

Proposta de Implementação de um Sistema Texto-Fala (TTS) Personalizado

Rodrigo Carpegianni, Maximiliano Zambonato Pezzin

Curso de Sistemas de Informação – Universidade do Contestado (UnC)
Rua Victor Sopena, 1000, Bairro Salete – 89.700-000 – Concórdia – SC – Brasil

r_carpegianni@yahoo.com.br; ermanopezzin@yahoo.com.br

Abstract. *In this Work we will demonstrate the possibility of if developing a System based on the Synthesis of Fala (TTS) personalized, for the job of this technique in the solution of problems of our daily one, as well as making possible the application of this technique in new contexts. In this work the viability of development of personalized system, the necessary requirements, the criteria to be will be boarded boarded, and parameters to be reached. This System will work with a bank of voice of the one of the final user, and also it would continue with the idea to make with that the same it groups automatically fonemas basic, necessary to the formation of words and phrases, in intention to reproduce a text, in form of sound.*

Resumo. *Neste Trabalho demonstraremos a possibilidade de se desenvolver um Sistema baseado na Síntese da Fala (TTS) personalizado, para o emprego desta técnica na solução de problemas do nosso cotidiano, bem como viabilizando a aplicação desta técnica em novos contextos. Neste trabalho será abordada a viabilidade de desenvolvimento de sistema personalizado, os requisitos necessários, os critérios a serem abordados, e parâmetros a serem alcançados. Este Sistema trabalhará com um banco de voz do usuário final, e também prosseguiria com a idéia de fazer com que o mesmo agrupe automaticamente os fonemas fundamentais, necessários à formação de palavras e frases, no intuito de reproduzir um texto, em forma de som.*

1. Introdução

Devido ao grande potencial, mencionado e comentado por diversos estudiosos e grupos de pesquisa, as técnicas de text to speech (TTS), vem a algum tempo atraindo interesses, não somente de Universidades (Grupos de Pesquisa), mas também de grandes empresas. Diversas técnicas já foram desenvolvidas e aplicadas, sempre procurando encontrar um modelo que resulte em melhores resultados, ou qualidade do som reproduzido.

2. Banco de Unidades de Texto-Fala

A diferença de um texto escrito para um texto falado, são duas, a primeira reside no fato de que o texto escrito é apresentado ao leitor pronto e acabado, enquanto o texto oral é construído na presença do ouvinte. De fato, o texto escrito, em sua forma definitiva, não deixa pistas de seu processo de construção, e o texto oral, por ter um caráter improvisado, permite ao interlocutor assistir a todo o seu processo de criação.

2.1. Técnicas Abordadas

Serão abordadas as técnicas de pré-definição de palavras chaves, definição de fonemas básicos (mais utilizadas), concatenação de fonemas que constituirão o BD, e estas seriam armazenadas em forma de arquivos WAV, para que posteriormente sejam utilizadas na formação de palavras e frases “genéricas”.

2.2. Geração de Banco de Unidades

O sistema proposto neste trabalho tem por objetivo a junção de termos básicos com a finalidade de criar frases e palavras.

Para isso, o sistema proposto seria dividido em etapas, sendo a primeira à constituição do banco de voz e reprodução de áudio, como é demonstrado com a figura a seguir, onde a constituição do banco de unidades de voz é do próprio usuário, sendo necessário que o usuário fale cada uma das palavras, e também, realize o fracionamento e armazenamento, de cada termo fracionado, isto de acordo com as regras escolhidas para utilização no processo de reprodução do TTS.

2.3. Fracionamento

Destacando, que a opção pela realização do fracionamento, que é parte do trabalho por nós proposto, será executado de forma manual, o que proporciona algumas desvantagens, com relação ao tempo e também o volume de trabalho empregado nesta tarefa, mas apresentará vantagens significativas relacionadas à qualidade final da palavra reproduzida, independente de entonação, velocidade de fala e questões relacionadas a sotaques de linguagem.

2.4. Critérios

Os critérios utilizados no desenvolvimento do Banco de Unidades dependem das técnicas que serão empregadas e da qualidade em si que se pretende alcançar (Braga 2003). A princípio, este sistemas de TTS tem o propósito de gerar a sua componente oral, o que emerge à necessidade de estreitar a relação entre o conhecimento sobre as línguas fornecido pela lingüística e o seu processamento e integração num sistema informático pela engenharia de software.

3. Processamento de Texto

O princípio de um sistema de conversão de texto-fala, parte de um texto escrito, este deverá interpretar o “texto” selecionado, e após através de regras e procedimentos introduzidos no sistema, converter o mesmo em voz, com semelhança o mais próximo do natural do ser humano. (Braga 2003)

4. Regras de Conversão

A constituição das línguas se dá por três tipos de sons, de acordo com as características da sua produção: consoantes, vogais e semivogais. Cada língua seleciona uma lista de sons que, são combinados segundo regras e de acordo com certos contextos (Braga 2003).

Em português existem 19 fonemas consonânticos (Mateus 1975), e mais 18 consoantes gráficas, isto nos mostra que uma transcrição direta de letras para sons se torna um pouco difícil.

5. Teste Realizado com Algoritmo Experimental (Software Experimental)

Através de testes realizados com um algoritmo experimental, podemos considerar o resultado viável, com relação ao processo de desenvolvimento ou geração de arquivo. Visto que o tempo inferior a meio segundo possibilitará a reprodução contínua e ininterrupta do texto analisado, é claro que quando chegamos ao ponto “qualidade” de reprodução do texto, nos deparamos com a necessidade de exclusão e regravação de alguns termos.

6. Conclusões

Atualmente as dificuldades para se fazer um sistema TTS confiável, e que tenha um funcionamento razoável, podemos dizer que é um pouco difícil ou complicado, isto por que nos deparamos com diversas dificuldades antes não percebidas ou talvez não imaginadas, como por exemplo, o povoamento do BD, este realizado pelo próprio usuário, também o fracionamento e armazenamento dos arquivos WAV, e não esquecendo as dificuldades encontradas na própria língua portuguesa, que é bastante complexa.

Referências

- Braga, Daniela (2003), Processamento Lingüístico Aplicado à Síntese da Fala, Instituto Politécnico do Porto, Portugal.
- Kafka, Sandra G. (2002), Utilização de Segmentos Transacionais Homorgâmicos em Síntese de Fala Concatenativa, Laboratório de Circuitos e Processamento de Sinais UFSC <http://linse.ufsc.br>.
- Seara, Rui (2003), Parâmetros Lingüísticos Utilizados para a Geração Automática de Prosódia em Sistemas de Síntese de Fala, Laboratório de Circuitos e Processamento de Sinais UFSC <http://linse.ufsc.br>.
- Yeon-Jun Kim (2002) Automatic Segmentation Combining An HMM-Based Approach And Spectral Boudary AT&T Labs, USA, <http://tcts.fpms.ac.be/synthesis>
- Conkie, Alistair; Beutnage, Mark; Syrdal, Ann (2002), Preselection of Candidate Units in a Unit Selection-based TTS Synthesis System, AT&T Labs, NJ, USA <http://www.research.att.com/projects/tts>
- Syrdal, Ann K. and Alistair D. Conkie, Data-Driven Percentually Based Join Costs, AT&T Labs – Research, Florham Park, USA