Sistema de Reconhecimento de Voz – Aplicabilidade

Luis Gustavo de Carvalho Uzai
Prof. Mauricio Duarte

Tecnologia em Informática para a Gestão de Negócios
Faculdade de Tecnologia de Garça (Fatec)
Caixa Postal 17400-000 – Garça - SP – Brasil

uzai_ff@hotmail.com

maur.duarte@gmail.com

Abstract. *This article describes the use of voice recognition in the implementation of routines. The system is able to process continuing voices and it returns the immediate response through the program. The commands and its functions are provided by the users, adjusting them to their practical needs. The system was developed for the Windows environment, using C #. NET language, a tool that is part of Microsoft Visual Studio 2010 Express Edition and the complementary package Microsoft SpeechSDK 5.1.*

Resumo. Este artigo descreve o uso do reconhecimento de voz na aplicação de rotinas. O sistema é 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. O sistema foi desenvolvido para ambiente Windows, utilizando a linguagem C#.NET, ferramenta que faz parte do Microsoft Visual Studio 2010 Express Edition e o pacote complementar Microsoft SpeechSDK 5.1.

1.Introdução

As interfaces homem-computador estão cada vez mais complexas e robustas, necessitando cada vez mais de melhores dispositivos de hardware e uma integração de software que seja de igual teor de inovação. Para estas interfaces deverão ser criados modos de interação de fácil operação, já que por meio do recurso de fala o usuário poderá ter maior liberdade para execução de outras tarefas que exijam a manipulação de entradas de forma convencional e ainda, com o uso de sistemas de síntese de fala o usuário poderá receber informações de forma direta e objetiva.

Há uma constante necessidade de sistemas para melhorar a usabilidade de um computador, como por exemplo, facilitar tarefas, agilizar operações ou melhorar a qualidade de determinados eventos efetuados com um computador. Sistemas são desenvolvidos freqüentemente com objetivo de cumprir as mais variadas tarefas requisitadas pelo usuário, porém, a grande maioria demanda um tempo para que aprendam a utilizá-los, o que para usuários inexperientes pode se tornar uma tarefa penosa e, mesmo para usuários experientes, alguns comandos poderiam se tornar mais práticos e eficazes por um sistema direto de recepção por comandos de voz, por exemplo.

Em grande parte, usuários portadores de necessidades especiais encontram dificuldades em operar o computador. Um dos objetivos deste projeto é o desenvolvimento de um sistema para auxiliar o manuseio de um computador, trazendo facilidades ao executar diversas funções apenas pela entrada de áudio através do microfone. Espera-se com isso a economia de tempo para o usuário. Outro objetivo do sistema é a possibilidade da leitura de arquivos do tipo texto convertendo-os para voz.

O principal foco do projeto é o estudo do reconhecimento de voz através do computador. O projeto visa especificar uma plataforma de hardware e implementar um software para síntese, identificação e reconhecimento de voz, baseado nas técnicas de processamento digital de sinais a partir de dados estatísticos e verificação, para isso é necessário realizar um estudo das técnicas de reconhecimento de padrões de voz. A síntese funciona como uma transformação de um texto em som, enquanto o reconhecimento faz o papel inverso. A leitura de texto deve ser de forma automática e para qualquer texto. O sistema deve ser capaz de realizar duas etapas: primeiro, fazer uma análise do texto para saber o que ele deve falar e, segundo, fazer a produção do som propriamente dito.

> As interfaces via voz estão rapidamente se tornando uma necessidade. Em um futuro próximo, sistemas interativos irão fornecer fácil acesso a milhares de informações e serviços que irão afetar de forma profunda a vida cotidiana das pessoas. Hoje em dia, tais sistemas estão limitados a pessoas que tenham acesso aos computadores, uma parte relativamente pequena da população, mesmo nos países mais desenvolvidos. São necessários avanços na tecnologia de linguagem humana para que o cidadão médio possa acessar estes sistemas, usando habilidades de comunicação naturais e empregando aparelhos domésticos, tais como o telefone.
> Sem avanços fundamentais em interfaces voltadas ao usuário, uma larga fração da sociedade será impedida de participar da era da informação, resultando em uma maior extratificação da sociedade, agravando ainda mais o panorama social dos dias de hoje. Uma interface via voz, na linguagem do usuário, seria ideal

> pois é a mais natural, flexível, eficiente, e econômica forma de comunicação humana. (YNOGUTI, 1999)

O projeto demonstra a criação de um sistema de suporte para reconhecimento de voz, possibilitando respostas instantâneas, substituindo comandos realizados manualmente por automatizados. O presente trabalho concentra-se prioritariamente na descrição das ações visando o desenvolvimento dos recursos necessários para implantação.

Como resultado, será realizada uma análise do tempo de resposta dos comandos executados pelo sistema, a fim de avaliar o aproveitamento do sistema ao reconhecer diferentes padrões de voz. Também será avaliado o desempenho do sistema ao executar o sistema de reconhecimento, determinando assim, as taxas de erros, necessárias na otimização do projeto.

Para o desenvolvimento do sistema vários recursos de softwares foram necessários. Entre eles, pode-se destacar: o uso de classes importadas do sistema que tem integração direta com o pacote de aplicativos de reconhecimento SpeechSDK; utilização de um pacote externo para desenvolvimento da interface do software e, as outras funções internas à interface de desenvolvimento Microsoft Visual Studio.

2. Desenvolvimento do Sistema

O sistema foi desenvolvido no mais recente software de desenvolvimento .NET da Microsoft, a ferramenta Microsoft Visual Studio 2010 Express Edtion, possibilitando o uso do NET Framework 4.0 que possui maior interação com uso de softwares externos a ferramenta principal de desenvolvimento e a possibilidade de desenvolvimento de aplicativos em paralelo, facilitando a programação. O pacote principal para o recurso de reconhecimento de voz utilizado foi Speech SDK 5.1. O pacote oferece recurso para transição de texto para fala e de reconhecimento de voz e é estruturado da seguinte forma: na criação de uma gramática, reconhecimento da gramática, manipulação da entrada de áudio, manipulação da saída de áudio, conversão de texto em áudio e fragmentação da entrada de áudio.

Como o padrão de voz do Windows é no idioma inglês (Mike), ele precisou ser substituído para acoplar o reconhecimento de palavras da língua portuguesa, assim, o

pacote de voz Raquel (Raquel Voice Brazilian Portuguese) foi utilizado. Ele permite que as respostas por comandos sejam precisas e entendíveis de forma clara e reduzindo o delay (atraso de som em transmissões) a praticamente zero.

O sistema apresenta uma interface amigável, tendo como início uma tela de boas vindas, configurável pelo usuário, que captura o horário no momento do login, a data, o nome do usuário (informações colhidas pelo registro do sistema operacional), e uma saudação definida pelo usuário (por exemplo : Bom Dia) para compor as configurações do sistema desenvolvido. O sistema foi dividido em vários componentes, cada um correspondendo a um determinado tipo de operações.

Implementou-se uma classe para o reconhecimento das palavras composta por vários métodos. O primeiro método recebe por parâmetros a palavra chave (que é a palavra que deverá ser pronunciada) e o comando que será executado (linha de comando que define a ação a ser executada quando a palavra chave for pronunciada), ambos fornecidos pelo usuário. Os próximos métodos implementam o suporte necessário para inicializar a fala, analisar as hipóteses de contexto, reconhecer o contexto, reconstruir a gramática com base no contexto, reconhecer comparar se a entrada do usuário é a mesma pré-programada para executar a ação, disparar o comando inserido previamente pelo usuário e desabilitar o processo de reconhecimento de fala, para evitar redundância.

O software do sistema de conversão de texto em áudio foi desenvolvido na terceira opção do menu principal, indicada por *leitura*. O usuário deverá inserir o texto que deseja converter no recipiente de texto exposto. Após a digitação do texto, o usuário acionará o botão de leitura e, com isso, o texto é capturado pelo sistema fonético Raquel em português, respeitando acentuações, espaços, vírgulas e a norma culta da língua portuguesa e convertido para voz. Foi implementado um botão para interromper o processo de fala, caso o usuário julgue necessário. Também é possível pelo sistema importar arquivos no formato de texto para o recipiente e, repetindo-se o processo de leitura, o sistema poderá gerar um arquivo de áudio (.wav) e o sistema garante que o arquivo de áudio gerado tenha as mesmas propriedades do texto lido no recipiente.

O sistema também contempla uma opção onde o usuário poderá realizar as configurações desejadas, como por exemplo: a tela de boas vindas e os comandos padrões

no sistema. Na opção de "*Design*" o usuário poderá configurar a interface de interação do sistema. O sistema ainda possui uma opção de "*Ajuda*" onde o usuário poderá esclarecer algumas de suas dúvidas sobre o sistema.

3. Interface do Sistema

A interface do sistema foi baseada no padrão do pacote de aplicativos Microsoft Office 2007 implementada com o auxílio da ferramenta de desenvolvimento Krypton (Krypton Toolkit 4.3). O sistema foi dividido em seis menus, nomeados respectivamente: Inicio, Comandos, Leitura, Configuração, Design, Ajuda, separados de acordo com suas funções.

Na figura 1 é representado a segunda opção do menu do sistema, "*Comandos*", onde o usuário pode definir as palavras chaves a serem inicializadas por comandos de voz, assim como a os comandos que deverão ser executados. Através do botão *adicionar*, a instrução é armazenada e, marcando-se a opção *ativar*, o respectivo comando definido ficará sempre ativo. Os comandos são instruções pré definidas, também acionadas por palavras chaves, as instruções serão selecionadas pelo usuário, que deve escolher entre as já pré-programadas, por exemplo : Comando "*Ctrl + C*" para a palavra chave "*Copiar*".

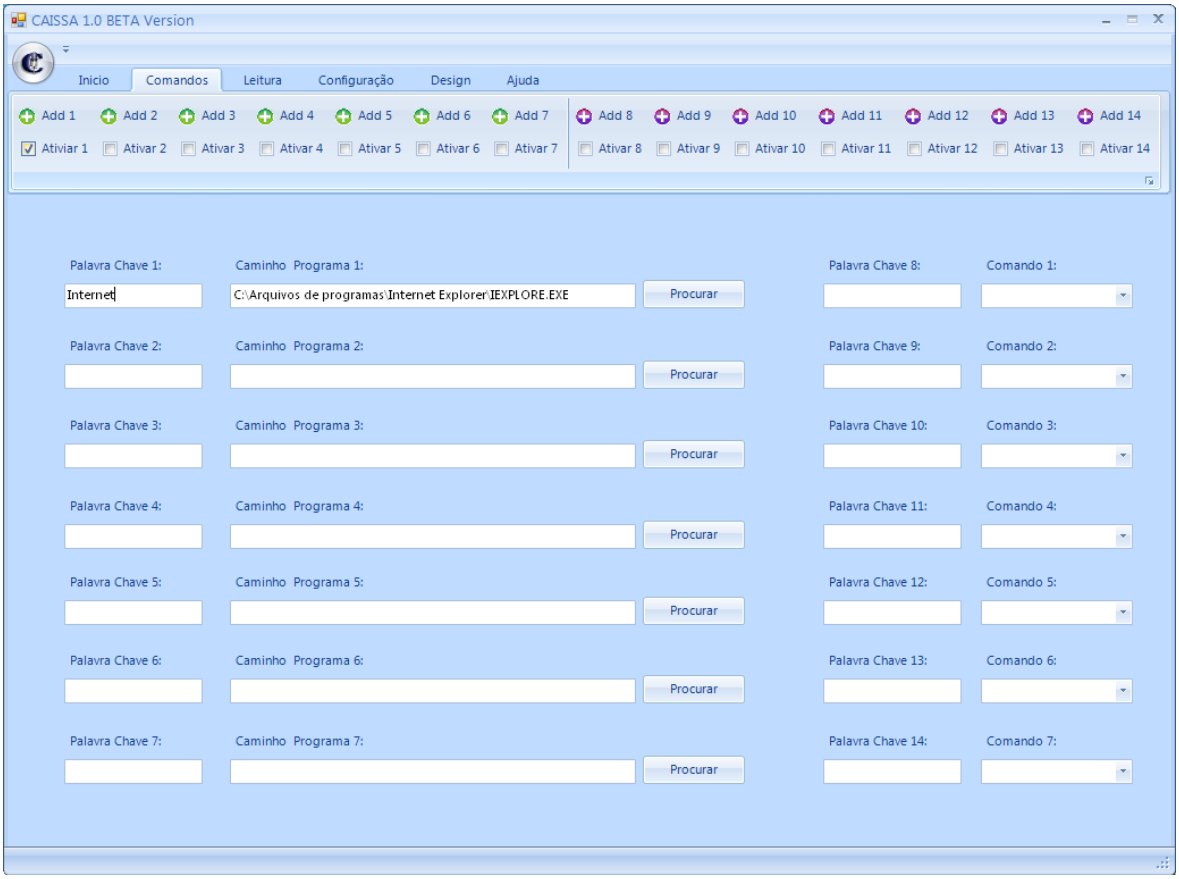

Figura 1. Software CAISSA 2º aba Comandos

Na figura 2 é representado a terceira opção do menu, a *Leitura*, por onde o usuário pode digitar um texto para ser lido pelo sistema (botão Ler), parar a execução da leitura (botão Parar) ou importar um arquivo de texto (botão Procurar) com a mesma finalidade. Também é possível exportar a leitura em um arquivo de áudio .wav (botão Exportar), o recipiente de texto só é visível nesta opção.

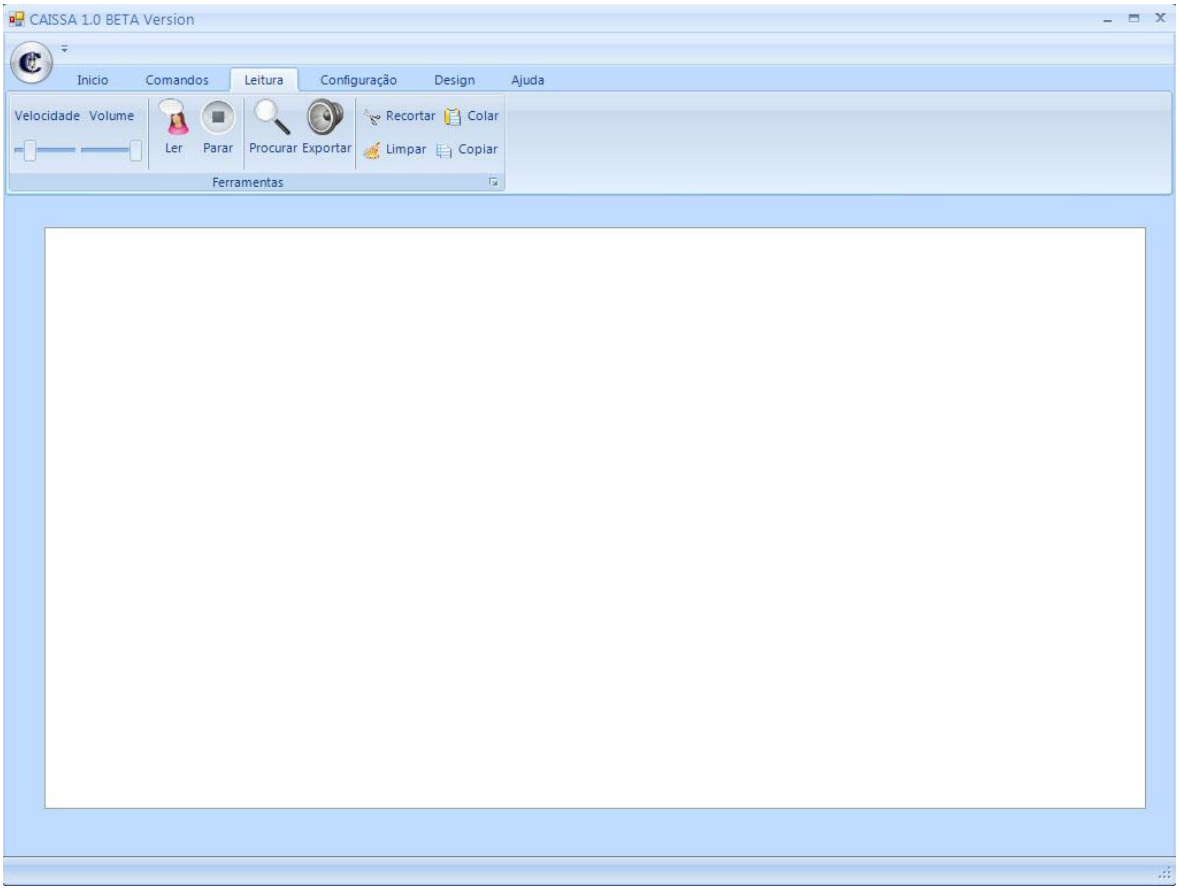

Figura 2. Software CAISSA 3º aba Leitura

A interface do sistema é configurável, na figura 3 é representado a quinta opção do sistema, a de *Design*, nesta opção o usuário tem a possibilidade de configurar o visual do sistema, selecionando uma das nove opções possíveis, respectivamente: Padrão Azul, Padrão Prata, Padrão Negro, Moderno Azul, Moderno Prata, Moderno Negro, Faísca Negro, Clássico Azul e Sistema. A configuração escolhida pelo usuário se torna a padrão e é alterada em todas as opções do sistema.

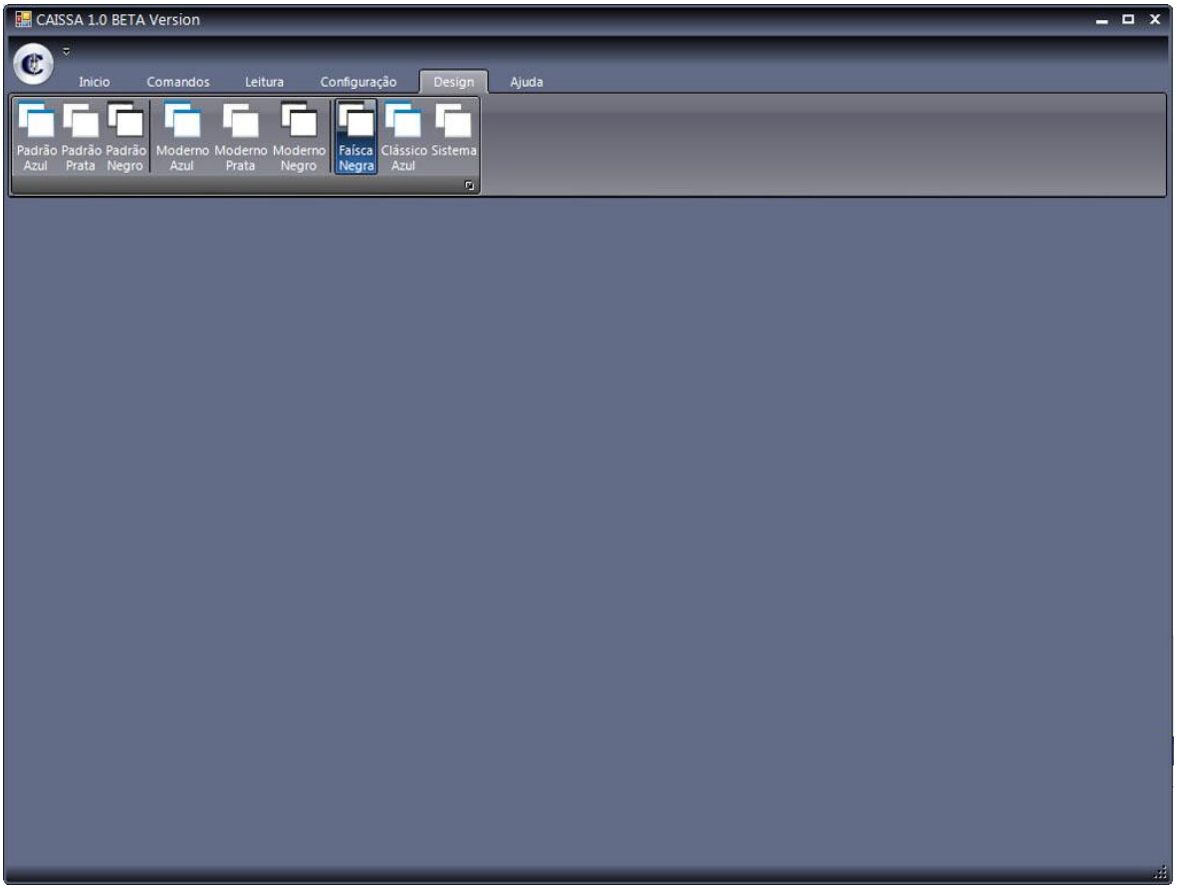

Figura 3. Software CAISSA 6º aba, Design – Faísca Negra

4. Resultados

Teste Realizado	Nº de tentativas	% de aproveitamento
Leitura de textos	50	100
Execução dos Atalhos	32	100
Reconhecimento de monossílabas	100	98
Reconhecimento de palavras curtas	100	96
Reconhecimento de palavras longas	100	90
Reconhecimento de palavras múltiplas	100	84
Reconhecimento de palavras simultâneas	100	72
Duplicidade no Reconhecimento	100	63

Os testes deixaram mostraram que o sistema tem performance aceitável no reconhecimento de voz, para execução de suas rotinas, porém o sistema não obteria desempenho necessário para se tornar aplicável em um sistema de ditar a voz, pela alta taxa de erro na duplicidade de palavras (37% aproximados) e taxa de falha no reconhecimento de palavras simultâneas (28% aproximados) , porém não apresentando queda no desempenho sendo utilizado na língua portuguesa, em relação ao mesmo sistema em língua inglesa, em quesito de reconhecimento para execução de tarefas, onde palavras individuais são usadas para acionar os comandos o sistema se mostrou otimizado.

5.Conclusão

O presente trabalho abordou o desenvolvimento de uma série de recursos tendo em vista a criação de um sistema de reconhecimento de fala como também conversão de texto em áudio, em português brasileiro. A identificação de palavras deste idioma foi possível graças ao sistema de Gramática e Fala substituto Raquel, sem perda de aproveitamento no reconhecimento como na conversão de texto em áudio.

Na confecção da base de dados pôde-se perceber que, em fala contínua, mesmo sendo produzida a partir da leitura de um texto, as coarticulações são bastante fortes.
Ainda, a variação de pronúncia e de ritmo de uma mesma palavra devido ao sotaque, nível de educação, e outros fatores é bastante grande. Todos estes fatores contribuem para tornar mais difícil o problema de reconhecimento de fala contínua com independência do locutor. Neste sentido, as técnicas de adaptação ao locutor são de grande importância no sentido de minimizar a amplitude destas variações para os sistemas de reconhecimento.

Foi constatada a possibilidade de um sistema de reconhecimento em português, sem perda de desempenho, a taxa de erro se manteve a mesma do sistema com gramática em inglês (media de 37%), o uso de palavras chaves semelhantes, pode aumentar significantemente a taxa de erro, a conversão de texto em áudio, teve aproveitamento plenamente satisfatório, considerando a norma culta da língua portuguesa.

O sistema é um protótipo que obteve com sucesso o desempenho necessário na aplicação das rotinas programadas, baixa taxa de delay e totalmente em português. Com base nesse protótipo será desenvolvido um sistema gerencial de apoio a decisão, analisando a estrutura utilizada para aprimorar o processo de reconhecimento, criando um

sistema com maior contato com o usuário, mais fácil acesso e de igual confiabilidade, desenvolvendo um sistema que auxilie de forma significativa o processo gerencial.

Referências Bibliográficas

CodeProject (2010).http://www.codeproject.com.

Microsoft (2010). http://msdn.microsoft.com/pt-br/library/

Microsoft (2010) “Microsoft Speech SDK API” .http://www. microsoft.com/downloads/en /details.aspx?FamilyID=5e86ec97-40a7-453f-b0ee-6583171b4530&displaylang=en

Microsoft (2010). http://social.msdn.microsoft.com/forums/pt-br/vscsharppt/.

YNOGUTI, Carlos Alberto (1999). Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov. Universidade Estadual de Campinas. pagina 1