

# *Opportunità applicative del riconoscimento automatico del parlato: il futuro della ricerca*



**Piero Cosi**



**ISTITUTO DI SCIENZE E TECNOLOGIE DELLA COGNIZIONE  
SEDE DI PADOVA - "FONETICA E DIALETTOLOGIA"**

Via Martiri della Libertà, 2 - 35137 Padova (Italy)

e-mail: [piero.cosi@pd.istc.cnr.it](mailto:piero.cosi@pd.istc.cnr.it)

www: <http://www.pd.istc.cnr.it>



**" La comunicazione vocale uomo-macchina "**

Incontro con De Mori (Université d'Avignon, Francia)

Martedì 11 marzo febbraio 2008 - Ore 10,30

Centro Congressi Palazzo Rospigliosi, Sala delle Statue

Via XXIV Maggio, 43 – Roma



# ISTC – SPFD CNR



<http://www.pd.istc.cnr.it/LUCIA>



<http://www.pd.istc.cnr.it/FESTIVAL>

+ 20.000 download

**SONIC  
SPHINX  
CLSU Toolkit**

ISTITUTO DI SCIENZE E TECNOLOGIE DELLA COGNIZIONE  
SEDE DI PADOVA - "FONETICA E DIALETTOLOGIA"

[www.pd.istc.cnr.it](http://www.pd.istc.cnr.it)

 Consiglio Nazionale delle Ricerche



**Associazione Italiana  
di Scienze della Voce**

***AISV***



**Associazione Italiana di Scienze della Voce**  
**[www.AISV.it](http://www.AISV.it)**

# Possible Interactions

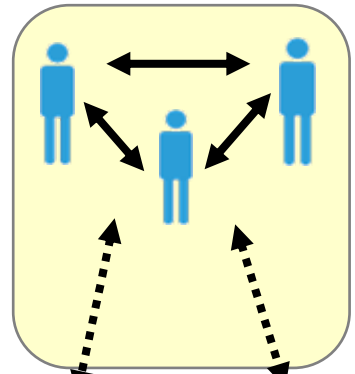
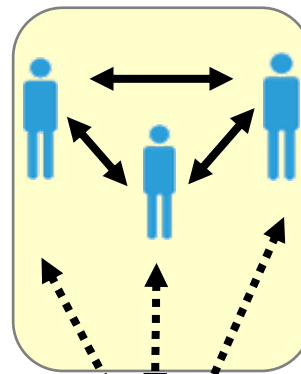
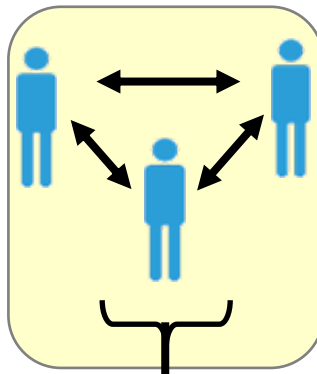
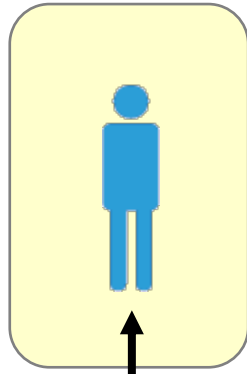
monologue

interactions

conversations

interactive  
conversations

multiple dialogs



System

User



# *How to Improve ASR Performance?*

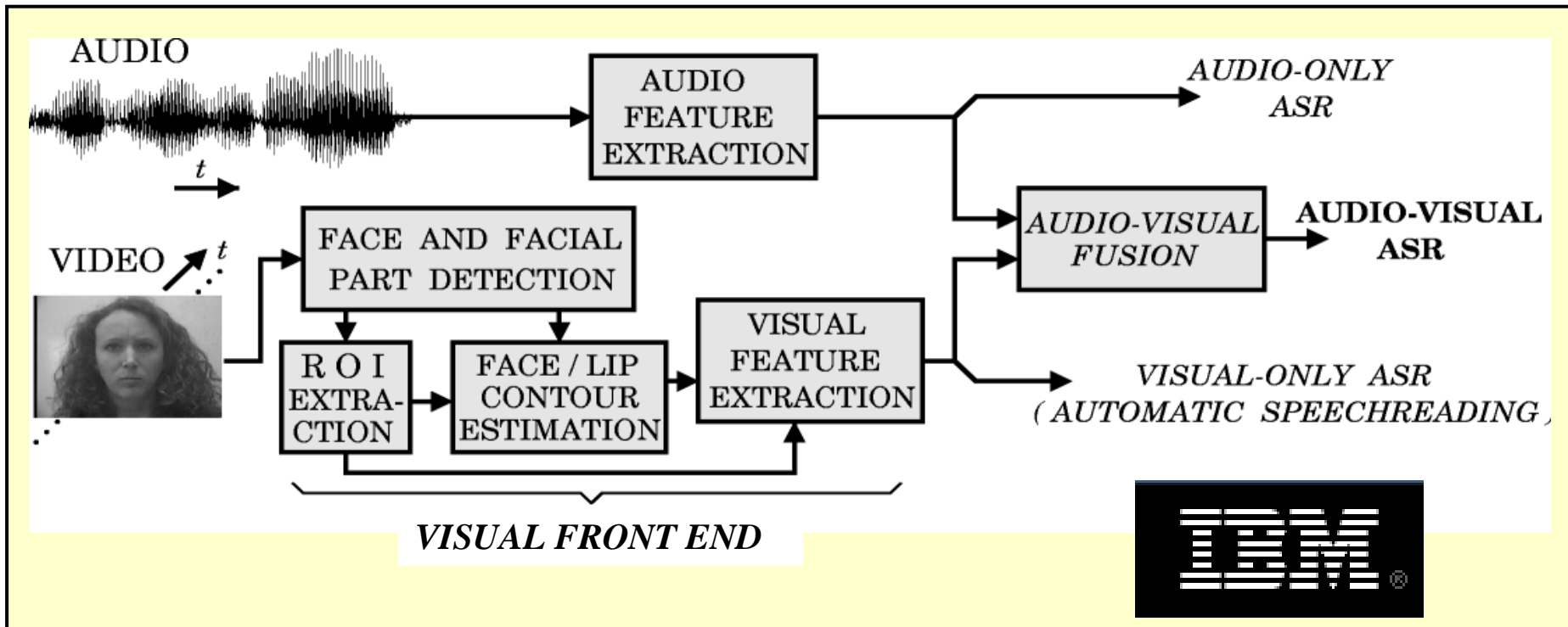
There are a number of issues that impact the performance of an automatic speech recognition (ASR) system:

- **Type of Speech** (read, continuous, spontaneous, ...)
- **Type of Channel** (microphonic, telephonic, ... )
- **Speaker Characteristics**
- **Speaking Rate**
- **Noise**
- **Vocabulary**
- .....

*Before deploying applications....  
be sure to offer **GOOD** performance!  
remember **DICTATION** (IBM, DRAGON...)*

***WE HAVE TO STUDY MORE!***

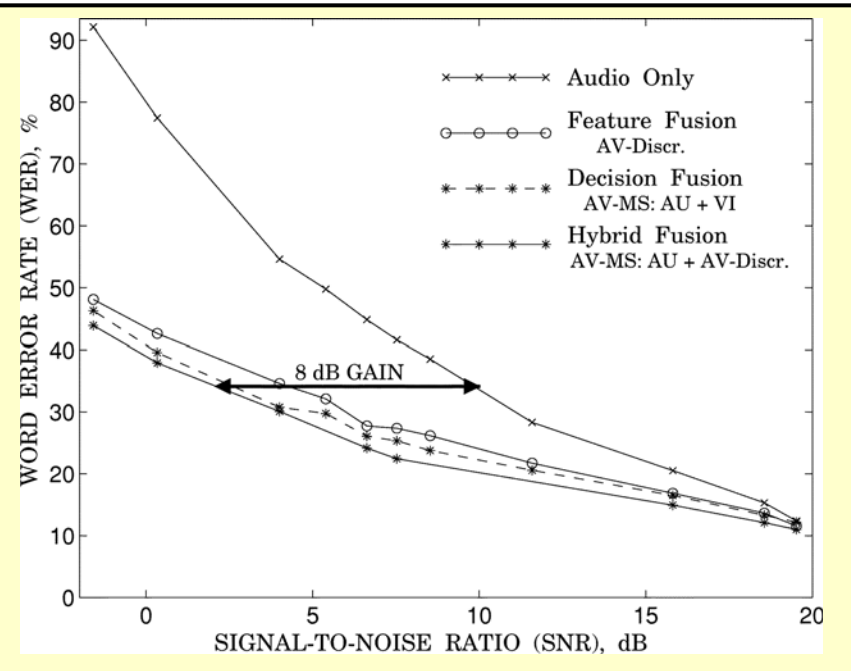
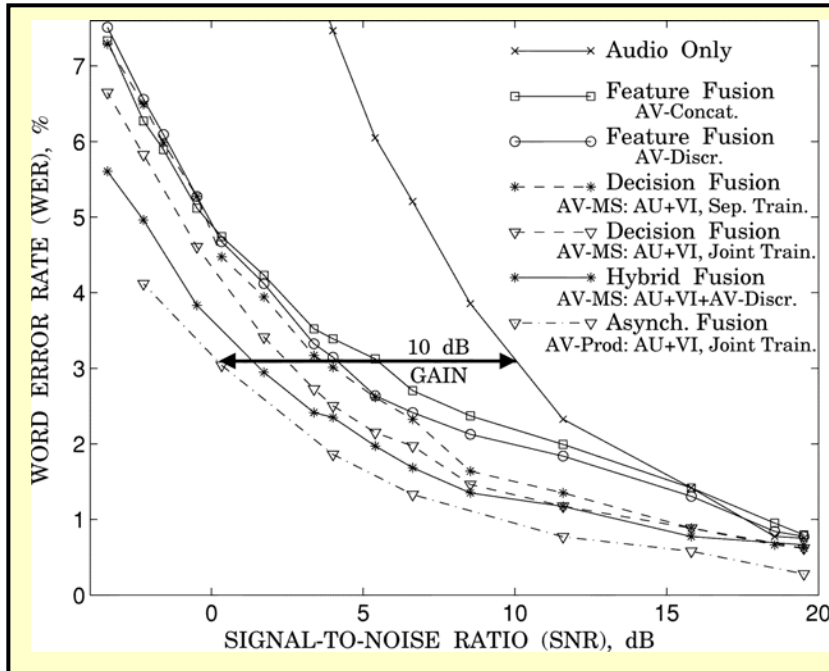
# Audio-Visual ASR



G.Potamianos, C.Neti, G. Gravier, A. GARG, and A.W. Senior, "Recent Advances in the Automatic Recognition of Audiovisual Speech",

Proc. IEEE, Vol.91, N.9, Sept. 2003, pp.1306-1326.

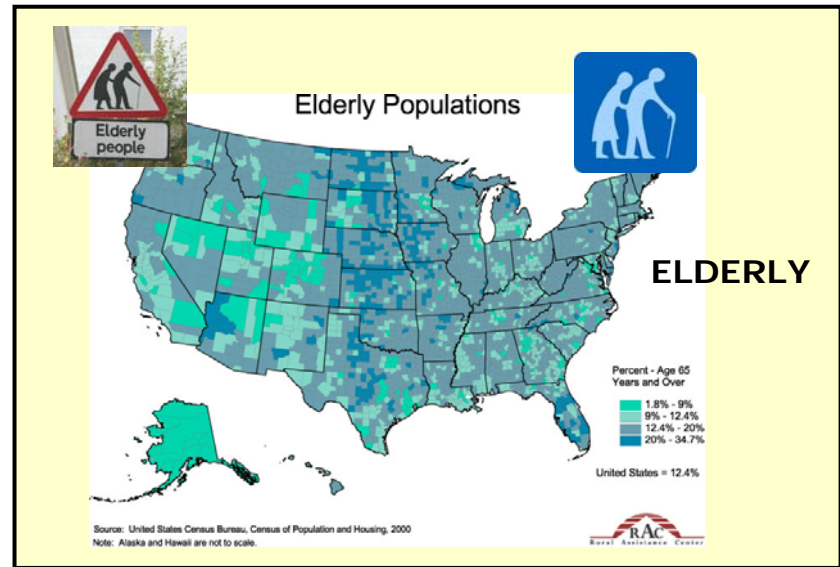
# Audio-Visual ASR



Audio-only and audiovisual ASR on the studio-DIGIT database test set using a number of integration strategies, namely, feature fusion, the state-synchronous two-stream HMM (decision fusion), the state-synchronous three-stream HMM (hybrid fusion), and the state-asynchronous product HMM (asynchronous decision fusion). In all cases, WER, %, is depicted vs. audio channel SNR. The effective SNR gain using the product HMM is also shown, reported with reference to the audio-only WER at 10 dB. All HMMs are trained in matched noise conditions.

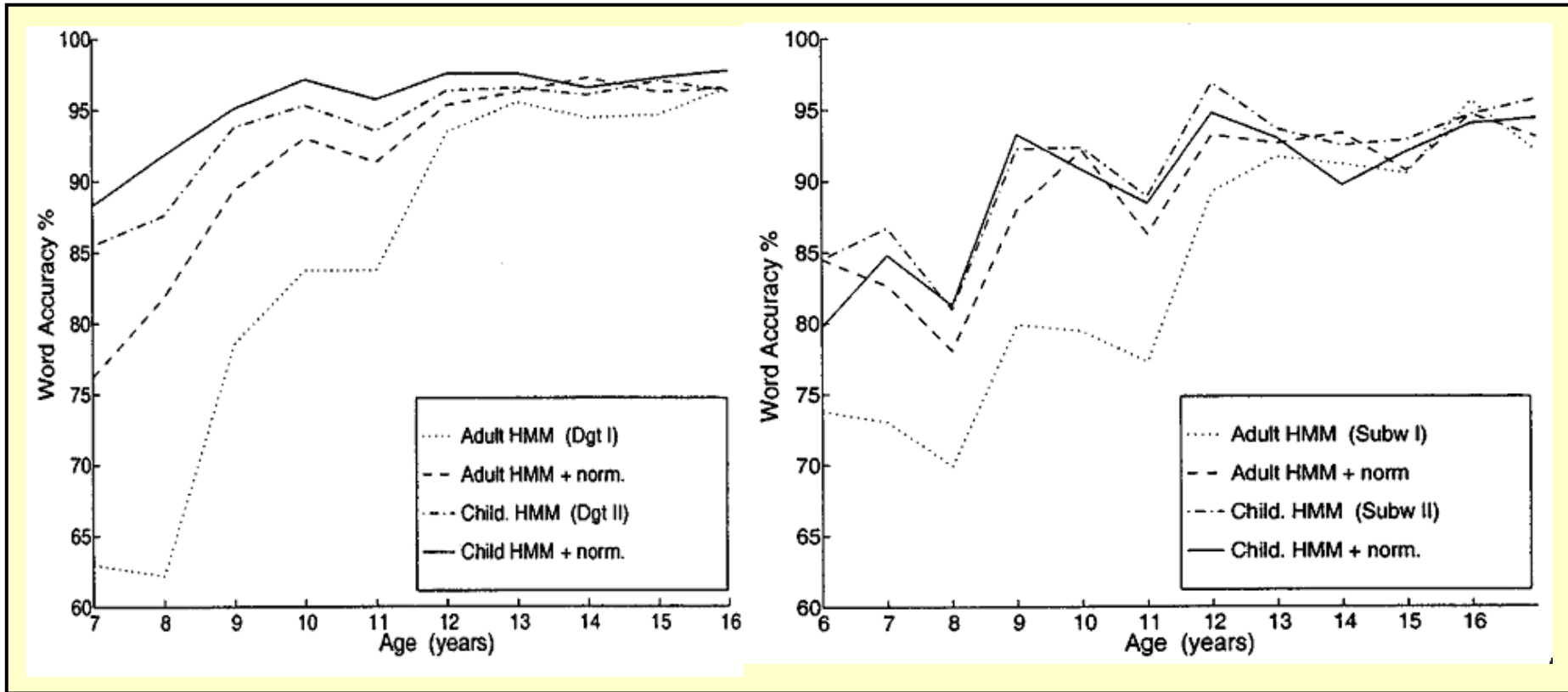
Audio-only and audiovisual WER, %, on the studio-LVCSR test set using discriminant feature fusion, as well as two-stream HMMs for decision and hybrid fusion. All models are trained in matched noise conditions.

# How to Broaden the ASR Target Population?





# ASR for KIDS



**Word accuracy (%) versus speaker's age using HMMs trained from children ("Child. HMM") or adult ("Adult HMM") speaker population with ("norm.") and without speaker normalization for: (a) connected digit task and (b) command and control task.**

S.Narayanan, A.Potamianos, "Creating Conversational Interfaces for Children", IEEE Trans. SAP, Vol.10, N,2, Feb 2002, pp. 65-78.


# ASR Applications ..... at School



[www.tmos.org](http://www.tmos.org)



**CSLU**

 OGI SCHOOL OF SCIENCE & ENGINEERING  
OREGON HEALTH & SCIENCE UNIVERSITY

[cslu.cse.ogi.edu](http://cslu.cse.ogi.edu)

The Center for Spoken  
**Language**  
Research

[cslr.colorado.edu](http://cslr.colorado.edu)

# *Current Virtual Human R&D Activities at CSLR*

CSLR has developed eight programs that use virtual humans. Each of these programs is currently under development, and being tested with human subjects. Each has been developed in close collaboration with "domain experts"-reading researchers, teachers and/or clinicians who have developed treatments that have demonstrated to be effective in the laboratory, classroom or clinic.

**Foundations to Literacy**, to teach children to read and learn from text.

**ICARE**, to assess students' reading skills and identify reading challenges.

**Flying Volando**, to teach language, literacy, math, science and social studies to English language learners.

**Workforce Previews**, to provide individuals with cognitive disabilities information about job opportunities.

**LSVT VT**, to automate portions of the LSVT speech and voice treatment for individuals with Parkinson disease.

**ORLA**, to teach reading, speech and language generation and comprehension to individuals with aphasia.

**AphasiaScripts™**, to enable individuals with aphasia to design, learn and practice daily conversations

**Sentactics**, to enable individuals with aphasia to comprehend and produce speech and language



# ASR Applications ..... at School



Language Technologies Institute

## Project LISTEN

Literacy I nnovation that Speech  
Technology ENables

A Reading Tutor that Listens



# “ILT” – Italian Literacy Tutor

Interactive Book

## “Interactive Books

### Flower Hunter

Chapter 1

## Il Cacciatore di fiori

di Debbie Darr Valdez illustrazioni di Shirlev Beckes

Piero era intento ad osservare le foreste di pini delle montagne della Georgia. Adesso il paesaggio non gli appariva bello come prima. Raccolse un bastone per attizzare il fuoco. Il padre di Piero era ferito. Era disteso a terra con una gamba rotta fasciata.

"Amos, pazzo di un mulo!" gridò Piero all'animale che pascolava. "Se non fossi scappato da quell'orso papà non sarebbe caduto!"



C'erano poche città e niente dottori sulle montagne della Georgia nel 1765.

Piero e suo padre avevano riassetato l'osso rotto, ma la guarigione sarebbe stata lunga.

Piero e suo padre erano andati in cerca di piante. Il padre di Piero era un botanico, cioè un esperto di piante. Lo Cherokee lo chiamava "Cacciatore di Fiori". Anche Piero voleva essere un cacciatore di fiori.

"Figliolo, voglio che tu guardi attentamente ogni fiore," il padre aveva detto a Piero durante una precedente escursione.

"Alcune piante non hanno fiori!" aveva risposto Piero.

"Giusto," aveva detto il padre. "Quell'abete non ha fiori. E' una conifera, ovvero una pianta che fa pigne. Ha pigne dure di color marrone come il caffè."



Default Mode

Word



Video

(13yrs)

(7yrs)

Change to mode

LOSE



# *Engaging Interactive Simulations*

## **SPEECH RECOGNITION**

People often badly mispronounce language, and blame the software when it misrecognizes their speech. AI technology combines grammar-based and "garbage" speech models to determine the probability of when the learner's speech is correct and when it isn't, and provide them with meaningful, immediate feedback

## **DIALOG MODELS**

The use of AI to balance two competing goals: allow the system to recognize, properly interpret and react to a large subset of the language; and restrict the possible spoken input so interactions train specific language and culture skills

## **BELIEVABLE VIRTUAL HUMANS**

AI methods generate virtual humans who can choose believable courses of action such as complaining, cooperating, making requests and answering questions; exhibit believable physical behavior adapted to dynamically changing social contexts; and express rich communicative acts that combine appropriate speech and gestures

## **LEARNER MODELS**

Each correct or incorrect use of relevant linguistic, cultural and task skills provides probabilistic evidence of mastery of that skill. Although the evidence by itself does not distinguish between guessing an answer, making an unconscious mistake, or the speech recognizer's misinterpretation of the learner's response, learner models can quickly identify the trainee's mastery level

## **CONTENT AUTHORIZING**

Authoring tools allow authors to create the rich content representations required by AI-based products, and perform AI-based processing themselves. For example, one tool can propose phonetic transcriptions for utterances written in the foreign language's standard orthography

# Training, Education, Entertainment, Advertising And Healthcare

## FOREIGN LANGUAGE AND CULTURE TRAINING

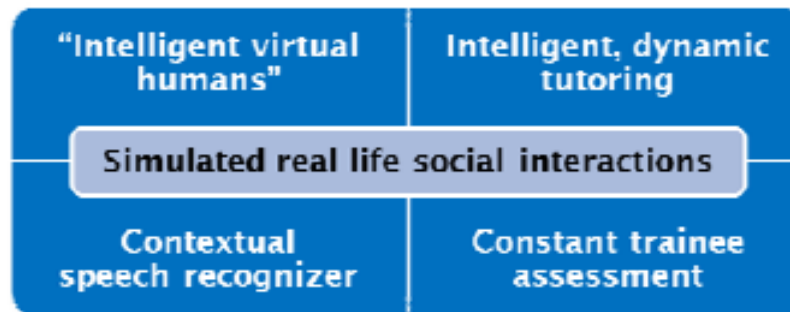
Learn foreign languages and cultures by playing game-based 3D simulations. Available for corporate, government, education, personal, entertainment and defense applications.

## SIMULATIONS OF SOCIAL COMMUNICATION

Role-play or learn interpersonal communication with virtual-world simulations of real-life social interactions. Available for training, education, entertainment, advertising, and healthcare applications.

## SCIENCE AND TECHNOLOGY

Technologic and pedagogic innovations in interactive simulations, intelligent tutoring systems, artificial intelligence, human-computer interaction, and speech recognition.



W. Lewis Johnson

Center for Advanced  
Research in Technology  
for Education at the USC  
Information Sciences  
Institute

<http://www.alelo.com/index.html>

alelo™

<http://www.tacticallanguage.com/index.html>



# Virtual-world Simulations of Real-Life Social Communication



**Tactical**  
**PASHTO**  
Language & Culture

**Tactical**  
**IRAQI**  
Language & Culture



<http://www.tacticallanguage.com/index.html>

**Tactical**  
**LANGUAGE**  
**& CULTURE**  
Training System  
TACTICAL LANGUAGE TRAINING LLC  
an edteco company

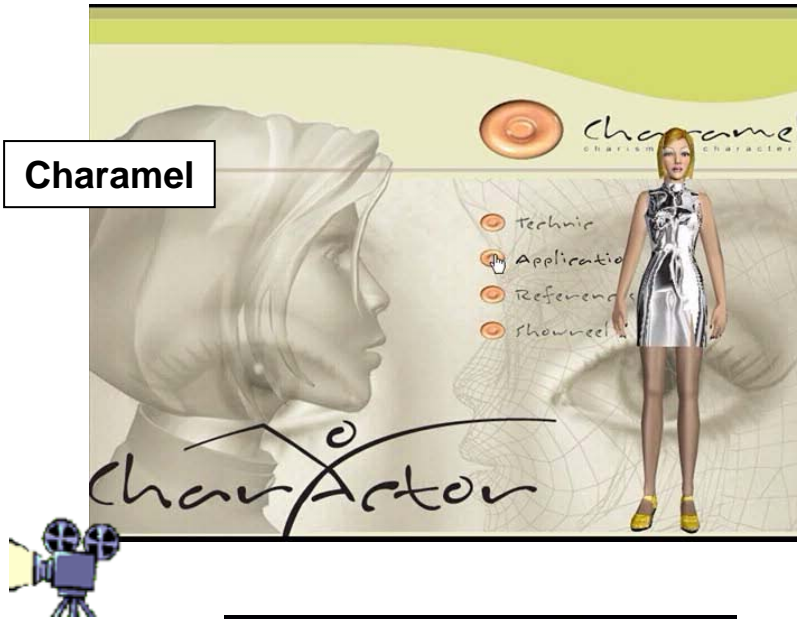


# *Animation Quality!*



SOFTIMAGE | FACE ROBOT™  
[www.softimage.com](http://www.softimage.com)

# ASR Applications ... Communication



Charamel



SitePal

**PC, web, TV, telephone ...**

# ASR Applications ... Communication

hotelbook.com  
stay independent

reservations: US 1-800-4-HOTELS; UK +44 (0) 200 604 2018  
Welcome to hotelbook.com, The Official Site of Independent Hotels™

home find hotel world map my booking about us special offers travel guides my hotels

Home > search results

view as list view thumbnails view with map destination info

You searched for hotel(s) in London from 6th of April, 2007 for 2 night(s).

sort by: highest star rating lowest price alphabetical A-Z print page

**Commodore Hotel** ★★★★★ £119.00 - £292.00 view rates  
58 Lancaster Gate, Hyde Park, London, United Kingdom  
This Victorian Townhouse hotel is located on a quiet residential square, opposite Hyde Park and in the heart of London (10 minutes from Oxford Street) which offers... [more](#)  
[view gallery](#) [view details](#) [view rates](#) [offers](#) [print](#) [save](#) [send](#)

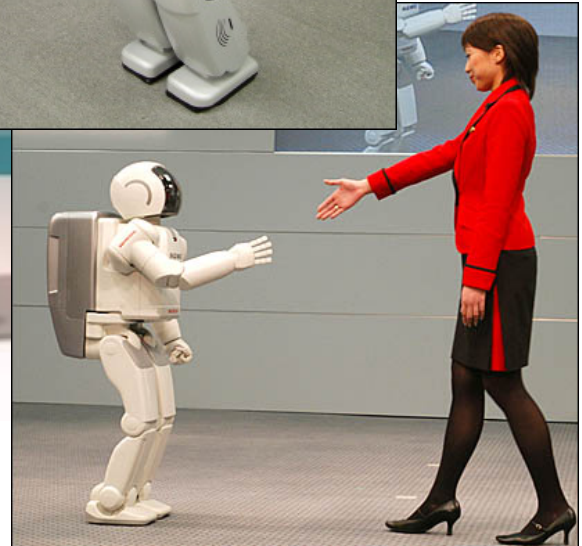
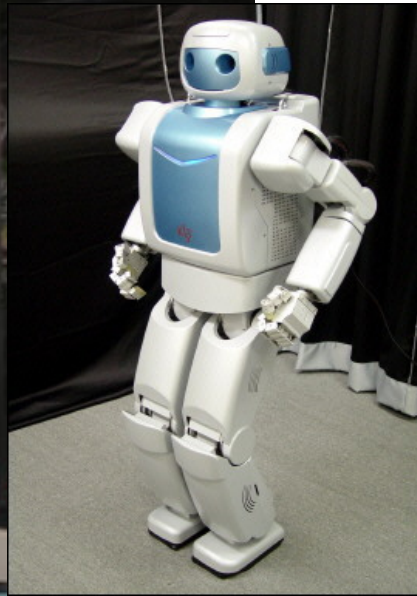
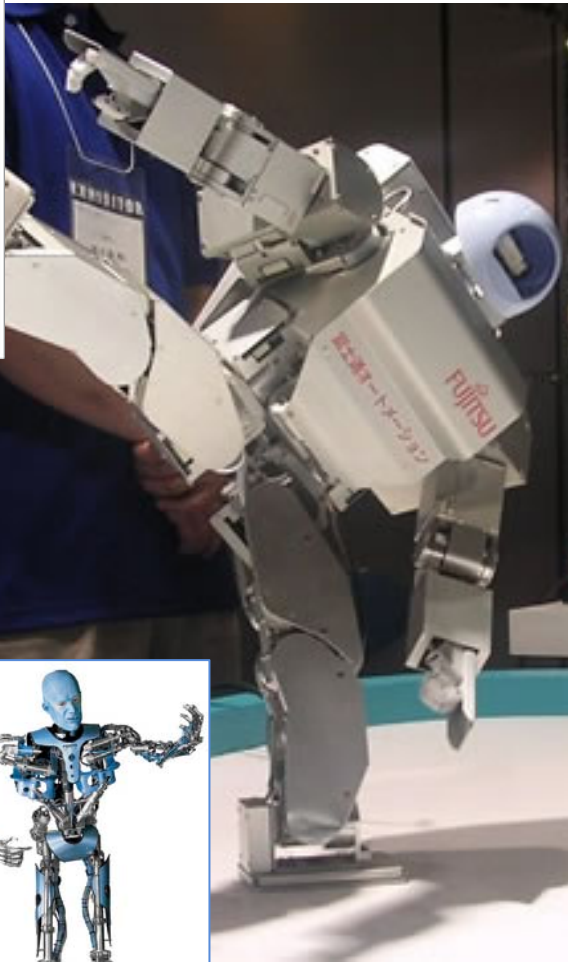
**Westminster Hotel** ★★★ £75.00 - £115.00 view rates  
16 Leinster Square, Bayswater, London, London, United Kingdom  
The hotel is located in the cosmopolitan area of Bayswater, which is on the fringes of Notting Hill. There is a wide choice of restaurants, bars, cafes, shops, ... [more](#)  
[view gallery](#) [view details](#) [view rates](#) [offers](#) [print](#) [save](#) [send](#)  
things to do: [Sea Beauty](#)

**Plaza on The River, Club and Residence** ★★★★★ £265.00 - £1,955.00 view rates  
18 Albert Embankment, London, United Kingdom  
Brand new, contemporary 5-star all-suite residence offering state-of-the-art facilities coupled with outstanding service to complement your London lifestyle. Striking... [more](#)  
[view gallery](#) [view details](#) [view rates](#) [offers](#) [print](#) [save](#) [send](#)  
things to do: [Golf](#) [Sea Beauty](#)

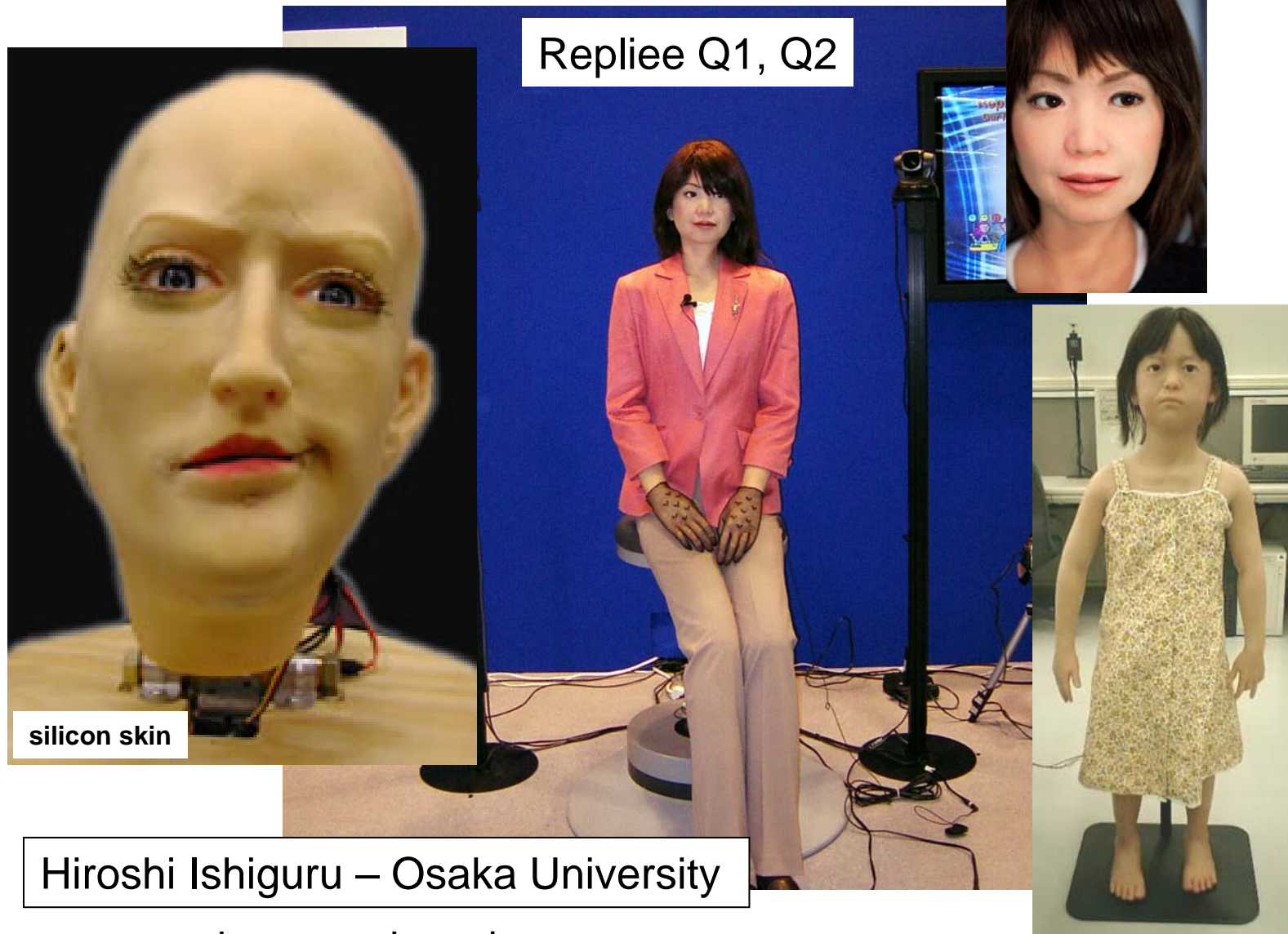
YOUR ASSISTANT close  
  
select new option  
view saved hotels  
special offers  
repeat  
Ask me to find out our best offers...

H-care

# *ASR Applications ... Robots*



# ASR Applications ... Androids



Repliee Q1, Q2



silicon skin

Hiroshi Ishiguro – Osaka University

Garrie