aracne

Classificazione Decimale Dewey

— 378.007 (23.) EDUCAZIONE SUPERIORE. EDUCAZIONE, RICERCA, SOGGETTI CONNESSI

Thema

— Soggetto: JNM. Istruzione superiore e universitaria, istruzione avanzata
— Qualificatori: 3MR. XXI secolo, 2000-2100

PIERO COSI

UNA VITA DA RICERCATORE

"CON UN PO' DI NOSTALGIA MA... ANCHE NO"

aracne

©

ISBN
979–12–218–2490–2

PRIMA EDIZIONE
ROMA 5 AGOSTO 2026

a Laura, Lucia, Carlo
e alla mia meravigliosa mamma Ida,
che ormai ci guarda dal cielo,

che mi hanno supportato…
e a dire il vero sopportato in tutti questi anni

> What if what you do to survive kills the things you love.

BRUCE SPRINGSTEEN, *Devils and Dust*

INDICE

15 *Ringraziamenti*

17 *Prefazione*

19 *Introduzione*

25 CAPITOLO I
Un po' di storia

31 CAPITOLO II
LPC

37 CAPITOLO III
Pitch Extraction "SIFT"

39 CAPITOLO IV
Trascrizione grafema–fonema

43 CAPITOLO V
IBM – "Oberlech Summer School"

47 CAPITOLO VI
Aneddoti divertenti

51 CAPITOLO VII
Finalmente all'estero

57 CAPITOLO VIII
Finalmente in Canada (McGill University)

67 CAPITOLO IX
Reti neurali e modelli uditivi

69 CAPITOLO X
ICASSP 1988

71 CAPITOLO XI
First INNS Annual Meeting, 1988

75 CAPITOLO XII
Segmentazione automatica e modelli uditivi

77 CAPITOLO XIII
È ora di tornare

79 CAPITOLO XIV
INFOVOX

85 CAPITOLO XV
Toni, Unipd e i laureandi

95 CAPITOLO XVI
A proposito di concorsi

97 CAPITOLO XVII
Bruce Springsteen ("The Boss")

101 CAPITOLO XVIII
Modello di Lyon
18.1. Introduzione, 102 – 18.2. Sistema di analisi e sintesi, 103 – 18.3. Conclusioni, 108.

111 CAPITOLO XIX
GFS (AIA)

115 CAPITOLO XX
AISV e GSCP

121 CAPITOLO XXI
ISCA

123 CAPITOLO XXII
Portland (Oregon)

129 CAPITOLO XXIII
Italian "CSLU Speech Toolkit"

145 CAPITOLO XXIV
TTS (Text To Speech)
24.1. La storia della sintesi vocale, 147 – 24.2. Approcci di sintesi TTS a confronto, 150 – 24.2.1. *TTS parametrico*, 150 – 24.2.2. *TTS concatenativo*, 152 – 24.2.3. *TTS statistico parametrico (SPSS)*, 154 – 24.3. Analisi del testo, 155 – 24.4. Previsione dei parametri, 157 – 24.4.1 *Previsione dei parametri basata su HMM,* 157 – 24.4.2 *Previsione dei parametri basata su DNN*, 158 – 24.5. Sintesi vocale basata su VOCODER, 159 – 24.6. TTS basato sull'apprendimento profondo (DL), 162 – 24.7. TTS FESTIVAL, 164 – 24.7.1. *TTS FESTIVAL in Italiano*, 168 – *27.7.1.1. Corpus dei difoni*, 168 – *24.7.1.2. Analisi del testo e altri moduli linguistici*, 172 – *24.7.1.3. Analisi prosodica – durata e intonazione (fo)*, 175 – 24.7.1.3.1. Metodo 1 (Regole esplicite), 175 – 24.7.1.3.2.

Metodo 2 (Statistico – Cart–Tree), 178 – 24.8. TTS Mary–TTS, 181 – 24.8.1. *Mary–TTS in Italiano*, 183 – *24.8.1.1. Moduli di elaborazione del linguaggio naturale*, 185 – 24.8.1.1.1. Set di allofoni, 186 – 24.8.1.1.2. Lessico e regole LTS, 186 – 24.8.1.1.3. Espansione dei numeri, 187 – 24.8.1.1.4. Etichettatore (Tagger) di parti del parlato, 188 – 24.8.1.1.5. Disambiguazione della pronuncia degli omografi, 188– 24.8.1.1.6. Regole TOBI, 189 – *24.8.1.2. Costruzione delle voci HMM basata su Corpus*, 189 – 24.8.1.2.1. Selezione degli script, 190 – 24.8.1.2.2. Registrazione dei parlanti, 191 – 24.8.1.2.3. Generazione della voce HMM (stima dei modelli HMM), 191.

195 CAPITOLO XXV
Animazione Facciale (Facce Parlanti)
25.1. Baldi, Baldini, 1965 – 25.2. Coarticolazione nella sintesi vocale multimodale, 200 – 25.3. Greta, 222 – 25.4. Lucia, 217 – 25.4.1. *Introduzione*, 219 – 25.4.2. *Interface*, 224 – *25.4.2.1. Elaborazione dei dati bimodali*, 228 – 25.4.2.1.1. TRACK – ricostruzione 3D delle traiettorie dei marker, 230 – Innovazioni introdotte con TRACK, 235 – TRACK: configurazione del modello di riferimento, 237 – TRACK: ricostruzione delle traiettorie dei marker, 238 – TRACK: configurazione della corrispondenza marker–FAP, 240 – 25.4.2.1.2. OPTIMIZE – ottimizzazione dei parametri articolatori, 242 – 25.4.2.1.3. AP–MANAGER E MAVIS – estrazione e visualizzazione dei parametri articolatori, 248 – *25.4.2.2. Animazione da XML e da testo scritto*, 251 – *25.4.2.3. Animazione da file audio*, 254 – *25.4.2.4. FACE–PLAYER e EMOTION–PLAYER*, 255.

263 CAPITOLO XXVI
Boulder (Colorado)
26.1. Colorado Literacy Tutor (CLT), 263.

273 CAPITOLO XXVII
ASR (Automatic Speech Recognition)

275 CAPITOLO XXVIII
ALIZ–E

281 CAPITOLO XXIX
Interspeech 2011 – Firenze

289 CAPITOLO XXX
LUCIA WebGL

291 CAPITOLO XXXI
MIVOQ
31.1. MIVOQ–PTTS, 293– 31.1.1. *MIVOQ–PTTS – Infrastruttura WEB*, 296.

301 CAPITOLO XXXII
Interspeech 2016 (San Francisco)

303 CAPITOLO XXXIII
Gamification

305 *Conclusioni*

307 *Bibliografia*

345 *Appendice*

RINGRAZIAMENTI

Lo so, è banale, ma è la sincera verità. Nulla di quello che sarà descritto in questo libro, sarebbe stato possibile senza la costante benevolenza dei miei figli, Lucia e Carlo, e certamente della pazienza e del sostegno di mia moglie Laura, a cui inevitabilmente mi sento di aver sottratto parte del mio tempo, che forse avrebbe dovuto essere maggiormente dedicato a loro. Spero non me ne vogliano, ma la passione per un'attività come quella della ricerca a volte ti prende fino al punto di trascurare, anche se spero che questo non sia stato vero al 100%, altre attività sociali e familiari altrettanto ricche di passione ed interesse.

Un ringraziamento particolare deve essere rivolto a tutti i fantastici collaboratori che ho incontrato nella mia carriera, studenti, professori, ricercatori, colleghi ed amici, che hanno con me contribuito a sviluppare un'attività di ricerca stimolante e spero utile alle prossime generazioni di ricercatori nel campo delle tecnologie vocali, soprattutto per la nostra meravigliosa lingua.

Vorrei anche poter ringraziare le Istituzioni, come il CNR, che mi hanno sopportato in questi anni anche se non sempre supportato. Ma si sa, la ricerca in Italia è spesso un "optional" e spesso le soddisfazioni sono esclusivamente personali.

PREFAZIONE

Questo libro o "racconto" vuole essere una semplice descrizione della mia "vita da ricercatore", in tutto e per tutto, nel senso che, a parte qualche digressione scientifica, evidenziata dal fondo grigio delle pagine in cui è raccontata, descrivo essenzialmente come sono passati questi quarant'anni della mia esperienza personale di ricercatore appassionato della mia attività nel mondo delle tecnologie vocali, ovviamente trascurando qualche dettaglio. Di conseguenza, per una lettura più discorsiva e scorrevole, le parti evidenziate dal fondo grigio possono essere tranquillamente tralasciate da un lettore non totalmente interessato alla parte scientifica di questo racconto.

Se qualcuno fosse invece anche interessato ai dettagli scientifici dei miei lavori, oltre alle parti su fondo grigio, allegate a questo racconto, al libro ho voluto includere anche un DVD (ordinabile separatamente) con tutta la mia bibliografia personale, dal 1984 al 2020, compresi alcuni software e programmi che ho realizzato in questi anni assieme ai miei fantastici collaboratori, in modo da poter ripetere, se qualcuno fosse interessato, i risultati delle mie ricerche.

INTRODUZIONE

Più di cinquant'anni fa, il 2 aprile 1968, avevo quasi 12 anni, usciva in anteprima a Washington lo storico «2001: Odissea nello spazio», pietra miliare della fantascienza. Diretto da Stanley Kubrick. Il film si basa sul romanzo «La sentinella» scritto da Arthur C. Clarke e vinse l'oscar per i migliori effetti speciali visivi, indice del fatto che aveva rappresentato al meglio la possibile evoluzione del nostro mondo. Non solo esteticamente. Ci sono infatti alcuni dispositivi o alcune innovazioni che poi abbiamo davvero finito per veder realizzate, come ad esempio:

- il *tablet*: i protagonisti della "Missione Giove" in «2001: Odissea nello Spazio», li usano per leggere le notizie e anche per giocare a scacchi; assomigliano tantissimo a quelli che siamo abituati a vedere oggi;
- gli *assistenti digitali*: il sogno di costruire un computer parlante è molto più vecchio degli anni '80 e, sicuramente, il computer parlante più impresso nell'immaginario collettivo è HAL, acronimo di *Heuristically programmed ALgorithmic Computer*. "HAL 9000" è il personaggio più famoso di «2001: Odissea nello Spazio», ed è l'intelligenza artificiale che guida il super

computer di bordo sulla nave spaziale Discovery; ancora ne devono fare di strada Siri, Google Assistant, Alexa e Cortana per poter arrivare ai suoi livelli: parla con voce umana, ha emozioni ed è in grado di pilotare l'astronave, ma anche lei commette errori.

Figura 1. "Immagine di un computer simile ad HAL 9000 (immagine creata mediante IA)".

- le *videochiamate*: di sicuro non c'era Skype, comunque nel film viene effettuata una prima videochiamata con cui il dottor Floyd chiama la figlia da un improbabile computer.
- gli *schermi in aereo*: nel 1968 chi viaggiava in aereo di sicuro non aveva a disposizione l'intrattenimento di cui possiamo godere oggi; non c'erano gli schermi personali, magari touch, attaccati al sedile del passeggero davanti su cui vedere gli ultimi film usciti, i documentari o le serie tv.; mentre in «2001: Odissea nello Spazio» già avevano pensato a come ammazzare il tempo durante il volo;
- le *stazioni spaziali*: la prima stazione spaziale è stata lanciata in orbita nel 1971 e si chiamava *Saljut* 1 ed era stata progettata dall'Unione Sovietica; voleva essere la risposta, in piena Guerra Fredda, all'atterraggio