



Progetti nell' area del Trattamento
Automatico della Lingua (TAL)

Andrea Paoloni

Attività FUB in ambito TAL

- ❑ Forum TAL
- ❑ Passaporto per l'Italia corso di italiano in e-learning
- ❑ IDEM software di riconoscimento del parlante
- ❑ TRASCRI sistema di trascrizione in ambito giudiziario
- ❑ EMOVO corpus di parlato emotivo

Il Forum TAL

- Il Forum permanente sul TAL viene costituito in data 25 luglio 2002.
- La prima riunione ufficiale si è tenuta in data 9 settembre 2002.

Obiettivi del Forum

- ❑ Monitoraggio delle realtà coinvolte nel TAL;
- ❑ Sviluppo di strumenti linguistici innovativi;
- ❑ Ampliamento del mercato e dell'industria nazionale;
- ❑ Promozione di investimenti pubblici e privati;
- ❑ Studio dell'evoluzione del TAL;
- ❑ Uso della tecnologia della lingua italiana all'estero.

Memברי del Forum

- CNIPA
- CRUI
- Expert System
- Fondazione Ugo Bordoni
- CNR Istituto di Linguistica Computazionale
- CNR Istituto di Scienze e Tecnologie della Congizione
- Loquendo
- Ministero delle Attività Produttive
- Ministero delle Comunicazioni
- Ministero della Giustizia
- Ministero dei Beni Culturali
- MIUR

Attività svolte

- Libro Bianco sul TAL
- Mostra di Firenze
- Sito web: www.forumtal.it
- Il TAL in Biblioteca
- Partecipazioni: Forum PA , TALEP
- Conferenza TAL2006
- Evento langTech 2008

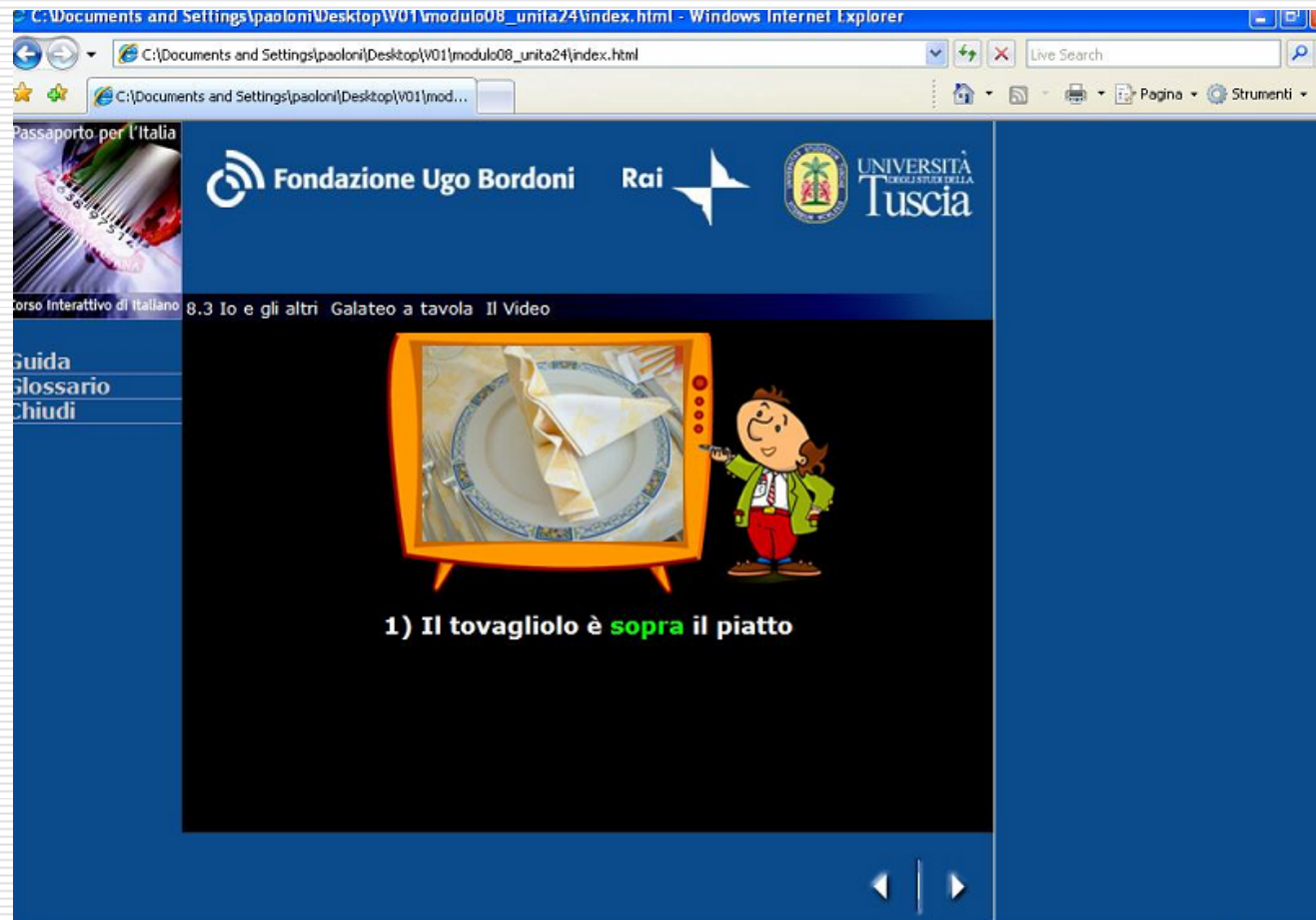
31 - passaporto per l'Italia

- ❑ corso di Italiano via internet, basato su contributi audiovisivi forniti dalla RAI
- ❑ fruito a partire da qualsiasi piattaforma di distribuzione
- ❑ dedicato in particolare ai cultori della nostra lingua residenti all'estero o in Italia
- ❑ autosufficiente (non integrativo di lezioni in presenza)
- ❑ fruibile in modo asincrono (in qualsiasi momento)

Caratteristiche di 3I

- ❑ è conforme allo standard SCORM 1.2
- ❑ guida lo studente alla competenza nell'uso delle strutture linguistiche secondo una prospettiva funzionale,
- ❑ fa riferimento agli standard europei per l'apprendimento della lingua in e-learning;
- ❑ fa uso di materiali multimediali di qualità utilizzando registrazioni delle Teche RAI;
- ❑ al termine del corso l'utente è in grado di superare con successo un esame con certificazione coerente con il syllabus europeo.

31 – Passaporto per l'Italia

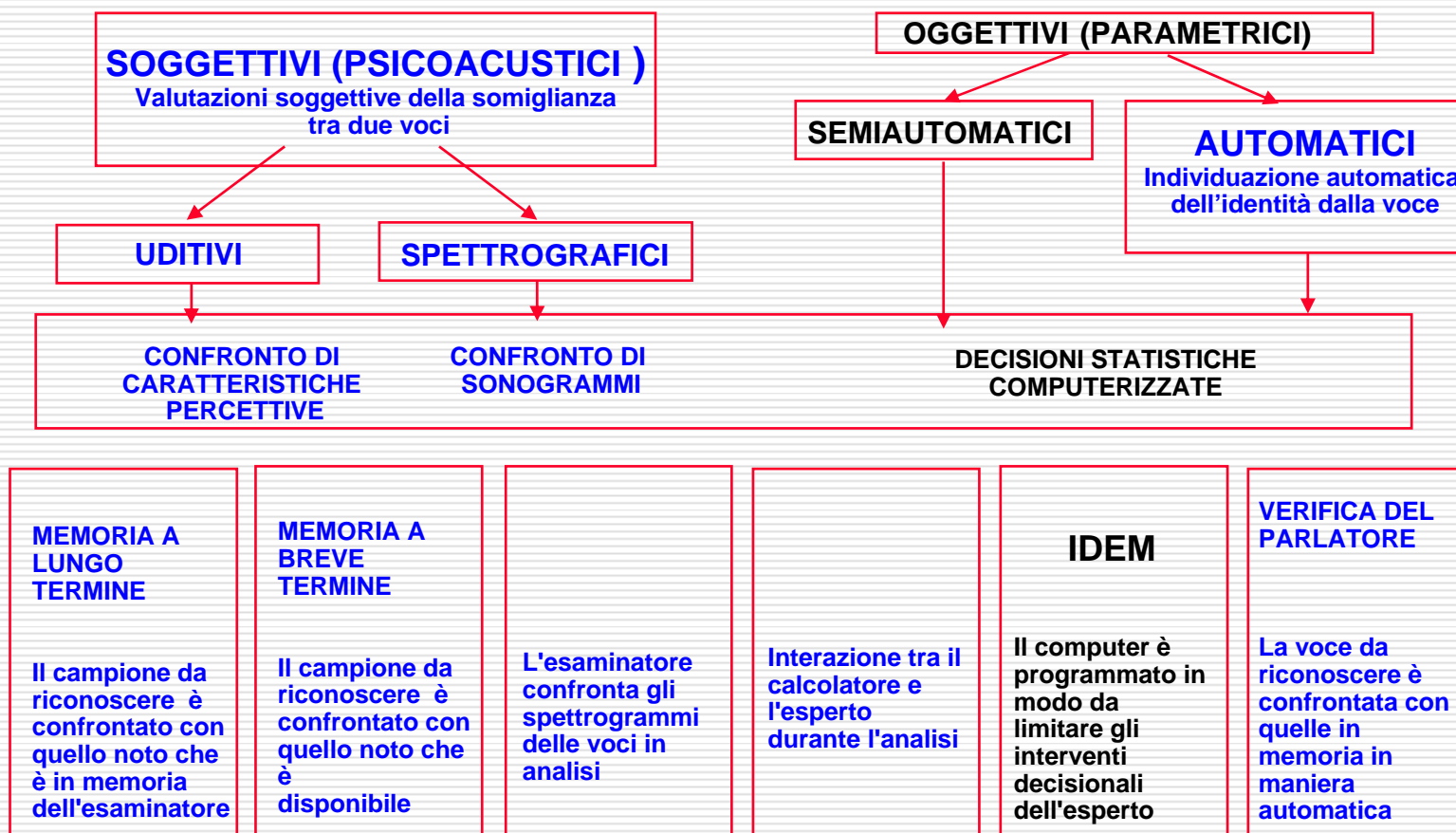


Incontro con Renato De Mori sul
tema "Comunicazione uomo -
macchina" 11 marzo 2008

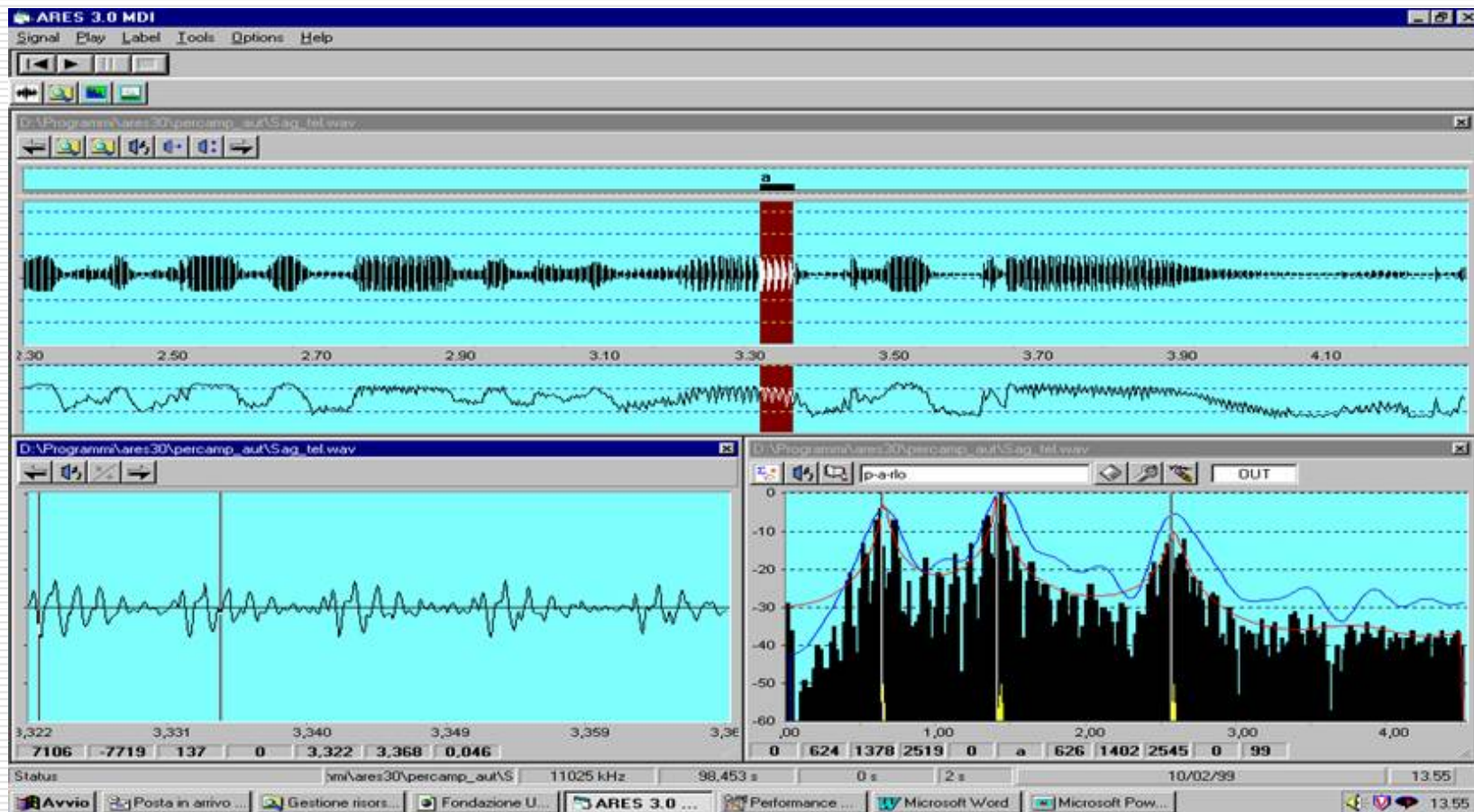
IDEM

Sistema semiautomatico per il riconoscimento del parlante ideato dalla Fondazione Ugo Bordoni ed adottato dal RaCIS (Raggruppamento Carabinieri Investigazioni Scientifiche) e dalla Gendarmeria Nazionale Argentina

Metodi di identificazione



Misura delle frequenze formanti



Incontro con Renato De Mori sul
tema "Comunicazione uomo -
macchina" 11 marzo 2008

SPREAD elaborazione statistica

Calcola:

$$LR = \frac{p(E | H_0)}{p(E | H_1)}$$

sulla base delle grandezze W e P

Dove:

W matrice di covarianza intraparlante

P matrice di covarianza interparlanti

Valore della prova

interpretazione Bayesiana

$$\frac{P(H_0)}{P(H_1)} \times \frac{P(E|H_0)}{P(E|H_1)} = \frac{P(H_0|E)}{P(H_1|E)}$$

Probabilità a priori

A cura del Tribunale

**Rapporto di
verisimiglianza(LR)
A cura del Perito**

Probabilità a posteriori

TRASCRI Rendicontazione automatica di sedute giudiziarie

Obiettivo

- Realizzazione di una piattaforma di trascrizione automatica di sedute giudiziarie in grado di permettere la pubblicazione del verbale multimediale (audio + trascrizione sincronizzata) entro la giornata in cui si svolge la seduta

Problematiche aperte

- Le condizioni di ripresa non sono sufficienti a garantire una separazione ottimale delle voci e ad evitare effetti di variazione del volume del suono
 - Esempio Processo Kalid min. 6
- Il modello del linguaggio deve essere adattato al “gergo” giudiziario e i nomi propri pronunciati più di frequente devono essere resi noti al sistema
 - Esempio Processo Rapetti min. 4,50
- Talvolta il mancato riconoscimento di piccoli segmenti di audio provoca uno stravolgimento del significato del testo
 - Esempio Processo Rapetti min 58,35
- Inflessioni dialettali o straniere molto marcate provocano malfunzionamenti
 - Esempio Processo Kalid min. 31

Corpus EMOVO

6 EMOZIONI

(gioia, disgusto, paura, rabbia,
sorpresa, tristezza, neutro)

6 ATTORI

3 maschi
3 femmine

14 FRASI

FRASI NONSENSE

*('La casa forte vuole col pane'
'La forza trova il passo e l'aglio rosso')*

FRASI LESSICALMENTE NON EMOTIVE

ELENCHI

(DI NUMERI, DI ALIMENTI)

FRASI BREVI

*('Gli operai si alzano presto'
'I vigili sono muniti di pistole'
'La cascata fa molto rumore')*

FRASI LUNGHE

*('Un attimo dopo si è
incamminato, ed è inciampato'
'Ora prendo la felpa di là, ed esco per
fare una passeggiata')*

FRASI INTERROGATIVE

*('Porti con te quella cosa?'
'Sabato sera cosa farà?')*

Incontro con Renato De Mori sul
tema "Comunicazione uomo -
macchina" 11 marzo 2008

Impieghi del corpus EMOVO

- Analisi spettroacustiche (*analisi micro- e macroprosodiche, di voice quality ecc. eventualmente finalizzate allo sviluppo di sistemi di sintesi di parlato emotivo*)
- Analisi di decoding (*test di riconoscimento maggiormente articolati, esperimenti con mascheramento del segnale ecc. eventualmente finalizzati allo sviluppo di sistemi di riconoscimento automatico del parlato emotivo*)
- Analisi interculturali (*confronti dei dati estrapolati da EMOVO con quelli relativi ad altri database di altre lingue*)
- Comparazioni fra materiale recitato e materiale naturale (*"this would be one of the highest priorities in the field" Klaus Scherer*)