**Resumo Expandido**

| | | |
|---|---|---|
| **Título da Pesquisa**: Desenvolvimento de Ferramenta de conversão de texto em fala | | |
| **Palavras-chave**: Conversão Texto-Fala. Processamento textual. Microsoft. System.Speech. Síntese de texto. | | |
| **Campus**: São João Evangelista | **Tipo de Bolsa**: PIBIT | **Financiador**: IFMG |
| **Bolsista** (**as**): Daniela Couto de Oliveira | | |
| **Professora Orientadora**: Karina Dutra de Carvalho Lemos | | |
| **Área de Conhecimento**: Educação Inclusiva. Computação | | |

**Resumo**:

Nesse projeto realizou-se o desenvolvimento de uma ferramenta para conversão texto-fala capaz de obter informações textuais e transformar tais informações em áudio. A conversão texto-fala ou TTS (*Text-To-Speech*) são sistemas capazes de converter uma linguagem escrita em fala. Essa possibilidade de armazenar informações na forma escrita, fornecendo saídas através de voz, amplia o uso de computadores para diversas aplicações como acesso a bancos de dados através de telefone, sistemas de correio eletrônico em correio por voz, e principalmente a facilidades providas para pessoas com deficiências vocais e visuais, através de máquinas de auxílio à fala e máquinas de leitura para cegos. Diante do exposto, objetivou-se com o trabalho o desenvolvimento de uma ferramenta que permitiu a conversão de um texto irrestrito em fala, permitindo o armazenamento dos arquivos em áudio para que possam ser acessados posteriormente. O sistema foi desenvolvido utilizando a ferramenta de programação *Microsoft Visual Studio 2010 Professional,* a linguagem de programação C# e a biblioteca *System.Speech* responsável pelo processo de sintetize de texto em fala. Para concretização do processamento textual em áudio de forma eficiente realizou-se teste como finalidade de obter os principais resultados da eficácia das fundamentais operações do sistema. O sistema foi apto a processar voz contínua e retorna a resposta imediata pelo programa. Os comandos e suas funções são fornecidos pelo usuário, adequando-se a sua necessidade prática.

## INTRODUÇÃO

A utilização da linguagem escrita tem predominado em sistemas computacionais desde o surgimento dos primeiros computadores, por ser a forma eficiente ao transmitir e armazenar informações. (CHBANE, 1994). Porém, com o avanço da tecnologia e com a crescente tendência de interfaces homem-máquina amigáveis, o uso da linguagem falada em computadores torna-se cada vez mais conveniente, na medida em que a fala é uma forma mais natural de comunicação.

Dados estatísticos de Chaponis, citados por Young e Fallside (1989), mostraram que a comunicação da informação através da voz em situações de interação homem-máquina é em média duas vezes mais eficiente do que qualquer outra forma de comunicação. Ainda segundo eles, a maior eficiência da comunicação verbal, apoiada na crescente evolução das técnicas de processamento de sinais digitais, tem

feito com que sistemas de compreensão da fala e síntese de voz difundam-se cada vez mais como meios de entrada e saída de informações em computadores.

Chbane (1994) descreve que “Atualmente os sistemas de compreensão da fala estão restritos a algumas aplicações especiais, enquanto que os sistemas de síntese de voz vêm sendo largamente utilizados”. Tais sistemas apresentam particular importância, pois proporcionam a produção de voz a partir de um texto de entrada, unindo a eficiência do armazenamento de dados na forma escrita com a comunicação através da fala. Esses sistemas possuem diversas abrangências permitindo, por exemplo, acesso a bancos de dados através de telefone, como os sistemas de consulta de saldo bancário, atualmente bastante difundidos. Merece destaque ainda, as facilidades advindas do uso desses sistemas para pessoas com deficiências vocais e visuais, através de máquinas de auxílio à fala e máquinas de leitura para cegos, pois a maioria das interfaces entre computadores e humanos funciona através da linguagem textual, mas este tipo de interação para pessoas cegas torna-se um grande desafio.

Dentre os benefícios que a fala sintetizada apresenta destaca-se a naturalidade e agilidade de sistemas computacionais, a interação com sistemas de telecomunicações e a acessibilidade aos deficientes visuais. Considerando tais benefícios surge o seguinte questionamento: “Como essa interatividade torna-se possível já que a principal forma de interação homem-máquina é através da escrita?”. A resposta a esta indagações é o uso dos sistemas de conversão texto-fala ou TTS (*text-to-speech*). Luiz Latsch (2002, p. 22) afirma que “Um conversor TTS é um sistema que a partir de um texto irrestrito, produz fala sintetizada correspondente à leitura”.

Avanços consideráveis têm sido feitos nesta área de aplicações TTS, visando tornar estes sistemas cada vez mais próximos da fala natural. Leandro Gomes ressalta que “A meta principal destes sistemas é a maximização da inteligibilidade da fala sintetizada”. (GOMES, 1998, p. 4). Entretanto, existem algumas dificuldades no desenvolvimento de um sistema de conversão texto-fala que não estão diretamente ligadas às técnicas de conversão, pois surgem cada vez mais novas tecnologias e novas linguagens que permitem o desenvolvimento de *softwares* das mais variadas formas, porém esse fato pode dificultar a obtenção de informações de quaisquer sistemas computacionais, para então converter estas informações em fala.

Em virtude do exposto, o objetivo do trabalho foi criar uma ferramenta de conversão texto-fala, permitindo o armazenamento dos arquivos em áudio para que os mesmos possam ser acessados posteriormente. Para isto o sistema realiza processamento dos textos, digitados ou importados de dentro do sistema, em áudio em tempo de execução com a ajuda de um sintetizador de fala desenvolvido por terceiros, fornecendo a possibilidade do usuário salvar os textos, convertidos em forma de áudio para acessos posteriores. Os processamentos dos textos estão disponíveis nos idiomas inglês, espanhol e português.

**METODOLOGIA:**

Para realização da ferramenta de conversão de texto em fala, tornou-se necessário inicialmente realizar levantamento dos requisitos que a ferramenta envolveria. Para isso estabeleceu os requisitos funcionais (RF) e não funcionais (RNF) da ferramenta. Após a etapa de levantamento dos requisitos da ferramenta, realizou-se criação dos diagramas de caso de uso, com a finalidade de estabelecer as principais funcionalidades da ferramenta proposta. Para auxílio na implementação do sistema, realizou-se os diagramas

de classes, com a finalidade de identificar as principais classes que compõem a ferramenta e a forma como estas estão estruturadas e interligadas.

Posteriormente, iniciou-se o desenvolvimento do sistema utilizando a ferramenta *Microsoft Visual Studio* 2010 *Professional.* Inicialmente, tornou-se necessário a instalação da biblioteca *System.Speech.* Essa biblioteca possibilitou a transição de texto em fala, reconhecimento da gramática, manipulação da entrada de áudio, manipulação da saída de áudio, conversão de texto em áudio e fragmentação da entrada de áudio.

Após os procedimentos apresentados no parágrafo anterior, efetivou-se a instalação dos pacotes de voz que seriam interpretados pelo sistema. Como padrão, o pacote de voz do Sistema Operacional (SO) *Windows* é no idioma inglês. Os três pacotes de voz dos SO mais utilizados são: *Microsoft Zira Desktop, Microsoft Mike Desktop, Microsoft Ana Desktop.* Dentro estes pacotes de voz, a que apresenta melhor inteligibilidade em termos de transcrição texto-fala é o pacote *Microsoft Zira Desktop*, diante disso adotou-se a voz no sistema.

Para o processo de desenvolvimento da ferramenta de conversão de texto-fala foi necessário adicionar a referência *System.Speech* no projeto. Dentro do *namespace* responsável pelo sintetizador de voz, a classe fundamental para síntese é a *Speech. Synthesizer.* Dessa forma, tornou-se necessário instanciar a classe e utilizando seu método *Speak.* Tais procedimentos são responsáveis pela sintetize do texto em fala.

O sistema apresenta uma interface de acordo com padrões estabelecidas pelos padrões de *software*, com *menu* de informações necessários ao usuário. Dentro da aplicação, o usuário deverá inserir ou exportar o texto que deseja realizar a conversão de texto em fala.

Após a digitação do texto, o usuário acionará o botão de leitura e, com isso, o texto é capturado pelo sistema fonético em português, inglês, e espanhol. Nesse processamento, respeitam-se acentuações, espaços, vírgulas e a norma culta da língua portuguesa. Na Figura 1 abaixo é ilustrado a tela principal do sistema.

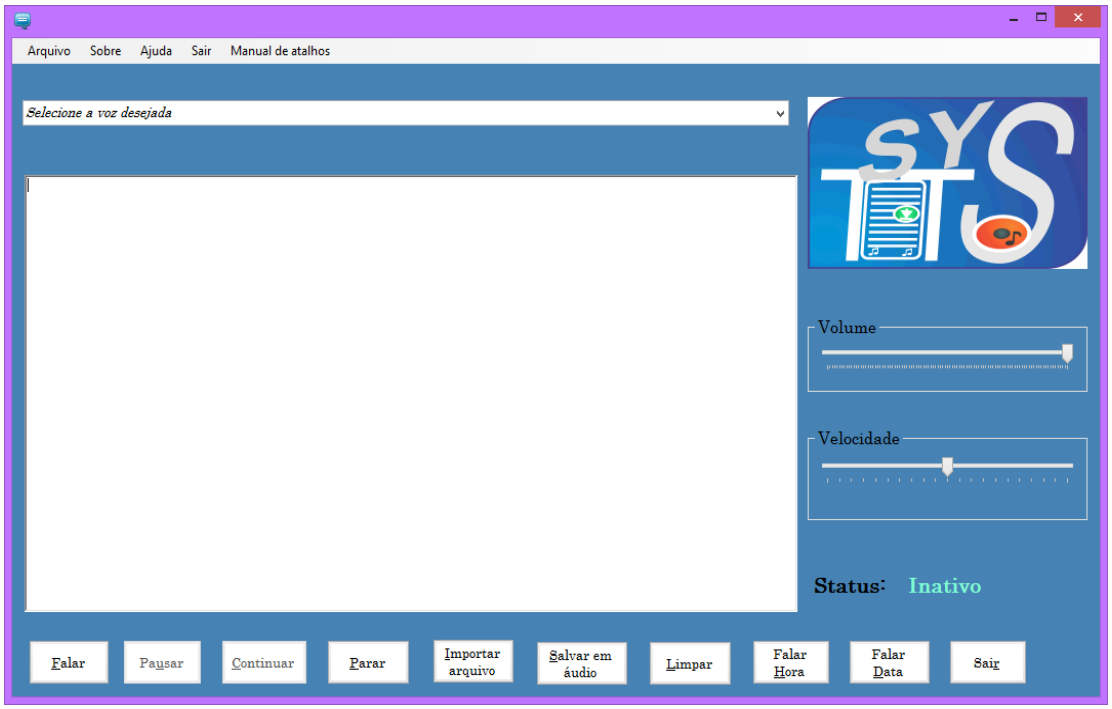


Figura 5 – Interface do sistema. **Fonte**: Própria do sistema

**RESULTADOS E DISCUSSÕES:**

A ferramenta desenvolvida conseguiu obter informações textuais e sintetizar estas informações conforme proposto. O fato da ferramenta utilizar-se de tecnologias de síntese de voz desenvolvidas por terceiros possibilitou a utilização de tecnologias mais apuradas, que tornaram a pronúncia mais natural, tanto no idioma português como no idioma inglês e espanhol. Para transmissão da fala, instalaram-se inúmeros pacotes de vozes empregadas apenas para testes de segmentação. Porém essas não apresentaram bons resultados quando a inteligibilidade, coerência e pronúncia para transcrição dos textos no idioma português brasileiro. Diante disso optou-se pela voz *Nuance Raquel Brasilian Female.* A maioria das dificuldades enfrentada nesse procedimento foi fazer com que SO reconhecesse a voz instalada.

O processamento textual aplicado pela ferramenta mostrou-se capaz de melhorar o resultado final da leitura. Nesse processo, o sistema apresenta diferenciação na entoação da pontuação, ou seja, há uma diferenciação na sintetize de um ponto de interrogação e de ponto final.

Para concretização do processamento textual em áudio de forma eficiente realizou-se teste como finalidade de obter os principais resultados da eficácia das fundamentais operações do sistema. Nestes testes procurou-se obter dados sobre aproveitamento da ferramenta quanto a leitores dos textos, a inteligibilidade das vozes no idioma português brasileiro, a execução dos atalhos nas interações com *software* leitores de tela, o processamento textual em diversos aspectos entre outros testes. Os dados destes testes estão descritos no Quadro 1.

| Teste Realizado | Nº de tentativas | % de aproveitamento |
|---|---|---|
| Leitura de textos | 40 | 100 |
| Inteligibilidade das vozes em português | 50 | 95 |
| Execução dos atalhos | 35 | 100 |
| Processamento de monossílabas | 50 | 98 |
| Processamento de palavras curtas | 50 | 96 |
| Processamento de palavras longas | 50 | 90 |
| Processamento de palavras múltiplas | 50 | 100 |
| Processamento de pause, resume e parar | 50 | 100 |
| Processamento simultâneo de texto | 30 | 50 |
| Processamento de textos em áudio | 50 | 100 |
| Interação como *software* leitores de tela | 30 | 75 |
| Importação de arquivo.txt | 30 | 95 |
| Exportação de arquivo gerando em áudio com mesma formatação do processamento textual | 30 | 100 |

Quadro 1 – Teste com o sistema. **Fonte**: Elaborado pelos autores

Os testes mostraram que o sistema tem performance aceitável no processamento de voz, na geração dos arquivos em áudio, no processamento textual em todas as amplitudes. Porém, quanto ao

processamento de texto simultâneo, confirmando, o que já afirmado anteriormente, a ferramenta não se mostrou apta. Além de atender a um maior número destas características, a ferramenta desenvolvida destaca-se pela qualidade da voz sintetizada e pela sua capacidade de gerar os arquivos processados em áudio.

**CONCLUSÕES:**

Os resultados obtidos ao término do projeto são satisfatórios. A ferramenta foi capaz de realizar a síntese dos textos em fala em tempo real nos idiomas português, inglês e espanhol, aumentando assim sua utilidade prática. O estudo dos sistemas de conversão texto-fala possibilitou um melhor entendimento de como estes sistemas devem funcionar e como limitar o escopo da ferramenta desenvolvida.

O desenvolvimento da ferramenta tornou-se possível através da utilização da biblioteca *System.Speech*. Esta biblioteca foi capaz de prover todos os recursos necessários para a síntese de voz. Os objetivos propostos foram atendidos visto que a ferramenta é capaz de solicitar ao usuário qual voz ele deseja para realizar o processamento textual. É possível também obter informações textuais externos através da importação de arquivo .txt e armazenar os arquivos em áudio para que os mesmos possam ser acessados posteriormente.

Em relação às técnicas de programação, é sugerido que o sistema seja rigorosamente analisado, conforme uma técnica específica UML (*Unified Model Language*), permitindo melhor documentação, modularização, atualização, manutenção e outras vantagens que um sistema bem documentado e estruturado permite.

Diversos estudos podem ser realizados futuramente como estudo mais aprofundado a respeito da modelagem da entonação e implementação de um modelo de geração automática da curva da frequência fundamental. Para isso também será necessário realizar a implementação de um analisador morfológico e sintático para enfatizar ou não grupos de palavras e demarcar as fronteiras prosódicas. Também será necessário um estudo da extração da semântica do texto, dado que a entonação é uma característica que acompanha a intenção do autor ao falar a frase, ou de algum sentimento (entusiasmo, decepção entre outros) que ele queira passar juntamente com a informação.

Após estudos acera sobre sistemas TTS, conclui-se que a ferramenta desenvolvida pode ser aprimorada envolvendo modelagem prosódica, maior naturalidade da fala e inserção de outros idiomas. Alguns pontos da ferramenta podem ser melhorados e outras funcionalidades ainda podem ser implementadas em projetos futuros.

**REFERÊNCIA BIBLIOGRÁFICA:**

CHBANE, Dimas T. **Desenvolvimento de um sistema para conversão de textos em fonemas no idioma português**. 1994. 125 f. Dissertação (Mestrado em Engenharia) - Escola Politécnica, Universidade de São Paulo, São Paulo. Disponível em: <http://www.linodecampos.net/textos/disdimas.pdf>. Acesso em 15 de Janeiro de 2013.

YOUNG,S.J.; FALLSIDE,F. Speech Synthesis from Concept: A Method for Speech Output from Information Systems. **Journal of Acoustical Society of America**, v. 66, n. 3, p. 685-95, Sept. 1989.

DUARTE, Mauricio; UZAI, Gustavo. **Sistema de Reconhecimento de Voz – Aplicabilidade.** 2008.

GOMES, Leandro C. T. **Sistema de conversão texto-fala para a língua portuguesa utilizando a abordagem de síntese por regras**. 1998. 107 f. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas. Disponível em < http://www.bibliotecadigital.unicamp.br/document//?code=vtls000132102&fd=y > Acesso em:16 de Março de 2013.

LATSCH, Vagner L. **Um sistema de conversão de texto-fala para Windows.** 2002. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) – Escolas Engenharia, Departamento de eletrônica e de computação, Universidade Federal do Rio de Janeiro. Disponível em: <www02. Lps.ufrj.br/~sergioln/theses/dsc05vagnerlatsch.pdf>. Acesso em 20 de outubro de 2012.