Universidade Federal de Santa Catarina
Programa de Pós-graduação em
Engenharia de Produção

A UTILIZAÇÃO DO MICROSOFT SPEECH SDK PARA O RECONHECIMENTO DE VOZ

Yuri Ademir Bernardi

Florianópolis
2003

A UTILIZAÇÃO DO MICROSOFT SPEECH SDK PARA O RECONHECIMENTO DE VOZ

Universidade Federal de Santa Catarina
Programa de Pós-graduação em
Engenharia de Produção

A UTILIZAÇÃO DO MICROSOFT SPEECH SDK PARA O RECONHECIMENTO DE VOZ

Yuri Ademir Bernardi

Dissertação apresentada ao
Programa de Pós-Graduação em
Engenharia de Produção da
Universidade Federal de Santa Catarina
Como requisito parcial para obtenção
Do titulo de Mestre em
Engenharia de Produção

Florianópolis
2003

Yuri Ademir Bernardi

A UTILIZAÇÃO DO MICROSOFT SPEECH SDK PARA O RECONHECIMENTO DE VOZ

Esta dissertação foi julgada e aprovada para obtenção do grau de Mestre em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

Florianópolis, 17 de setembro de 2003

_____

Prof. Edson Pacheco Paladini, Dr.

Coordenador

BANCA EXAMINADORA:

_____

Profa. Edis Mafra Lapolli, Dra.

Orientadora

_____

Profa. Ana Maria B.Franzoni, Dra.

_____

Profa. Lia Caetano Bastos, Dra.

Agradecimentos

Ao Senhor Deus, por todas as coisas que percebemos e por aquelas que também não percebemos, pelos meus pais, pela família, e pelo trabalho, obrigado pela vida.

Aos meus pais, Ademir e Marilda pelo esforço, dedicação, educação e, ensinamentos que me fizeram com que eu vencesse meus desafios.

A minha esposa Aliciane, que me incentivou e foi uma grande aliada neste desafio, e que soube entender os momentos que estive ausente da sua companhia para realizar este trabalho, te amo.

Ao meu irmão Victor, que sempre me apoiou e ajudou nas horas difíceis e alegres.

A meus cunhados Viviane e Jorge que me ajudaram a ter melhores condições para realização do meu estudo.

Agradeço também a todos os amigos que compartilharam comigo os bons e maus momentos durante este período em Florianópolis e Blumenau.

A professora Edis pela orientação do trabalho e por todas as oportunidades de desenvolvimento que me proporcionou durante o período do mestrado.

Sumário

Lista de Figuras

Lista de Tabelas

Resumo

A necessidade do ser humano por novos meios de comunicação, tanto entre seus semelhantes, quanto na relação do homem com a máquina, nos traz novos desafios. Este trabalho avalia o “kit” da Microsoft de reconhecimento de fala. Esta ferramenta propõe facilidades para projetos que necessitam utilizar o reconhecimento de voz.

Neste estudo são abordados vários conceitos, podendo assim o leitor ter uma boa conceituação sobre os tópicos descritos. Os principais assuntos relacionados na revisão bibliográfica são o reconhecimento de padrões, inteligência artificial e a voz.

Finalmente apresenta-se o Microsoft Speech SDK e suas características. Após são apresentadas as avaliações e conclusões obtidas sobre o uso da ferramenta utilizada.

Palavras Chaves: Voz, Reconhecimento de fala, Microsoft Speech SDK, Reconhecimento de padrões.

Abstract

The human being need for new communication means so much among its fellow creatures as in the relationship man machine this brings us new challenges. This work evaluates the kit of Microsoft of speech recognition. This tool proposes means for projects that need to use the voice recognition.

In this study several concepts are approached, being able to not like this the reader to have a good concept on the described topics. The main subjects related in the bibliographical revision are the recognition of patterns, artificial intelligence and the voice.

Finally it introduces it Microsoft Speech SDK and its characteristics. After healthy presented the evaluations and conclusions obtained on the use of the used tool.

Key words: voice, speech recognition, Microsoft Speech SDK, recognition of patterns.

1 INTRODUÇÃO

1.1 Contextualização

Os crescentes avanços tecnológicos fazem com que o reconhecimento de fala seja um campo de estudos fascinante e ao mesmo tempo desafiador, em que através do uso da voz, busca-se extrair da fala as informações relevantes para a realização do reconhecimento.

Atualmente o campo de pesquisa envolvendo reconhecimento de fala é muito grande, envolvendo várias áreas, tais como jogos, editores de texto, uso como extensão de habilidades para deficientes físicos, atividades domésticas simples, tais como ligar e desligar aparelhos, sistemas de reconhecimento de fala para possibilitar diálogos inteligentes com a máquina, e uso em pesquisas médicas sobre a voz e órgãos relacionados, sistemas de segurança com utilização de sons e reconhecimento de voz.

1.2 Justificativa e Importância do Trabalho

Existem idéias de projetos que envolvem reconhecimento todos os dias. Estes trabalhos em geral são iniciados com a finalidade de se desenvolver teses, monografias ou dissertações. Com o desenvolver destes trabalhos pode-se observar que a iniciativa através dos métodos científicos para soluções que envolvem o reconhecimento ou identificação de objetos de qualquer natureza são grandes.

Aqui será analisada a proposta de uma nova tecnologia que está sendo oferecida como um software auxiliar a outros que permite aos desenvolvedores de software abstraírem-se da complexidade do reconhecimento de fala. Desta forma criam-se novas oportunidades para as tecnologias que buscam novas alternativas nesta área. Com isto tem-se o ganho de facilidades de desenvolvimento, redução de custos em projetos, flexibilidade em mudanças, adaptações de projetos e a praticidade de uma ferramenta maleável.

Após o levantamento de informações sobre os produtos disponíveis no mercado envolvendo reconhecimento de fala, tais como o Voice Pilot da empresa Americana Microsoft, o Voice Blaster da empresa Covox, e do VoiceType da IBM, e suas características, pode-se concluir que ainda existe a necessidade pela criação de novos produtos nesta área, pois os já existentes são caros, exigem hardware e software compatíveis, e fazem com que exista uma constante dependência de fabricantes externos.

Este trabalho busca avaliar e analisar o Microsoft Speech SDK que é uma das tecnologias no mercado para o auxilio ao desenvolvimento de soluções com o reconhecimento de voz. Existe um grande “vazio” em produtos para a área de reconhecimento de voz, principalmente no Brasil, pois a maioria das referências que se encontra são dos Estados Unidos e paises da Europa.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo deste trabalho é analisar e avaliar a ferramenta para o reconhecimento de voz que a Microsoft disponibiliza no mercado gratuitamente. Com a utilização desta tecnologia, pode-se abstrair a complexidade do reconhecimento de voz e dispensar mais tempo ao trabalho fim. Juntamente será analisada a eficiência do Microsoft Speech SDK 5.1 que irá trabalhar como o integrador entre o usuário e o modelo proposto.

1.3.2 Objetivos Específicos

Têm-se como objetivos específicos:

- Classificar e reconhecer palavras;
- Realizar aquisição do som através de uma interface analógica-digital (A/D) para reconhecimento em tempo real;
- Avaliar o kit de desenvolvimento Microsoft Speech SDK 5.1.

1.4 Procedimentos Metodológicos

De acordo com o objetivo geral e a finalidade deste trabalho, esta pesquisa se caracteriza como descritiva, pois descreve e avalia uma nova tecnologia emergente na área de reconhecimento. Também pode ser considerada uma pesquisa aplicada, devido a utilizar e propor novas tendências em aplicações práticas e soluções de reconhecimento de fala.

Quanto aos meios para atingir o objetivo geral, utiliza-se nesse trabalho o levantamento bibliográfico e aplicação de simulações dentro do modelo proposto. Desta forma pode-se conhecer mais sobre a área de reconhecimento, inteligência artificial e esta nova tecnologia que se propõe a estudar nesta jornada. Para a validação dos testes, análise e validação da ferramenta empregada aqui, utilizou-se o modelo proposto.

1.5 Delimitações do Trabalho

- ✓ O agente relativo à fonética está direcionado ao sistema padrão de fonética internacional.
- ✓ A ênfase deste trabalho é dada em relação à utilização do Microsoft Speech SDK independente de uma língua em específico.
- ✓ Um outro fator é o contexto onde é feita a aplicação da técnica.
- ✓ Algumas delimitações são apresentadas no desenvolvimento do texto, à medida que são abordados os assuntos teóricos no decorrer dos capítulos.

1.6 Estrutura do Trabalho

A fim de atender aos objetivos, além deste primeiro capítulo, este trabalho está estruturado como mostrado esquematicamente na figura 1.

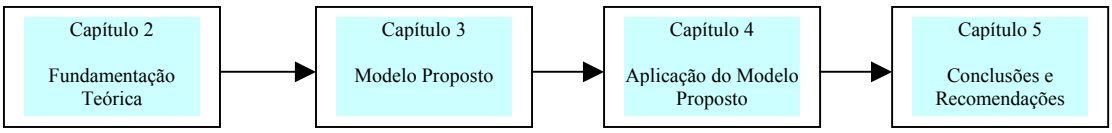

Figura 1 : Estrutura do trabalho

- No capítulo 2 será apresentada a fundamentação teórica necessária para o desenvolvimento deste trabalho. Serão levantados tópicos como reconhecimento de padrões, o kit de desenvolvimento da Microsoft Speech SDK 5.1 (SDK), ambiente Delphi.

- No capítulo 3 apresenta-se o modelo proposto, suas definições, características. Através do modelo será feita a avaliação da ferramenta se é viável ou não utilizar a mesma como fonte alternativa de tecnologia.

- No capítulo 4 apresenta-se o desenvolvimento do trabalho, aplicando-se as definições do modelo proposto. Relato dos problemas encontrados durante o processo de construção do software, os testes com seus respectivos comentários.

- No capítulo 5, apresenta-se às conclusões e recomendações para futuros trabalhos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Inteligência Artificial

O comportamento inteligente, envolve percepção e argumentos, aprendendo, comunicando, e agindo em ambientes complexos. Inteligência Artificial (IA) tem como uma de suas metas o desenvolvimento de máquinas que podem fazer estas coisas como também os humanos podem, ou possivelmente até melhor. Outra meta de IA é entender este tipo de comportamento se acontece em máquinas ou em humanos ou outros animais.

No início dos anos setenta, a IA já oferecia técnicas robustas para o desenvolvimento de aplicações práticas envolvendo sistemas inteligentes. Técnicas de representação de conhecimento, como quadros e redes semânticas, aliadas à tecnologia dos sistemas de produção formavam a base para a construção de ferramentas para o desenvolvimento de sistemas especialistas. Estas ferramentas são baseadas, analogamente à grande parte da pesquisa em IA da época, na visão de um sistema inteligente composto por apenas um centro de controle, um foco de atenção e uma única base de conhecimento (BITTENCOURT, 2001).

Para Durkin (1994), IA é um campo de estudo na ciência da computação que procura uma meta de fazer um computador argumentar de maneira similar aos humanos.

Para Bremer (1998), outro ponto que existe para compensar as deficiências da IA clássica dizem respeito à Inteligência Artificial Distribuída (IAD) na construção de agentes inteligentes. Isto diz respeito à organização de estruturas, estratégias de solução de problemas, e mecanismos de cooperação e coordenação por alcance de distribuição.

Segundo Bittencourt (2001), a Inteligência Artificial Distribuída é uma das áreas que mais se desenvolveram nos últimos anos. Esta área estuda o conhecimento e as técnicas de raciocínio que podem ser necessárias ou úteis para que agentes computacionais participem de sociedades de agentes.

Soluções de problemas distribuídos tem como objetivo global desenvolver técnicas de raciocínio e representação de conhecimento necessários para que nodos, contendo solucionadores de problemas interligados em uma rede francamente acoplada, cooperem efetivamente para solucionar um problema distribuído complexo. Entre seus objetivos genéricos pode-se citar:

- Aumentar a eficiência através do paralelismo
- Aumentar o número de tarefas realizáveis através do compartilhamento de recursos
- Aumentar a confiabilidade (tarefas duplicadas por diferentes métodos)
- Diminuir a interferência entre tarefas evitando interações inúteis

Algumas técnicas da IA são mencionadas por Tafner (1995), tais como Redes Neurais que se baseiam exatamente na neurotransmissao ocorrida nos animais para lançar suas bases de fundamentação. Nesta linha, faz-se uma analogia entre células nervosas vivas e o processo eletrônico. Os Sistemas Especialistas também são famosos na IA, procuram representar o conhecimento do especialista através de regras de produção (se... então....). Esse conhecimento armazenado é acionado por máquinas de inferência, cujo objetivo é deduzir algum novo conhecimento das regras embutidas. Esse método tem sido mais aplicado na área de diagnóstico médico e de sistemas que proporcionam linguagens naturais. Outra técnica que pode-se acrescentar divulgada é chamada Algoritmos Genéticos, baseada no processo de “seleção natural” definido por CHARLES DAWIN. A teoria desenvolve, basicamente o processo da natureza onde os mais fracos se extinguem. Os algoritmos genéticos estão sendo bastante utilizados para otimização de funções dentro do processo de programação e produção industrial, demonstrando resultados favoráveis.

Entre as técnicas encontra-se a Lógica Fuzzy, também conhecida como Conjuntos Difusos. A Lógica Fuzzy consiste em aproximar a decisão computacional da decisão humana. Isto é feito de forma que a decisão de uma máquina não se resuma apenas a um “sim” ou um “não”, mas também tenha decisões “abstratas”, do tipo “um pouco mais”, “talvez sim”, e outras tantas variáveis que representam as decisões humanas (TAFNER ,1995).

Para Barreto (1999), existem vários níveis de inteligência a serem usados na solução de um problema. Pode-se dizer que, para resolver um problema, é

necessário ter algum conhecimento do domínio do problema e utilizar alguma técnica de buscar a solução. Se o conhecimento for total sabe-se a solução e não é necessário uso de inteligência. Se não se dispõe de conhecimento, o problema é insolúvel pois não se conhece o enunciado!

A seguir descreve-se algumas áreas onde se aplica a IA.

- Base de dados inteligente, uma base de dados inclui fatos organizados de diversas maneiras para facilitar seu armazenamento e seu acesso;
- Problemas de decisão, tais como investimentos de mercado, decisões de qual jogada fazer em um jogo;
- Diagnóstico - processo de encontrar um defeito em um sistema;
- Interpretação – processo de associar significado a um conjunto de dados. Assim interpretação esta ligada intimamente ao reconhecimento de padrões e compreende dois problemas fundamentais: Determinação das interpretações válida para cada caso. E determinação e implementação da função que associa o conjunto de dados possíveis ao conjunto de interpretações válidas.
- Reparação uma vez feito um diagnóstico e o defeito do sistema foi encontrado, a etapa seguinte é corrigir este defeito. A reparação pode se tornar um problema muito difícil pois freqüentemente implica em uma seqüência de ações, cada uma produzindo um certo efeito sobre o mundo exterior que por sua vez modifica a seqüência de ações.
- Monitoração é diagnosticar em tempo real. Normalmente interpretam-se sinais vindos do mundo exterior e produz-se um alarme quando uma intervenção se faz necessária.
- Raciocínio sobre o mundo físico - técnicas de IA podem ser usadas para construir um modelo intuitivo do mundo físico e simular seu comportamento.

2.2 Voz

Para Amorim (1972), a voz é um elemento da linguagem; é a produção, que o ser humano faz, de sons, através das cordas vocais. É o elemento sonoro da

comunicação. Segundo Amorim (1981), a voz como todo som é resultado da vibração das partículas da matéria, portanto possui qualidades como: intensidade, altura, timbre, emissão e ressonância. Cada som que sai da boca é composto de dois períodos transitórios e de um período de estabilidade. O período de estabilidade chama-se vogal e os períodos transitórios, chamam-se consoantes (MATRAS, 1995).

Existem várias razões para adicionar a fala como outra modalidade na comunicação entre homem máquina, tais como:

- A fala é a maneira mais natural da comunicação humana. A maior parte da população não esta familiarizada com a digitação.
- Fala é a mais eficiente maneira de comunicação: a média humana de palavras está entre 120 e 250 palavras por minuto. Enquanto que um digitador bem treinado digita entre 100 e 150 palavras por minuto (AINSWORTH,1988).
- Para muitas pessoas é possível interagir apenas pela fala com a máquina.
- Em muitas aplicações é necessário, ou pode-se fazer melhor o trabalho utilizando a fala ao invés das mãos. Por exemplo, a operação de rádio ou telefone em um carro através da fala contribuiria muito a segurança.

Conforme Rahim (1994), a voz pode ser descrita como uma seqüência de sons os quais são gerados onde o fluxo de ar rompe ou perturba o aparato vocal, isto é, LIPS, JAW, TONGUE, VELUM AND LARYNX. Também conhecida como articulação vocal.

Conforme Tafner (1996), o reconhecimento de fala não é exatamente um problema fácil de ser resolvido. Tafner inclui o reconhecimento de fala juntamente com outras dezenas de itens em uma lista que considerou como problemas difíceis de serem resolvidos.

Quando é possível a existência de uma comunicação simples, fica igualmente simplificado a expressão daquilo que se deseja. A comunicação homem máquina em linguagem natural é interessante pois desperta nos usuários uma atração em torno da maneira como a comunicação com a máquina é realizada. Portanto há uma certa comodidade, por parte do usuário, no trabalho com sistemas que possuem uma interface que torna possível o uso da linguagem natural (KLABUNDE, 1996).

Uma das barreiras encontradas no reconhecimento da fala é a de armazenar as informações em forma de dados para este fim, ou seja, as informações de entrada devem ser transformadas e manipuladas antes do armazenamento da informação em forma de dados (KLABUNDE, 1996).

No reconhecimento da fala existem duas abordagens em relação ao tratamento da fala: o método global e o método analítico (HUGO,1992). O método global, também conhecido como reconhecimento de palavras, utiliza a técnica de reconhecimento de forma a comparar, globalmente, a palavra a reconhecer com as diversas formas de referência armazenadas. Este método mostra-se insuficiente no tratamento de grandes vocabulários e da fala contínua. Por esta deficiência, torna-se necessário o método analítico, que consiste em segmentar uma mensagem de fonemas, meia-sílabas, sílabas, etc. Após feita esta segmentação, a mensagem deve ser reconstruída considerando os aspectos léxicos, sintáticos e semânticos.

No tratamento da fala devem ser levados em consideração três aspectos: o reconhecimento, a compreensão e a interpretação da fala (BRUNS, 1995).

O reconhecimento da fala por computador é dividido em quatro fases distintas: a busca de sinais sonoros, o pré-processamento dos sinais, o processamento dos sinais obtidos e a verificação dos resultados (BRUNS, 1995). A fase da busca de sinais sonoros (palavras) consiste em capturar o som de um meio externo (sinais análogos) e transforma-los numa informação que o computador possa processar (sinais digitais). O pré-processamento de sinais é uma forma de pré-processar ou pré-ajustar os sinais que serão processados, com o objetivo de otimizar este processamento e garantir maior sucesso nesta operação. A fase de processamento utiliza os sinais pré-processados para reconhecer as palavras. Este processamento pode ser considerado a mais demorada das fases, pois não consiste apenas em organizar os padrões de palavras apresentadas, mas também o aperfeiçoamento de uma estrutura que torne possível a distinção destes padrões quando comparados com outros padrões não identificados ou não conhecidos. Nesta fase de processamento de sinais pode ser utilizada a técnica de redes neurais artificiais. A fase de verificação dos resultados, é feito uma comparação da saída produzida pela fase

de processamento e a saída desejada, para verificar se o resultado obtido foi equivalentemente ao resultado desejado.

2.3 Reconhecimento de Padrões

Um dos componentes que freqüentemente é necessário no desenvolvimento de sistemas inteligentes é o reconhecimento de padrões. O RP, é a ciência que trata da descrição ou classificação (reconhecimento) de medidas. Não há dúvidas que o RP se constitui em uma tecnologia útil e importante, que vem se desenvolvendo rapidamente com interesses e participações interdisciplinares. RP não se reduz a uma forma de aplicação mas apresenta um leque de abordagens possíveis que podem ou não ser relacionadas entre si (SCHALKOFF,1992).

Para Barreto (1999), o reconhecimento de padrões é uma tarefa geralmente desempenhada muito melhor usando as capacidades cognitivas do homem do que executando um algoritmo. Esta capacidade é altamente desenvolvida nos seres humanos e em muitos animais. Por exemplo, seres humanos são excelentes no reconhecimento de rostos, músicas, a caligrafia de alguém conhecido, etc. Cães são excelentes em reconhecer odores e gatos são capazes de sentir o humor de pessoas fugindo daquelas que exprimem características agressivas. Isso pode ser atribuído a um sistema bastante desenvolvido de reconhecimento de padrões. Observa-se ainda que nos casos citados a facilidade de reconhecer padrões depende dos padrões a que o indivíduo foi exposto anteriormente.

Por outro lado os esforços para fazer computadores baseados no conceito de instrução, têm encontrado sérias dificuldades, sendo o problema de reconhecimento de padrões de solução difícil.

2.3.1 Definição de Reconhecimento de Padrões

Para Ross, (1995) o reconhecimento de padrões pode ser definido como um processo de identificar estruturas nos dados por comparações com estruturas

conhecidas; as estruturas conhecidas são desenvolvidas através de métodos de classificação.

Em Barreto (1999), pode-se encontrar a seguinte definição: “O Reconhecimento de Padrões é o processo de identificar objetos através da extração de suas características a partir de dados sobre objeto”.

Ainda se encontra a seguinte definição em Klir (1995): "O reconhecimento de padrões pode ser definido como um processo pelo qual buscam-se estruturas nos dados e classificam-se estas estruturas dentro de categorias tais que o grau de associação é maior entre as estruturas da mesma categoria e menor entre as categorias de estruturas diferentes. As categorias relevantes, são usualmente caracterizadas por estruturas prototípicas derivadas da experiência do passado. Cada categoria pode ser caracterizada por mais de uma estrutura prototípica".

2.3.2 Reconhecimento de padrões para classificação

Barreto (1999) nos da uma breve explicação de como funciona esquematicamente um reconhecedor de padrões. O transdutor é munido de um sensor que traduz a energia suporte de informação sobre o objeto em sua forma original (ex. foto-elétrica ou células de retina se informação visual visual, terminações nervosas do ouvido interno ou microfone se informação sonora) para outra capaz de ser processada (neurotransmissores e sinais elétricos de sistema biológico ou elétricos de circuitos artificiais). O processamento inclui geralmente uma primeira fase em que atributos relevantes são selecionados para processamento e este processamento age como uma função, associando ao valor de um conjunto de atributos relevantes um elemento de um conjunto de padrões possíveis, o qual é apresentado como resposta do classificador.

O paradigma mais comum de aprendizado no caso de reconhecimento de padrões é o supervisionado, associado a uma rede direta multi-camadas. Devido a sua disponibilidade, a regra de retro-propagação é freqüentemente usada bem como suas variantes. Entretanto bons resultados são obtidos também com o aprendizado competitivo tipo redes de Kohonen. Este último é principalmente interessante quando não se sabe quantas classes possíveis existem a identificar, o que não é o caso do reconhecimento de padrões.

A estrutura de um sistema típico de reconhecimento de padrões é mostrada na figura 2. Note que ela consiste em um sensor (uma câmera por exemplo), um mecanismo de extração de características (algoritmo), e um algoritmo de descrição ou classificação (dependendo da aplicação). Complementarmente, é usual que alguns dados que já tenham sido classificados ou descritos estejam disponíveis para treinar o sistema (este conjunto de dados é chamado conjunto de treinamento) (SCHALKOFF,1992).

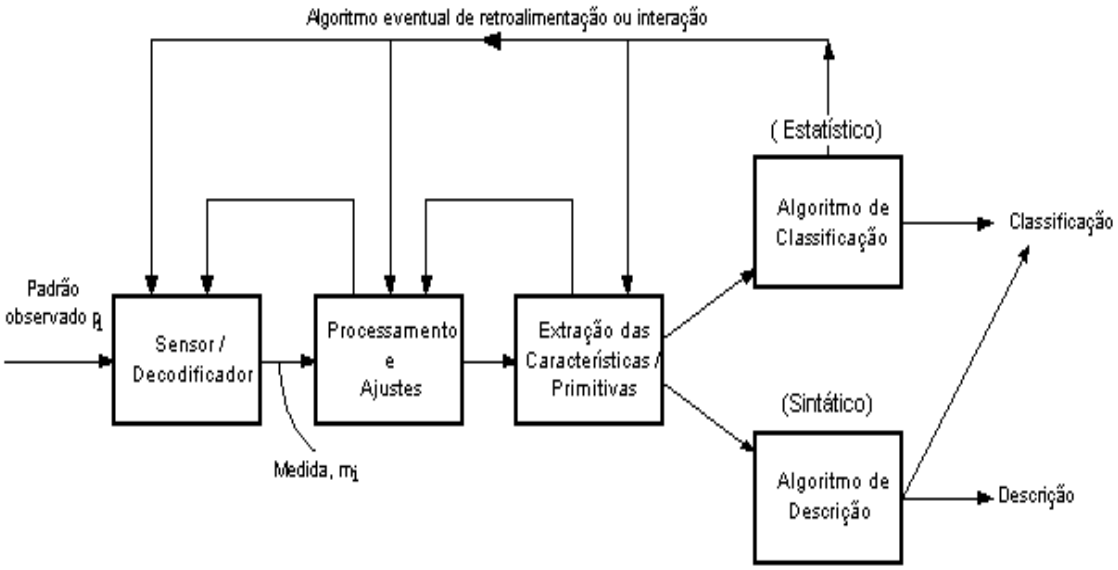

Figura 2 : Sistema típico de reconhecimento de padrões

O reconhecimento de caracteres é uma aplicação bem sucedida de redes neurais desde o Perceptron de Rosenblat. Muitos programas incluem alguma forma de reconhecimento de caracteres como programa de demonstração. Este tipo de reconhecimento foi também abordado com bastante sucesso por técnicas baseadas em manipulação simbólica, mas o reconhecimento de letras em posições diversas o uso de caracteres distintos o reconhecimento de letras manuscritas, por exemplo, continuam a ser problemas que a abordagem simbólica encontra dificuldades em resolver.

O reconhecimento de faces é uma área mais complexa, mas ela segue procedimentos semelhantes ao do reconhecimento de letra. No entanto este tipo de problema envolve uma grade muito mais fina o que aumenta consideravelmente a quantidade de neurônios da rede (BARRETO,1999).

2.4 O Microsoft Speech SDK

O Microsoft Speech SDK (SDK) é um kit de desenvolvimento de software para construção de agentes de fala e aplicações para o Microsoft Windows. A interface de programação de aplicações de fala (SAPI) reduz de forma substancial o código requerido para que se possa utilizar em um projeto o reconhecimento de voz e a transformação de textos em fala. Desta forma tenta-se fazer com que a tecnologia de fala torne-se mais accessível e robusta para as novas aplicações que irão surgir (Microsoft, 2002).

Este kit de desenvolvimento suporta OLE Automation o que permite a portabilidade do kit para as linguagens ou ambientes de programação que possuem suporte a estas características.

Existem dois tipos básicos de máquinas SAPI que são o sistema de Texto para Fala (TF) e o Reconhecimento de Voz (RV). O sistema de TF sintetiza o texto e arquivos para serem ouvidos usando vozes sintéticas. O reconhecedor de fala converte a voz humana falada no áudio para texto e arquivos. A figura 3 mostra como é a estrutura de programas que utilizam ou interagem com o SDK.

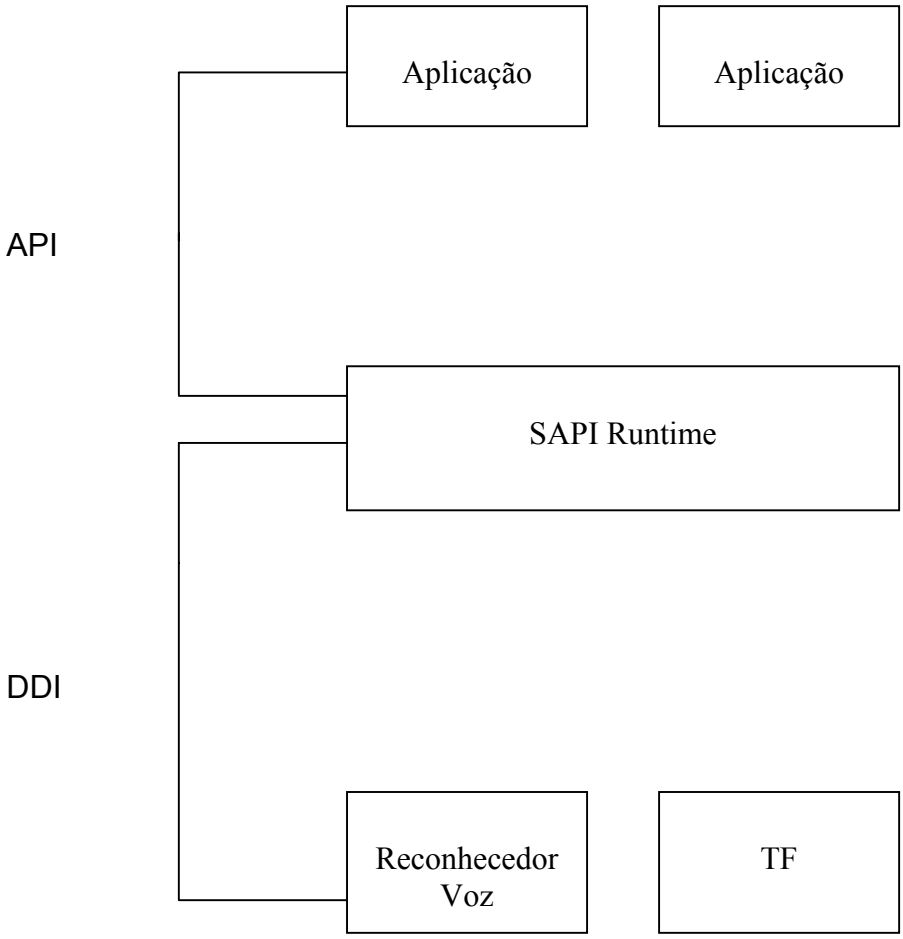

Figura 3 : Estrutura do SDK
Fonte: Microsoft

SAPI controla um grande número de aspectos de um sistema de fala, tais como:

- Controle de entrada de áudio se for de um microfone, arquivos e converte dados de áudio para o formato válido do SAPI
- Carrega arquivos de gramática (se foi criado dinamicamente ou criado a partir de dados que já existiam na memória)
- Compila padrões SAPI XML em formato de gramática, conversão de formatos gramaticais.

- Compartilhamento de reconhecimento através de múltiplas aplicações utilizando agentes compartilhados
- Retorna resultados e outras informações para a aplicação através de notificação de métodos
- Armazena dados e encaminha os resultados para análises posteriores

O reconhecedor de voz realiza as seguintes tarefas:

- Utiliza a interface de gramática do SAPI e carrega os ditados
- Realiza o reconhecimento
- Retorna a partir do SAPI informações sobre a gramática e o estado das alterações
- Gera reconhecimento e outros eventos para retornar informações a aplicação

A seguir estão descritos os vários tipos de interfaces que estão disponíveis:

- Interface de Compilação de Gramática - Esta interface é um compilador de agentes que permite definir, criar e testar novos agentes dentro do SAPI 5.1, antes da integração com a aplicação.
- Interface Léxica - Está interface define a maneira que as aplicações e agentes podem acessar os recursos léxicos. Quando uma aplicação léxica é adicionada ao sistema todas passam a compartilhá-la.
- Interface de recursos - Permite manipular os arquivos de voz e as interfaces léxicas. Esta interface dá as condições para que um desenvolvedor compreenda os conceitos de token (Token é um objeto que representa um recurso), suas categorias e a utilização das mesmas.
- Interface de Reconhecimento de voz - Habilita as aplicações para controlar os agentes de reconhecimento de voz.

O SDK é projetado em interfaces para ser modular. No desenvolver do trabalho se torna necessário a utilização de cada uma das interfaces. Podem ser utilizadas independentemente uma das outras. Para utilizar o SDK, primeiramente deve-se selecionar o agente de reconhecimento. Este agente define as características principais que poderão ser utilizadas na tarefa de

reconhecimento. Enumeram-se algumas delas como o suporte à lista de palavras, número máximo de palavras, utilização do treinamento de fala, tentativa de reconhecer palavras desconhecidas. Após a utilização do agente de RF define-se a gramática dentro do sistema e as regras que auxiliam o SDK durante o processo de reconhecimento. Com a utilização da interface léxica pode-se fazer o treinamento da fonética de palavras ou alterar a forma com que uma silaba será pronunciada. O agente léxico é o recurso do SDK que trabalha com as definições da fonética.

Conforme Productions (2003), quando o Agente utiliza a palavra inteira ele compara o sinal de áudio de entrada com o sinal de uma palavra modelo pré-gravada. Está técnica utiliza muito menos processamento do que a técnica de sub-palavras, mas requer que cada palavra que se deseja reconhecer seja pré-gravada.

Utilizando-se o reconhecimento através da técnica de sub-palavras, o agente procura pelas sub-palavras durante o processo de autenticação das mesmas, usualmente são fonemas. Neste tipo de pesquisa o reconhecimento de padrões é feito sobre essas sub-palavras. Esta técnica utiliza mais processamento do que o padrão de palavra inteira mas requer menos armazenamento de palavras pré-gravadas.

Na dependência de locutor o agente requer que o usuário treine a sua própria fala no reconhecedor, para que seja possível fazer a analise de voz posteriormente. Em geral o treinamento envolve uma série de frases pré-selecionadas. Para cada nova palavra deve-se realizar o treinamento.

O agente dependente de locutor pode trabalhar sem treinamento, mas o percentual de acerto é abaixo de 95 por cento e não melhora até que o usuário faça treinamento. Esta técnica requer menos processamento mas é frustrante para muitos usuários, por que o treinamento é tedioso e o tempo de treinamento pode levar de cinco minutos a várias horas.

Fala adaptativa neste tipo de treinamento o agente treina a si próprio para fazer o reconhecimento. O nível de acerto em geral inicia em torno de noventa porcento, mas sobe o nível de acerto após algumas horas de treinamento.

Fala Independente o agente inicia com a eficiência acima de 95 por cento para muitas pessoas (com palavras sem acento). A maioria dos agentes

independentes de fala tem treinamento ou habilidade de adaptação que aperfeiçoa a capacidade de acerto por alguns poucos pontos percentuais. Os programas de fala independentes necessitam de capacidade computacional muito maior do que os programas dependentes de usuário.

2.4.1 Gramática

Para a Microsoft (2002), a gramática define as palavras que uma aplicação pode reconhecer. O RF é baseado na gramática. Uma aplicação pode realizar o RF usando três diferentes tipos de gramática. Cada gramática utiliza-se de uma estratégia diferente para reduzir a possibilidade de sentença a ser reconhecida, aumentando assim as chances de assertividade durante o processo de RF.

Existem três tipos de gramática os quais estão relacionados abaixo.

A) Gramática em formato arquivo texto

Este tipo de gramática que é definido em arquivo texto tem o seu formato similar a arquivos ".ini". Este arquivo é consistido de um número de seções. Cada seção é identificada por uma nova linha com o nome da seção entre colchetes "[]". As seções têm um número de valores, identificadas pelo nome do valor, seguido do sinal de igual "=" e finalmente pelo valor. Comentários podem ser acrescentados no arquivo. Para isto basta no início da linha acrescentar ponto e virgula ";" ou duas barras "//". A vantagem do uso de listas, são que aplicações podem facilmente setar novos parâmetros para a sua lista em tempo de execução sem ter que recompilar a gramática. No anexo 4 mostra-se um exemplo de uma gramática onde é possível fazer a troca de canais de uma televisão.

B) Gramática livre

Aqui são utilizadas regras de uso, que predizem as próximas palavras, ou seja, reduz o número de candidatos para avaliar e para reconhecer a próxima palavra. Para iniciar o uso deste tipo de gramática no RF, a aplicação primeiro deve ativar uma regra específica dentro da gramática. Isto coloca a regra dentro do nó inicial do reconhecimento. Em outras palavras para a regra ser utilizada, deve ser colocada no "Start" de regra. Este tipo de padrão requer uma lista de

todas as regras que podem ser ativadas. Mas estas listas devem optar por ter sempre um número mínimo de regras, pois desta forma, a eficiência da gramática será aumentada, crescendo assim, diretamente a precisão do reconhecimento.

Esta ferramenta é muito poderosa devido à habilidade de recursão (RECURSION), mas também pode se tornar muito complexa. A eficiência de uma máquina de RF deteriora quando é utilizada a gramática com muita recursão (quando é utilizado vários níveis), ou muita perplexidade, essencialmente o número de possíveis escolhas em um determinado nodo em um caminho de reconhecimento.

C) Gramática de ditado

Uma gramática de ditado define um contexto para falar, que identifica o assunto do ditado, o tipo de estilo de idioma que é esperado, e que ditado foi acabado no passado.

Uma gramática de ditado não contém informação sobre o modelo de idioma (a máquina tem aquela informação) nem específica tudo das palavras que podem ser faladas. Específica palavras só incomuns ou grupos esperados de palavras.

Existem algumas características que podem influenciar a gramática durante o seu processamento. O peso da gramática é um destes fatores. Utilizado a cada transação para alterar a semelhança de certos pesos iniciados. Este peso é uma probabilidade e a faixa de valores permitida está dentre 0 e 1. Os pesos com valores 0, sempre serão interpretados como transições impossíveis de passarem a ser reconhecidos. Por definição, as gramáticas não possuem pesos definidos, para cada transição o peso será 1 dividido pelo número de transições fora do estado de precedência (PRODUCTIONS, 2003).

O atributo de confidência requerida, utiliza-se para tornar mais fácil o reconhecimento, isto é, se o reconhecimento será aceito ou rejeitado.

2.4.2 Regras

Cada gramática pode conter uma ou mais regras. Podem ser de alto nível, indicando que podem ser ativados para o reconhecimento. Cada uma tem um estado inicial, os quais são conectados por vários tipos transição:

- Uma palavra de transição indicando uma palavra para ser reconhecida;
- Uma regra de transição indicando uma referencia para uma sub-regra;
- Transições especiais que caracterizam ditados especiais dentro de um contexto.

2.4.3 Contexto Livre de Gramática.

As referências para sub-regras podem ser recursivas, podem referenciar a elas próprias, um ao outro de forma direta ou indiretamente.

Notificações de regras chamam o agente para informar quando regras são adicionadas, alteradas ou removidas. Existem cinco ações que são tomadas na realização das regras:

- Novas regras podem ser adicionadas;
- Regras podem ser removidas;
- Regras podem ser ativadas;
- Regras podem ser desativadas para o reconhecimento;
- Regras podem ser invalidadas, o qual é uma maneira utilizada para editar através de uma aplicação e deste modo o agente precisa fazer uma leitura nova do conteúdo da regra.

A figura 4 mostra o agente de reconhecimento de fala com suas especificações utilizadas neste trabalho.

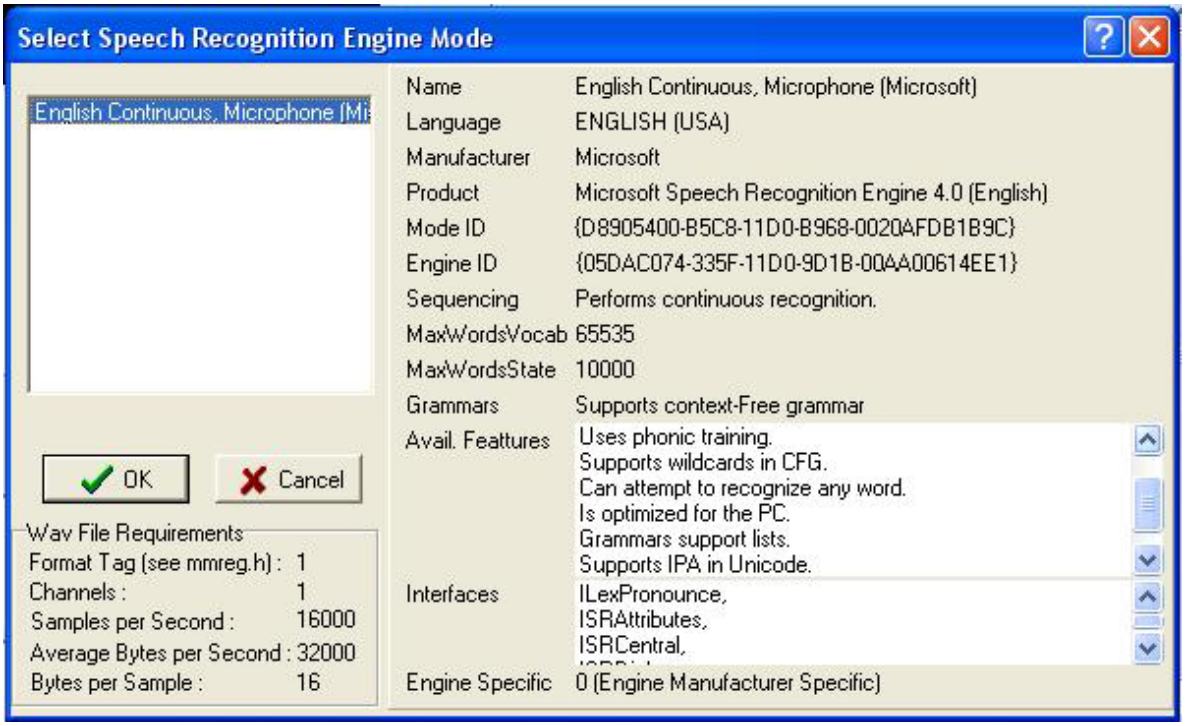

Figura 4 : Tela das configurações do Agente de Reconhecimento de Fala

2.4.4 Automação

Automação (Automation) é parte do protocolo COM. Descreve aplicativos para o servidor, expõe uma interface para aplicativos cliente, e como os clientes podem controlar a programação com o servidor. Aplicativos clientes são chamados de Automation Controllers (controladores de automação). Controladores de Automação podem ser aplicativos ou bibliotecas de vínculo dinâmico escritos em qualquer linguagem que suporte Automação (MICROSOFT, 2001).

A Microsoft disponibiliza em seus softwares como o Pacote de softwares do Office e o Speech SDK recursos de automação. Desta forma esses aplicativos podem ser utilizados de forma individual ou como aplicativos servidores. A utilização destes poderosos servidores oferece grande poder de processamento de palavras, banco de dados, numeração decisiva e gerenciamento de capacidades de contato, nivelando o código existente o qual já foi utilizado por outras pessoas (KIMMEL, 2001).

3 MODELO PROPOSTO

3.1 Introdução

Através deste modelo, analisa-se a viabilidade da utilização do Microsoft SDK como uma ferramenta auxiliar no desenvolvimento de software em relação à fala. Optou-se pelo SDK devido a sua facilidade de integração com outros softwares, e a vantagem dos desenvolvedores se abstraírem da complexidade do reconhecimento da fala.

O SDK é lançado pela Microsoft no mercado com suporte aos idiomas Inglês, Japonês e Chinês. Como não existe um modelo prévio para o idioma português, este trabalho buscará a formulação de um modelo base no idioma português. O que dará suporte ao SDK através da criação da gramática, fonema, análise léxica e finalmente o reconhecimento da fala.

Para realizar o reconhecimento, uma aplicação necessita de uma origem de áudio para a entrada da fala, um agente para processar a fala, e uma ou mais gramáticas para permitir ao agente com as listas de palavras ou regras que determine o que possa ser reconhecido.

3.2 Especificação da proposta

O modelo a ser utilizado está baseado nas seguintes estruturas:

- Agente
- Definição de gramática
- Definição de regras
- Definição de atributos
- Definição léxica
- Treinamento e Reconhecimento

A figura 5 mostra a estrutura do modelo a ser utilizada. Na figura apresenta-se a seqüência lógica que é necessária para se chegar ao reconhecimento.

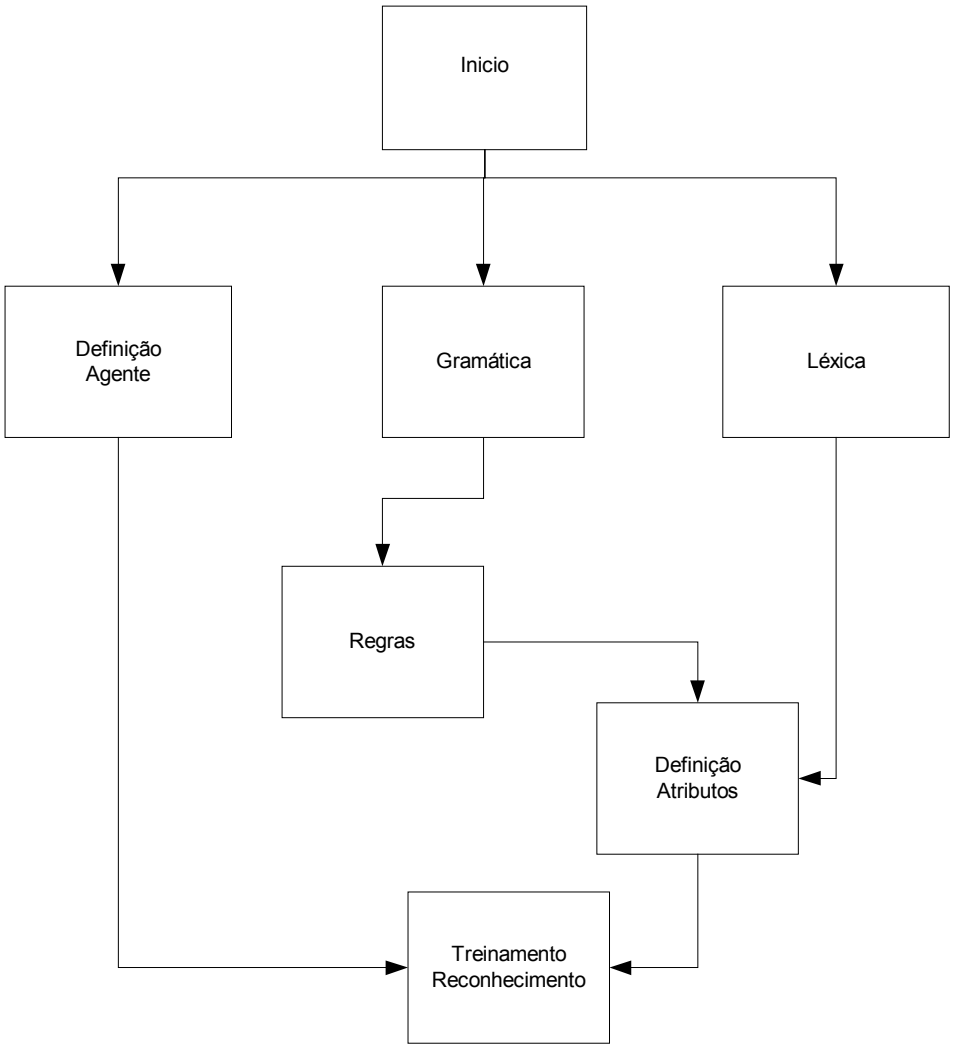

Figura 5 : Estrutura do modelo

Chegou-se ao modelo mostrado na figura 5 como sendo a seqüência natural a ser seguida para se trabalhar com o SDK. Pode-se utilizar também outras formas de trabalho com o SDK.

3.3 Agente

Os agentes são os conjuntos de interfaces que serão utilizadas na montagem do modelo. A avaliação e utilização destes agentes são o objeto de estudo deste trabalho. O SAPI é fornecida pela Microsoft de forma gratuita na intenção de alavancar a utilização do reconhecimento de voz.

3.4 Definição de gramática

A definição de gramática é uma das partes do reconhecimento de voz que envolve a decisão de qual palavra que pode ter sido eventualmente falada. No caso de ditados a gramática pode ser utilizada para auxiliar e identificar algumas palavras que são parecidas com a que foi falada.

Na figura 6 mostra-se a sintaxe de uma gramática que reconhece cores.

```
<GRAMMAR LANGID="809">

 <!—DEFINIÇAÕ DE CONSTANTES -->
 <DEFINE>
  <ID NAME="RID_start" VAL="1"/>
 </DEFINE>

 <!—DEFINIÇÃO DE REGRAS -->
 <RULE NAME="start" ID="RID_start" TOPLEVEL="ACTIVE">
  <L>
   <P>red</P>
   <P>blue</P>
   <P>green</P>
  </L>
 </RULE>
</GRAMMAR>
```

Figura 6: Sintaxe de gramática

A definição de gramática é feita via arquivos textos nos quais pode-se definir qual a gramática e as regras que podem ser utilizadas.

3.5 Definição de regras

Regras são elementos utilizados pelos agentes para restringir as possibilidades de escolha de palavras ou sentenças durante o processamento. O uso dos agentes de gramáticas com regras para controlar os elementos de sentença e construção, usando listas de reconhecimentos predeterminadas ou frases escolhidas formam a base do vocabulário do agente de RV.

3.6 Definição de atributos

Os atributos são as propriedades que estão direta ou indiretamente ligadas as configurações do agente ou da aplicação envolvida.

Segue abaixo alguns exemplos de atributos que podem ser utilizados:

- Ajuste de velocidade do sinal. Valor entre 0 e 100
- Percentual de tempo do processo que o agente espera para usar durante a fala constante
- Threshold – indica o nível de certeza da resposta
- TimeOut Incompleto – tempo em milisegundos que o agente espera antes de descartar uma frase por que o usuário parou de falar.

Existem vários outros tipos de atributos que são utilizados pelo SDK. E que conforme forem utilizados no desenvolvimento do trabalho serão explicados.

3.7 Definição léxica

O controle léxico permite ao usuário definir a pronúncia correta de uma palavra no agente de pronúncia léxica. Isto deve ser definido para cada linguagem.

3.8 Treinamento e Reconhecimento

O treinamento se dará a partir de um conjunto de frases a serem faladas. Estas palavras deverão ser repetidas várias vezes no intuito de se chegar ao treinamento correto e após o reconhecimento.

4 DESENVOLVIMENTO DO PROTÓTIPO E APLICAÇÃO DO MODELO PROPOSTO

4.1 Introdução

Uma vez definido no capítulo 3 o modelo a ser seguido, neste capítulo apresentam-se algumas simulações para colher os resultados desejados neste projeto. As simulações foram utilizadas com o intuito de avaliar e analisar o SDK.

Para cada conjunto de simulação que utilizou-se neste trabalho foi necessário a criação de estruturas de gramáticas conforme visto anteriormente.

Os conjuntos de dados foram escolhidos com o objetivo de direcionar os dados de forma a seguir alguns tipos de aplicações práticas realizadas no cotidiano. Nestes conjuntos, tentou-se realizar testes que criassem algumas situações reais e práticas.

4.2 Aplicativo para Teste

Chama-se neste contexto aplicativo para teste o software que foi desenvolvido e utilizado para auxiliar nos testes para realizar o trabalho. Este aplicativo foi criado em forma de protótipo com o intuito de apenas ser utilizado para sanar as dificuldades e necessidades que surgiram durante o desenvolvimento deste trabalho. O principal foco no desenvolvimento, foi criar as opções necessárias para que fosse possível a utilização do SDK. Obteve-se um bom grau de qualidade e requisitos necessários para que este trabalho conseguisse ser validado.

O aplicativo foi desenvolvido em Delphi 7 e utilizou-se de componentes de automação para realizar a ligação entre o aplicativo e o SDK. A própria Borland (empresa proprietária do Delphi 7) fornece alguns componentes que auxiliam e facilitam a integração entre o usuário e o SDK. O modelo de gramática pode ser utilizado em qualquer editor de texto que gere arquivo no formato de texto (TXT). Por exemplo, a figura 7 mostra uma gramática que foi compilada e está pronta para ser utilizada.

Figura 7 : Tela de uma gramática válida

Para a realização dos testes, utilizando-se da pronúncia de palavras no idioma português, foi definido como padrão de fonética o modelo internacional. Utilizou-se este padrão pois ele segue um padrão base para os fonemas parecido com o do idioma português. No anexo 1 existe uma relação dos fonemas internacionais.

A figura 8 mostra um breve exemplo do aplicativo que foi desenvolvido para que fosse possível a utilização do SDK.

Figura 8 : Tela principal do Aplicativo Teste

A figura 9 mostra a tela que a Microsoft disponibiliza para fazer algumas definições de reconhecimento de fala. Nesta tela é possível definir novas configurações de fala de acordo com o perfil do locutor.

Figura 9 : Propriedades de fala

A figura 10 mostra a tela que a Microsoft disponibiliza para fazer o treinamento do reconhecimento de fala. Aqui é feito o treinamento para que o reconhecedor apreenda a identificar as características da voz do locutor.

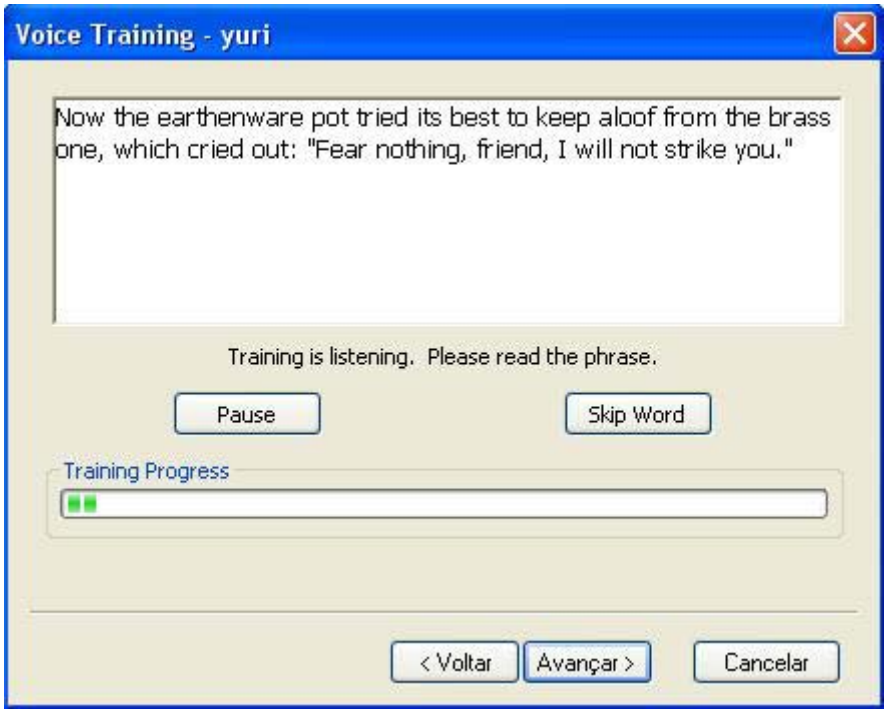

Figura 10 : Treinamento para o reconhecimento

A figura 11 mostra a tela onde é possível realizar as alterações de pronúncia das palavras, além de permitir a alteração de pronúncia permite determinar a alteração dos fonemas de uma palavra.

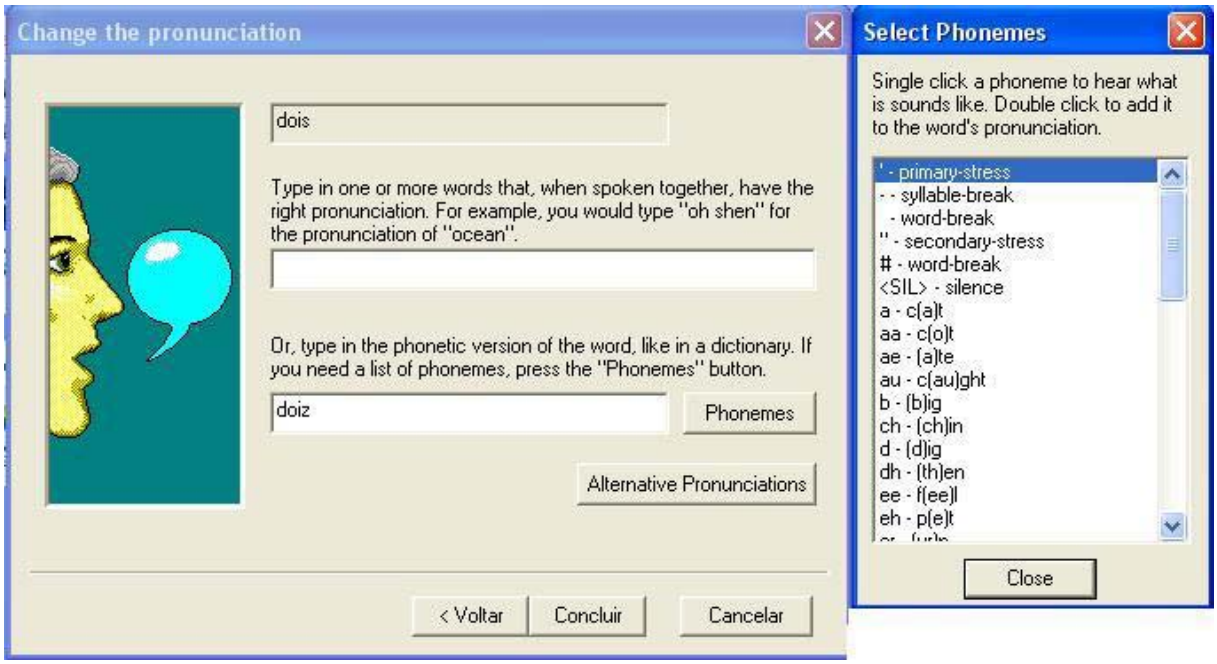

Figura 11 : Alteração de Pronúncia

4.3 Reconhecimento de Números

Este teste foi realizado com a intenção de fazer o reconhecimento de números. Para isto foi utilizado o modelo do agente fornecido pela Microsoft. No primeiro passo realizou-se os testes com números e pronúncias na língua inglesa. Após os testes com este modelo realizou-se testes com um modelo ao qual a pronúncia dos números fosse realizada em português.

Tabela 1 : Amostra de acertos dos números

Número	Inglês (%)	Fonema Internacional (%)
Um	100	98
Dois	89	70
Três	94	46
Quatro	92	43
Cinco	98	60
Seis	98	80

No anexo 2 encontra-se um exemplo de gramática que pode ser utilizado como base para números,datas, horas, meses. O anexo 2 também foi aplicado na avaliação.

Tabela 2 : Amostra de acertos do Anexo 2

Palavra	Inglês (%)	Fonema Internacional (%)
Março	93	85
Quinze	89	70
um	100	98
Maio	92	70
Junho	90	75
Seis	98	80

A tabela 2 mostra a utilização de três regras para o reconhecimento. A regra dos meses do ano, que identifica as palavras, março,maio, junho. A regra de

números entre 10 e 19 identificado pelo número quinze e finalmente a regra que identifica os números de 1 a 9.

4.4 Reconhecimento de Palavras

Este teste foi realizado com o objetivo de relacionar palavras definidas em uma gramática e durante o treinamento foi adicionado novas palavras com a intenção de analisar a adequação de novos fatos a serem reconhecidos.

Este teste foi realizado com a intenção de fazer o reconhecimento de palavras de uma lista de compras. Nesta lista de compras a pessoa fala qual o produto que deseja. Conforme o reconhecimento acontece, poderia-se montar uma lista de pedidos do cliente.

Tabela 3 : Amostra de acertos de palavras

Palavra	Inglês (%)	Fonema Internacional (%)
Omo	85	75
Kiwi	70	70
Sal	86	46
Uva	60	46

No anexo 3 encontra-se um exemplo de gramática que pode ser utilizado como base para uma lista de compras de supermercado.

Tabela 4 : Amostra de acertos das Bebidas

Palavra	Inglês (%)	Fonema Internacional (%)
Pepsi	90	80
Coca	90	75
Cerveja	92	60
Vinho	90	70

Tabela 5 : Amostra de acertos da Padaria

Palavra	Inglês (%)	Fonema Internacional (%)
Bolo	90	82
Rosca	88	70
Pão	95	90
Leite	93	75
Cuca	88	83

Tabela 6 : Amostra de acertos das Verduras

Palavra	Inglês (%)	Fonema Internacional (%)
Nabo	85	80
Picles	80	80
Alface	78	65

Conforme eram modificadas ou acrescentadas novas regras o percentual de acerto variava muito. Nas palavras em língua inglesa a assertividade sempre se manteve maior do que as utilizadas em português. A variação dos fonemas apresentou-se como um fator determinante no reconhecimento de fala. Isto se deve a forte variação que existe em relação à pronúncia das palavras da língua inglesa e portuguesa. A maior dificuldade nesta etapa foi sempre ter que adequar os fonemas de forma com que uma palavra em português pudesse ser reconhecida. Como a ferramenta utilizada permite alterar ou associar os fonemas, nas palavras em português alteraram-se os fonemas de forma que pudessem ser identificados na língua portuguesa.

Os testes somente foram realizados e validados após terem sido efetuados todos os treinamentos necessários para que o reconhecedor estivesse otimizado para realizar a melhor performance possível.

Para chegar aos percentuais de acerto demonstrados nas tabelas anteriormente citadas utilizou-se o seguinte procedimento.

A primeira fase de testes foi realizada somente com palavras na língua inglesa. As palavras foram faladas de forma intercalada entre si cem vezes cada uma. Os acertos e erros foram apresentados, e com os valores calculados a

partir destas anotações pode-se chegar às tabelas com as respectivas margens de acerto. Após estes testes adotou-se os mesmos procedimentos para o idioma português.

Estes foram alguns dos tipos de testes que foram aplicados para fazer a analise do Microsoft Speech SDK.

5 CONCLUSÕES E RECOMENDAÇÕES PARA FUTUROS TRABALHOS

5.1 Conclusões

Neste trabalho foi possível verificar algumas das características que tornou o SDK como uma das novas alternativas para a utilização do reconhecimento de voz.

Pode-se citar como uma vantagem neste tipo de ferramenta à utilização das gramáticas como delimitador de palavras que podem ser aceitas no contexto. Com isto consegue-se restringir a um menor número de possibilidades de uma palavra ser selecionada como a alternativa correta no momento do reconhecimento.

A falta de um agente que utiliza especificamente o vocabulário português é uma das deficiências apresentadas no SDK. Mostrou-se mais eficiente o SDK no reconhecimento de palavras da língua inglesa. Algumas características que tornam a língua inglesa mais fácil de ser reconhecida no SDK que a portuguesa se deve às facilidades de se montar gramáticas no SDK, estruturas fonéticas pré-definidas e as estruturas das palavras.

Em relação ao objetivo de classificar e reconhecer palavras pode-se dizer que o SDK apresenta uma boa performance principalmente quando a “língua” utilizada é o Inglês. Para o português existem algumas deficiências que devem ser pesquisadas com melhor nível de detalhamento.

Na comparação entre os resultados obtidos houve maior assertividade nos testes efetuados com números. Os números apresentam variações fonéticas menores do que as palavras. Estas além de possuírem grandes variedades de fonemas exigem que o locutor faça o treinamento antes de iniciar o procedimento de reconhecimento. Nas tabelas de amostras de dados demonstradas no decorrer deste trabalho, pode-se observar que nas comparações utilizadas entre as palavras em inglês a assertividade foi maior do que as palavras pronunciadas no idioma português. O SDK tem como padrão no agente léxico fonemas da língua inglesa já definidos. Desta forma existe maior facilidade para o

desenvolvimento de aplicações para este idioma. Entre as duas simulações apresentadas anteriormente no capitulo 4, a maior dificuldade encontrada no trabalho foi a adequação da pronuncia as palavras em português.

Características como: sotaque das pessoas, a voz está diferente da forma com que foi efetuado o treinamento da fala, ou ambiente com ruído pode interferir diretamente no resultado do reconhecimento de fala.

O Software é capaz de reconhecer palavras independentemente de locutor, mas não apresentou resultados satisfatórios quando existia uma variação de locutores.

A aquisição do som é feita de forma satisfatória embora o ideal é ter um ambiente sem nenhum som para ter melhores resultados durante a utilização da ferramenta (SDK).

Quanto ao SDK mostrou ser um bom software auxiliar para as pessoas que desejam utilizar o reconhecimento de voz em suas aplicações. Mas é necessário estudo mais aprofundado para se ter e criar condições de trabalho que sejam seguras. Como não se tem um fator de certeza coerente durante o reconhecimento, deve-se ter cuidado ao desenvolver uma aplicação com este tipo de ferramenta. Pode-se obter resultados imprevisíveis dependendo do grau de utilização desejado.

Os testes realizados neste trabalho estiveram voltados para o reconhecimento de comandos de voz. Não se teve a intenção de gerar ditados contínuos para se avaliar a ferramenta.

5.2 Recomendações para Futuros Trabalhos

Como sugestão para trabalhos futuros, pode-se realizar uma pesquisa para a definição de agentes que trabalhem com o idioma português. Isto não é uma tarefa fácil de se realizar pois, além da complexidade da formação de várias palavras também tem-se o sotaque das várias regiões do Brasil e dos outros paises que falam português.

Coloca-se como sugestão à utilização de aplicações que utilizem o reconhecimento de comandos através da voz. Pois é uma forma mais simples de

se poder utilizar tecnologias deste tipo e com um nível de segurança mais aceitável.

Referências Bibliográficas

Ainsworth, W.A. **Speech Recognition by Machine, volume 12 of IEE Computing Series.** Peter Peregrinus LTDA, London, 1988.

AMORIM, Antônio. **Fonoaudiologia geral**. São Paulo: Livraria Pioneira, 1972.

AMORIM, Antônio. **Fundamentos científicos da fonoaudiologia.** São Paulo: Ciências Humanas, 1981.

BARRETO, Jorge Muniz. **Inteligência Artficial noLimiar do Século XXI.** Florianópolis: PPP Edições, 1999.

BITTENCOURT, Guilherme. **Inteligência Artificial Ferramentas e Teorias.** Florianópolis: Editora da UFSC, 2001.

BREMER Walter e ZARNEKOW, Rudiger e WITTING, Hortmt. **Intelligent Software Agents Foundations and Applications.** Dresden: Springer, 1998.

BRUNS, Fábio A. Trabalho de conclusão de curso. Curso de Ciências da Computação. **Protótipo de reconhecimento de palavras através da fala.** Blumenau: FURB,1995.

DURKIN, John. **Expert Systems Design and Development.** Prentice-Hall, 1994.

HUGO, Marcel. Trabalho de Conclusão de Curso. Curso de Ciências da Computação. **Especificação de um protótipo de software comandado por voz.** Blumenau: FURB, 1992.

KIMMEL, Paul. **Desenvolvendo Aplicações em Delphi 6.** Rio de Janeiro: Editora Ciência Moderna LTDA, 2001.

KLABUNDE, Charles Cristiano. Trabalho de conclusão de curso. Curso de Ciências da Computação. **Aplicação do modelo de Rede Neural RBF-Fuzzi ARTMAP para o reconhecimento da fala.** Blumenau: FURB, 1996.

KLIR, G. J. e YUAN, B. ***Fuzzy sets and fuzzy logic - theory and applications*.** Prentice Hall, 1995.

MATRAS, Jean-Jaques. **O som.** São Paulo: Universidade Hoje, 1995.

MICROSOFT **Help Microsoft Speech SDK version 5.1**, Microsoft Corporation , 2002.

MICROSOFT **Microsoft Speech API Documentation** Microsoft Corporation, 2001.

PRODUCTIONS, O & A **Why Should I Use Speech in My Application** Disponivel em: http://www.o2a.com/framesPrimer.htm acessado em: 25/03/2003.

RAHIM, Mazin G. **Artificial Neural Networks for Speech Analysis/Sunthesis.** Londres : Chapman & Hall, 1994.

SCHALKOFF, Robert J., **Pattern Recognition: Statistical, Structural and Neural Approches**, John Wiley & Sons, Inc., 1992.

TAFNER, Malcon Anderson. **Reconhecimento de palavras faladas isoladas usando redes neurais artificiais.** Dissertação submetida à Universidade Federal de Santa Catarina para obtenção do Grau de Mestre em Engenharia. Florianópolis: UFSC, 1996.

TAFNER, Malcon A. e XEREZ, Marcos de e FILHO, Ilson W. Rodrigues. **Redes Neurais Artificiais Introdução e Princípios de Neurocomputação.** Blumenau: Editora EKO, 1995.

Anexo 1 – Fonemas definidos no Padrão Internacional

SYM	Example	PhoneID
-	syllable boundary (hyphen)	1
!	Sentence terminator (exclamation mark)	2
&	word boundary	3
,	Sentence terminator (comma)	4
.	Sentence terminator (period)	5
?	Sentence terminator (question mark)	6
_	Silence (underscore)	7
1	primary stress	8
2	secondary stress	9
aa	F<u>a</u>ther	10
ae	C<u>a</u>t	11
ah	C<u>u</u>t	12
ao	D<u>o</u>g	13
aw	F<u>ou</u>l	14
ax	<u>A</u>go	15
ay	B<u>i</u>te	16
b	<u>B</u>ig	17
ch	<u>Ch</u>in	18
d	<u>D</u>ig	19
dh	<u>Th</u>en	20
eh	P<u>e</u>t	21
er	F<u>ur</u>	22
ey	<u>A</u>te	23
f	<u>F</u>ork	24
g	<u>G</u>ut	25
h	<u>H</u>elp	26
ih	F<u>i</u>ll	27
iy	F<u>ee</u>l	28
jh	<u>J</u>oy	29
k	<u>C</u>ut	30
l	<u>L</u>id	31
m	<u>M</u>at	32
n	<u>N</u>o	33
ng	Si<u>ng</u>	34

ow	G<u>o</u>	35
oy	T<u>oy</u>	36
p	<u>P</u>ut	37
r	<u>R</u>ed	38
s	<u>S</u>it	39
sh	<u>Sh</u>e	40
t	<u>T</u>alk	41
th	<u>Th</u>in	42
uh	B<u>oo</u>k	43
uw	T<u>oo</u>	44
v	<u>V</u>at	45
w	<u>W</u>ith	46
y	<u>Y</u>ard	47
z	<u>Z</u>ap	48
zh	Plea<u>s</u>ure	49

Anexo 2 - Exemplo de Gramática para Números

A Gramática descrita neste anexo serve como exemplo de uma gramática que pode ser definida para se utilizar como um exemplo de uma lista de compras em supermercado

```
[Grammar]
langid=1033
type=cfg
[<Start>]
<Start> = call "PhoneNum=" <PhoneNumber>
<Start> = the year is "Year=" <Year>
<Start> = the month is "Month=" <Month>
<Start> = the date is "Date=" <Date>
<Start> = the time is "Time=" <Time>
<Start> = it costs "Dollars=" <Dollars>
<Start> = the extension is "Digits=" <Digits>
<Start> = speak the fraction "Fraction=" <Fraction> now
<Start> = a natural number is "Natural=" <Natural>
<Start> = an ordinal number is "Ordinal=" <Ordinal>
<Start> = an integer is "Integer=" <Integer>
<Start> = a floating point number is "Float=" <Float>
<Start> = it happened in the "Plural=" <PluralNumber>
<Start> = a single digit "Digit=" <0..9>
```

Anexo 3 - Exemplo de Gramática

A Gramática descrita neste anexo serve como exemplo de uma gramática que pode ser definida para se utilizar como um exemplo de uma lista de compras em supermercado.

```
[Grammar]
langid=1033
type=cfg

[Lists]
=Padaria
=Bebida
=Verdura
=Fruta
[<Start>]
<Start>= [opt] <Verdura> [opt] <Bebida> [opt] <Padaria>

[Padaria]
=Bolo
=Rosca
=Pão
=Leite
=Cuca

[Verdura]
=Alface
=Cenoura
=Nabo
=Picles

[Fruta]
=Laranja
=Uva
=maracujá
=kiwi
=pêssego

[Bebida]
=Skol
= Cerveja
=Bavaria
=Coca-cola
=Colonia
=Antártica
=Budweiser
=Pepsi
=Vinho
```

Anexo 4 - Exemplo de Gramática para Troca de Canais

Abaixo segue o exemplo do conteúdo de um arquivo texto contendo a gramática.

```
// Comentários
// Está gramática mostra o uso de listas para troca de canais de TV.
// O <MostrarTV> and <Canal> tem a lista dos nomes.

// Try saying:
//   I want to watch CNN
//   Switch to MSNBC
//   ABC please

[Grammar]
langid=1033
type=cfg

[Lists]
=TVShow
=Channel

[<Start>]
<Start>= "TVShow=" [opt] <Junk> [opt] <Watch> <ShowOrChannel> [opt] <JunkEnd>
<ShowOrChannel>=<TVShow>
<ShowOrChannel>=<Channel>

<Watch>=watch
<Watch>=change to
<Watch>=go to
<Watch>=switch [opt] to
<Watch>=change the channel to
```

<Watch>=see
<Watch>=turn to

<Junk>=please
<Junk>=could you
<Junk>=I want [opt] to
<Junk>=I wanna
<Junk>=computer
<Junk>=uh

<JunkEnd>=please
<JunkEnd>=now

// The TVShow and Channel list entries can be automatically sent to
// the grammar by calling AutoList
[TVShow]
=Babylon five
=Seinfeld
=sixty minutes

[Channel]
=C N N
=N B C
=C B S
=A B C
=M S N B C
=Discovery
=Nickelodian
=A and E