

2. INTERFACES NATURAIS

2.1 Introdução

Para que se possa compreender melhor como se processa a interação entre o homem e o computador, deve-se procurar entender o processo da comunicação humana.

2.1.1 Comunicação e Linguagem

As pessoas se comunicam para comandar, interrogar, responder, prometer, convencer as outras pessoas. Todas as pessoas quando têm uma idéia e pretendem compartilhá-la com outras - não importando seu objetivo - utilizam-se da comunicação. Comunicar é compartilhar um modelo, torná-lo comum [FIS87].

Para que haja a comunicação é necessário um veículo, uma forma de as duas entidades comunicantes compartilharem o mesmo modelo. "Uma linguagem é um conjunto de signos e símbolos que permitem um grupo social de se comunicar e facilita o pensamento e as ações dos indivíduos" [FIS87]. Thro [THR91] define linguagem como sendo "um modo de comunicação estruturado e inteligente que é falado, escrito ou passado por sinais."

Os seres humanos utilizam-se principalmente da linguagem natural para se comunicarem. Para Savadovsky [SAV88], "a linguagem natural é uma das formas mais humanas de manifestação externa da atividade mental... Em particular, a linguagem é um meio muito rico para comunicação entre as pessoas." "Linguagem natural é a comunicação estruturada e inteligente entre pessoas. Ela consiste de sons organizados; vocabulário; estruturas, tais como alfabetos ou outras representações simbólicas; gramática ou sintaxe; significado dependente da estrutura ou semântica; e métodos de interpretação daquilo que é ouvido ou lido [THR91].

2.1.2 Interface homem-computador

Além de um veículo para a comunicação, as duas entidades devem possuir meios de se comunicar. Em um sentido amplo, uma interface é um dispositivo que serve de limite comum a várias entidades comunicantes, as quais se exprimem em uma linguagem específica a cada uma. Para que a comunicação seja possível, o dispositivo deve assegurar a conexão física entre as entidades e efetuar as operações de tradução entre os formalismos existentes em cada linguagem. Uma vez que a comunicação esteja estabelecida, a interação (ação recíproca) pode ocorrer entre as entidades [COU90]. Para Thro [THR91], uma interface é um local para encontro ou interação.

No caso da interface homem~computador, a conexão entre estas duas entidades se realiza entre a imagem do sistema (ou seja, sua manifestação externa) e os órgãos sensoriais-motores do usuário; a tradução se efetua entre os formalismos do sistema e os do usuário [cougo].

Barthet [BAR88] afirma que há uma linguagem de interação permitindo ao usuário, por meio de um vocabulário e de uma sintaxe, expressar as operações que ele deseja efetuar à máquina. Estas operações manipulam comandos ou dados.

O vocabulário desta linguagem de interação pode ser formado de diferentes modos:

- códigos numéricos;
- códigos mnemônicos;
- palavras da língua portuguesa;
- palavras desconhecidas;
- pictogramas.

Porém, segundo Scapin [apud BAR88], a utilização de palavras da língua natural do usuário (no caso, a língua portuguesa) facilita a interação, pois ele utilizará uma linguagem mais intuitiva.

2.2 Interface em Linguagem Natural

2.2.1 Diálogo homem-máquina

A língua, seja escrita ou falada, é o meio de comunicação entre o homem e a máquina [PIE87]. Até os anos 70, tinha-se um "mosteiro" de especialistas em informática para manipular o computador; homens que dominavam a utilização de linguagens especializadas/específicas no uso do computador. Na década de 80, com o advento e expansão de microcomputadores e suas redes, o usuário final foi colocado frente-a-frente com a máquina e seus dados. Desta forma, linguagens mais naturais, bem como meios de acesso mais naturais eram prementes.

Em adição ao exposto, Feigenbaum coloca que "a riqueza das nações será amanhã essencialmente a informação e o conhecimento" [apud PIE87]. Esta informação e conhecimento tem como suporte (veículo) básico nada mais que linguagem natural, visto que grande parte do conhecimento da humanidade encontra-se armazenado (geralmente escrito) nesta forma. Daí advém a necessidade de compreender esta linguagem, seja de forma escrita ou oral.

Pierrel [PIE87] e Anick [ANI93] entendem que com o crescimento de tamanho e importância das bases de dados, os computadores necessitam ter meios de interpretar a linguagem natural. Desta forma um usuário pode mais facilmente encontrar algum argumento de pesquisa na base de dados, sem se preocupar com sua especificação exata - seja em termos de comandos de busca, seja em palavras a

pesquisar - permitindo-lhe realizar suas pesquisas através do vocabulário que lhe é conhecido, sendo o computador responsável por oferecer-lhe sinônimos e ajuda direcionada ao argumento.

Scapin [apud BAR88] afirma que vários profissionais de informática estimam que a utilização da linguagem natural é o que melhor pode se oferecer ao usuário em termos de interface. Contudo para haver comunicação é necessário haver diálogo, senão se reduz a uma simples transmissão de informações que não satisfaz ao ser humano [PIE87]. Um exemplo desta situação é o comportamento de um correspondente ao telefone face uma secretária eletrônica. Com exceção das pessoas acostumadas a este tipo de aparelho, o correspondente se encontra despreparado, ao ponto de esquecer às vezes a mensagem que gostaria de transmitir ou de tentar provocar um diálogo para facilitar sua comunicação. Porém o diálogo ocorre normalmente por comunicação em multicanais [PIE87], ou seja, utilização da fala, de gestos, de olhares, etc. Fischler e Firschein [FIS87] denominam de linguagem do corpo (*body language*) todos os componentes não-verbais que compõem um diálogo, mas que podem afetar o sentido da comunicação pelo canal verbal.

2.2.2 Entradas/Saídas Naturais

A utilização de linguagem natural não garante que a interface seja natural. Isto é, fazer com que o usuário possa digitar seus comandos de acordo com seu vocabulário coloquial facilita seu acesso ao computador, porém oferecer-lhe uma interface através da qual ele consiga dar entrada a esta mesma linguagem por voz ou escrita manual (ou ambos) seria mais próximo ao modo comum dele comunicar-se.

Crane e Rtschev [CRA93] apontam que a utilização de reconhecimento de voz e de escrita manual derrubarão as barreiras que teclados, mouses e GUIs impõem à comunicação natural com o computador. Vários fabricantes de sistemas dizem que pode se esperar para ver computadores combinando alguma forma de entrada por voz e caneta dentro de poucos anos. Raj Reddy, diretor da *School of Computer Science* da Universidade de Carnegie Mellon e pesquisador há mais de 30 anos na área de fala, prediz que computadores pessoais utilizarão entradas por voz e caneta dentro de cinco anos [CRA93]. E com a miniaturização cada vez mais acentuada dos computadores, será possível pensar em computadores que podem ser usados como roupas ou chapéus, ou seja, tão pequenos e leves que são facilmente transportados e com entrada e saída por voz, através de um microfone direcional e de pequenos fones de ouvido, por exemplo.

Cada método de comunicação com um computador tem suas vantagens e desvantagens [TEB95]. Um mouse ou joystick não pode ser facilmente usado para digitação e um teclado ou a fala não são ideais para manobrar um cursor na tela. Muitos dispositivos podem ser associados a controle além de suas funções principais de posicionamento ou entrada de dados. Rudnický [RUD93] apresenta o seguinte quadro (Quadro 1) que aloca tarefas a modos de entrada. Alguns modos de entrada atendem atividades particulares melhor que outros, conforme as indicações: B representa que o modo de entrada é uma boa escolha para a atividade; A representa que o modo de entrada é meramente adequado; e I representa que o modo de entrada é inadequado.

Quadro I: Atividades versus modos de entrada [RUD93]

"Caneta, mouse, teclado e voz coexistirão no *desktop* do futuro" prediz Lempesis [MEZ93], afirmando que nenhuma opção de entrada dominará. O usuário selecionará seu dispositivo de entrada baseado na aplicação e nas suas preferências pessoais. Tebbutt [TEB95] vislumbra a utilização de várias opções de entrada, como reconhecimento de voz, canetas, telas sensíveis ao toque e outros estranhos aparelhos de realidade virtual, porém ele crê que nenhum se sobressairá como único, sendo a combinação de vários deles a forma de entrada em um futuro próximo.

Meisel [MEI93] afirma que um microfone conectado a um computador pessoal pode possuir várias finalidades, entre elas a de armazenar recados, a de verificar a identidade do locutor para fins de segurança e a de reconhecer palavras na fala. Destas funções, o reconhecimento de voz é a que tem maior potencial para fundamentalmente mudar o modo de interação com o computador: "A tecnologia está produzindo uma nova interface homem-computador".

"O objetivo da interação homem-computador é, ou deveria ser, prover uma interface tão natural quanto possível. De fato, a solução perfeita seria aquela na qual o usuário nem percebesse a utilização de uma linterface ... Então, talvez a falta de uma interface seja o Nirvana dos usuários de computador. expõe Tebbutt [TEB95].

Outros modos de interação homem-máquina podem ainda ser listados, como telas sensíveis ao toque, luvas (*datagloves*), sistemas de câmeras que captam gestos, sistemas que controlam o movimento dos olhos, etc. Todos estes podem ser formas de entrada/saída que proveriam os multicanais citados por Pierrel e a linguagem do corpo citada por Fischler e Firschein, no tópico 2.2.1.

2.3 Interfaces de voz

2.3.1 Introdução

As línguas naturais são principalmente faladas [MAI85]. Uma das razões históricas para a predominância da fala é que, visto ser o homem um animal que trabalha, é vantajoso utilizar a audição para a comunicação, deixando os demais sentidos livres para exercerem outras atividades, que podem ou não ter funções comunicativas [MAI85]. Rich [RIC88] ainda destaca que a linguagem escrita é uma invenção recente e ainda desempenha um papel menos crítico que a fala, na maioria das atividades.

Focalizando, então, a linguagem falada, destacam-se duas vantagens na sua utilização em um sistema de comunicação homem/máquina [PIE87]:

- a fala é indispensável em algumas situações, para substituir outros canais de comunicação, como a relação piloto-avião ou para deficientes físicos;

- a fala pode ser mais eficaz que outros modos de comunicação. Em comparação à comunicação escrita, a fala permite melhores performances em questões de tempo de resolução de problemas e emissão de mensagens.

Allen [ALL90] afirma que a utilização da fala humana como meio de entrada/saída expandiria enormemente o uso de sistemas em computador. Ele explica que os sistemas têm a tendência de se tornarem cada vez mais complexos, concluindo que, facilitando a interação homem-máquina, haveria uma maior utilização destes sistemas. "Reconhecimento de fala é um dos pontos-chave do cliente de negócios" diz Bob McBreen, gerente de produto para a Microsoft *Windows Sound System*. Ele acredita que o reconhecimento de fala será parte integrante da computação no futuro [MEI93].

2.3.2 Abordagens de reconhecimento de fala

Pierrel [PIE87] apresenta duas grandes abordagens para o reconhecimento. Estas abordagens não representam uma classificação rígida, podendo até ser complementares entre si a fim de solucionar o problema de reconhecimento.

- Método global e método analítico

O método global, também conhecido por reconhecimento de palavras, utiliza basicamente as técnicas de reconhecimento de formas para comparar, globalmente, a palavra a reconhecer das diversas formas de referência armazenadas.

O tratamento acústico preliminar é bem simples: a mensagem a identificar é considerada como uma forma atômica, sem problemas de segmentação. o ponto delicado é a melhor forma de representação da palavra.

Esta abordagem global se revela insuficiente se pretendesse tratar com grandes vocabulários ou fala contínua. Se faz, então, necessário adotar a abordagem analítica, que consiste em segmentar a mensagem em constituintes elementares (fonemas, meia-sílabas, sílabas, etc). Após identificados estes últimos, reconstituir enfim a frase pronunciada por etapas sucessivas: léxica, sintática, etc.

No quadro 2 são observadas aplicações dos dois métodos em relação à entrada da fala.

[Quadro 2: Diferentes tipos de reconhecimento \[PIE87\]](#)

- Reconhecimento x Compreensão

Uma segunda classificação foi proposta pelos pesquisadores do projeto ARPA, que introduziram o

termo Compreensão da fala (*speech understanding*) em oposição ao termo Reconhecimento de fala (*speech recognition*) .

Reconhecimento de fala consiste no reconhecimento de fonemas, sílabas, palavras para formar a mensagem original, como foi pronunciada. Como exemplo deste tipo tem-se as máquinas de ditar e editores de texto por fala.

Compreensão de fala baseia-se no entendimento do senso, do significado da mensagem, visando fazer com que o sistema execute algo. Para tal, são aceitos eventuais erros. Utilizando o exemplo de Verhaeghe [VER92]: em uma aplicação Windows, poderia ser falado "*Please start ... euh... spreadsheetll*". O comando é reconhecido como *start application spreadsheet*" ignorando algumas palavras (*please*) e subentendendo outras (*application*), pois o entendimento se dá por certas palavras-chave (*start spreadsheet*).

Esta distinção foi feita pois notou-se que a habilidade de um sistema em responder inteligentemente à fala era um critério mais significativo para a avaliação de sistemas de fala. Além disso, acreditou-se que o sinal de fala era uma pobre fonte de informação e que o conhecimento do contexto de uma pronúncia era essencial para seu reconhecimento e interpretação com sucesso [ALL90].

2.3.3 Reconhecedor automático de fala

Antes de se conhecer um sistema computadorizado de reconhecimento de fala, far-se-á uma verificação superficial de como isto ocorre naturalmente, ou seja, em seres humanos.

Para que haja comunicação é necessária a existência de dois personagens: o locutor e seu interlocutor, ou ainda, o emissor e o receptor da mensagem.

O emissor produzirá uma mensagem fazendo com que determinada idéia que possua seja transformada em sons, através do comando de nervos motores do aparelho fonador, ou seja, o emissor tem uma idéia e codificada em símbolos que são transmitidos ao receptor [THO81].

O receptor decodifica estes símbolos em um *código interno* (idéia) , ou seja, o receptor perceberá a mensagem, através de nervos sensoriais do seu aparelho auditivo, procurando transformar os sons recebidos na idéia original. A comunicação pode ser considerada boa caso haja um isomorfismo entre os estados internos de idéia tanto do emissor quanto do receptor [THO81].

Figura 1: Codificação e decodificação sucessivas na comunicação oral [PIE87].

Um reconhecedor automático de fala (*automatic-speech Recognition ASR*) será sempre o receptor da mensagem. Ele fará a percepção das ondas sonoras da mensagem e executará algum processamento procurando "captar a idéia" do emissor.

Figura 2: Configuração típica de um ASR [DAS92].

Uma configuração típica de um ASR pode ser vista na figura 2, onde há um microfone (fazendo as funções dos nervos sensoriais do aparelho auditivo) conectado a um sistema ou dispositivo (hardware +software) ligado ao computador.

Este dispositivo consiste basicamente em:

- hardware capaz de transformar as ondas sonoras captadas (sinal analógico) em números (sinal digital) para algum processamento pelo software;
- software que capte a massa de dados numéricos enviada pelo hardware, e reconheça alguma palavra ou execute algum comando.

No desenvolvimento e aprimoramento deste software é que se concentra o esforço de milhares de pesquisadores em todo o mundo. Em linhas gerais, pode se dizer que cabe a este software [ALL90] [BRI90]:

- análise do sinal e extração de parâmetros: a quantidade de bits por segundo gira na faixa de 50.000 nos sistemas com boa qualidade de recepção. Esta é uma massa muito grande de dados para ser tratada. Logo, devem ser aplicados meios de redução ou extração desta informação, sem perder as características do sinal representativo da mensagem. Para tal, são utilizadas várias técnicas, entre elas: transformada discreta de Fourier, banco de filtros, densidade de passagem por zero, etc;
- determinação do ponto-final da fala: determinar quando os dados recebidos não significam mais a fala propriamente dita, mas apenas ruído do ambiente, ou mesmo o silêncio;
- normalização em frequência e tempo: a necessidade de normalização em frequência origina-se do fato de que cada usuário-locutor possui um timbre e entonação diferentes, fazendo com que as frequências para um mesmo fonema sejam diferentes. A normalização em tempo advém das diferentes velocidades com que pode ser dita uma palavra;
- reconhecimento: após terem sido efetuados os passos anteriores (na ordem apresentada ou não), algum modo de identificação deve ser aplicado. Várias técnicas são utilizadas, como: comparações com dicionários, regras de produção, programação dinâmica, modelo escondido de Markov (*hidden Markov model*) e mais recentemente, redes neurais.

2.3.4 Aplicações do reconhecimento de fala

A imaginação é o limite para definir aplicações para este novo tipo de interface. Contudo, há que ser lembrado as atuais restrições existentes para estas aplicações. A primeira e mais importante delas é a capacidade: de memória e de processamento. As técnicas atualmente utilizadas não podem evoluir muito

mais em razão das restrições de capacidade. Isso acarreta em soluções hoje existentes que possuem contexto bem definido, ou seja, área de aplicação bem definida com vocabulário específico deste domínio, tendo como consequência uma pequena quantidade de palavras, para que possam ser obtidos resultados em tempo real.

Pierrel [PIE87] coloca que é vantajoso restringir o universo do diálogo, pois desta forma não será necessário tanto conhecimento a respeito de sintaxe, semântica, prosódia, etc. diminuindo a possibilidade de ambigüidades e simplificando o processo de coleta de informações.

Fischler e Firschein [FIS87] lembram que normalmente o vocabulário utilizado pelas pessoas é bem menor do que se pode imaginar. os quadros 3 e 4 demonstram isto. Eles são para o inglês, porém são semelhantes para português, francês ou qualquer outra língua natural. O quadro 3 mostra que o vocabulário normal de um adulto consiste de cerca de 10% das palavras de um dicionário resumido.

Quadro 3: Tamanho do vocabulário empregado por várias fontes. Kendratov [apud FIS87].

No quadro 4 pode-se observar que com um vocabulário de 3000 palavras, consegue-se reconhecer 90% das palavras em um página de texto comum. Com um vocabulário de 1000 palavras consegue-se o mesmo índice de reconhecimento para palavras faladas.

Quadro 4: Frequência do uso de palavras na linguagem escrita e oral. Kendratov [apud FIS87].

Os sistemas atualmente existentes trabalham com uma quantidade que varia de algumas palavras, passando por centenas, chegando até 50.000 nos sistemas mais sofisticados [MEI93]. O Tangora, da IBM, é um sistema tipo máquina de escrever comandada por voz, o qual trabalha com as 20.000 palavras mais comuns em ambiente de escritório. Contudo possui uma outra restrição: o ditado deve ocorrer de forma mais pausada. Esta tecnologia é conhecida como entrada de fala isolada ou discreta, em contraposição à entrada de fala contínua. Porém Tebbutt [TEB95] lembra que o teclado padrão QWERTY foi projetado para que as pessoas não conseguissem datilografar muito rapidamente, o que travaria o sistema mecânico das máquinas de escrever da época. Desta forma, a pequena pausa, porém não-natural, realizada entre as palavras durante o ditado para um sistema de reconhecimento de fala pode ser encarada como normal, visto que a demanda da tecnologia tem precedentes sobre os desejos do usuário.

Mesmo obedecendo às restrições comentadas, várias aplicações poderiam ser destacadas:

- em linhas de produção para descartar peças falhadas através de comandos vocálicos para movimentar a esteira de produção;
- na mesma idéia de movimentar esteira, manuseio de bagagens aéreas (de mão ou não);

- em situações mãos e olhos ocupados , como por exemplo, para o motorista discar um telefone em um carro em movimento, ou um piloto dentro da cabine de um avião;
- auxílio a deficientes físicos;
- em caixas bancários automáticos, para solicitar saldos, extratos, etc;
- operações por telefone, como consultas de saldos, de notas em colégio, etc
- preparo de relatórios médicos ou odontológicos durante ou após a consulta.

2.4 Histórico dos trabalhos em reconhecimento de fala

Pierrel [PIE87] relata que contrariamente à síntese de fala, cujos primeiros trabalhos datam do século XVII, foi preciso esperar o meio deste século para aparecerem as primeiras experiências em reconhecimento.

Conforme Pierrel [PIE87], por volta de 1930, o americano R.J.Wensley construiu o TeleVox,, primeiro autômato capaz de receber ordens por telefone e executar alguns movimentos correspondentes.

Os primeiros sistemas realmente de reconhecimento de voz aparecem somente em 1950: Daves, em 1952, apresenta um sistema inteiramente de cabos capaz de reconhecer os dez números pronunciados por um locutor. Sistema este aperfeiçoado em 1958 para uma versão que aceita diversos locutores; Olson e Belar, em 1956, propuseram um sistema ambiciosamente chamado de máquina de escrever fonética, capaz também de reconhecer uma dezena de palavras; Denes, em 1958, define um sistema em duas etapas, no qual a primeira etapa realiza um reconhecimento puramente acústico que, na segunda, será refinado pela utilização de conhecimentos lingüísticas.

Em 1960, a aparição dos métodos numéricos e a utilização do computador dão uma nova dimensão a estas pesquisas. Em 1966, sistemas em laboratório conseguem identificar corretamente de trinta a cinquenta palavras ditas por diferentes pessoas. Estas experiências eram baseadas na comparação das formas das palavras. No início dos anos 70, a programação dinâmica otimiza este tipo de abordagem.

Em 1968 um importante passo é alcançado, Alter e Reddy verificam a utilidade das informações lingüísticas no reconhecimento da fala. Contudo, os primeiros trabalhos a utilizarem esta realização são os de Vicens, em 1969, e Tubach, em 1970. Estas pesquisas sobre sistemas de reconhecimento de palavras isoladas continuam a progredir em variados eixos.

Paralelamente surgem as primeiras investigações no tratamento de fala contínua, que vão seguindo fracamente até 1973, quando é publicado o relatório Newell, fortemente influenciado pela Inteligência

Artificial, para relançar estes estudos no quadro de um projeto americano financiado pela ARPA (*Advanced Research Projects Agency*) de 1971 a 1976. Os objetivos deste projeto definiam que o sistema deveria aceitar a fala contínua, aceitar um grande número de locutores cooperativos, trabalhar num ambiente calmo, com um bom microfone, compreender um vocabulário de mil palavras, utilizar uma sintaxe artificial no escopo de uma tarefa precisa, responder em tempo próximo ao real (*a few times real time*) em uma máquina de cem MIPS (milhões de instruções por segundo).

Este projeto faz surgir várias pesquisas em reconhecimento de fala contínua. Muitos sistemas foram propostos. Julgado positivo, o projeto ARPA serviu de catalisador de pesquisadores neste domínio [PIE87].

Sistemas como DRAGON (Baker em 1975), HEARSAY (Lesser, Fennel, Erman e Reddy em 1975) e HARPY (Lowerre em 1976) trabalhavam com um discurso contínuo de um único usuário e um vocabulário de até 1.000 palavras, obtendo taxas de acerto entre 84 e 97%. O TANGORA da IBM (citado no tópico 2.3.5) tem uma versão lançada na metade da década de 80, a qual sacrificava a fala contínua para atingir uma taxa de acerto de 97% para um vocabulário de 20.000 palavras. Um sistema criado nos Laboratórios Bell em 1987 para o reconhecimento de dígitos para números de telefone, também obteve 97% de precisão. o SPHINX (Lee e Hon em 1988) foi o primeiro sistema a alcançar alta precisão (96%) em fala contínua, independente do locutor e em tempo real com vocabulário de 1.000 palavras [RIC93].

Não apenas nos Estados Unidos estas pesquisas vêm sendo realizadas. Ao longo da década de 80, várias são as empresas que se lançaram em busca do reconhecimento de fala, bem como muitos centros de pesquisa. Pode-se citar, entre outras [PIE87]:

- EUA - IBM e SRI (*Stanford Research Institute*) ;
- Japão - NEC (*Nippon Electronic Company*), o projeto PIPS (*Pattern Information Processing Signal*), NTT (*Nippon Telegraph and Telephon Public Corporation*) e várias universidades;
- Europa - Alemanha, Itália, Inglaterra, França, etc.

No final da década de 80, Teuvo Kohonen da Universidade de Tecnologia de Helsinki desenvolveu uma máquina de escrever por voz utilizando a combinação de várias tecnologias disponíveis: processamento de sinal digital, sistemas baseado em regras e redes neurais. A máquina obteve alguns bons resultados. Foi testada utilizando-se casos extremos de conversão fala-texto: vários locutores, fala contínua e grande vocabulário. Com um treinamento de apenas 100 palavras por locutor, a máquina chegou a uma taxa de 92 a 97% de acerto nas conversões, com delay de 1/4 de segundo de resposta. São bons resultados, mas não ainda suficientes para aplicações comerciais. Contudo, o casamento de várias tecnologias permitiu a visualização de um futuro promissor [CAU92].

Em 1994, vários artigos proclamam a chegada do reconhecimento de voz no mercado. Fritz

[FRI94] escreve que o reconhecimento de fala finalmente ganha algum respeito, prevendo que até o ano de 1999 esta área deve movimentar uma cifra em torno de US\$ 1 bilhão. E testes conduzidos pela Seybold Publications [SEY94] consideraram os sistemas *Personal Dictation System* da IBM para OS/2 e *Dragon Dictate* da Dragon Systems para MS-DOS como impressionantes, pois apesar dos problemas detectados quando da existência de sons semelhantes ("to many people" e "too many people") a disponibilização dos resultados imediatamente permite ao usuário fazer qualquer correção necessária.

2.5 Considerações Finais

É cada vez maior a necessidade do homem interagir de forma mais natural com o computador, ou genericamente, com as máquinas. Modos de entrada/saída naturais vêm sendo vislumbrados para prover esta comunicação, que acarreta em várias facilidades e vantagens.

A fala é um dos meios de comunicação mais naturais entre os seres humanos. Assim, é de suma importância a aquisição e utilização de tecnologias para torná-la realidade nas interfaces de computador. Ao longo das últimas décadas vários trabalhos foram desenvolvidos, porém ainda não há a possibilidade de uma utilização plena de tal forma de interação.

O estudo apresentado neste capítulo demonstra que houve uma significativa evolução nos resultados de pesquisas desenvolvidas para alcançar estas tecnologias. O aumento da capacidade dos computadores - processamento e memória - contribuiu decisivamente para estes avanços.

Contudo, vale ressaltar que a comunicação humana não ocorre apenas através da fala e sim através de multicanais, formados pelo conjunto dos sentidos. Grandes avanços científicos e tecnológicos ainda estão reservados para estas áreas de pesquisa.

