

# Survey sobre Reconhecimento de Fala

Vitor M. N. Fernandes

Cadeira de Interfaces Multimodais Inteligentes (IMMI – 2004/2005),  
Mestrado em Engenharia Informática e de Computadores (MEIC),  
Instituto Superior Técnico (IST)  
Av. Rovisco Pais, 1000 Lisboa  
vmnf@{yahoo.com, mega.ist.utl.pt}

## Sumário

*O reconhecimento de fala é uma área onde ainda existem possibilidades de melhoria mas que neste momento já atinge uma performance muito boa. O documento mostra o funcionamento genérico de um sistema de reconhecimento desde a captura do sinal áudio voz pelo microfone até às palavras ditas. São descritas as técnicas e modelos utilizados no reconhecimento, como o processamento de sinal, a programação dinâmica, os modelos de Markov não observáveis, a modelação acústica e a modelação da linguagem. Apresentam-se as principais áreas de aplicação como o Reconhecimento Automático de Fala (ASR). O estado da arte mostra as potencialidades actuais desta tecnologia que já permite taxas de erro abaixo dos 10%. A secção trabalho futuro mostra o que ainda falta fazer como atingir a total independência do orador (ou adaptativo) e da aplicação, reconhecer em qualquer local e qualquer tipo de discurso.*

## Palavras-chave

*Reconhecimento de Fala, Fala, Tecnologia para Reconhecimento de Fala, Reconhecimento Automático de Fala.*

## 1. INTRODUÇÃO

Este documento está organizado em várias secções: reconhecimento de fala, aplicações, o estado da arte e direcções futuras.

O Reconhecimento de Fala consiste em extrair o texto das palavras faladas a partir de um sinal de voz. Na secção 2 apresentam-se as várias técnicas e modelos que podem ser utilizados neste reconhecimento. De entre as várias técnicas apresentada destaco o processamento de sinal, a programação dinâmica, os modelos de Markov não observáveis, a modelação acústica e a modelação da linguagem.

A secção aplicações está dividida nas várias áreas em que o reconhecimento de fala intervém. É possível aplicar esta tecnologia: a) na criação de interfaces de comando por voz em que sejam estes discretos ou contínuos, b) na realização de ditados, c) na identificação do emissor para autenticação e d) na criação de condições de acessibilidade para pessoas com deficiência.

Na figura 1 verifica-se a tendência claro de melhoria na redução do número de erros ao longo do tempo.

A última secção direcções futura mostra o caminho que ainda falta percorrer para tornar o reconhecimento de fala numa tecnologia robusta e eficiente.

O estado da arte mostra a evolução dos sistemas de reconhecimento de fala e o estado actual da tecnologia em já permite erros inferiores a 10%<sup>1</sup> o que é considerado aceitável [Baker03].

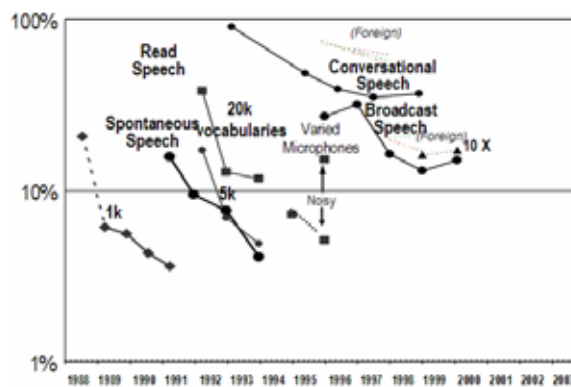


Figura 1: Número de erros [%] vs. anos de 1988 a 2000

<sup>1</sup> Em leitura de texto, ditado ou teleprompter. Em conversação natural é mais complicado garantir esta qualidade. On TV and in movies, characters never say "Uh, what? I didn't get that last thing you said" [Kotelly03].

## 2. RECONHECIMENTO DE FALA

### 2.1 Uma solução genérica

Uma solução genérica para a questão do reconhecimento de fala passa pela criação de vários módulos, cada um responsável por melhorar a detecção do módulo anterior [Picone95].

A solução genérica apresentada é constituída por 6 módulos em que a saída do anterior é a entrada do seguinte (figura 2).

Os módulos constituintes são: o modelo do sinal e os geradores de hipóteses do som, das sílabas, das palavras, das frases e da acção a realizar.

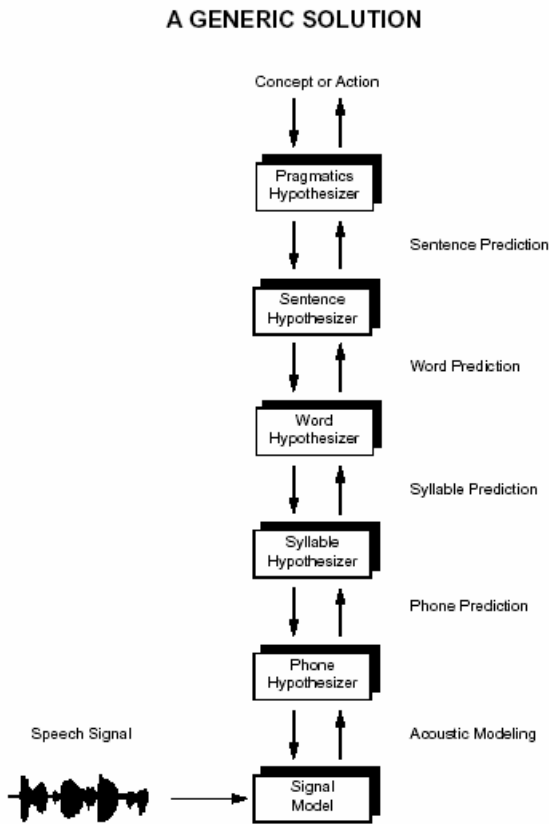


Figura 2: Uma solução genérica para reconhecimento de fala

O reconhecimento de padrões pode fazer-se através de Markov não observáveis que leva ao reconhecimento dos símbolos contidos no sinal (figura 2).

Um dos problemas principais do reconhecimento de fala é quando para os símbolos reconhecidos não é possível desambiguar. Esta situação é frequente quando o início e o final da frase são semelhantes (figura 3) [Baker03, Picone95].

Na situação da figura 4 para as regiões que se sobrepõem é necessário utilizar o contexto para desambiguar. Como já foi referido isso pode não ser possível quando o início e o final das frases são semelhantes.

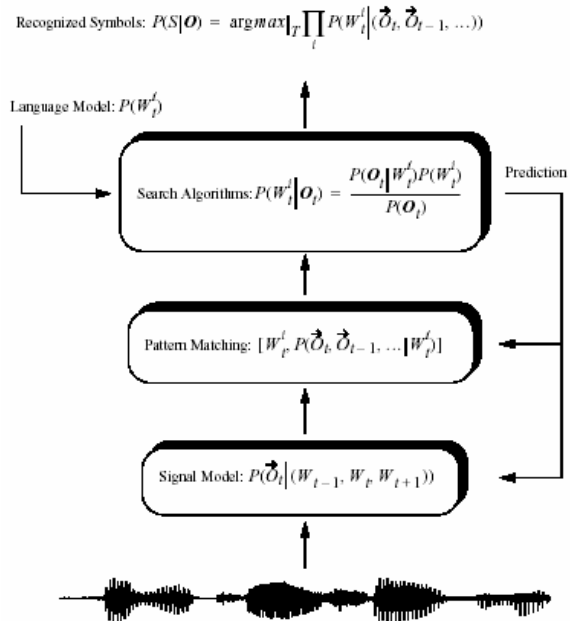


Figura 3: Reconhecimento de padrões baseado em modelos de Markov não observáveis

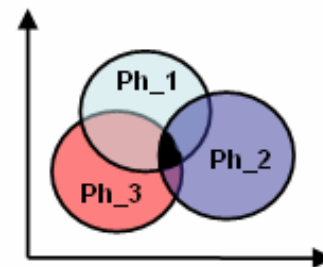


Figura 4: Reconhecimento de padrões baseado em modelos de Markov não observáveis

A análise do sinal áudio é também uma variável importante pois a mesma fala capturada por diferentes microfones gera sinais diferentes [Picone95] (figura 5).

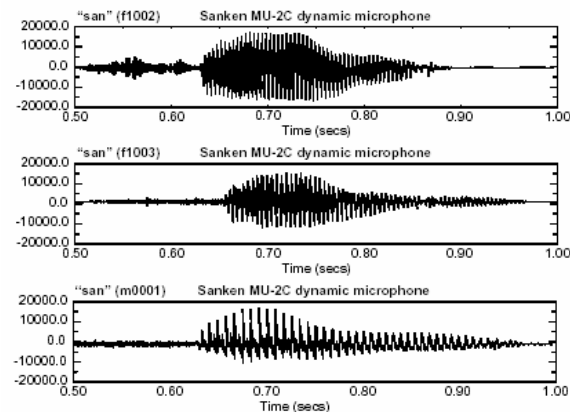


Figura 5: Sinal áudio para a mesma palavra gravado por diferentes microfones

**2.2 Arquitecturas para reconhecimento de fala**

**2.2.1 Aproximação teórica à comunicação**

Formulação baysiana para o reconhecimento de voz cujo objectivo é o de minimizar o número de erros cuja abordagem é a de maximizar  $P(W|A)$  sendo  $P(W|A) = P(A|W) \cdot P(W) / P(A)$  sendo  $P(A|W)$  o modelo acústico e  $P(W)$  o modelo da linguagem [Bakker03].

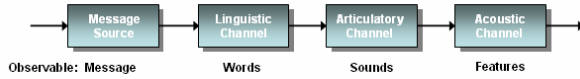


Figura 6: Aproximação teórica à comunicação

**2.2.2 Incorporação de múltiplas fontes de conhecimento**

O sinal é convertido numa sequência de vectores de características baseados em características temporais e espectrais do sinal.

Os modelos acústicos representam partes de palavras (fonemas) como uma máquina de estados finita em que os estados modelam a estrutura espectral e as transições.

O modelo da linguagem prevê o próximo conjunto de palavras e controla quais as hipóteses a considerar.

A pesquisa é fundamental pois terá de investigar várias combinações de palavras e descobrir qual a sequência mais provável [Bakker03].

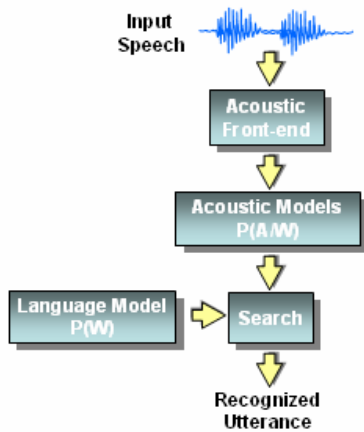


Figura 7: Incorporação de múltiplas fontes de conhecimento

**2.3 Modelação acústica**

**2.3.1 Extracção de características**

Incorpora conhecimento da natureza dos sons da fala em medidas de características. Usa modelos rudimentares da percepção humana.

Tipicamente usa 512 amostras a 16KHz ou seja uma janela de 30ms para a realização da análise. A análise consiste na medida no domínio da energia espectral [Bakker03].

**2.3.2 Modelos de Markov não observáveis**

Os modelos acústicos codificam a evolução temporal das características do sinal (espectro). Distribuições Gaussianas são utilizadas para assimilar as variações no emissor, pronúncia e sotaque.

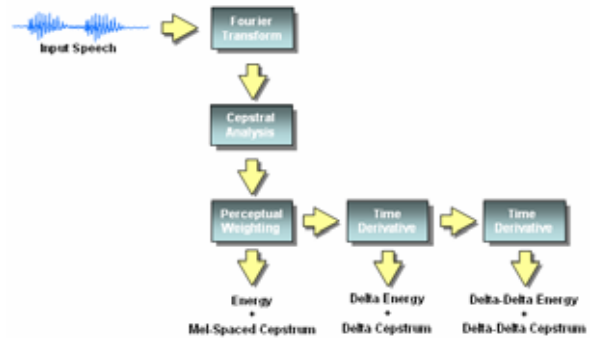


Figura 8: Extracção de características

O modelo dos fonemas é uma estrutura da esquerda para a direita. O salto de estados (*time-warping*) e múltiplos caminhos (pronúncias alternativas) são características destes tipos de modelos [Bakker03].

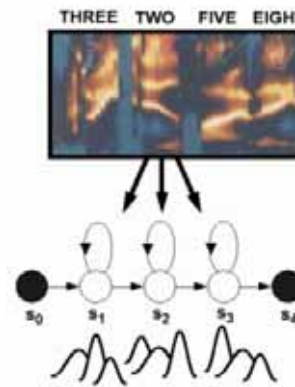


Figura 9: Modelos de Markov não observáveis

**2.3.3 Estimação de parâmetros**

Transcrição ao nível das palavras. Modelação em círculo fechado supervisionada. Algoritmos são usados para melhorar os parâmetros estimados. São geradas árvores de decisão que em conjunto com o conhecimento da língua são usados para melhorarem os parâmetros e a complexidade do sistema [Bakker03].

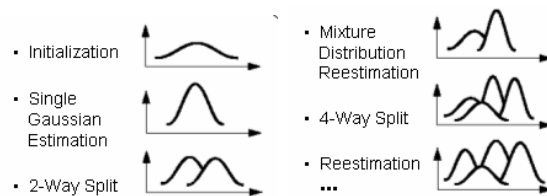


Figura 10: Estimação de parâmetros

Supondo que se detectou  $c\_s\_$  quais as letras que originam palavras (um pouco como a roda da fortuna). Nesta situação pode ser “casa”, “caso”, etc (nos dois exemplos dados o ‘c’ e o ‘s’ têm exactamente o mesmo som). De seguida a escolha da melhor hipótese está directamente ligada à probabilidade da sua ocorrência.

**2.4 Modelação da Linguagem**

**2.4.1 N-Grams**

As probabilidades de ocorrência de determinados N-Grams determinam a resposta do reconhecedor [Bakker03].

ker03]. (N corresponde ao número de sons na palavra e/ou expressão – tabelas 1 a 3)

Mais comuns	“I”, “and”, “the”, “you”, “a”
Menos Comuns	“Abraham”, “Alastair”, “Acura”

Tabela 1: Unigrams (SWB)

Mais comuns	“you know”, “yeah SENT!”, “!SENT um-hum”, “I think”
Menos Comuns	“raw fish”, “moisture content”, “Reagan Bush”

Tabela 2: Bigrams (SWB):

Mais comuns	“!SENT um-hum SENT!”, “a lot of”, “I don’t know”
Menos Comuns	“you have parents”, “you seen Brooklyn”

Tabela 3: Trigrams (SWB):

2.4.2 Integração da língua natural

Os constrangimentos da língua natural podem ser facilmente incorporados. Tem como dificuldades a falta de pontuação e o espaço de procura. [Picone95,Baker03]

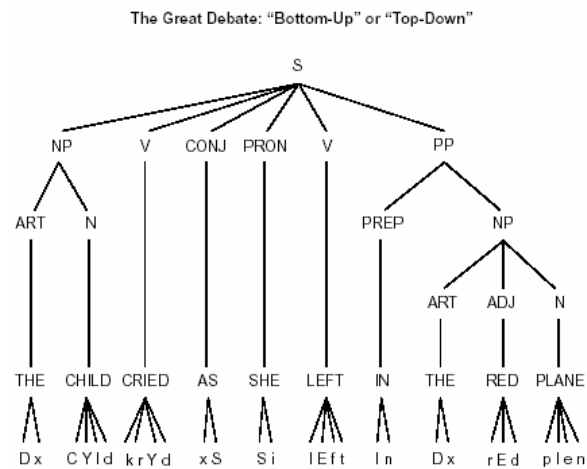


Figura 11: Integração da língua natural

2.4.3 Programação dinâmica

A programação dinâmica é uma técnica que permite, neste caso concreto, determinar qual o caminho mais provável através de um grafo dirigido com pesos nas transições [Bakker03].

2.4.4 Palavras que se intersectam podem ser dispendiosas

Como os limites das palavras não ocorrem no discurso espontâneo, nós temos de seguir os sons que estendem os limites das palavras [Bakker03].

2.5 A pesquisa usa os recursos disponíveis de forma intensiva

Existem vários métodos de pesquisa como: Time Synchronus (Viterbi DP), State Synchronus (Baum-Welch), Stack Decoding (n-Best, Best-First, A\*) e esquemas

híbridos como: Forward-Backward, Multipass e Fast Matching. [Picone95]

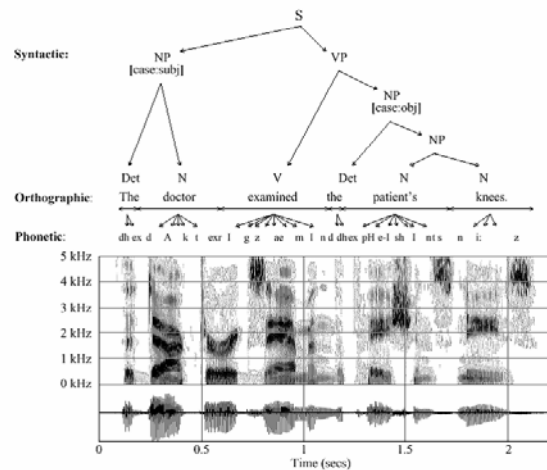


Figura 12: Integração da língua natural

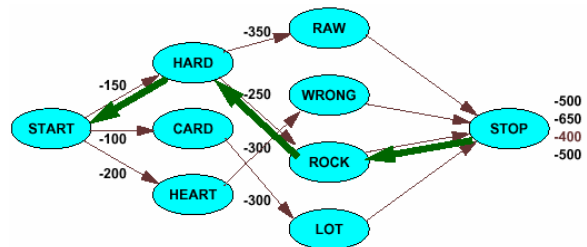


Figura 13: Programação dinâmica

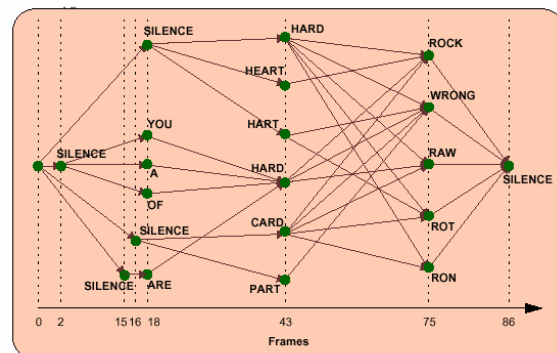


Figura 14: Programação dinâmica

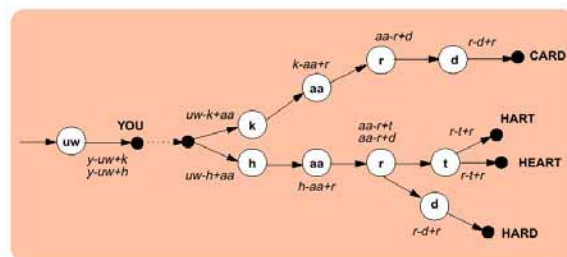


Figura 15: Palavras que se intersectam

São apresentados os requisitos de memória e de tempo de CPU para os vários elementos constituintes do sistema [Bakker03].

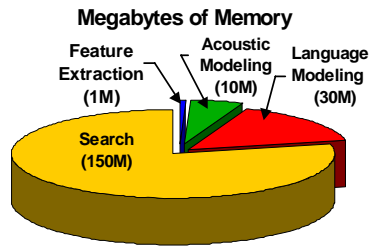


Figura 16: Memória

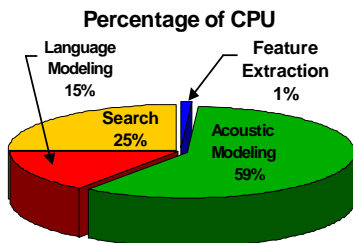


Figura 17: Percentagem da UCP

### 3. APLICAÇÕES

#### 3.1 Especialização dos sistemas

Os sistemas de reconhecimento de fala podem ser otimizados para várias situações de utilização segundo 5 vectores: independência do orador, portabilidade do sistema, tamanho do vocabulário, discurso contínuo e microfone remoto (desconhecido) ou local [7].

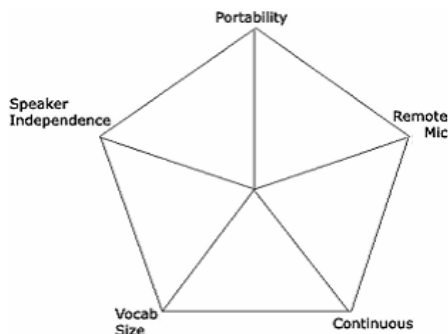


Figura 18: Linhas de especialização da tecnologia do Reconhecimento de Fala Automático (RFA)

#### 3.2 Reconhecimento de fala contínuo

##### 3.2.1 Reconhecimento de fala

Existem no mercado várias aplicações que permitem a realização do reconhecimento de fala.

Da IBM o Via Voice / Voice Server Applications que permite independência em relação ao emissor, reconhecimento de comandos contínuos, vocabulário grande. Da Philips que permite ditados, reconhecimento via telefone, controlo por voz (SpeechWave, VoCon SDK, chip-sets).

Existem ainda outras empresas que têm trabalhado sobre o reconhecimento de fala como a Dragon Systems / Ler-

nout & Hauspie (L&H Voice Xpress™), Microsoft's SAPI, NeuVoice, WATSON (AT&T Advanced Speech Products Group), Whisper engine (Microsoft Research Speech Technology Group), IN CUBE recognizer (Command Corporation), DECIPHER - EduSpeak - DynaSpeak (SRI International's STAR Lab), Mac OS.

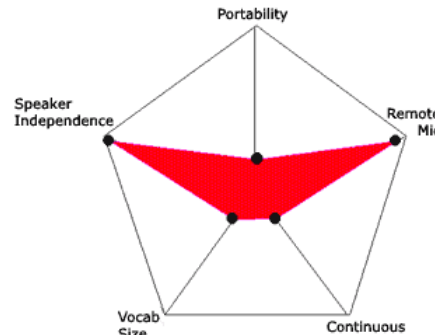


Figura 19: Motor RFA otimizado para aplicações telefónicas

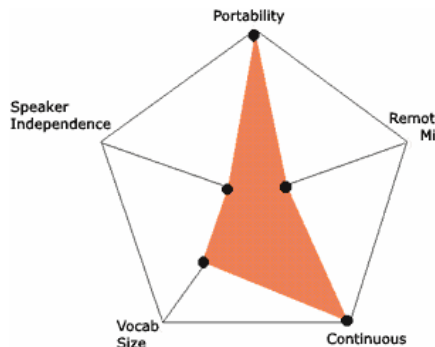


Figura 20: Motor RFA otimizado para aplicações industriais

##### 3.2.1.1 Atendimento telefónico - O sistema Jupiter

O sistema Jupiter permite realizar o reconhecimento de fala sobre telefone.



Figura 21: O sistema Júpiter

De seguida apresento a transcrição de um exemplo de utilização do sistema desenvolvido pelo MIT, que demonstra ser um sistema com um índice de reconhecimento elevado e com recuperação de erros.

**Jupiter:** Hi. Welcome to Jupiter, the M.I.T. Lab for Computer Science Weather Information System. This call is being recorded for system development. You may hang up or ask for help at any time. How can I help you?

**Human:** What is the forecast for Boston today?

**Jupiter:** In Boston this afternoon, partly sunny and windy. High around 50. Tonight, clear, lows around 35. What other information can I give you?

**Human:** Will it rain tomorrow?

**Jupiter:** To my knowledge, the forecast calls for no rain in Boston tomorrow. Is there anything else?

**Human.:** How about Seattle?

**Jupiter:** To my knowledge, the forecast calls for no rain in Seattle tomorrow. Is there something else you would like to know?

**Human:** Is it sunny anywhere in the Caribbean?

**Jupiter:** Sorry, I am not sure what you said.

**Human:** Is it sunny anywhere in the Caribbean?

**Jupiter:** I am sorry, I may not know the city you are asking about. Try asking for the state or country. For example, what cities do you know about in Massachusetts?

**Human:** What is the temperature in Montego Bay?

**Jupiter:** In Montego Bay today, high 86 and low 73. Is there something else?

**Human:** Good-bye.

### 3.2.2 Ditados

Porquê escrever quando pode falar? Esta é a questão principal no reconhecimento de fala em ditados. Com precisões que rondam os 99% (informação comercial do Dragon Naturally Speaking [12]) será então um “novo paradigma” para a escrita. As vantagens claras surgem na comparação do ritmo de escrita (120 palavras digitadas / minuto contra 160 palavras escritas / minuto).

Existe outra versão o Dragon Dictate que faz uma pausa após reconhecer cada palavra (ver secção sobre acessibilidade).

Outros sistemas existentes: IBM ViaVoice, RealSpeak e Philips Dictation Systems.

### 3.2.3 Tradução em tempo real

Citação do Estado da Nação do presidente Clinton em Janeiro de 2000: “These kinds of innovations are also propelling our remarkable prosperity... Soon researchers will bring us devices that can translate foreign languages as fast as you can talk... molecular computers the size of a tear drop with the power of today’s fastest supercomputers.” [Baker03].

Nesta vertente podemos considerar sem intervenção humana e apenas através da voz (via telemóvel ou telefone fixo) consultas a bases de dados, tradução para todas as línguas (português-chinês e vice-versa), identificação por voz entre outras.

Uma aplicação recente (2005) foi a de reduzir a barreira linguística e melhorar as comunicações no Iraque (“Real-time speech translation decreases language barriers and improves communications in Iraq. Linguistic data collection and software programming are increasing warfighters’ abilities to interact effectively with Iraqis.” [11]).

### 3.2.4 Conversação através de canais ruidosos

Detecção de voz através de canais ruidosos, como seja o canal telefónico, onde surge o ruído de fundo, flutuações na velocidade da fala, palavras cortadas, hesitações, ruídos vocais entre outros.

Mesmo nas condições acima citadas em 6 anos (desde 1998 até 2003) conseguiu-se uma redução dos 100% para os 30% no WER – Word Error Rate [Baker03].

Sistemas que funcionam via telefone: Nortel Networks, SpeechWorks, Nuance, Fonix, Locus Speech Corporation, Vocalis, Philips Speech Processing offers, BaBel Technologies SA.

#### 3.2.4.1 Voz sobre o Telefone

Esta é também uma área onde se tem verificado um interesse crescente de modo a automatizar o atendimento a clientes, por exemplo.

Os pioneiros nesta área foram a AT&T Bell Labs. Outras companhias como a Nuance, a Philips e a SpeechWorks têm estado activa nesta área já há alguns anos. A IBM disponibiliza um conjunto de serviços como o pedido de notícias, páginas de Internet, correio electrónico e outras informações de forma automática.

Destacam-se as seguintes aplicações da Philips:

- SpeechPearl® que tem um vocabulário vasto (200 000 palavras) para reconhecimento de língua natural;
- SpeechMania® que simula um operador humano verdadeiro;
- SpeechWave™ que tem a limitação de um vocabulário pequeno (centenas de palavras) mas que está disponível em dezenas de linguagens;
- Voice ReQuest que permite fazer o encaminhamento de chamadas entre extensões.

### 3.2.5 Indexação de Notícias TV

A indexação deste tipo de informação vídeo levanta várias dificuldades.

O léxico tem muita informação em palavras pouco frequentes, a modelação acústica varia em função da transmissão de estúdio vs. na rua, uma adaptação contínua do modelo da língua em função do tempo e a transmissão de notícias em língua não nativa.

#### 3.2.5.1 SSNT - Summarization of Broadcast News Services

Trabalho realizado no L2F (INESC-ID) – SSNT [9] que permite através do reconhecimento de fala transcrever as notícias do telejornal, indexá-las e recuperá-las através de um motor de pesquisa.

## 3.3 Reconhecimento de fala discreto

### 3.3.1 Atendimento a clientes automático

Esta é outra área onde os sistemas de reconhecimento de fala estão a evoluir.

Destaca-se o atendimento automático para e-banking, e-mail, consultas ao estado do tempo, etc.

### 3.3.2 Consultas em motor usando fala

Trabalho realizado no L2F (INESC-ID) – Tumba [10] que permite a realização de consultas para páginas Web portuguesas através de voz. (Existe a possibilidade de sintetizar as palavras reconhecidas.)

### 3.3.3 Sistemas embebidos

Nos sistemas embebidos destaco os das empresas: Natural Speech Communication, VoxTec, Rubidium Ltd.'s Dialog Engine, ART, Sensory Inc., Fluent Speech Technologies, Images, Voice Signal, voice INTER connect .

## 3.4 Interfaces de comando por voz

### 3.4.1 “Put that there” e “Put that where? Voice and Gesture at the graphics Interface”

São certamente as referências mais importantes na área da utilização do reconhecimento de fala numa interface de comando por voz e gesto ([Bolton80] e [Billinghamurst98] respectivamente).

### 3.4.2 Adição de voz à interface WIMP

Existem vários trabalhos e no projecto NavMI [Fernandes05] o autor verificou que conseguia ter taxas de erro inferiores a 5% na interface de comando num sistema robusto através da utilização 2 modos de utilização do reconhecedor da Loquendo [13].

## 3.5 Identificação

A verificação de identidade não se deve confundir com o reconhecimento de fala pois enquanto uma reconhece o que se DIZ a outra reconhece QUEM diz [16]. Sistemas que realizam este tipo de verificação têm protecções contra ataques como o spoofing através de testes liveness

## 3.6 Acessibilidade

### 3.6.1 Dificuldades motoras

No caso de não ser possível digitar uma possibilidade é falar [14]. Para tal as aplicações de ditado podem ser usadas bem como aplicações que permitem o dizer uma palavra de cada vez para quem tenha também dificuldade ao nível da fala. A aplicação Dragon Dictate será uma das mais indicadas pois pode aprender palavras a partir de sons de treino.

Outro trabalho muito interessante é o The Vocal Joystick: A Voice-Based Human-Computer Interface for Individuals with Motor Impairments [Bilmes05] que permite a realização de um controlo analógico utilizando voz como se esta fosse um joystick.

### 3.6.2 Dificuldades auditivas

Ferramentas como os TTS (Text-To-Speech ou leitores de ecrã) podem em conjunto com o reconhecimento de fala permitir uma interacção boa para pessoas com dificuldades visuais ou mesmo invisuais. A integração sugerida na referência [14] é uma das soluções possíveis para a criação de um ambiente deste tipo com p.e. o Jaws for Windows [15].

## 4. O ESTADO DA ARTE

### 4.1 Factores e evolução dos sistemas de reconhecimento de fala

Existem vários factores que condicionam o reconhecimento de fala como seja o ruído ambiente, o estilo de conversação, o emissor e a complexidade.

De seguida apresenta-se uma imagem que permite ter uma ideia da evolução dos sistemas de reconhecimento de fala desde 1985 até ao próximo ano de 2005 (figura 22).

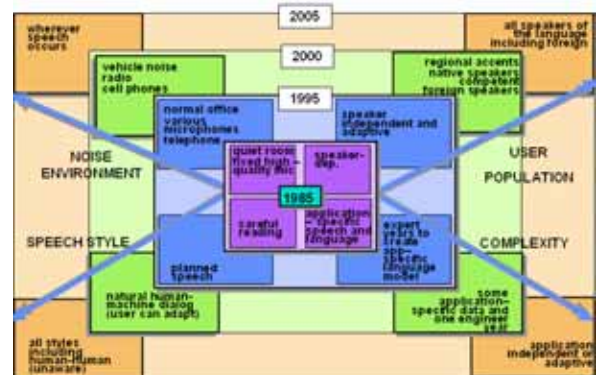


Figura 22: Factores condicionantes e evolução dos sistemas de reconhecimento de fala

Da imagem apresentada verifica-se que a expectativa para 2005 será o reconhecimento de fala em qualquer tipo de ambiente, para todos os tipos de emissores – inclusive estrangeiros e para todo o tipo de discurso.

#### 4.1.1 Segmentação do sinal de fala

Apresento as conclusões de um trabalho sobre segmentação do sinal áudio de voz em função da quantidade de informação presente no sinal (entropia) [Waheed00].

Este trabalho permite melhorar significativamente a qualidade do reconhecimento em ambientes ruidosos. Com melhorias que vão de 9% para SNR de 15dB até 15,6% para SNR de 5dB.

Este tipo de abordagem permite ter mais ruído no ambiente.

#### 4.1.2 Questões relevantes sobre reconhecimento de fala

Existe um conjunto de questões que foram analisadas por engenheiros e cujo resultado se apresenta de seguida [Huckvale96]:

- O discurso real é mais variável do que o pensado;
- Um detalhe muito elevado é irrelevante;
- O desempenho de baixo nível é pior do que o pensado;
- Os emissores apenas são semelhantes a eles próprios;
- A consequência da obscuridade leva à ambiguidade;
- Segmentar o conhecimento e não a representação;
- A mediocridade está presente em tudo;
- A unidade é a força;

- i) Avaliar e não analisar;
- j) O reconhecimento é apenas uma otimização.

## 5. DIREÇÕES FUTURAS

A evolução desde os anos 60 levou-nos da análise bases de dados analógicos, *time-warping* dinâmico e neste milénio aos modelos de Markov não observáveis [Bakker03].

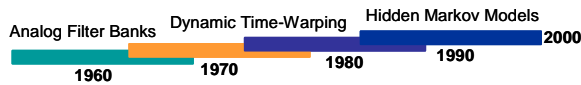


Figura 23: Direções futuras

No futuro esperam-se melhorias em várias áreas de modo a chegarmos a sistemas quase infalíveis.

Alguns desafios a considerar são:

- Discriminação vs. representação;
- Generalização vs. memorização;
- Modelação da pronúncia;
- Modelação da linguagem centrada nos humanos.

## 6. REFERÊNCIAS

### 6.1 Artigos científicos

- [Bakker03] E.M. Bakker, Seminar Speech Recognition 2003, *LIACS Media Lab - Leiden University*, 2003.
- [Billinghurst98] Billinghurst, Mark. Put that where? Voice and Gesture at the graphics Interface. *Human Interface Technology Laboratory, University of Washington, 1998*.  
<[http://portal.acm.org/ft\\_gateway.cfm?id=307730&type=pdf&coll=GUIDE&dl=ACM&CFID=48796827&FTOKEN=72480303](http://portal.acm.org/ft_gateway.cfm?id=307730&type=pdf&coll=GUIDE&dl=ACM&CFID=48796827&FTOKEN=72480303)>
- [Bilmes05] Bilmes, Jeff; Li, Xiao; Malkin, Jonathan; Kilanski, Kelley; Wright, Richard; Kirchoff, Katrin; Subramanya, Amarnag; Harada, Susumu; Landay James A.; Dowden, Patrícia; Chizeck, Howard, The Vocal Joystick: A Voice-Based Human-Computer Interface for Individuals with Motor Impairments, *Dept. of Electrical Engineering, Dept. of Linguistics, Dept. of Computer Science, Dept. of Speech & Hearing Science, University of Washington, Seattle, WA*, Julho de 2005.  
<<https://www.ee.washington.edu/techsite/papers/documents/UWEETR-2005-0007.pdf>>
- [Bolton80] Bolton, Richard A.. Put-that-there. *Architecture Machine Group, Massachusetts Institute of Technology*, 1980.  
<[http://www.media.mit.edu/speech/papers/1980/bolt\\_SIGGRAPH80\\_put-that-there.pdf](http://www.media.mit.edu/speech/papers/1980/bolt_SIGGRAPH80_put-that-there.pdf)>
- [Cettolo03] Cettolo, Mauro, Brugnar, Fabio, Federico, Marcello, Advances in the Automatic Transcription of Lecture, 2003.
- [Fernandes05] Fernandes, Vitor, Navegador Multimodal para Imagens (NavMI), *Instituto Superior Técnico – INESC-ID, Julho 2005*,  
<<http://www.vmnf.net/vmnf/ist/meic/cadeiras/12-immi/projecto/>>.

[Huckvale96] Huckvale, Mark, 10 things Engineers have Discovered about Speech Recognition.

[Jelinek02] Jelinek, Fred, Speech Recognition, *IBM - 2002*,  
<<http://ska.ai-depot.com/paper/node4.html>>

[Kanedera97] Kanedera, Noburu, Arai, Takayuki, Hermansky, Hynek, Pavel, Misha, On the Importance of Various Modulation Frequencies for Speech Recognition.

[Kotelly03] Kotelly, Blade, The Art and Business of Speech Recognition. *Ubiquity, An ACM IT Magazine and Forum, January 2003*.

[Picone95] Picone, Joseph, Fundamentals of Speech Recognition – A Short Course.

[Robbins04] Robbins, Christopher A., Speech and Gesture Based Multimodal Interface Design, April 2004.  
<<http://ska.ai-depot.com/paper/node4.html>>

[Tsai99] Tsai, Mitchell, Reiher, Peter, Popek, Jerry. Operating System Services for Managing Context.

[Waheed00] Waheed, Khurram, Weaver, Kim, Salam, Fathi M. A Robust Algorithm for Detecting Speech Segments using an Entropic Contrast

### 6.2 Páginas Web

- [1] Seminar Speech Recognition 2003,  
<<http://www.liacs.nl/home/erwin/speechrecognition.html>>
- [2] Speech – Ready for Mainstream,  
<<http://www.proxenus.com/whitepapers/speech-mainstream.htm>>
- [3] Voice Recognition Software – Comparison and Recommendation,  
<<http://www.io.com/~hcexres/tcm1603/achtml/rtecomx7c.html>>
- [4] Speech Recognition Engines,  
<<http://slydiman.compass.com.ru/eng/links/isreng.htm>>
- [5] System Requirements for Dual-Engine Automated Transcription,  
<<http://www.custumspeechusa.com>>
- [6] TypeWell – Speech Recognition,  
<<http://www.tywell.com/speechrecog.html>>
- [7] TechOnline – Speech Recognition Technology Backgrounder,  
<[http://www.techonline.com/community/ed\\_resource/feature\\_article/20044](http://www.techonline.com/community/ed_resource/feature_article/20044)>
- [8] Commercial Speech Recognition (2003),  
<<http://www.speechstudio.com/commercial/speech.htm>>
- [9] SSNT - Summarization of Broadcast News Services (2005),  
<<http://www.l2f.inesc-id.pt/projects/SSNT/SSNT-en.php>>
- [10] www.tumba.pt,  
<<http://www.l2f.inesc-id.pt/projects/tumba.html>>
- [11] Real-time speech translation decreases language barriers and improves communications in Iraq

- <<http://www.jfcom.mil/newslink/storyarchive/2005/pa062005.htm>>
- [12] Dragon NaturallySpeaking 8.0,  
<<http://www.freedomofspeech.com/fos/dnsdetails.html>>
- [13] Loquendo,  
<<http://www.loquendo.com>>
- [14] Speech Technology for Disabilities,  
<<http://www.speechtechnology.com/products/disabilities/index.html>>
- [15] Jaws for Windows,  
<<http://www.speechtechnology.com/products/disabilities/jaws.html>>
- [16] Speech recognition / Speaker biometrics, Markowitz J., Consultants,  
<<http://www.jmarkowitz.com/ask.html>>