



Survey

(<http://www.vmnf.net/vmnf/ist/meic/cadeiras/12-immi/survey/>)

Survey sobre Reconhecimento de Fala

Vitor M. N. Fernandes

vmnf@yahoo.com

IST/DEI: MEIC: IMMI

(v1.00 - 2005/07/21)



Agenda

1. Introdução
2. Reconhecimento de fala
3. Aplicações
 - 3.1 Especialização dos sistemas
 - 3.2 Reconhecimento de fala contínuo
 - 3.3 Reconhecimento de fala discreto
 - 3.4 Interfaces de comando por voz
 - 3.5 Identificação
 - 3.6 Acessibilidade
4. O estado da arte
5. Direcções Futuras
6. Referências



1. Introdução

Reconhecimento de fala:

Área onde ainda existem muitas possibilidades de evolução
No momento actual (2005) já atinge uma performance aceitável

Apresentar uma solução genérica de um sistema de reconhecimento desde a captura do sinal áudio voz pelo microfone até às palavras ditas.

Ver quais as técnicas e modelos utilizados no reconhecimento
Principais áreas de aplicação como o Reconhecimento Automático de Fala (ASR).

Potencialidades actuais desta tecnologia que já permitem taxas de erro abaixo dos 10%.



2. Reconhecimento de fala

2.1 Uma solução genérica

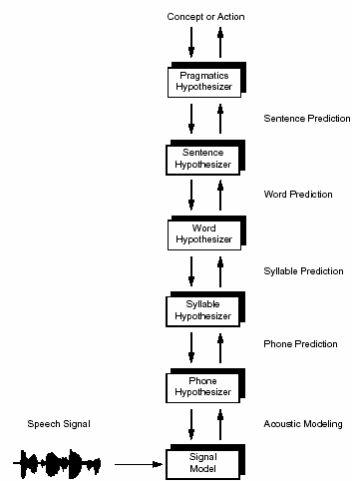
Vários módulos:

- A entrada de cada módulo é a saída do módulo anterior

Deteção de ...

- Frases
- Palavras
- Sílabas
- Sons
- Sinal analógico

A GENERIC SOLUTION

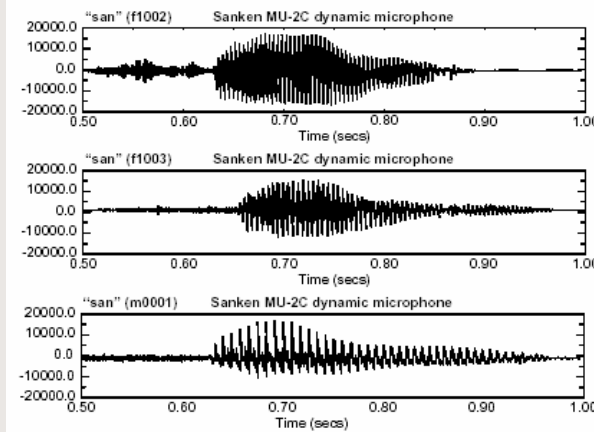




2. Reconhecimento de fala

2.1 Uma solução genérica

O sinal é muito variante em função da tecnologia (microfone) utilizado



Vitor M. N. Fernandes
IST/DEI: MEIC: IMMI

5
Julho 2005

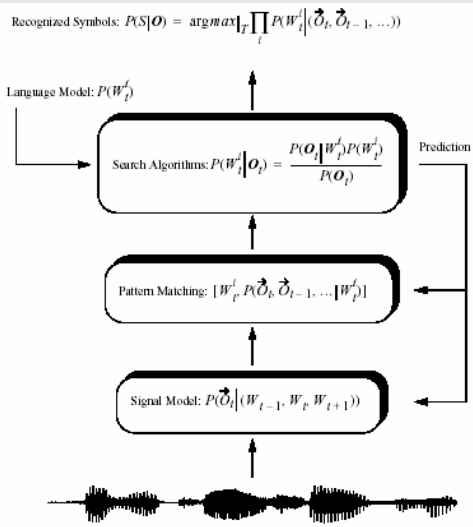
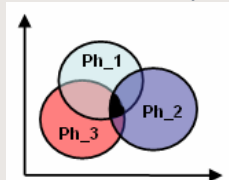


2. Reconhecimento de fala

2.1 Uma solução genérica

Reconhecimento de padrões baseado em modelos de Markov não observáveis

A desambiguação pode ser uma tarefa complexa



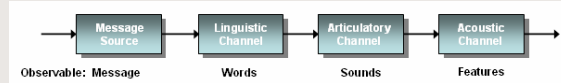
Vitor M. N. Fernandes
IST/DEI: MEIC: IMMI

6
Julho 2005

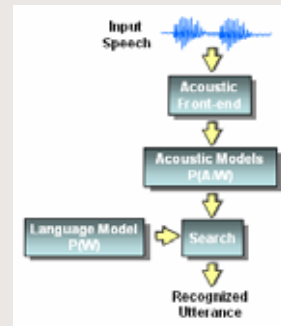


2. Reconhecimento de fala

2.2 Arquitecturas para reconhecimento de fala Aproximação teórica à comunicação

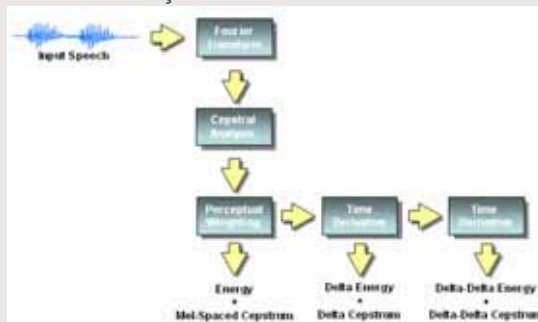


Incorporação de múltiplas fontes de conhecimento

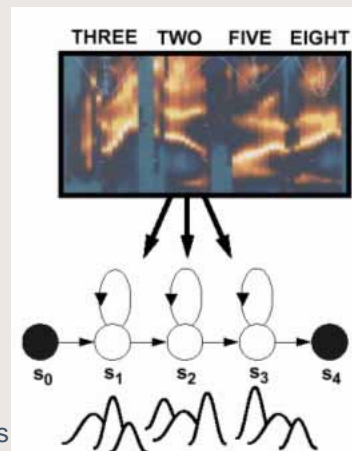


2. Reconhecimento de fala

2.3 Modelação acústica Extracção de características



Modelos de Markov não observáveis

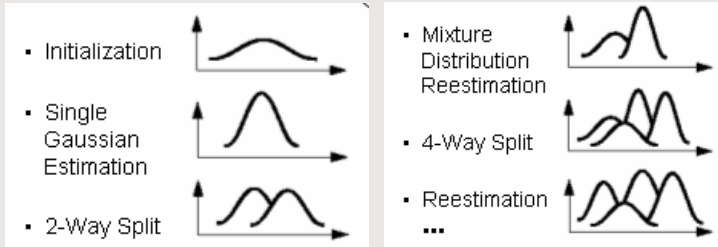




2. Reconhecimento de fala

2.3 Modelação acústica

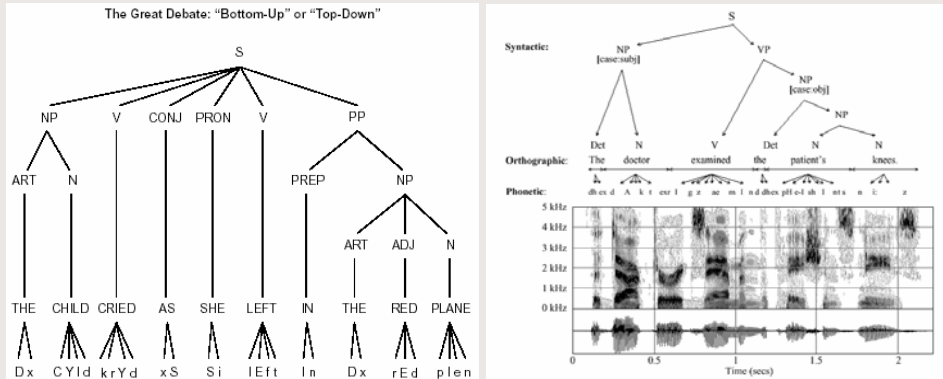
Modelação de parâmetros



2. Reconhecimento de fala

2.4 Modelação da linguagem

Gramáticas

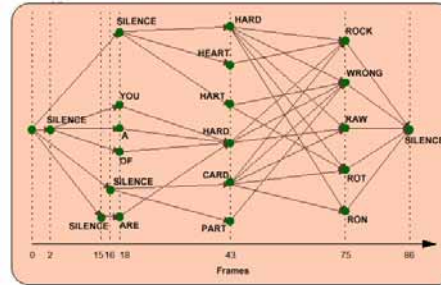
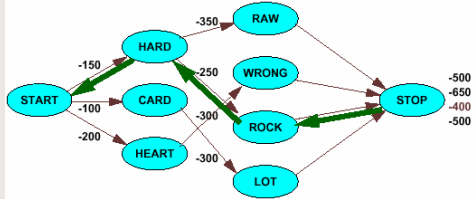




2. Reconhecimento de fala

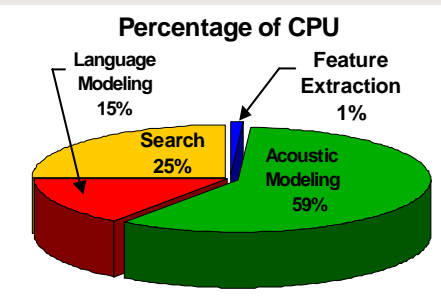
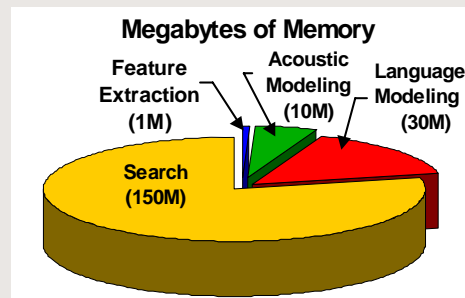
2.4 Modelação da linguagem

Grafos, pesos e caminhos – Programação dinâmica



2. Reconhecimento de fala

2.5 A pesquisa usa os recursos disponíveis de forma intensiva



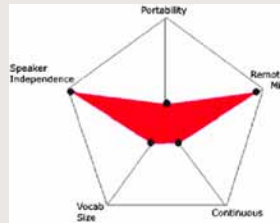


3. Aplicações

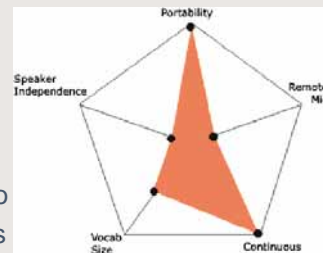
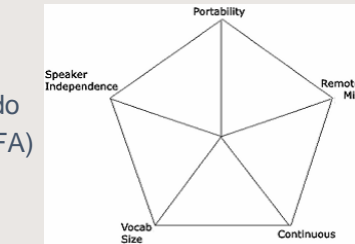
3.1 Especialização dos sistemas

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

Motor RFA
Otimizado
para
aplicações
telefónicas



Motor RFA otimizado
para aplicações industriais



3. Aplicações



3.2 Reconhecimento de fala contínuo – Sistema Jupiter

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. Aplicações

3.2 Reconhecimento de fala contínuo

Ditados

IBM ViaVoice, RealSpeak e Philips Dictation Systems.

Tradução em tempo real

Em utilização no Iraque para traduzir Inglês – Iraquiano

Conversação através de canais ruidosos

Nortel Networks, SpeechWorks, Nuance, Fonix,
Locus Speech Corporation, Vocalis, Philips Speech Processing offers,
BaBel Technologies SA

Indexação de Notícias TV

SSNT - Summarization of Broadcast News Services (L2F – INESC-ID)



3. Aplicações

3.3 Reconhecimento de fala discreto

Atendimento a clientes automático

e-banking, e-mail, consultas ao estado do tempo,...

Consultas em motor usando fala

Tumba (L2F – INESC-ID)

Sistemas embebidos

Natural Speech Communication, VoxTec,
Rubidium Ltd.'s Dialog Engine, ART, Sensory Inc.,
Fluent Speech Technologies, Images, Voice Signal,
voice INTER connect .



3. Aplicações

3.4 Interfaces de comando por voz

“Put that-there” e

“Put that where? Voice and Gesture at the graphics Interface”

Voz e gesto na interface de comando

Adição de voz à interface WIMP

NavMI

3.5 Identificação

Sistemas de identificação biométricos por voz

BIOPersona™



3. Aplicações

3.6 Acessibilidade

Dificuldades motoras

Dragon Dictate

The Vocal Joystick

Dificuldades auditivas

Jaws (leitor de ecrã) + sintetizador de voz (TTS)

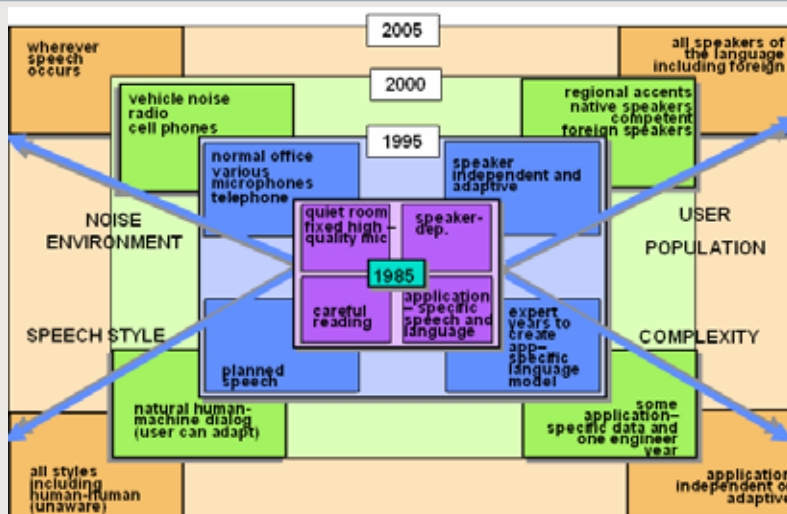


3. Aplicações

- "Talk to me, baby" (2005): A Honda e a IBM melhoram o software de reconhecimento de voz
 - Opção por \$2,000 para o
 - 2005 Acura MDX sport-ute e
 - Honda Odyssey minivan
 - Utilizado para
 - Pedidos e respostas de navegação
 - Localizar restaurantes
 - Controlo do rádio
 - Controlo da climatização
- A Honda usa o reconhecimento de voz da IBM desde 2003



4. O estado da arte





4. O estado da arte

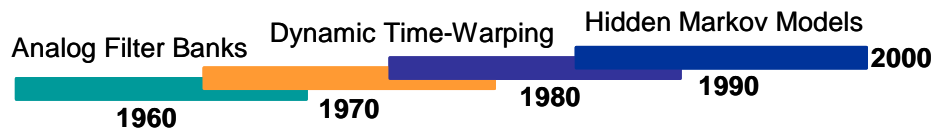
■ Comparativo IBM Via Voice vs. Dragon NaturallySpeaking

Operating system	Software used	Dictating into:	First script	% accuracy
Windows XP	Via Voice 9	Speak Pad	Alice in Wonderland	96.4
Windows XP	Via Voice 10	Speak Pad	Alice in Wonderland	98.4
Windows XP	Dragon NS 6	Dragon Pad	Alice in Wonderland	97.2
Windows XP	Dragon NS 7 Pro	Dragon Pad	Alice in Wonderland	98.8
Windows XP	Dragon NS 8 Pro	Dragon Pad	Alice in Wonderland	98.9

- O comparativo foi realizado utilizando apenas uma pessoa em 2005 para as duas das mais utilizadas aplicações de reconhecimento de fala (ditado)
- Resultados podem diferir se repetidos!



5. Direcções futuras



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

Alguns desafios a considerar são:

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



6. Referências - Artigos

- [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&CFTOKEN=72480303>
- [Bilmes05] Bilmes, Jeff; Li, Xiao; Malkin, Jonathan; Kilanski, Kelley; Wright, Richard; Kirchoff, Katrin; Subramanya, Amarnag; Harada, Susumu; Landay James A.; Dowden, Patricia; 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/UWEEETR-2005-0007.pdf>>
- [Bolton80] Bolton, Richard A.. Put-that-where. *Architecture Machine Group, Massachusetts Institute of Technology*, 1980.
<http://www.media.mit.edu/speech/papers/1980/bolt_SIGGRAPH80_put-that-where.pdf>
- [Cettolo03] Cettolo, Mauro; Brugnara, Fabio; Federico, Marcello. Advances in the Automatic Transcription of Lecture, 2003.
- [Fernandes05] Fernandes, Vitor, Navegador Multimodal para Imagens, *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://sika.ai-depot.com/paper/node4.html>>
- [Kanedera97] Kanedera, Noburu, Arai, Takayuki, Hermansky, Hyeon, 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://sika.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

Vitor M. N. Fernandes
IST/DEI: MEIC: IMMI

23
Julho 2005



6. Referências – 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/acchtml/recomx7c.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 Background, <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>>

Vitor M. N. Fernandes
IST/DEI: MEIC: IMMI

24
Julho 2005



Survey

(<http://www.vmnf.net/vmnf/ist/meic/cadeiras/12-immi/survey/>)

Survey sobre Reconhecimento de Fala

Vitor M. N. Fernandes

vmnf@yahoo.com

IST/DEI: MEIC: IMMI

(v1.00 - 2005/07/21)