo de síntese do framework FIVE e a integração ao avatar educação para reconhecer o serviço proposto no MaryTTS.

A primeira fase da integração foi realizado uma instalação do sistema Mary ao servidor Web, com isso esse serviço passou a ser disponibilizado no servidor do avatar educação, com as configurações de endereço e a sua porta especificada pelo sistema Mary. O servidor Web passou a executar os serviços cliente e servidor do sistema Mary e disponibiliza esse serviço em múltiplas plataformas. Conforme mostrado na Figura 3.

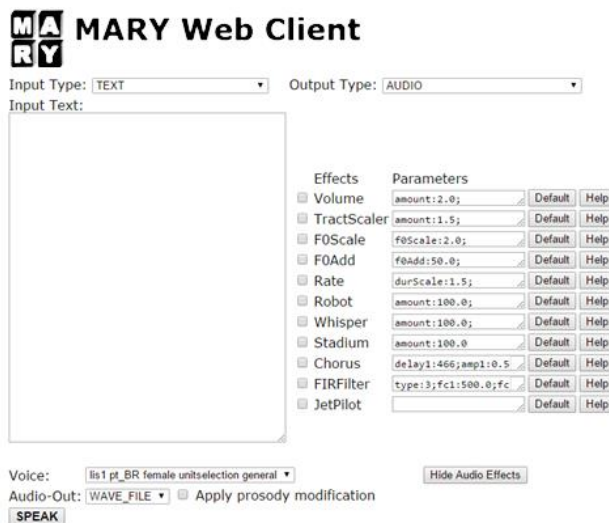


Figura 3: Interface do servidor Web MaryTTS

Com o servidor Mary executado por um serviço na WEB, tornando assim, o seu cliente disponível por meio de uma interface gráfica. Caso deseja realiza a geração da voz sintética é necessário solicita os seguintes parâmetros: tipo de entrada de texto e saída de áudio, tipo de voz masculino ou feminino e a sua respectiva técnica seleção de unidade ou baseada em Modelos Ocultos de Markov e o

tipo de áudio nos formatos de áudio WAV, AIFF ou WMA.

A segunda fase ocorreu a modificação do módulo de síntese do framework FIVE. Sendo assim, realizado a importação da biblioteca do Mary ao repositório FIVE CORE. Com a execução do Mary disponível no FIVE foi necessário adaptar os seus parâmetros disponibilizado pelo Mary, para assim, fica compatível ao módulo de síntese de fala do FIVE.

As modificações para reconhecer a execução do motor de síntese ocorreu no módulo FIVE API, para isso foram utilizadas as configurações do motor síntese HTS, como forma de manter as características da arquitetura do FIVE [22]. Com isso foi possível executa o motor Mary no FIVE com uma solicitação de áudio sintético, contendo como entrada um texto. A geração desse áudio ocorreu por duas formas pela geração do arquivo de síntese de fala nos formatos WAV ou pela execução do áudio da síntese de fala.

Com a execução do motor de síntese do Mary compatível a arquitetura do FIVE foi necessário realizar a conexão dos parâmetros utilizado no servidor do Mary ao módulo FIVE CORE. Essa conexão entre o servido que está executando na WEB ocorreu com a estruturação dos parâmetros do servidor, para assim reconhece o seu cliente no FIVE.

Com isso o FIVE CORE passou a executa informações do servidor Mary, no qual foi realizado a configuração com o caminho do seu endereço Web e a sua porta específica, os caminhos do texto e a o seu respectivo áudio. além das configurações padrões do motor de síntese do Mary: a entrada do texto e a saída do áudio, o idioma utilizado português do Brasil e o geração do arquivo de áudio no formato WAV como mostrado a Figura 4.

```
String serverHost = System.getProperty("server.host", "vocallab.com.br");
int serverPort = Integer.getInteger("server.port", 59125);
MaryClient mary = MaryClient.getMaryClient(new Address(serverHost, serverPort));
String text = phrase;
String locale = "pt_BR";
String inputType = "TEXT";
String outputType = "AUDIO";
String audioType = "WAVE";
String defaultVoiceName = speaker;
ByteArrayOutputStream baos = new ByteArrayOutputStream();
mary.process(text, inputType, outputType, locale, audioType, defaultVoiceName, b
AudioInputStream ais = AudioSystem.getAudioInputStream(
    new ByteArrayInputStream(baos.toByteArray()));

AudioSystem.write(
    ais, AudioFileFormat.Type.WAVE, new File(enginePath + File.separator + "
    + File.separator + fileName + ".wav"));
```

Figura 4: codificação com as configurações do Mary Cliente

A terceira fase foi realizado a integração do serviço disponibilizado no FIVE do Mary ao Avatar Educação. Sendo realizado uma reformulação do módulo Avatar Vocálico *Middleware*, que realizar uma comunicação com

framework FIVE. Com essas mudanças o Avatar Vocálico, consegue executa o áudio no FIVE, utilizado os mecanismos HTS [25] e o mecanismo de síntese de fala MaryTTS [23], e a executa desses mecanismos a depende da solicitação realizada pelo Plugin Avatar educação, no qual pode enviar solicitação ao motor HTS ou ao motor Mary, com os parâmetros frase, voz e o arquivo de áudio. Como mostrado na Figura 5.

```
String enginePath = args[0];
String phrase = args[1];
String speaker = args[2];
String fileName = args[3];
int type = Integer.parseInt(args[4]);

Synthesizer synthesizer = new Synthesizer(enginePath);
if (type == 1) {
    synthesizer.synthesizeMary(phrase, speaker, fileName);
} else if (type == 2) {
    synthesizer.synthesizehts(phrase, speaker, fileName);
}
```

Figura 5: Codificação modificada do Avatar Vocálico

A arquitetura modificada do Avatar Educação passou a executa os dois motores de síntese de fala. Como mostrado na Figura 6. Com isso o mecanismo de síntese HTS necessita dos parâmetros com a descrição do texto e o tipo de áudio masculino ou feminina. Com essas especificações definidas ocorre a síntese no motor HTS (*HMM-based Speech Synthesis Toolkit*) e a geração do seu respectivo áudio [26].

Já o mecanismo de síntese MaryTTS é uma sistema que possui uma execução de serviço entre o seu cliente e o um servidor web [18]. Com a possibilidade receber solicitação do Avatar Vocálico e gerar áudio utilizando as técnicas baseada em Seleção de unidade e baseado em HMM, possibilitando assim a gerações de áudio e dos visemas e realizar a geração das novas animação.

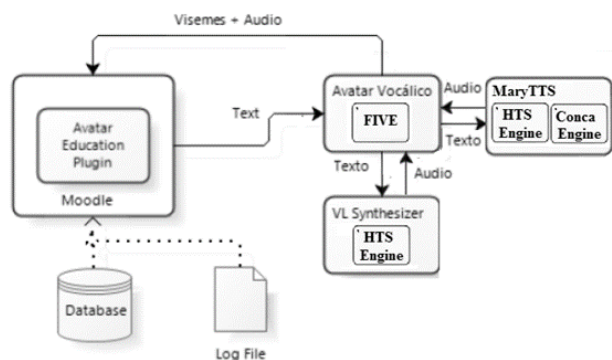


Figura 6: Avatar Educação modificado

4 Resultados e discussão

Os resultados obtidos com a integração do mecanismo de síntese de fala MaryTTS ao Avatar educação foram relevantes. Com essa integração do mecanismo de síntese de fala ficou disponível na Web um serviço de síntese de voz, com a possibilidade de gerar vozes sintética utilizando as técnicas seleção de unidade e baseada em HMM, além da modificação no dicionário da língua português Brasileiro.

Com esse serviço web sendo executado no servidor do Avatar Educação, além de utiliza essas vozes no ambiente virtual de aprendizagem, existe a possibilidade de executa em qualquer sistema WEB e na plataforma Android. Além disso, o servidor WEB está sendo executado como um modulo cliente no framework FIVE, e com isso ser utilizado em múltiplas plataformas. Conforme mostrado na Figura 7.

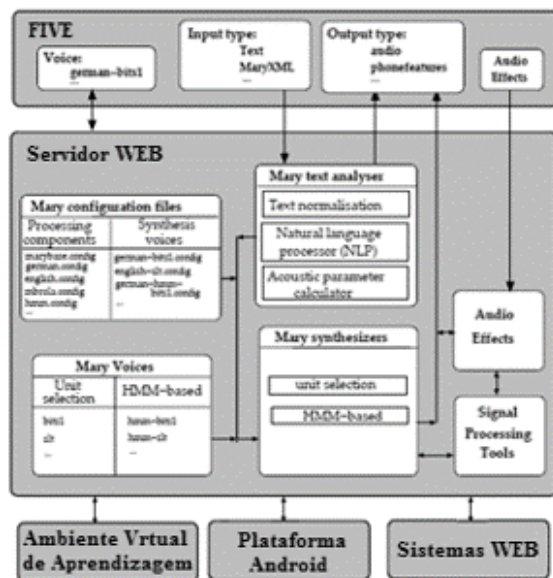


Figura 7: Arquitetura de integração das ferramentas

Na integração do sistema MaryTTS por uma aplicação do FIVE ao plugin Avatar Educação foi observado como resultado a possibilidade de alteração das vozes geradas pela animação do Avatar. Sendo assim, possível atualiza o repositório do servidor Mary que está como serviço no servidor do Avatar Educação. Podendo assim, altera a depende da necessidade do Avatar Educação o tipo de voz masculino ou feminino e a sua técnica seleção de unidade ou baseada em HMM.

5 Conclusão

Este artigo realizou a integração do mecanismo de síntese de fala do MaryTTS para o Avatar Educação. Para isto foi realizado um agregação dos módulos cliente e servidor do MaryTTS ao módulo CORE do FIVE e com instanciação no módulo API do FIVE. Depois de realizado todo o encapsulamento das classes, foi alterado o módulo API do FIVE para que o mesmo reconheça os novos serviços de voz e os forneça às aplicações que realiza instanciação desse motor de síntese de fala. Na sequência, foi realizada uma adaptação do *middleware* Avatar Vocálico para que ele fosse capaz de optar pelas vozes produzidas pelo mecanismo HTS ou pelo mecanismo MaryTTS.

Acreditamos que essa integração, possibilitou a utilização do Avatar educação com a geração de vozes produzidas com as técnicas baseada em seleção de unidade e baseada em HMM, além da construção de novas com o mecanismo de síntese MaryTTS e a sua execução com a solicitação do Avatar educação. Desse modo, concluímos que essa integração possibilitou utiliza o Avatar educação com a execução de diversas geração de áudio com as técnicas disponibilizada no MaryTTS. Como trabalhos futuros integrar novos motores e serviços de fala, assim como criar um serviço web para o HTS, para que as mudanças na arquitetura do FIVE fiquem mais consolidadas.

Referências

- [1] M. L. P. Flôres, L. M. R. Tarouco. Diferentes tipos de objetos para dar suporte a aprendizagem. *Revista Renote*, 6(2), 2008.
- [2] L. M. A. Santos. A Inserção de um Agente Conversacional Animado em um Ambiente Virtual de Aprendizagem a partir da Teoria da Carga Cognitiva. Tese de Doutorado, Universidade Federal do Rio Grande do Sul, Jan 2009.
- [3] D. C. B. Marcelo, D. S. M. Ana Paula, R. B. A. Zulatto. Educação a distância online. Autêntica, 2013.
- [4] A. M. A. Maciel, R. L. Rodrigues, E. C. B. Carvalho. Desenvolvimento de um Assistente Virtual Integrado ao Moodle para Suporte a Aprendizagem Online. In *Proceedings of Simpósio Brasileiro de Informática na Educaçã*, Mato Grosso do Sul, páginas 06-10, 2014.
- [5] L. M. A. Santos. A Inserção de um Agente Conversacional Animado em um Ambiente Virtual de Aprendizagem a partir da Teoria da Carga Cognitiva. Tese de Doutorado Universidade Federal do Rio Grande do Sul, 2009.

- [6] A. M. A. Maciel. Investigação de um Ambiente para o Desenvolvimento Integrado de Interface de Voz, Tese de Doutorado Universidade Federal de Pernambuco, dez, 2012.
- [7] K. Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi, K. Oura. Speech synthesis based on hidden Markov models. In Proceedings of the IEEE, 101(5): 1234-1252, 2013.
- [8] D. Sousa, S. Levi, A. M. A. Maciel. A Portability Evaluation of Brazilian Portuguese voices produced with MARY TTS. In *Systems, Signals and Image Processing*, 2014.
- [9] M. Schroder, J. Trouvain. O sistema de síntese MARY alemão text-to-speech: Uma ferramenta para pesquisa, desenvolvimento e ensino. *International Journal of Speech Technology*, v 6, n. 4, p. 365-377, 2003.
- [10] P. Taylor. Text-To-Speech Synthesis. Universidade Cambridge, 2009.
- [11] V. L. Latsch. Desenvolvimento de um sistema de conversão texto-fala com modelagem de prosódia. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Jan 2011.
- [12] R. S. Maia. Speech Synthesis and Phonetic Vocoding for Brazilian Portuguese based on Parameter Generation from Hidden Markov Models. Tese de doutorado do instituto de tecnologia Nagoya, 2006.
- [13] Y. Tabet, M. Boughazi. Speech Synthesis Techniques: A Survey. In: Proceedings of International Workshop on Systems, Signal Processing and their Applications, Argel, Argélia, 2011.
- [14] A. J. Hunt, A.W. Black. Unit Selection in a Concatenative Speech Synthesis System Using a Large Speech Database. In: International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 373-376. IEEE, Atlanta, Georgia, 1996.
- [15] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, T. Kitamura. Speech parameter generation algorithms for HMM-based speech synthesis. In *Acoustics, Speech, and Signal Processing*, 3(2): 1315-1318, 2000.
- [16] R. L. Rodrigues, A. M. A. Maciel, E. C. Carvalho. Desenvolvimento de uma ferramenta para a produção de mídias utilizando personagem animado com síntese de voz. In *Anais do Simpósio Brasileiro de Informática na Educação*, 2012.
- [17] A. Maciel, E. Carvalho. FIVE - Framework for an Integrated Voice Environment. In *International Conference on Systems, Signals and Image Processing*, Rio de Janeiro, 2010.
- [18] The MARY Text-to-Speech System, <http://mary.dfki.de>, nov. 2015.
- [19] I. Couto, N. Neto, V. Tadaiesky, A. Klautau, R. Maia. An Open Source HMM-based Text-to-Speech System for Brazilian Portuguese. In *7th International Telecommunications*, Manaus, 2010.
- [20] P. Stavropoulou, D. Tsonos, G. Kouroupetroglou. Language Resources and Evaluation for the Support of the Greek Language in the MARY Text-to-Speech. In *17th International Conference on Text, Speech and Dialogue*, Springer International Publishin, Czech Republic, 2014.
- [21] M. Charfuelan, S. Pammi, I. Steiner. MaryTTS Unit Selection and HMM-based Voices for the Blizzard Challenge 2013. In *Proc. of the Blizzard Challenge*, , Spain, 2013.
- [22] S. Pammi, M. Charfuelan. HMM-based sCost Quality Control for Unit Selection Speech Synthesis. In *Blizzard Challenge Workshop, Proc. of the Blizzard Challenge*, 2013.
- [23] M. Schroder , J. Trouvain. The German text-to-speech synthesis system MARY: A tool for research, development and teaching. In *International Journal of Speech Technology*, 6(4):365-377, 2003.
- [24] A. Maciel, CARVALHO, E. Carvalho. Integration and evaluation of an HMM-based Text-To-Speech System to five. In: *Systems, Signals and Image Processing*, 2012
- [25] HTS, HMM-Based Speech Synthesis System, <http://hts.sp.nitech.ac.jp>, Abr, 2014.
- [26] HTK, extensões com parâmetros do HTK, <http://htk.eng.cam.ac.uk/extensions>, Jun, 2015.