Quaderni di informatica 19

gna di tecnologia ed applicazioni degli elaboratori elettronici – Anno X - Numero 1 - 1983

Honeywell
Honeywell Information Systems Italia

FPDM-64

Honeywell

Honeywell Information Systems Italia

Quaderni di informatica

Rassegna di tecnologia ed applicazioni degli elaboratori elettronici
Anno X - Numero 1 - 1983

Sommario

Nuove tecniche costruttive per grandi elaboratori

Un grande elaboratore deve essere... il più piccolo possibile. Per risolvere questa apparente contraddizione occorrono nuove soluzioni strutturali, quali il "micropackaging", nonché nuove tecniche di raffreddamento del sistema.
F. FILIPPAZZI

L'informatica nel lavoro d'ufficio

L'automazione sta ormai entrando in tutte le tipiche attività d'ufficio. L'articolo esamina questa nuova prospettiva, fornendo un'analisi documentata e quantitativa degli aspetti più rilevanti.
G. OCCHINI

Il riconoscimento automatico della voce

Il riconoscimento della voce è da tempo oggetto di studio, ma i risultati pratici finora ottenuti sono stati piuttosto deludenti. Sta però nascendo nei laboratori industriali una nuova generazione di dispositivi che imprimerà una svolta nell'interazione vocale tra l'uomo e la macchina.
T. EDMAN

L'intelligenza artificiale ed il gioco degli scacchi

Da tempo il gioco degli scacchi è un banco di prova della "intelligenza" dell'elaboratore elettronico. Su questo tema fa il punto l'Autrice dell'articolo, nella sua duplice veste di ricercatrice sull'Intelligenza Artificiale e di campionessa di scacchi.
B. PERNICI

Direttore Responsabile
F. Filippazzi

Comitato di Redazione
A. Cicu, C. Falcetti, P. Lupo,
M. Nobile, G. Occhini
G. Rapelli, F. Sala

Redazione
Honeywell Information Systems Italia
Centro di Ricerca e Progettazione
20010 Pregnana Milanese (Milano)

Stampa
tipolito Maggioni s.a.s.

Autorizzazione del Tribunale di Milano
n. 93 del 20 Marzo 1974

Nuove tecniche costruttive per grandi elaboratori

F. FILIPPAZZI
Honeywell Information Systems Italia

1. Introduzione

Un aspetto degli elaboratori che letteralmente "salta agli occhi" è la drammatica riduzione delle dimensioni. La stessa macchina che non molti anni fa richiedeva diversi pesanti armadi, sta oggi su un angolo di scrivania.
Originariamente, una fondamentale spinta alla miniaturizzazione venne dalle esigenze di settori specifici, in particolare quello spaziale dove il risparmio di ogni grammo di carico trasportato era importante. Successivamente, però, la miniaturizzazione venne vista non più come fine a se stessa ma come strada obbligata per migliorare praticamente ogni altro aspetto delle apparecchiature elettroniche, quali la velocità di operazione, l'affidabilità, la dissipazione di energia, nonché il costo.
A titolo di esempio, consideriamo la velocità di funzionamento. Schematizzato l'elaboratore come un insieme di elementi logici ("*gate*") tra loro interconnessi, consideriamo una coppia di tali elementi (fig. 1). Un segnale elettrico applicato all'ingresso dell'elemento G_1, prima di arrivare all'ingresso dell'elemento G_2 subisce due tipi di ritardi. Il primo, t_g, è il ritardo proprio dell'elemento logico, ed il suo valore è caratteristico della tecnologia impiegata. Il secondo, t_p, è il tempo impiegato dal segnale per percorrere la distanza che separa i due elementi considerati. Sul valore di t_p influiscono sia la lunghezza del percorso che i carichi derivati lungo di esso. È interessante confrontare tra loro le grandezze di questi due tipi di ritardo.
I tempi di commutazione t_g sono, per logiche veloci, dell'ordine del nanosecondo (da qualche ns per circuiti di tipo TTL a meno di un ns per quelli di tipo CML). Il valore unitario di t_p si può invece approssimare a circa 1 nanosecondo per ogni 10 cm di percorso. (Per confronto, la luce in 1 nanosecondo percorre 30 cm).
Pur nei limiti della schematizzazione adottata, quanto sopra fornisce una importante indicazione: se la distanza tra gli elementi logici è elevata, il tempo di comunicazione tra essi tende ad essere il fattore limitante la velocità del sistema. In altre parole, per costruire elaboratori più veloci occorre farli sempre più piccoli.

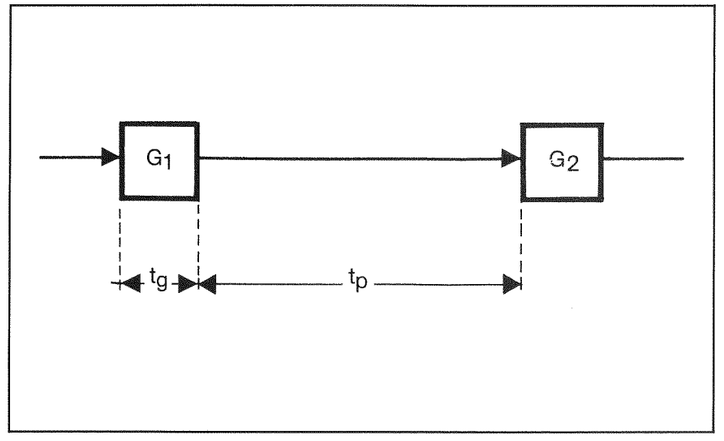


Fig. 1 - La velocità di un elaboratore è legata a due fattori fisici fondamentali: il ritardo t_g proprio degli elementi logici, e il ritardo t_p dovuto al packaging. Quest'ultimo tende a costituire il fattore preponderante.

Segue anche che se la miniaturizzazione è importante per elaboratori piccoli o medi, essa diventa essenziale per realizzare grandi elaboratori.

2. L'approccio fondamentale

La strada maestra della miniaturizzazione è costituita dalla integrazione circuitale, cioè dalla realizzazione di un grandissimo numero di elementi logici e delle loro interconnessioni in un "chip" di silicio di pochi millimetri di lato. A questa tecnica si deve in gran parte il fondamentale progresso dell'hardware degli elaboratori. Nel giro di neanche vent'anni si è infatti passati dal transistore isolato a chip estremamente complessi (VLSI = Very Large Scale Integration). In tutto questo periodo, il numero di elementi per chip è aumentato a ritmo costante, raddoppiando in pratica ogni anno (legge di Moore). Ciò è il risultato combinato della continua riduzione delle dimensioni dei dispositivi elementari e del parallelo aumento dell'area del chip.

Per ogni dato stato della tecnologia, il numero massimo di elementi per chip risulta fissato in pratica da considerazioni economiche. Infatti, sia le dimensioni dei dispositivi elementari che l'area del chip presentano, ad ogni dato momento, dei valori ottimali agli effetti del costo per funzione. Superando tali valori si ha un rapido aumento degli scarti di produzione con una conseguente crescita del costo unitario. La "resa" del processo di fabbricazione segue infatti una legge esponenziale del tipo e^{-DA}, dove D è la densità di difetti (legata al processo) ed A è l'area del chip.

Sul livello di integrazione influiscono anche altri fattori. Uno dei più importanti è la dissipazione di energia dei circuiti impiegati, ossia il calore sviluppato.

Sotto questo profilo, i circuiti si possono dividere in due classi fondamentali, caratterizzate rispettivamente dall'impiego di transistori unipolari (MOS) e di transistori bipolari.

La dissipazione dei primi è di un ordine di grandezza inferiore a quella dei secondi e ciò consente, grazie anche ad una maggior semplicità costruttiva, di ottenere livelli di integrazione parallelamente più elevati.

I circuiti MOS presentano però velocità di funzionamento assai inferiori, che ne limitano il campo di applicazione.

Per realizzare elaboratori di grande potenza occorre basarsi su logiche veloci, ossia di tipo bipolare (quali i circuiti TTL o, per prestazioni ancora più spinte, i circuiti CML), anche se ciò consente livelli di integrazione minori.

3. L'effetto del packaging

L'integrazione circuitale su larga scala è certamente un mezzo formidabile per ridurre le dimensioni fisiche di un elaboratore. Però, per prestazioni particolarmente spinte, come nel caso dei grandi elaboratori, occorre agire ulteriormente sul "packaging" del sistema.

Si consideri infatti lo schema classico di packaging. Esso presenta una gerarchia strutturale a più livelli, come mostrato in fig. 2.

Questa struttura trova la sua origine nelle esigenze pratiche di montaggio, di collaudo e di manutenzione delle macchine. Tali vantaggi vanno, però, in generale, a scapito della compattezza del sistema.

L'altissimo grado di miniaturizzazione realizzato nei chip di silicio viene in larga misura perso nel packaging del sistema. Basti dire che, tra il volume del chip "nudo" ed il volume equivalente del chip "montato" esiste un rapporto dell'ordine di diecimila. I chip contenuti in un armadio di elaboratore stanno tranquillamente nel palmo di una mano; tutto il resto è packaging.

La degradazione volumetrica subisce un importante incremento già al primo livello di packaging, cioè nell'incapsulamento del chip in un contenitore. Quest'ultimo assolve non soltanto un compito di protezione, ma anche una funzione geometrica, allargando a raggiera i punti terminali del chip in modo da farne un oggetto facilmente maneggiabile. Si tenga conto che la distanza tra i terminali del contenitore (circa 2,5 mm) è circa 200 volte maggiore dell'equivalente distanza sul chip.

Il contenitore risulta quindi molto più voluminoso del chip che racchiude. Questa dilatazione si propaga inoltre, ovviamente, alle connessioni del successivo livello di packaging, cioè alla "scheda" a circuito stampato che collega più chip tra loro.

Gli effetti di questo schema strutturale possono essere valutati in termini quantitativi. Con riferimento al secondo livello di packaging (cioè alla scheda), il diagramma di fig. 3 mostra le relazioni tra i fattori che limitano la velocità dell'insieme.

A maggior chiarimento, il diagramma è basato sulle seguenti assunzioni.

a) Il ritardo t_s a livello sistema è la somma del ritardo t_g proprio del gate e del ritardo t_p imputabile al packaging:

$$t_s = t_g + t_p \qquad (1)$$

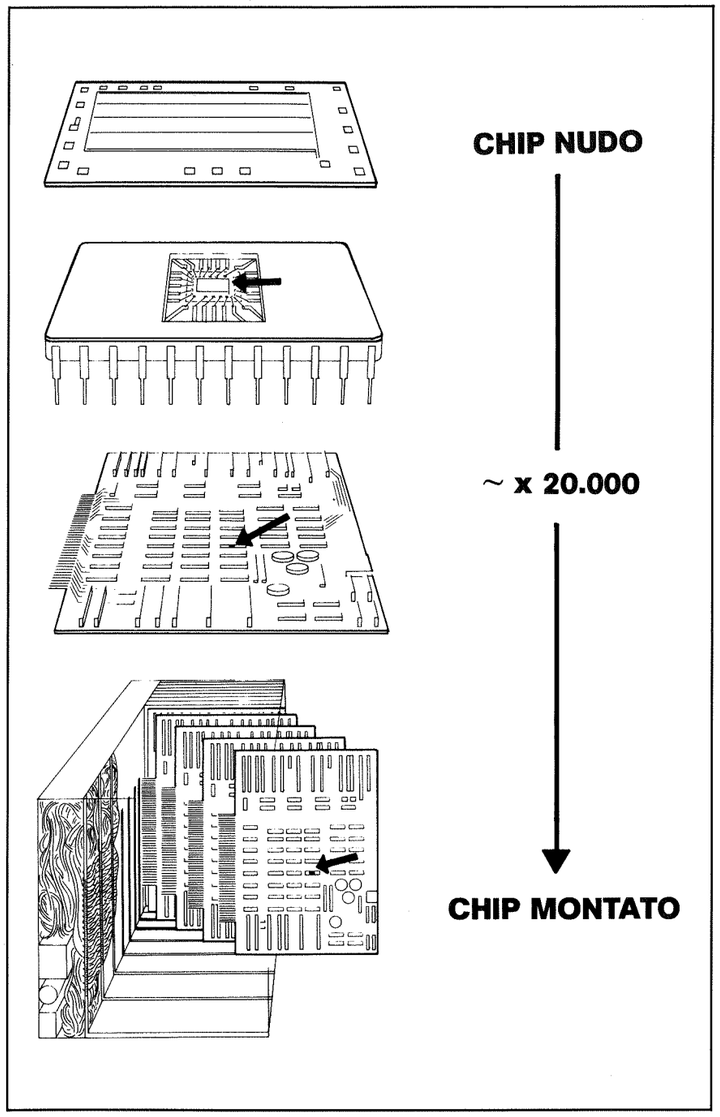
b) In base a considerazioni empiriche, si assume che il ritardo t_p è inversamente proporzionale alla radice quadrata della densità di gate sulla piastra (d_g):

$$t_p = k_1 (d_g)^{-1/2} \qquad (2)$$

c) Il prodotto "ritardo x potenza dissipata per gate" è una costante della tecnologia impiegata:

$$t_g \cdot p_g = k_2 \qquad (3)$$

Fig. 2 - Lo schema convenzionale di packaging. L'eccezionale compattezza realizzata a livello chip viene persa nelle successive fasi di assemblaggio. Il volume occupato da un chip montato è infatti oltre diecimila volte quello del chip "nudo". (Bibl. [5]).



Introducendo la densità di potenza per unità di area, d_p ($\equiv d_g \cdot p_g$), dalle relazioni precedenti si ottiene:

$$d_p = k_1^2 \cdot k_2 / t_g (t_s - t_g)^2 \qquad (4)$$

$$d_p = k_2 \cdot d_g / t_g \qquad (5)$$

I grafici di fig. 3 mostrano le relazioni (4) e (5), nel caso di $k_1 = 5$ (ns $\cdot$ in^{-1} $\cdot$ gate$^{-1/2}$) e $k_2 = 20$ (pJ $\cdot$ gate^{-2}).

Il diagramma permette, in particolare, di determinare per un dato tipo di gate quale densità di impaccamento occorre raggiungere per ottenere una richiesta velocità a livello sistema.

4. Dal packaging al micropackaging

Da quanto precede, per ottenere prestazioni più elevate diventa cruciale realizzare maggiori densità di impaccamento a livello scheda.

In questa ottica, è stata messa in discussione la necessità del primo livello di packaging, ossia dell'incapsulamento individuale del chip. È stata cioè studiata la possibilità di impiegare i chip "nudi", collegandoli tra loro direttamente con connessioni miniaturizzate, realizzate su un apposito substrato che sostituisce la scheda convenzionale (fig. 4).

Questa impostazione, denominata *micropackaging,* comporta un

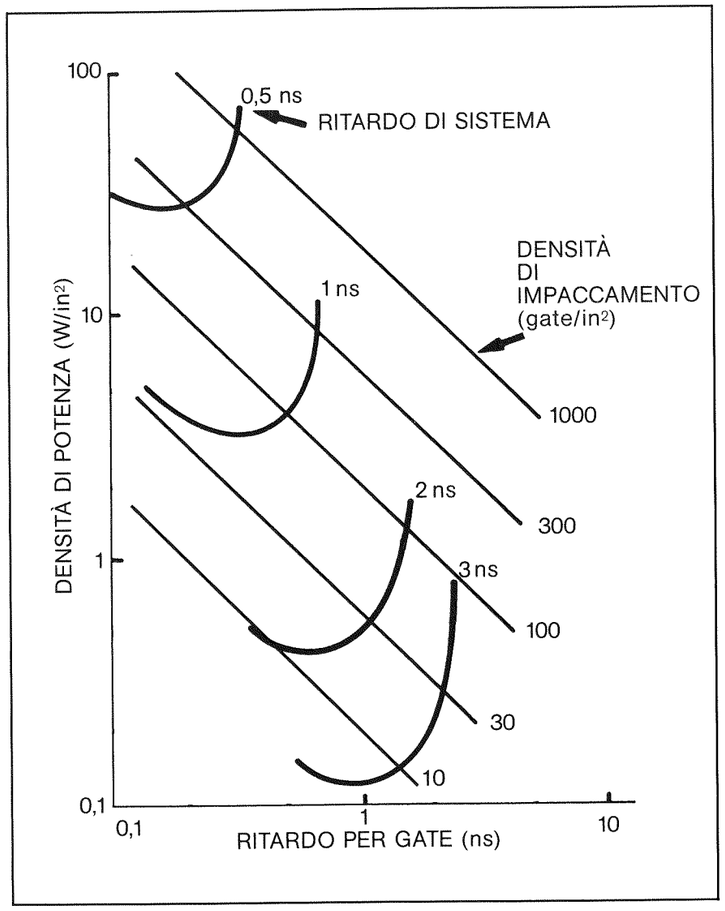


Fig. 3 - Relazione tra i vari fattori fisici che limitano la velocità di elaborazione. (Ci si riferisce al secondo livello di packaging).

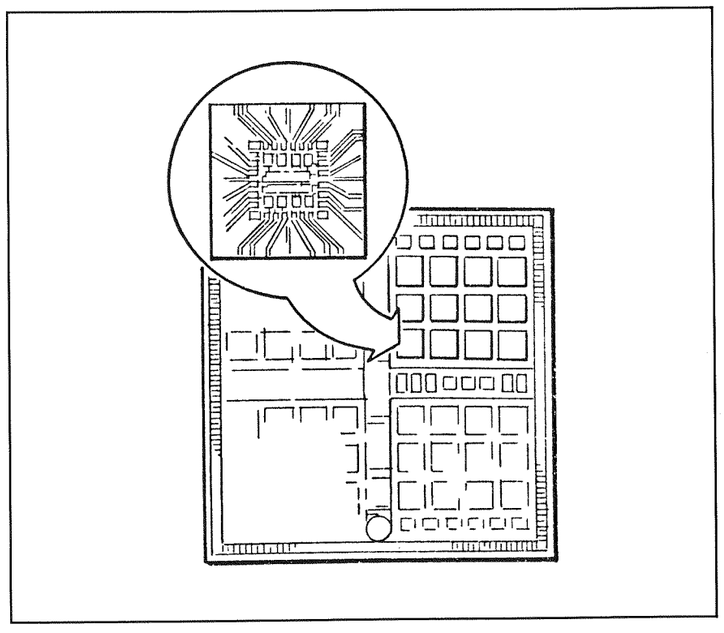


Fig. 4 - Il concetto di micropackaging implica l'eliminazione del primo livello di packaging tradizionale, cioè dell'incapsulamento individuale dei chip. Nel micropackaging, infatti, i chip sono montati "nudi" su un opportuno substrato, che provvede anche alla loro interconnessione.

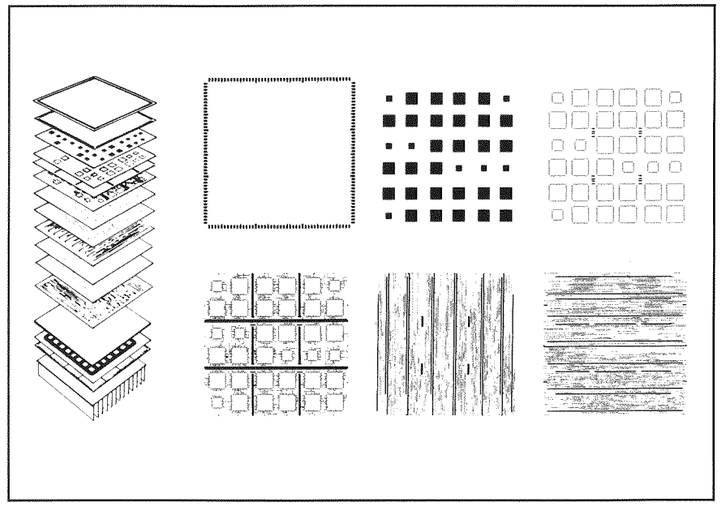
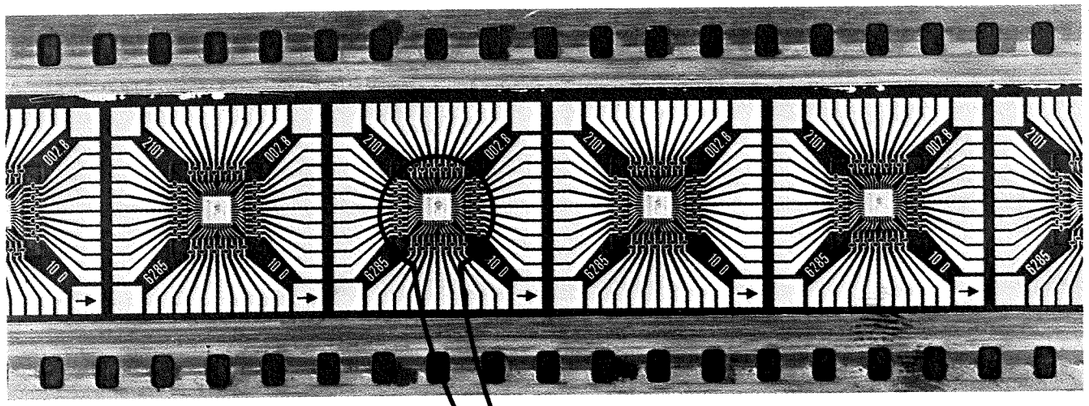


Fig. 5 - Le connessioni tra i chip del micropack sono realizzate mediante strati sovrapposti stampati serigraficamente sul supporto ceramico. Nell'esempio di figura, i due strati di percorsi perpendicolari realizzano le connessioni logiche tra i chip, mentre gli altri servono per distribuire l'alimentazione elettrica, per gli isolamenti e per i contatti di superficie.

radicale cambiamento dei materiali e dei processi costruttivi.
Il substrato non può infatti essere costituito dall'usuale materiale delle schede, essendo richieste caratteristiche chimico-fisiche, meccaniche ed elettriche decisamente superiori. Nella fattispecie occorre passare a materiali ceramici.
Analogamente, la realizzazione sul substrato di connessioni molto più dense e sottili comporta l'impiego di altri tipi di processi. La tecnica più appropriata risulta essere la "serigrafia", simile, anche se molto più sofisticata, a quella impiegata in certi tipi di stampa. Si realizzano in tal modo strati conduttivi e isolanti, alternati e sovrapposti, con delle "vie" verticali di comunicazione (fig. 5). Si rende inoltre necessario un diverso modo di proteggere i chip. Si richiede ora infatti una protezione collettiva, a livello dell'intero substrato. Il componente così realizzato ("micropack") deve avere una copertura ermetica, che però possa essere rimossa per permettere l'ispezione e l'accesso ai chip.
Bisogna inoltre prevedere il modo di asportare il calore generato dal micropack, nonché realizzare un adeguato sistema di connessione tra esso ed il successivo livello di packaging, cioè la scheda convenzionale.
A tutto ciò, occorre aggiungere una lunga lista di problemi indotti, che vanno dalla revisione dei criteri di progettazione logica ed elettronica, alla soluzione dei nuovi problemi posti al collaudo e alla manutenzione.
Non deve quindi stupire se il micropackaging ha richiesto anni di ricerche per la messa a punto e grossi investimenti per la sua utilizzazione industriale. Solo recentemente questa tecnica è stata effettivamente introdotta in nuovi modelli di elaboratori di grande o grandissima potenza.
La Honeywell ha svolto una attività di avanguardia nello studio e nella utilizzazione del micropackaging. Su questa tecnologia è basato l'hardware di due dei più recenti sistemi, il DPS7 ed il DPS88. L'approccio Honeywell, che presenta soluzioni originali, viene brevemente illustrato nel seguito.

5. La tecnica flip-TAB.

Un primo obiettivo nella realizzazione del micropackaging è l'adozione di un sistema che consenta un facile maneggio dei chip nudi ed un alto grado di automazione. Questo sistema è rappresentato dal TAB (Tape Automated Bonding), illustrato nella fig. 6.
Secondo questo approccio, i chip vengono montati su un nastro di materiale plastico (Kapton), sul quale sono realizzate delle apposite sagome conduttive a forma di "*ragno*".
A tale scopo, il nastro è costituito, oltre che dal supporto plastico, da un sottile foglio di rame ad esso incollato. Il rame viene asportato selettivamente con processo fotochimico, in modo da ottenere le sagome suddette.
Il chip viene collocato al centro della sagoma e le sue piazzole periferiche saldate automaticamente alle "zampe" del ragno. Queste diventano in tal modo i terminali del chip, espansi in modo da consentire una facile connessione alle apparecchiature di test.
In sostanza, il chip "viaggia" insieme col nastro su cui è montato, che costituisce il veicolo per trasferirlo agevolmente lungo le varie fasi di collaudo e di montaggio.
Ogni nastro porta generalmente uno stesso tipo di chip. Il nastro ha una larghezza di 35 mm e possiede dei fori laterali per l'avvolgimento/svolgimento su una bobina.
Parallelamente alla preparazione del chip, si provvede alla fabbricazione del substrato ceramico multistrato. La fig. 7 mostra uno di tali substrati. Si intravede in trasparenza il tracciato dei percorsi degli strati interni, mentre sulla superficie sono visibili le piazzole sulle quali verranno saldati i chip. I percorsi conduttori hanno una larghezza di 100 micron e uno spessore di 15 micron. (Per confronto, un capello ha un diametro sui 75 micron).
I chip montati su nastro ed il sup-



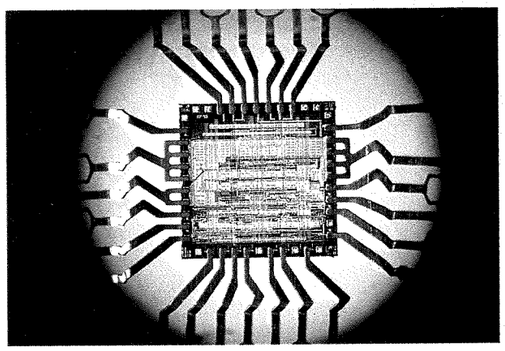


Fig. 6 - Fotografia di un tratto di nastro realizzato secondo il processo TAB (Tape Automated Bonding). Nell'ingrandimento, un chip saldato al "ragno" di rame.

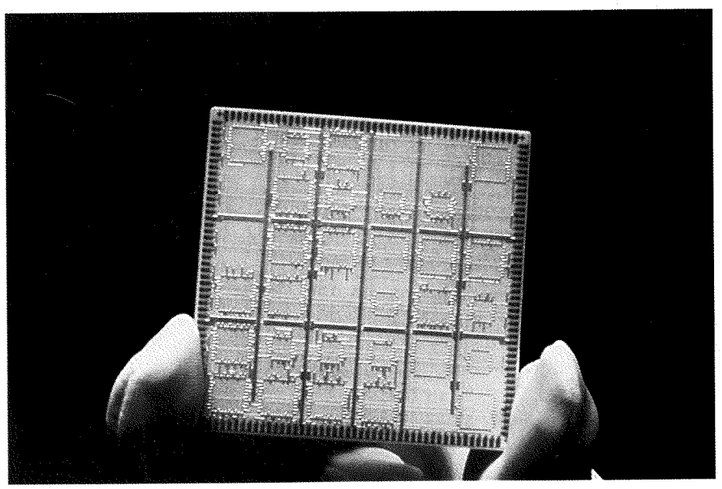
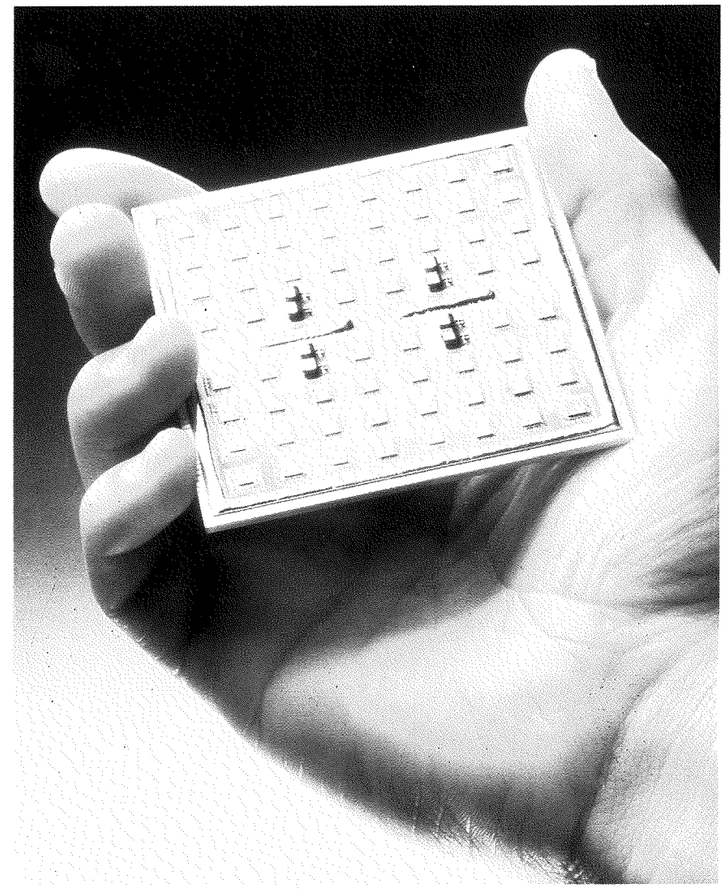


Fig. 7 - Un supporto ceramico multistrato finito. Si intravedono, in trasparenza, i tracciati degli strati interni. Sulla superficie sono visibili le piazzole cui vanno saldati i chip.

Fig. 8 - Un micropack completo di chip. La piastrina misura 8 x 8 cm, porta oltre 60 chip CML e dissipa 60 Watt. (Elaboratore Honeywell DPS 88).



porto ceramico multistrato convergono in una stazione di montaggio automatico. Un robot preleva un chip alla volta, scegliendo la bobina giusta, taglia la parte superflua delle "zampe" attorno al chip, sistema il chip nella posizione richiesta del substrato e infine lo salda a questo mediante l'applicazione di lame termiche. La fig. 8 mostra un substrato del DPS 88 completo di chip. Questo substrato ha dimensioni 8 x 8 cm, contiene una media di 60 chip e sostituisce due schede multistrato convenzionali di 30 x 30 cm ciascuna. Il micropackaging del DPS 88 presenta una soluzione originale per quanto riguarda il modo di montare i chip. Infatti, contrariamente al solito, questi sono montati sul substrato *con la faccia attiva in giù. (Flip-TAB).*

Come illustrato in fig. 9, il chip viene incollato al substrato faccia in giù mediante uno strato di epoxy, elettricamente isolante ma buon conduttore del calore. Successivamente, i terminali del chip sono saldati per termocompressione alle piazzole stampate sul substrato.
Il sistema di montaggio Flip-TAB presenta diversi vantaggi; in particolare protegge la faccia attiva del chip e offre un buon cammino termico per la dispersione del calore generato dal chip. Esso assicura inoltre un più accurato allineamento tra terminali del chip e piazzole del substrato ed elimina fattori critici del montaggio, quale lo spessore del chip.
Sul micropack viene alla fine collocato un coperchio metallico a tenuta ermetica (fig. 10). Oltre che provvedere alla protezione meccanica dei componenti racchiusi, il coperchio fornisce la tensione primaria al substrato.
I micropack vengono montati su schede a circuito stampato convenzionali. In fig. 11 è mostrato lo schema costruttivo di una unità logica del DPS 88. Su ogni scheda sono montati 8 micropack, e vi sono in totale 8 schede. Al piede, c'è l'alimentatore di tutta l'unità.

6. Il raffreddamento a liquido

Col micropackaging si realizza un drastico aumento della compattezza rispetto al packaging convenzionale. Ciò comporta però un problema e cioè una forte concentrazione del calore sviluppato.
Con l'impiego del micropackaging la densità di circuiti logici risulta infatti oltre dieci volte maggiore

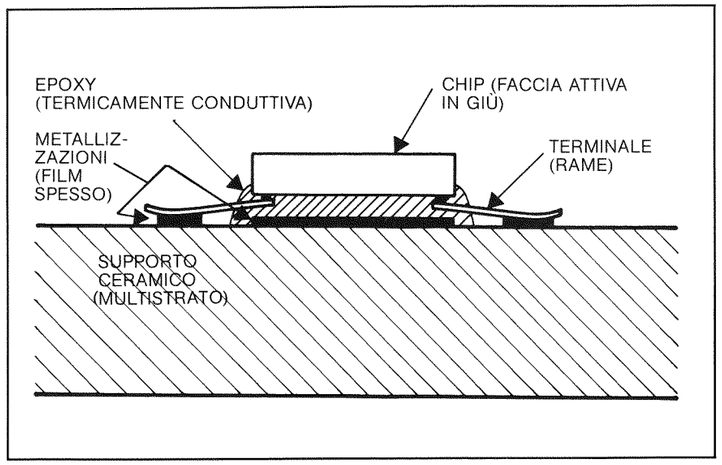


Fig. 9 - Nella tecnica *"Flip-TAB"* della Honeywell il chip è montato con la faccia attiva in giù. Esso viene incollato al supporto ceramico mediante un materiale (epoxy) elettricamente isolante ma buon conduttore del calore.

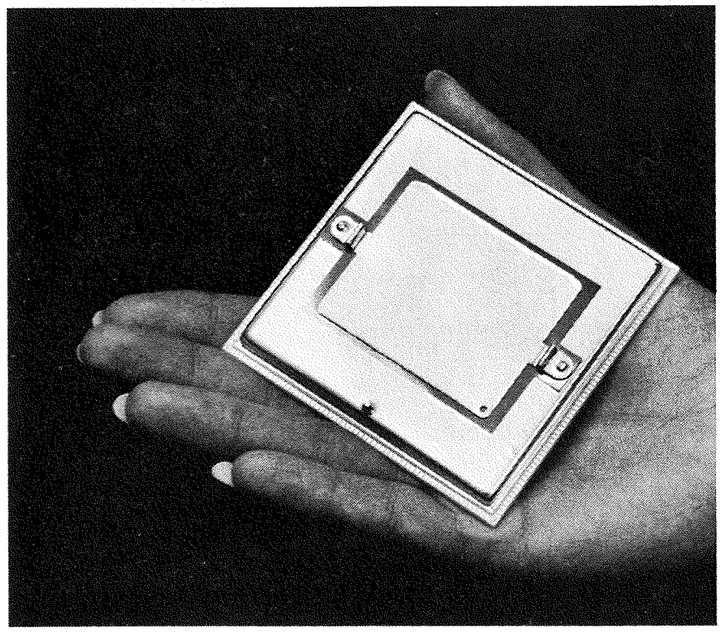
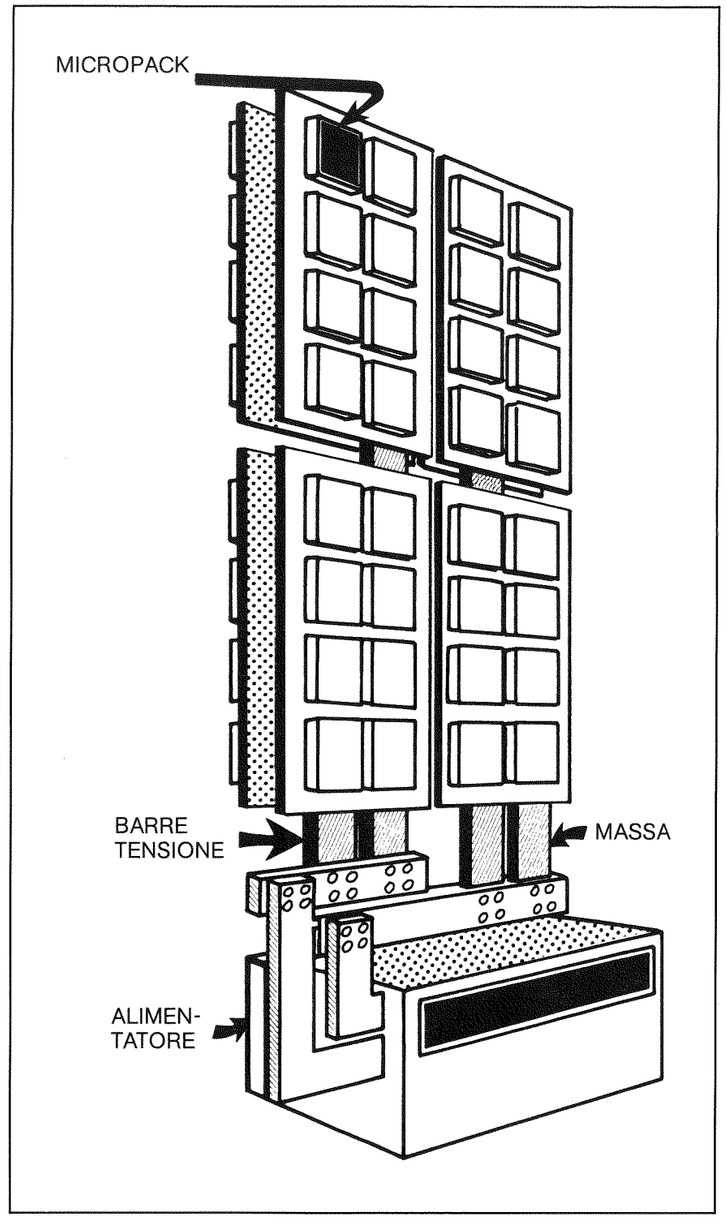


Fig. 10 - La foto mostra il micropack di fig. 8 completo di coperchio a tenuta ermetica. Oltre che per la protezione meccanica, il coperchio viene usato per fornire la tensione primaria al dispositivo.

di quella del packaging tradizionale. Un incremento analogo si ha pertanto nella densità di potenza dissipata. L'impiego di circuiti ad alta velocità, quale il CML, tende ad aggravare il problema, in quanto maggior velocità di commutazione implica maggiore dissipazione.
Per fissare le idee, un micropack del DPS 88 dissipa mediamente una potenza di 60 Watt. Tenendo conto delle dimensioni del dispositivo, l'asportazione del calore con i tradizionali sistemi ad aria diventa praticamente improponibile. Occorrerebbero ventilatori di grandi dimensioni e bisognerebbe montare sul micropack enormi radiatori; ciò condurrebbe a perdere per altra via la compattezza guadagnata col micropackaging. Oltre a ciò, il volume e la velocità dell'aria richiesti per il raffreddamento produrrebbero un livello di rumore inaccettabile.
La soluzione del problema va cercata in un'altra direzione e cioè nel raffreddamento a liquido che è di gran lunga più efficiente di quello ad aria.
Se questa è una scelta obbligata per macchine di grandissima potenza, quale il DPS 88, vi sono però molti modi per attuarla. Il sistema studiato dalla Honeywell si distingue sugli altri per la semplicità realizzativa.
La chiave di tale semplicità è la presenza nel micropack di un percorso di trasmissione del calore dal chip alla superficie posteriore

Fig. 11 - Una unità logica realizzata con micropackaging (Elaboratore Honeywell DPS 88). Vi sono in totale 64 micropack, montati su 8 schede. Al piede dell'unità, l'alimentatore relativo.



del substrato. Ciò è il risultato di alcune scelte di base nel progetto del micropack.
In primo luogo, il sistema di montaggio Flip-TAB, che permette la sottrazione di calore dall'intera area del chip.
In secondo luogo, i contatti di ingresso/uscita del micropack sono realizzati senza pin sporgenti, lasciando libera l'intera superficie posteriore del substrato.
In terzo luogo, lo strato di epoxy termoconduttivo usato nel processo Flip-TAB si estende oltre l'area del chip, abbracciando i terminali in rame del chip stesso e realizzando in tal modo un percorso secondario di conduzione del calore. Questo doppio cammino di dispersione termica elimina la necessità di altri contatti fisici/meccanici col chip.
Una volta che il calore raggiunge la superficie posteriore del micropack, esso viene asportato da un dispositivo denominato SLIC (Silent Liquid Integral Cooler), che realizza un contatto secco (cioè senza grasso) con il micropack.
Lo SLIC consiste di tre parti (fig. 12): un sottile diaframma di rame (il rame è un ottimo conduttore di calore), con superficie annerita, che aderisce al retro del micropack; una scatola di plastica stam-

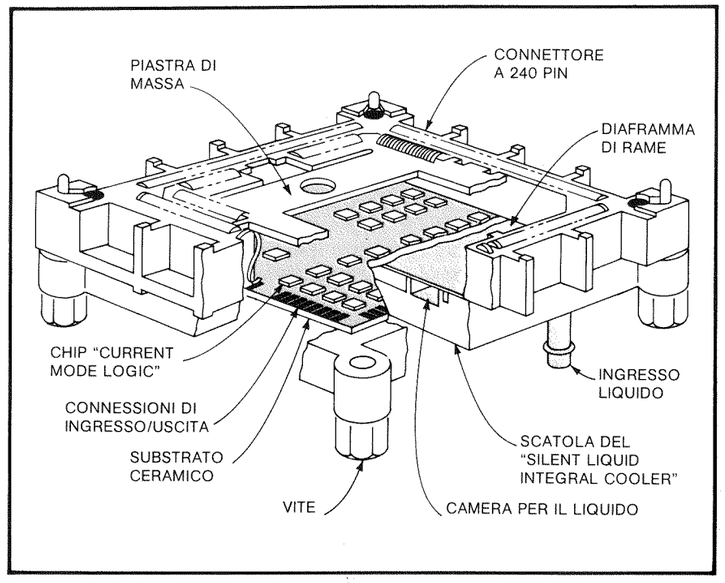


Fig. 12 - Un sistema di raffreddamento a liquido (Elaboratore Honeywell DPS 88). È costituito sostanzialmente da una scatola di plastica, chiusa da un lato con un sottile diaframma di rame. Quest'ultimo viene posto in contatto col retro del micropack. Nella scatola circola del liquido che asporta il calore generato. Questa soluzione (SLIC = Silent Liquid Integral Cooler) si distingue per semplicità ed efficienza.

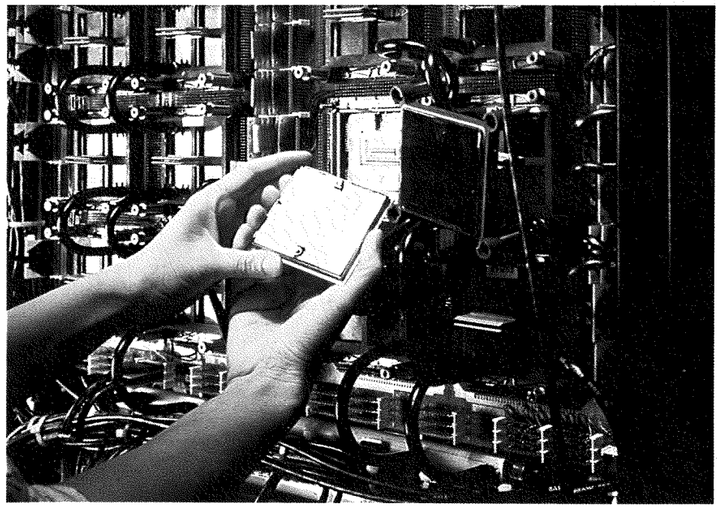


Fig. 13 - Fotografia del pannello posteriore di una unità logica dell'elaboratore DPS 88. Si può vedere un micropack smontato e sulla destra il relativo scambiatore di calore (SLIC). Il liquido di raffreddamento viene distribuito mediante tubi flessibili di plastica. Ciò consente una agevole manutenzione e, in particolare, di togliere un micropack senza interrompere il raffreddamento e quindi senza spegnere la macchina. La sostituzione di un micropack difettoso richiede circa 6 minuti.

pata con un percorso interno a serpentina; e infine due boccole di ottone per l'ingresso e l'uscita del liquido di raffreddamento. Le tre parti suddette sono incollate tra loro con epoxy.
Lo SLIC viene applicato al micropack mediante quattro viti; con questa operazione si preme anche il micropack nella sua sede, realizzando i relativi contatti elettrici, sia logici che di alimentazione.
Il liquido di raffreddamento è fornito ad ogni SLIC mediante tubi flessibili di plastica. Ciò consente una agevole manutenzione e, in particolare, di togliere un micropack senza interrompere il raffreddamento del sistema (fig. 13). La sostituzione di un micropack difettoso si effettua in circa 6 minuti, e non richiede di spegnere l'elaboratore.
Il liquido, che è acqua distillata con alcuni additivi, è tenuto in circolazione a bassa velocità e bassa pressione da una pompa. Il raffreddamento del liquido è realizzato a parte, lontano dall'elaboratore, mediante un convenzionale refrigeratore ad acqua o ad aria.
Il sistema descritto si distingue dalle soluzioni adottate in altri elaboratori per la semplicità e l'efficienza.
Un raffreddamento a liquido ben realizzato non presenta in sè problematiche d'uso particolari. Garantisce invece un livello di temperatura ottimale dei circuiti elettronici, che si traduce in una maggiore affidabilità globale del sistema.

7. Conclusioni

La realizzazione di elaboratori di grande potenza richiede una revisione sostanziale degli usuali schemi di packaging.
Il micropackaging costituisce una importante risposta a questo problema, consentendo di aumentare di almeno 10 volte la densità dei circuiti.
Il drastico aumento di compattezza realizzato col micropackaging implica una forte concentrazione di calore e quindi la necessità di passare dal tradizionale raffreddamento ad aria ad un sistema a liquido. Quest'ultimo, se ben progettato, non presenta in sè problematiche d'uso particolari; garantisce invece migliori condizioni di lavoro dei circuiti e quindi una maggiore affidabilità complessiva del sistema.

Bibliografia

[1] C. H. MCLVER *«Flip-TAB, copper thick films create the micropackage»* Electronics, Nov. 3, 1982.
[2] J. LYMAN *«Chip-carriers, pin-grid arrays change the printed-circuit-board landscape»* Electronics, Dec. 29, 1981.
[3] *«IEEE Workshop on Computer Packaging»* Monterey, Cal., May 13-15, 1981.
[4] T. CHIBA *«Impact of the LSI on High-Speed Computer Packaging»,* IEEE Transaction on Computer, April 1978.
[5] Le Scienze, Nov. 1978.
[6] Rapporti interni Honeywell sulla tecnologia del sistema DPS 88.
[7] J. LYMAN, *Supercomputers demand innovation in packaging and cooling»,* Electronics, Sept. 22, 1982.

L'informatica nel lavoro di ufficio

GIULIO OCCHINI

Honeywell Information Systems Italia
Milano

1. Premessa

Office automation e office augmentation, automazione ufficio e sua razionalizzazione, sono modi diversi di rappresentare un'unica realtà che si sta profilando all'orizzonte delle società industriali e quindi anche di quella italiana.

La prospettiva è tale da poter far compiere un salto senza precedenti alla produttività economica nazionale, ma anche, se non ben utilizzata, di creare situazioni di spreco tecnologico e aggravare le tensioni esistenti nel mondo del lavoro.

Le componenti sociali, metodologiche ed organizzative del fenomeno giuocano un ruolo così determinante che sarebbe ingenuo pensare di poterlo affrontare con mezzi e competenze esclusivamente tecnologiche: la stessa realtà dell'ufficio è per definizione così complessa e ricca di interconnessioni (formali e informali) da rendere assolutamente imprescindibile un approccio multidisciplinare.

In altre parole, se è in generale vero che qualsiasi processo di automazione non è mai il primo, ma piuttosto l'ultimo passo di una complessa attività di revisione e riprogettazione organizzativa che è compito specifico dei quadri direttivi indirizzare e controllare, quando il processo riguardi l'ufficio questo lo è in modo particolare per il numero delle persone coinvolte e per l'impatto profondo sulle loro modalità di lavoro e quindi sulla stessa struttura dei ruoli esecutivi, professionali e manageriali.

La automazione ufficio diventa così una variabile strategica di gestione della impresa, in uno con le risorse umane e con quelle finanziarie cui è strettamente collegata, e il modo con cui ne viene affrontata la pianificazione e la attuazione è in grado di condizionare la stessa strategia di sopravvivenza competitiva.

Ciò premesso, lo scopo di questo articolo è di:

- chiarire le ragioni per cui si ritiene che il processo di razionalizzazione e "informatizzazione„ dell'ufficio si porrà nel prossimo futuro come una scelta irrinunciabile da parte delle aziende e degli Enti che vorranno mantenere un ragionevole livello di efficienza nel settore in cui operano;
- analizzare, sia pure in modo sintetico, gli aspetti di differenzazione e di continuità nei riguardi delle applicazioni di informatica classica anche se può apparire paradossale definire "classica" una disciplina che non ha più di vent'anni di vita;
- fornire degli elementi di pianificazione per un intervento nell'ufficio, elementi fondati sia su considerazioni finanziarie (ritenute anche in questo caso determinanti nella scelta imprenditoriale) che di opportunità applicative;
- prospettare alcuni risultati di ricerche e di esperienze sviluppate negli Stati Uniti o altrove non tanto perchè si giudichino trasferibili così come sono alla realtà italiana, ma perchè significativi di approcci che possono fornirci orientamenti e interessanti stimoli di riflessione.

Lo spirito con cui verranno affrontati tali argomenti, al di là di quanto detto sulla necessità di un approccio interdisciplinare, è gia esplicito nello stesso titolo dell'articolo.

Utilizzeremo anche noi la dizione "automazione di ufficio„ ma parlando di "Informatica nel lavoro di ufficio„ vogliamo sottolineare due aspetti a nostro avviso determinanti per una corretta comprensione del tema.

Il primo si riferisce al fatto che, avendo già l'informatica una storia ventennale di applicazioni al lavoro di ufficio (o meglio, al lavoro impiegatizio), dal punto di vista tecnologico si tratta di estendere quelle che sono le sue "tradizionali„ aree di intervento; il secondo che proprio la dizione "automazione di ufficio„ si presenta inadeguata a descrivere la complessità del fenomeno che vogliamo trattare in quanto evocatrice di quei modelli di parcellizzazione del lavoro di fabbrica che, se trasferiti in un contesto assolutamente diverso sia per obiettivi che per struttura dei processi di lavoro, perdono qualunque significato concreto.

Ciò non toglie che continueremo ad adottare questa dizione in quanto ormai universalmente accettata per significare l'attuale stadio di sviluppo del processo di informatizzazione del lavoro di ufficio, processo peraltro in corso, da più decenni, nelle società occidentali.

Fig. 1 - Gli spostamenti nella occupazione

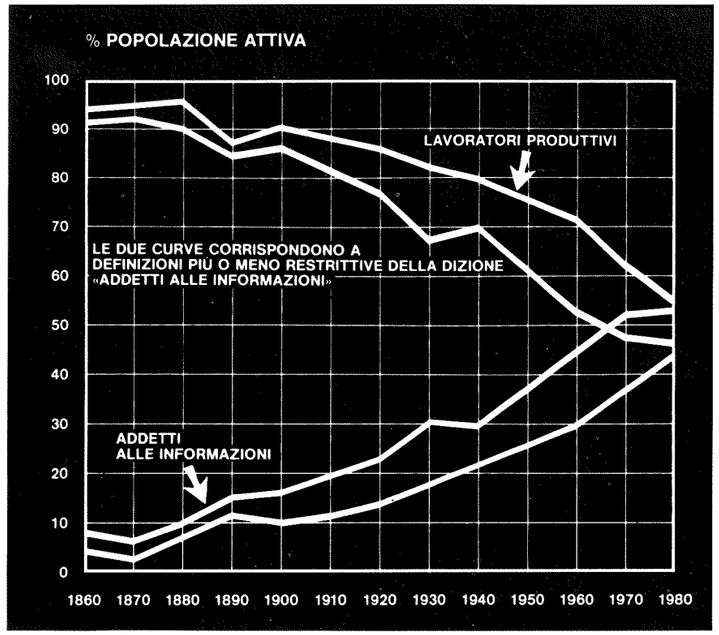


Fig. 2 - La velocità del mutamento

PAESE	1955	1976
FRANCIA	45.2%	51.1%
R.F.T.	32.6%	47.8%
U.S.A.	45.2%	67.5%
U.K.	45.1%	57.2%
SVEZIA	38.1%	58.4%

FONTE: FROST AND SULLIVAN

2. Le ragioni dell'automazione

Le motivazioni che sostengono la crescita esponenziale della domanda di automazione nel trattamento delle informazioni sono tutte concettualmente riconducibili al valore che l'informazione stessa assume nel mondo industriale avanzato e quindi all'importanza che la sua attività di gestione riveste.

In realtà nelle economie occidentali (e in quella giapponese) si sta assistendo da tempo a un progressivo spostamento della forza lavoro da attività di produzione di beni (lavoro operaio o di "blue collar„ secondo la dizione americana) ad attività di trattamento di informazioni (lavoro impiegatizio o di "white collar").

La fig. 1 rappresenta questo fenomeno così come si manifesta nelle economie (tra cui l'italiana) aderenti all'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE) e mostra come in tali economie sia già stata superata la quota del 50% della occupazione nel lavoro di trattamento informazioni.

Ma, fatto forse meno noto e più interessante, tale spostamento sta avvenendo in modo fortemento accelerato da circa venti anni come dimostra la fig. 2 in particolare nel caso della Svezia che certamente rappresenta per l'Europa un modello socioeconomico di riferimento.

Anche se non esistono, a conoscenza di chi scrive, statistiche analo-

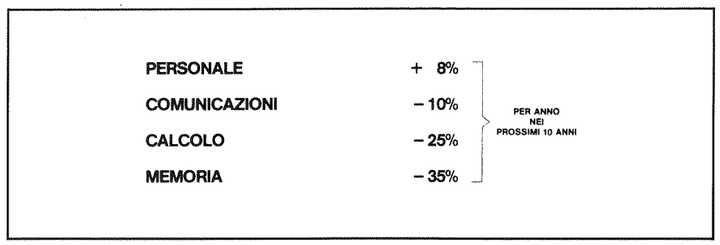


Fig. 3 - L'andamento dei costi

ghe, è chiaro che anche in Italia ci si trova di fronte ad andamenti del tutto confrontabili: le indicazioni del saggio di Sylos Labini [1] la esplosione delle attività terziarie e la crescente attenzione prestata al movimento dei quadri sono tutti elementi che lo comprovano.

Da parte di alcuni economisti si prospetta a questo riguardo la nascita di un settore a sè, il quaternario (o terziario avanzato), in cui la merce di scambio sia l'informazione, la competenza, il sapere [2].

Se questa è la prospettiva verso la quale stanno procedendo le nazioni industriali, ha un senso preciso domandarsi quali siano, a livello aggregato prima, e a livello aziendale poi, le componenti di costo delle attività di trattamento informazioni, la loro relativa incidenza sul costo totale e il loro andamento relativo nel tempo.

La fig. 3 mostra come, sempre con riferimento ai paesi dell'OCSE, le stime correnti assumano un netto incremento annuo in valore reale (al netto cioè dell'inflazione) per il costo del personale e (ricordiamolo, stiamo ragionando a parità di prestazioni) un ancor più netto decremento di quello delle apparecchiature di supporto all'uomo per il trattamento delle informazioni (nelle aree della archiviazione, elaborazione e comunicazione).

Questo secondo fatto non deve stupire considerando che tutti e tre i tipi di apparecchiature si riconducono alla tecnologia elettronica che sappiamo presentare, da oltre dieci anni, in uno con il processo di miniaturizzazione (cioè di contrazione delle dimensioni a parità di funzioni logiche), un andamento di riduzione costi che si sviluppa in modo proporzionale.

Semmai ci si può domandare perchè questa drastica diminuzione dei costi tecnologici trovi un riscontro molto limitato nella moderata riduzione del prezzo delle comunicazioni riportata in tabella (valida mediamente nei paesi OCSE) o addirittura dia luogo ad incrementi come nello specifico caso italiano.

Le ragioni di questa apparente anomalia sono molteplici ma tutte, in realtà, riconducibili alla considerazione che nei paesi europei (e in particolare in Italia), vigendo il regime di monopolio sulle comunicazioni, le tariffe vengono stabilite non soltanto in base al costo ma piuttosto in base a considerazioni politiche e di opportunità sociale.

C'è inoltre da aggiungere che, sempre nel caso delle comunicazioni, la tecnologia elettronica non ha ancora del tutto eliminato dal mercato quella elettromeccanica e quest'ultima presenta indubbiamente costi di gestione assai superiori (in Italia in particolare si può stimare che, alla data, meno del 5% delle centrali di commutazione sia di tipo elettronico).

3. Il costo del trattamento informazioni

Queste semplici considerazioni sull'andamento di costo delle componenti relative al trattamento delle informazioni conducono naturalmente ad una prima importante conclusione [3].

Si definisca come sistema informativo di un'azienda, di un Ente o addirittura di una nazione l'insieme delle attività di trattamento informazioni che vi vengono svolte.

Tenendo allora conto di quanto detto, si può immediatamente osservare come, in assenza di qualsiasi forma di automazione (cioè di adozione delle tecnologie di archiviazione, calcolo e comunicazioni elettroniche) nel corso di circa 8 anni il costo del sistema informativo venga ad essere moltiplicato per 2.14 a parità di volume di informazione trattato (vedi prima curva dall'alto della fig. 4).

In altre parole ci si trova dopo otto anni a fare esattamente le stesse cose (almeno quanto a volumi) ma con costi più che raddoppiati.

Se si introduce un indice di automazione (espresso come rapporto tra costo della tecnologia utilizzata e costo totale del sistema informativo) si riesce a limitare la crescità dei costi (vedi seconda curva dall'alto della stessa fig. 4) per arrivare nel caso del valore di tale indice pari a 0.5 ad annullarla mantenendo tali costi invariati nel tempo.

Per tassi di automazione superiori al 50% si riesce addirittura a ridurre il costo del sistema informativo sino a un livello minimo pari allo 0.16 di quello originario arrivando al limite del 100% di automazione.

Un tasso di automazione del 100% é ovviamente da considerarsi asintotico anche se valori dell'80%-90% non sono irrealizzabili in certi settori industriali (si pensi alle aziende a produzione continua) o in certe aree aziendali (si pensi al ciclo integrato ordini-materiali-produzione-fornitori in tante aziende a produzione discreta oppure alla gestione in linea delle operazioni di sportello realizzata presso un gran numero di Istituti di Credito).

Fig. 4 - Costo del sistema informativo e tasso di automazione

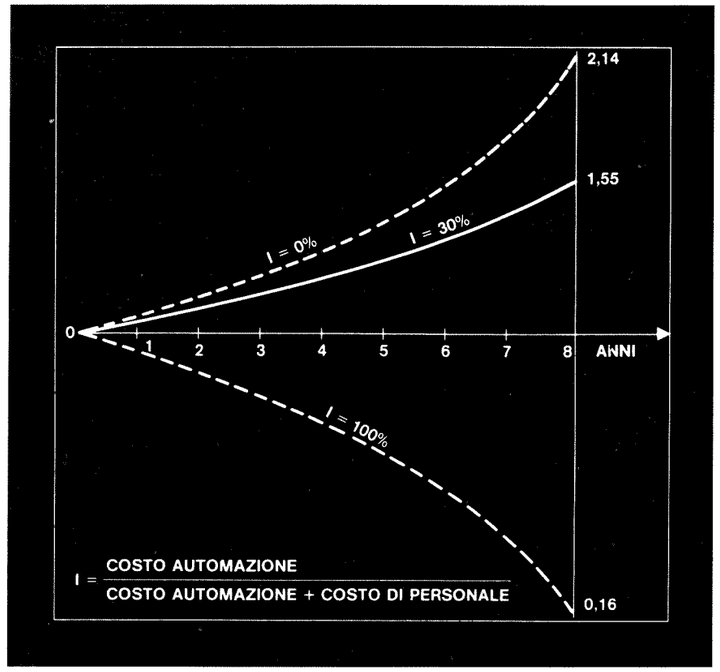
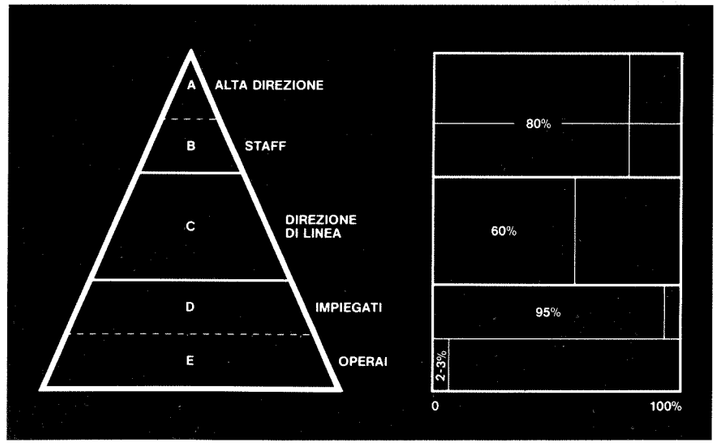


Fig. 5 - Aziende industriali: attività dedicata alla gestione informazioni



4. La situazione nelle aziende metalmeccaniche e finanziarie

Volendo rapportare queste considerazioni generali alla realtà di specifiche aziende e facendo ancora riferimento a [3], si assuma come primo esempio un campione appartenente al settore metalmeccanico.

Se si analizza la quantità di tempo dedicato al trattamento informazioni dai vari ruoli aziendali secondo la fig. 5 e quindi si rapporta questa percentuale al costo che aziendalmente tali ruoli assumono si ha, come illustrato nella figura 6, che circa il 35% dei costi di personale è da imputarsi al sistema informativo.

E ciò, nonostante il fatto che, in questo tipo di aziende, la popolazione operaia assorba più del 50% del costo del personale.

Se infatti prendiamo in considerazione le aziende finanziarie (o gli Enti amministrativi) dove l'informazione rappresenta la materia prima su cui esercitare l'attività di trasformazione e dove quindi è assente, per definizione, il lavoro operaio, la percentuale del costo del personale dedicato al sistema informativo sale all'86% (vedi fig. 7).

Queste valutazioni, che per la loro approssimazione vogliono avere un carattere puramente indicativo, raggiungono tuttavia lo scopo di rendere evidente la dimensione economica del problema informativo aziendale e quindi la necessità di affrontarlo puntando a migliorare sostanzialmente la produttività di coloro che vi sono addetti.

FASCE	% COSTI PERSONALE	% ATTIVITÀ GESTIONE INFORM.	% COSTO
A	5	80	4
B	7	80	6
C	25	60	15
D	10	95	9
E	53	2	1
	100		35

Fig. 6 - Costo di gestione informazioni nelle aziende industriali

FASCIA DI ATTIVITÀ	% SUL COSTO DI PERSONALE	% DI TEMPO DEDICATO ALLA GESTIONE INFOR.	COSTO DI GESTIONE INFOR.
IMPIEGATI	43	95	41
DIREZIONE OPERATIVA	25	80	20
ALTA DIREZIONE	32	80	25
TOTALI	100	100	86

Fig. 7 - Costo di gestione informazioni nelle aziende finanziarie e amministrative

Il rinunciare a porsi questo obbiettivo significa subire passivamente la lievitazione dei costi illustrata al punto precedente, senza ottenere in contropartita alcun tipo di miglioramento gestionale, il che vuol dire accettare ineluttabilmente la perdita della competitività di mercato.

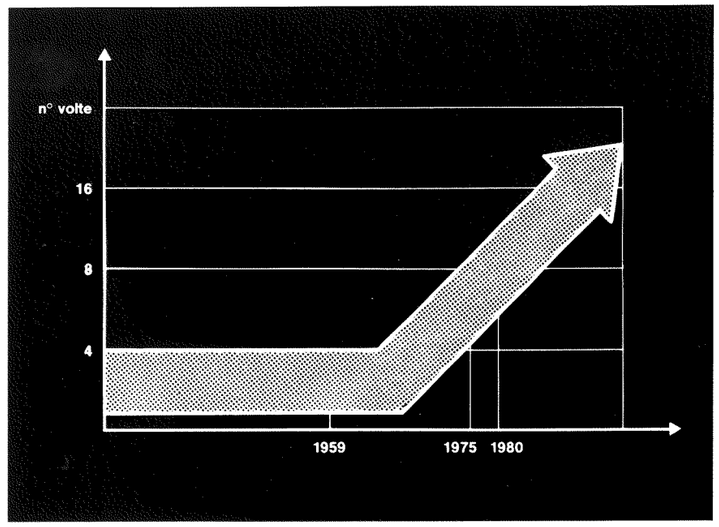
Naturalmente, nel porselo, e soprattutto nel perseguirlo, bisognerà tenere opportunamente conto del fatto che la produttività del lavoro di trattamento informazioni non è tanto questione di efficienza nell'eseguire certe procedure quanto piuttosto di continua verifica della congruenza di ciò che si fa con gli obbiettivi della organizzazione di cui si è parte, e quindi, come si usa dire, di efficacia, e quest'ultima non può non essere fortemente condizionata da fattori quali la motivazione personale, l'ambiente nel quale il lavoro si svolge, il tenore dei rapporti interpersonali, etc.

5. L'esplosione delle esigenze informative

Si è accennato al punto 2 alla importanza dominante che, l'informazione va assumendo nel mondo industriale avanzato.

Questo fenomeno si è in particolare manifestato negli ultimi 10-15 anni, anni in cui si è verificata una crescita esponenziale delle esigenze informative necessarie a gestire una qualsiasi organizzazione, oltre che dal punto di vista qualitativo anche da quello quantitativo (fig. 8).

Fig. 8 - La esplosione delle informazioni



Molteplici sono le ragioni che possono giustificare questo andamento destinato ad ulteriormente accentuarsi nei prossimi anni e che possono farsi risalire sia alla internazionalizzazione dei mercati come alla diffusa instabilità economica e politica, alla variabilità dei tassi di cambio come alla crisi delle fonti energetiche tradizionali.

In realtà tutte queste cause risultano, se sottoposte ad una analisi sufficientemente accurata, riconducibili ad una matrice concettuale unica esprimibile in questi termini:

l'informazione risulta essere la risorsa più conveniente per gestire, in situazioni di incertezza, una qualsiasi organizzazione che abbia finalità di efficienza economica e/o sociale.

L'alternativa alla informazione è infatti rappresentata dall'accantonamento di riserve (di materiale, di personale, di capitale) mobilitabili al soppraggiungere dell'evento imprevisto (e imprevedibile, dato il generale contesto di incertezza) per poterlo fronteggiare e superare.

Ma questa strategia, dati gli attuali costi di immobilizzo, è in pratica perdente dal punto di vista economico rispetto a quella di mettere in atto un sistema informativo sufficientemente puntuale e flessibile che consenta alla organizzazione di adattarsi dinamicamente al contesto nel quale si trova ad operare ed anzi traendo il massimo di opportunità proprio dalla situazione di incertezza.

Alla luce di questa considerazione assumono un valore diverso anche le considerazioni sviluppate nei punti precedenti. Non si tratta più infatti di minimizzare semplicemente i costi di trattamento delle informazioni, ma piuttosto di sviluppare una capacità informativa (che non può non avvalersi della automazione come supporto essenziale) in grado di mantenere o aumentare la competitività della organizzazione affinandone il processo decisionale.

6. Il problema della produttività e degli investimenti

La produttività, nella sua accezione comune, è una misura di efficienza che viene definita come la quantità di lavoro necessaria a produrre un determinato risultato sia esso quantitativo o qualitativo (nel secondo caso, la definizione richiede ovviamente di essere ulteriormente chiarita).

Se il risultato migliora senza che sia necessario aggiungere lavoro, la produttività aumenta; così come aumenta se lo stesso risultato può essere ottenuto con meno lavoro.

Le statistiche sulla produttività d'ufficio o non esistono o sono notevolmente controverse in quanto vengono ricavate per complemento, rispetto alla produttività economica generale, da quelle relative alla attività di produzione dei beni.

La fig. 9, ad esempio, riporta una stima ottenuta con questi criteri che può essere tuttavia assunta come indicativa del divario oggettivo che sussiste tra lavoro operaio e lavoro impiegatizio; per poter però attribuire un significato alla produttività di quest'ultimo non si può prescindere da alcune considerazioni che ne chiariscano il significato.

In effetti, se può avere un senso misurare, secondo la definizione classica sopra richiamata, la produttività impiegatizia a livello esecutivo (sia pure con le opportune e non immediate modifiche), per le attività di tipo direttivo e professionale tale definizione dà luogo a evidenti assurdità.

Un responsabile di settore vale di più se produce 10 documenti al giorno anzichè 5 o se passa più tempo in riunioni di un altro? Evidentemente no e di qui la necessità di rifarsi a misure di tipo sostanzialmente diverso.

Queste misure non possono che riguardare la cosiddetta "prestazione", cioè il grado di raggiungimento degli obbiettivi. La pianificazione marketing di una azienda, giusto per fissare un esempio, può essere misurata dal costo, dalla qualità e tempestività dei suoi piani, nonchè

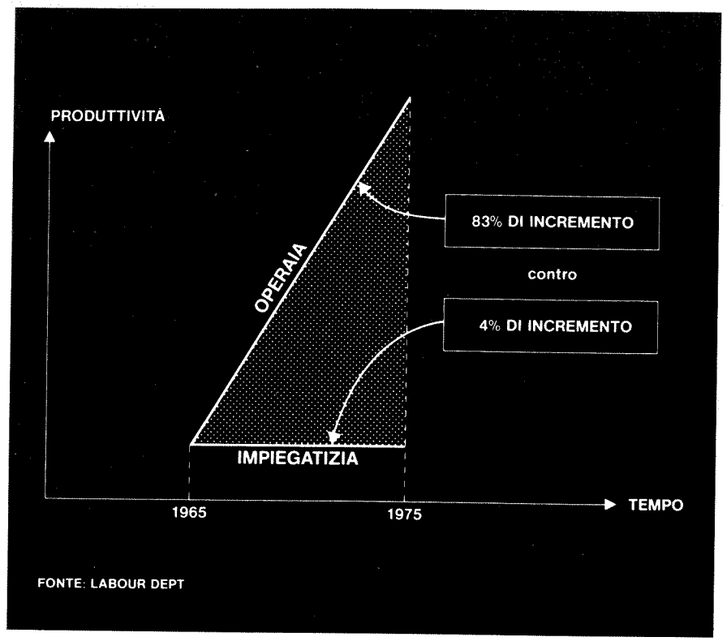


Fig. 9 - La produttività

SETTORE	AMMONTARE
AGRICOLTURA	50 K $
PRODUZIONE INDUSTRIALE	30 K $
TRATTAMENTO INFORMAZIONI	2.5-3 K $

FONTE: LABOUR DEPT

Fig. 10 - Gli investimenti per addetto (USA)

dalla sua capacità di operare in sinergia con i responsabili commerciali per implementarli.
Di fatto, anche per il lavoro esecutivo, la prestazione è la misura per eccellenza del rendimento di una persona, di un gruppo o di un settore aziendale.
In un certo senso, la prestazione è un concetto assai più ampio di quello di produttività e migliorare la prestazione significa ottenere risultati ben più efficaci ai fini degli obiettivi aziendali fissati di quanto non avvenga limitandosi ad aumentare la semplice produttività.

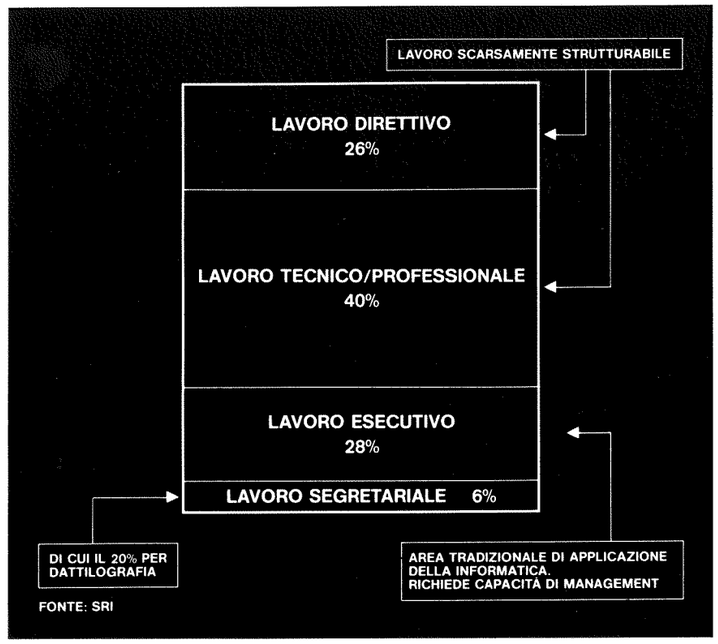
Nel lavoro esecutivo, si può talvolta semplificare il problema, limitandolo alla misura della produttività quantitativa, ma anche in questo caso non va dimenticato che tale produttività non è un valore di per sè ma deve essere finalizzata agli obiettivi della azienda.

Purtroppo, disponiamo di dati empirici relativamente scarsi per correlare le prestazioni dei livelli professionali medio-alti con la disponibilità di strumenti di automazione, e questo perchè l'informatica classica si è finora limitata ad insistere con le sue applicazioni sui ruoli esecutivi con esiti spesso più che soddisfacenti [4, 5].

Resta tuttavia la nota correlazione riportata in fig. 10 tra produttività comunque intesa e livelli di investimento nel settore primario, secondario e terziario.

Questa correlazione mette in effetti in luce una legge economica ricorrente: un settore nasce per rendere più produttivo, tramite investimenti di automazione, il settore che lo precede in termini di valore aggiunto del prodotto scambiato (nell'ordine agricolo o minerario, manifatturiero o di servizio); quindi espandendosi ed assorbendo sempre più risorse, diventa a sua volta causa di inefficienza del sistema.

Fig. 11 - Dove e come investire



Così la macchina a vapore fu inizialmente impiegata in agricoltura (cioè per rendere più efficace il settore primario) e il terziario nacque per rendere più efficace con la sua dotazione di servizi il settore manifatturiero o secondario.

Attualmente, è lo stesso terziario che deve essere reso più efficiente e ciò avverrà grazie allo sviluppo di quel settore la cui merce di scambio, come già accennato precedentemente, è costituita dall'informazione e dalla competenza, cioè dal settore che oggi si comincia a denominare come "quaternario" (o "terziario evoluto") e di cui l'informatica fa parte per definizione.

Il processo si sviluppa facendo lievitare gli investimenti per addetto in misura via via maggiore quanto più consolidato è il settore stesso anche in termini di capacità di valutazione della sua efficienza e gli anni che stiamo vivendo sono evidentemente quelli in cui i problemi sul tappeto riguardano il settore dei servizi informativi.

7. Dove investire

Si è parlato di investimenti in automazione, ma proprio alla luce delle considerazioni sulla produttività (o sulle prestazioni) fatte precedentemente, è necessario, nel caso del trattamento delle informazioni, identificare almeno in termini generali i ruoli cui ci si vuole prioritariamente rivolgere.

La fig. 11 mostra una ripartizione indicativa del costo del lavoro di ufficio in un azienda del secondario evoluto (al netto, in questo caso, della componente operaia) o del terziario.

Come si può constatare, poco più di un terzo del costo è concentrato sui ruoli esecutivi, mentre quasi due terzi è riferito ai ruoli professionali e direttivi.

Questa constatazione suggerisce immediatamente alcune considerazioni. La prima, che la quota relativamente minoritaria del costo del lavoro esecutivo possa essere interpretata come il risultato di 20 anni di investimenti di tipo informatico attuati nel settore: in effetti, come è noto, le applicazioni correnti di EDP si sono storicamente concentrate nella "automazione" del lavoro impiegatizio di carattere ripetitivo e dovrebbero quindi essere riuscite a ridurne l'incidenza sui costi totali.

La seconda, che nell'ambito del lavoro esecutivo e, più in particolare, di quello segretariale, l'attività di trattamento testi presenti un'incidenza sul totale di poco più dell'1% dei costi; ciò significa che se l'obiettivo della automazione ufficio è quello di un significativo recupero di produttività non è certo operando sull'attività di dattiloscrittura che si può pensare di conseguirlo.

Bisogna puntare invece all'attività professionale e direttiva che insieme costituiscono quasi i due terzi del costo totale e sulle quali sinora, salvo casi particolari, l'informatica non ha inciso se non in misura marginale.

D'altra parte il trattamento testi ha una sua indiscutibile importanza come mezzo di introduzione delle informazioni nel sistema di automazione ufficio; si può, in altre parole, sostenere che la dattiloscrittura stia all'automazione ufficio come il "data entry" sta alla elaborazione dati tradizionale.

Ma ciò non deve giustificare, proprio per le ragioni precedenti, la confusione tra investimenti in automazione ufficio e quelli in sistemi di dattiloscrittura che, o sono contestuali, o possono dar luogo a in-

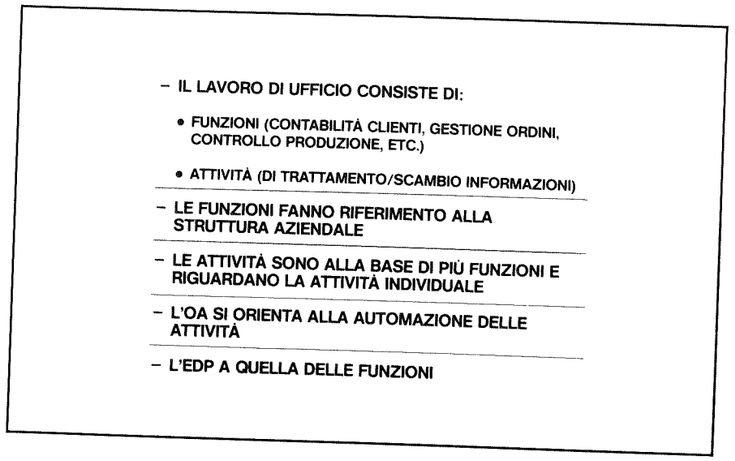


Fig. 12 - Office automation e data processing

compatibilità insuperabili e a frammentazioni del lavoro assolutamente controproducenti in termini di produttività ed efficacia.

8. Funzioni e attività

Individuati i ruoli cui indirizzare prioritariamente la automazione di ufficio si pone il problema di come, grazie ad essa, si possa migliorare la loro prestazione nel senso sopra discusso.

Allo scopo è necessario approfondire le caratteristiche del lavoro di trattamento informazioni e le sue modalità di automazione.

Lo sviluppo delle "tradizionali" applicazioni di informatica ha operato (vedi fig. 12) nel senso di "proceduralizzare" le varie funzioni aziendali che a tale processo si prestassero (le funzioni, cioè, cosiddette strutturabili) e che, per lo più, risultavano di pertinenza dei ruoli esecutivi (a titolo di esempio, basti ricordare, nel mondo industriale, le varie contabilità, la gestione ordini, quella della produzione e degli acquisti etc. mentre nel settore bancario, i diversi servizi operativi, dai conti correnti ai libretti di risparmio, dai titoli al portafoglio effetti, etc.).

Nel far questo, l'informatica ha implicitamente considerato come parte integrante di tali funzioni le attività elementari di archiviazione, reperimento e scambio di informazioni incorporandole così nelle relative procedure e quindi nei relativi programmi su elaboratore.

In realtà queste attività sono largamente indipendenti dalla applicazione in quanto, per definizione, rappresentano il supporto di qualunque funzione informativa che si svolga nell'ufficio; si può quindi pensare di enuclearle e normalizzarle, rendendole quindi disponibili ogni qualvolta risultino necessarie non solo per le applicazioni di tipo esecutivo ma anche per i compiti non strutturabili, e quindi non affrontabili con l'informatica "classica„.

La automazione di tali attività costituisce appunto uno degli obiettivi prioritari della automazione ufficio e uno degli elementi di differenziazione dalla informatica "classica„.

Un'altra considerazione riguarda la destinazione del servizio automatizzato.

Nel caso delle funzioni aziendali di tipo strutturato (oggetto, cioè, della informatica) i destinatari fanno parte, nell'ambito della struttura organizzativa della azienda, di gruppi (e/o di uffici) "omogenei„ quanto a obiettivi e modalità di lavoro (la procedura, appunto, di esecuzione della specifica funzione aziendale); nel caso delle attività elementari di trattamento informazioni invece, proprio per il loro carattere di supporto, l'utente lo può essere a titolo individuale, anzichè come appartenente a un determinato settore operativo, e l'impiego dello strumento di automazione può quindi diventare una sua scelta professionale di opportunità.

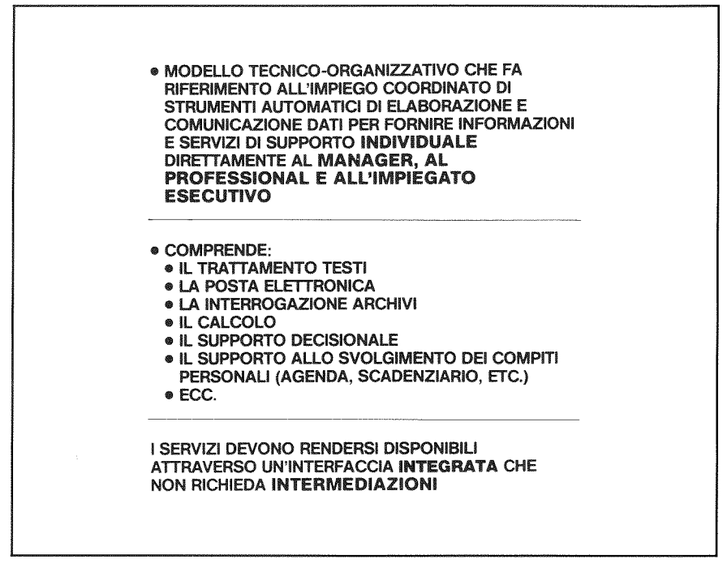
In altre parole, nelle applicazioni informatiche è implicita una componente di costrizione all'uso pena l'insuccesso delle applicazioni stesse (non è organizzativamente accettabile che le stesse procedure vengano eseguite da alcuni in modo manuale e da altri in modo automatizzato); in quelle di automazione ufficio invece, non può essere che l'utente a decidere sulla opportunità del loro impiego in funzione del compito che ha da svolgere e della sua professionalità.

Ciò significa che l'automazione ufficio va considerata in gran parte come un servizio di supporto allo svolgimento del lavoro oggi non strutturabile dello specialista e del dirigente ed è in questa accezione che può contribuire significativamente, come approfondiremo nel seguito, alla qualità della loro prestazione.

9. Alternative di automazione

Dovendo schematizzare per chiarezza di esposizione, si può affermare che gli approcci "tradizionali" alla automazione dell'ufficio si possano suddividere in due grandi categorie:

Fig. 13 - Definizione di automazione ufficio



a) approcci che adattano sistemi esistenti del tipo multifunzionale alle applicazioni di ufficio lasciando inalterata la interfaccia EDP

b) approcci che inseriscono apparecchiature specifiche per attività specifiche

Alla prima categoria appartengono le soluzioni realizzate tramite lo sviluppo di un opportuno "software„ su elaboratori gestionali di tipo multifunzionale di piccola, media o grande dimensione.

Tale "software„ può essere utilizzato mediante terminale e fornisce generalmente le funzionalità di trattamento automatico delle informazioni testuali, di loro archiviazione su memoria di massa e di loro messa a disposizione, tramite il servizio chiamato di "posta elettronica„ (electronic mail), a tutti gli interessati che, sempre tramite terminale, abbiano diritto ad accedervi per consultarle ed eventualmente modificarle.

Poichè sullo stesso elaboratore è stata attuata la automazione delle procedure informatiche della azienda, questo tipo di soluzione ha il grande merito di realizzare la unificazione, almeno fisica se non anche logica, del mondo "tradizionale„ della elaborazione dati con quello delle informazioni di ufficio; di tradurre in atto, cioè, quella complementarietà tra i due ambiti applicativi che, a nostro (ma non solo nostro) giudizio, è essenziale per qualsiasi strategia di sviluppo del sistema informativo aziendale.

I limiti di questa impostazione stanno nella sua scarsa possibilità di diffusione presso utenti che non accettino l'interfaccia di comando caratteristica dei sistemi di elaborazione dei dati; d'altra parte, trasformare tale interfaccia in modo da renderla di più facile utilizzo significherebbe appesantire eccessivamente il carico aggiuntivo di gestione del sistema con conseguente drastica riduzione delle sue prestazioni generali.

Se si riflette al fatto che una delle capacità essenziali della automazione di ufficio deve essere proprio quella di rivolgersi a personale assolutamente privo di competenza informatica, ci si rende conto del perchè tale soluzione non abbia sinora trovato che un modesto riscontro applicativo.

Alla seconda categoria appartiene invece tutto quell'insieme di apparecchiature di ufficio che ha lo scopo di rendere più efficienti certe specifiche attività di trattamento della informazione.

Così ad esempio, la macchina da scrivere elettronica per ridurre il costo segretariale di elaborazione dei testi, la fotocopiatrice programmabile per razionalizzare il processo di duplicazione, il telefono a tastiera e con memoria per accelerare la composizione dei numeri e il richiamo di quelli più frequentemente utilizzati, etc.

Le osservazioni che si devono fare su questa impostazione, quando sia fine a sè stessa e non inquadrata in un disegno più generale, sono di due tipi.

Prima di tutto, il recupero di efficienza che questi dispositivi consentono riguarda essenzialmente il ruolo segretariale che, lo si è già precedentemente sottolineato, incide in misura relativamente modesta sui costi del lavoro di ufficio (vedi fig. 11).

Ma, elemento forse più importante, tale ipotetico recupero lo si otterrebbe ulteriormente accentuando quella frammentazione dei compiti che rappresenta oggi la causa principale della inefficienza (e della frustazione) del lavoro impiegatizio; di conseguenza, anche se questa soluzione può consentire maggiore efficienza a livello della singola attività, al livello della mansione è in profondo contrasto con l'obiettivo prioritario di ricomposizione e integrazione dei compiti; va quindi considerata come concettualmente erronea e incapace di dare contributi che non siano marginali.

10. Cosa è l'automazione di ufficio

Alla luce della analisi sin qui svolta risulta chiaro come per "automazione di ufficio„ si debba intendere quanto specificato nella definizione di fig. 13 che tuttavia suggerisce alcune ulteriori precisazioni.

Il modello presenta certamente una componente tecnologica che lo rende oggi (a differenza di pochi anni fa) economicamente perseguibile, ma è soprattutto di tipo organizzativo proprio in quanto attiene alla razionalizzazione del lavoro di ufficio e quindi alla professionalità dei suoi addetti; ciò significa che gli aspetti di coinvolgimento e di assimilazione devono assumere una importanza prioritaria nella sua applicazione.

La destinazione individuale dei servizi è già stata messa in evidenza precedentemente e la priorità dei ruoli destinatari deriva dalla loro incidenza nel costo totale del lavoro di ufficio.

La funzionalità elencate includono non solo le attività di trattamento informazioni che abbiamo chiamato "elementari„ ma anche attività di carattere più evoluto la cui automazione può attuarsi esclusivamente in termini individuali in quanto caratteristiche della professionalità e dello stile dell'utente.

Si tratta per lo più di semplici procedure di analisi e calcolo su dati reperiti in archivio finalizzate al supporto del processo decisionale, alla pianificazione del lavoro e allo svolgimento dei compiti personali.

Va sottolineata, come conseguenza delle considerazioni svolte al punto precedente, la necessità di accedere ai diversi servizi tramite una interfaccia unificata che non eriga assurde barriere tra mondo della elaborazione dati e mondo dell'ufficio.

Infine il destinatario di tali servizi li deve poter utilizzare in modo diretto, senza intermediazioni di sorta, e questo è un altro elemento di differenziazione dall'informatica "classica„.

Ci si rende agevolmente conto di questa esigenza se si riflette al fatto che la ricerca di informazioni a supporto di una attività non strutturabile non può, per definizione, essere delegata in quanto la scelta delle informazioni stesse è intrinseca alla modalità individuali di svolgimento del compito; in altre parole, solo chi sa di quali informazioni abbisogna e perchè, è in grado di selezionare, tra le informazioni disponibili, quelle utili, nonchè di ovviare con altre alla eventuale carenza delle informazioni cercate.

La mancanza di ruoli di intermediazione, caratteristica dell'informatica "classica„, comporta una radicale revisione delle modalità di colloquio uomo-macchina; revisione che è effettivamente in corso ma sulla quale resta molto da fare anche in termini di ricerca teorica.

In particolare, oltre alla linguistica, la branca dell'informatica che è nota con il nome di "intelligenza artificiale„ è destinata a trovare in questa area, in uno con la psicologia cognitiva, il terreno ideale di applicazione.

11. La impostazione Honeywell

Torna a questo punto opportuno aprire una parentesi per accennare alle caratteristiche della attuale offerta Honeywell nel settore della automazione ufficio in quanto la si ritiene particolarmente rappresentativa di una impostazione coerente con le considerazioni sinora svolte.

I principi informatori di tale offerta sono:

- *Soluzioni ai problemi anticipabili (strutturati) e strumenti di supporto facilmente utilizzabili per quelli che non lo sono*
- *Integrazione al posto di lavoro delle funzioni di trattamento dei testi, dei dati, dei grafici, di archiviazione, di reperimento, di comunicazione, ecc.*
- *Ottimizzazione locale del sistema nei confronti della unità organizzativa cui è destinato (scelta delle sue dimensioni)*
- *L'ufficio come parte integrante della azienda e non come ente isolato*
- *Impiego dello strumento in prima persona da parte dell'utente*
- *Sicurezza sulle comunicazioni e sui documenti*

e, in quanto segue, vengono commentati brevemente.

Le attività di trattamento informazioni sono classificate in due categorie: quelle di tipo strutturabile (trattamento testi, archiviazione, reperimento, scambio di informazioni, etc.) e quelle che non lo sono.

Per le prime i sistemi OAS (Office Automation Systems) offrono soluzioni interamente pre-programmate e di utilizzo guidato attraverso processi di selezione "ad albero„ (colloquio "menu-driven„); per le seconde, degli strumenti di supporto (linguaggi "end-user„) delle cui modalità di impiego può facilmente impadronirsi anche chi non disponga di alcuna esperienza informatica.

La gamma OAS che comprende soluzioni che vanno dalla singola stazione di lavoro dedicata (specializzata cioè sul trattamento delle informazioni testuali) con capacità di elaborazione e memorizzazione locale (e tuttavia collegabile in rete), alla versione multi stazione e multifunzionale (utilizzabile cioè sia per applicazioni di automazione ufficio che per "tradizionali„ applicazioni di informatica), è intrinsecamente compatibile in quanto fondata su un'unica architettura di "hardware„ e di "software„.

Come tale la interfaccia del colloquio uomo-macchina è comune alla intera gamma e da qualsiasi stazione di lavoro (costituita, nella versione più elementare, da tastiera e schermo di visualizzazione), indipendentemente dalle caratteristiche tecnologiche e di collegamento, è possibile accedere in modo integrato alla totalità delle funzioni del sistema.

Al di là dalla dimensione del sistema complessivo, é importante, sia per il suo effettivo utilizzo che per ragioni economiche, che localmente la soluzione proposta sia ottimizzata sulle particolari esigenze della unità organizzativa in cui deve essere inserita.

Tale ottimizzazione, che consente appunto di dimensionare il sistema locale in modo da massimizzarne la resa economico/applicativa, è l'obiettivo più qualificante delle architetture di informatica distribuita.

Fig. 14 - Caratteristiche della stazione di lavoro "ideale"

CARATTERISTICHE	GIUDIZIO: ESSENZIALE
INTEGRAZIONE OA/DP	86%
CONTROLLO RISERVATEZZA	84%
POSTA ELETTRONICA	70%
ARCHIVIAZIONE	61%
DISPLAY GRAFICO	34%
CALENDARIO E SCHEDULATORE DI RIUNIONI	16%
DISPLAY A COLORI	8%

FONTE: IDC

Nel caso dell'OAS la architettura di riferimento è la "Distributed Systems Architecture„ (DSA) che adotta integralmente lo standard ISO (International Standards organisation) noto con il nome di "Open Systems Interconnection (OSI)„ [6].

L'importanza di riferirsi a una architettura sistemistica anche per una semplice sperimentazione è dovuta alla necessità di evitare, man mano che le applicazioni si estendono, qualsiasi cambiamento di impostazione in un'area così delicata dal punto di vista delle conseguenze sulla organizzazione del lavoro.

Cambiare impostazione significherebbe infatti alterare le modalità di utilizzo del sistema e questo comporterebbe nuovi investimenti in formazione dell'utente e in riprogettazione organizzativa ma soprattutto ingenererebbe frustrazione e scetticismo.

Queste considerazioni si attagliano in particolare alle comunicazioni intraufficio e interufficio che rappresentano il tessuto connettivo della azienda e che devono sottostare a specifiche discipline normalizzate a livello internazionale e facenti parte dello standard OSI.

La semplicità del colloquio uomo-macchina, la possibilità di ottimizzare il sistema sulle esigenze locali e la assoluta compatibilità delle varie configurazioni nel loro insieme, fanno si che l'utente possa utilizzare le varie funzionalità senza ricorrere a intermediazioni che, lo abbiamo già osservato, comportano, in questo campo, una sostanziale perdita di efficacia.

Infine, aspetto ultimo nella elencazione ma non assolutamente in ordine di importanza, le misure che oggi assicurano la riservatezza e affidabilità delle informazioni di ufficio devono essere garantite e, possibilmente, rafforzate dal processo di automazione.

Su questo argomento non ci si è soffermati in precedenza in quanto non lo si ritiene caratteristico della sola automazione ufficio ma piuttosto del processo generale di informatizzazione; tuttavia è chiaro che quando ci si indirizza ai ruoli direttivi e professionali il problema assume una importanza particolare e richiede quindi soluzioni particolarmente esaurienti.

Prima di chiudere la parentesi di questo paragrafo torna utile richiamare, a verifica della validità della impostazione sin qui seguita, alcuni risultati di una ricerca condotta recentemente (1981) dalla International Data Corporation (IDC) su un campione significativo delle maggiori aziende statunitensi [7].

Si sono scelti due specifici argomenti: la struttura della stazione di lavoro di ufficio "ideale„ secondo l'opinione dei quadri direttivi di tali aziende e la tipologia dell'utilizzatore di detta stazione.

Sul primo tema la fig. 14 dimostra come la possibilità di integrazione tra il mondo delle applicazioni informatiche (presumibilmente, già molto sviluppato nelle aziende del campione) e quello della automazione di ufficio sia giudicata la più importante tra le caratteristiche della stazione di lavoro "ideale„ seguita a ruota dalla capacità di garantire la sicurezza e riservatezza delle informazioni trattate.

Vengono poi alcune funzionalità di tipo più tecnico (prima fra esse la posta elettronica) con all'ultimo posto la capacità di visualizzazione a colori che é evidentemente ritenuta oggi eccessivamente costosa rispetto ai vantaggi forniti.

Sul secondo tema la fig. 15 è una incontrovertibile testimonianza del fatto che, negli Stati Uniti, ci si sta apprestando a fare dell'automazione di ufficio una infrastruttura di supporto al lavoro quotidiano dei quadri direttivi e professionali.

È ben vero che in USA si può contare su una cultura informatica assai più diffusa che in Europa, grazie in particolare all'esplosivo sviluppo del mercato degli elaboratori personali (nel solo 1985 è previsto ne vengano consegnate circa 2 milioni di unità per un valore di circa 10.000 miliardi di lire), tuttavia è risaputo che fenomeni del genere si sono sempre riprodotti anche nei paesi

DOMANDA: CHI DEI SEGUENTI RUOLI UTILIZZERÀ PERSONALMENTE LA STAZIONE DI LAVORO?

RUOLO	PERCENTUALE
PRESIDENTE	15%
AMM. DELEGATO	35%
CAPI DIVISIONE	70%
CAPI INTERMEDI	89%
SPECIALISTI	96%
SEGRETARIE	100%
SUPPORTO ESECUTIVO	83%

FONTE: IDC

Fig. 15 - Impiego previsto della stazione di lavoro

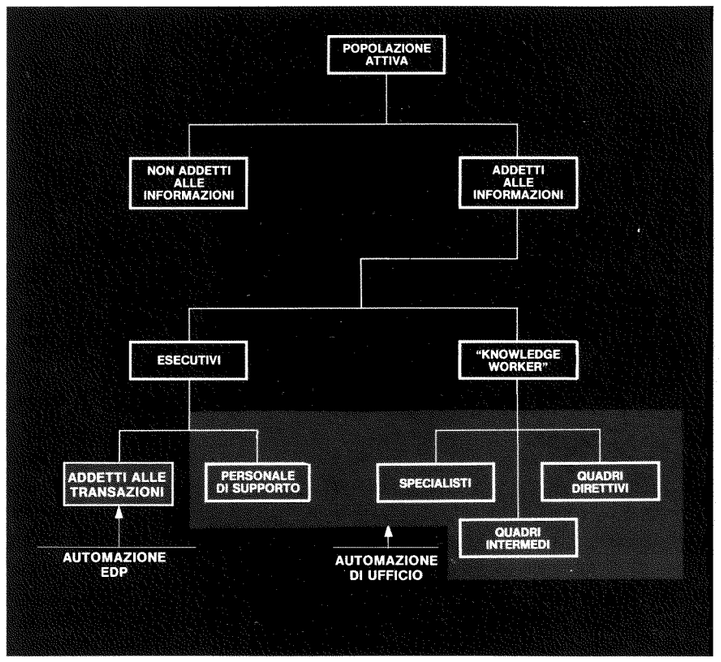


Fig. 16 - Ruoli, informatica e automazione di ufficio

europei con un ritardo variabile dai 2 ai 5 anni.

Alcune iniziative, purtroppo ancora episodiche, nel campo della riforma della scuola secondaria italiana sembrano andare in questa direzione.

12. Strategie di intervento

Chiarite le ragioni per cui si ritiene che il processo di razionalizzazione e automazione dell'ufficio sia destinato a porsi come scelta irrinunciabile per mantenere la competitività e analizzati alcuni degli aspetti di differenziazione e di continuità nei riguardi della informatica "classica„ ci si propone in questo e nel prossimo paragrafo di fornire alcuni elementi di pianificazione di un intervento di automazione ufficio.

Come già accennato nella introduzione, si terrà conto sia degli aspetti applicativi che di quelli economico-finanziari nella convinzione che questi ultimi debbano essere oggetto di particolare attenzione da parte dei quadri direttivi date le dimensioni dell'investimento richiesto a medio e lungo termine (a breve termine l'investimento può essere anche molto modesto se ci si limita a un esperimento di portata opportunamente ridotta).

Lo schema di fig. 16 riprendendo le

Fig. 17 - Profili per tipo di attività

ATTIVITÀ	SPECIALISTI %	QUADRI INTERMEDI %	QUADRI DIRETTIVI %
COMUNICAZIONI	38	54	64
CREAZIONE DOCUMENTI	15	11	8
ANALISI	9	7	4
LETTURA	4	8	6
ALTRE NON DIRETTAMENTE PRODUTTIVE	34	20	18

FONTE: BOOZ. ALLEN

Fig. 18 - Deviazione dalla stima nella distribuzione delle attività

ATTIVITÀ	% DEL TEMPO TOTALE		% DEVIAZIONE
	STIMATO	EFFETTIVO MEDIO	
COMUNICAZIONI	29.1	45.6	− 36
CREAZIONE DOCUMENTI	12.7	12.9	− 2
ANALISI	14.6	8.2	+ 78
LETTURA	9.1	7.9	+ 20
ALTRE NON DIRETTAMENTE PRODUTTIVE	34.5	25.4	+ 34

FONTE: BOOZ. ALLEN

considerazioni sviluppate ai punti 2 e 7, puntualizza i ruoli cui indirizzare il processo di automazione in senso lato.

Tra gli addetti alle informazioni, si distinguono gli impiegati esecutivi dai cosiddetti "knowledge worker„ vale a dire da quelle persone che nella azienda esercitano un ruolo che richiede un alto grado di professionalità (sia essa di tipo specialistico o manageriale).

Nell'ambito del lavoro esecutivo si possono ulteriormente distinguere gli "addetti alle transazioni„ vale a dire coloro che attuano determinate procedure gestionali del tipo di quelle già precedentemente richiamate, e il personale di supporto costituito essenzialmente da segretarie, dattilografe, archivisti, etc. che hanno il compito di rendere più "produttivo„ il lavoro degli "knowledge worker„.

Come è già stato più volte detto, il lavoro degli "addetti alle transazioni„ è quello su cui si è maggiormente insistito con l'automazione attraverso le applicazioni di informatica "classica„ mentre è sugli altri ruoli che si dovrebbe operare prioritariamente con la automazione di ufficio.

Sulla base di questa classificazione, la fig. 17 fornisce la ripartizione del tempo lavorativo (e quindi del costo) giornaliero degli specialisti e dei quadri direttivi a livello intermedio e a livello superiore.

Le cinque attività su cui è stata effettuata la ripartizione esauriscono logicamente, nel loro insieme, la to-

Fig. 19 - Esempio di recupero di efficienza

talità delle funzioni espletate dai tre ruoli indicati, come risulta chiaro dalla breve analisi che segue.

Nella attività di *comunicazioni* vanno comprese sia le riunioni (con gli eventuali tempi di spostamento) sia i colloqui telefonici.

Nella *creazione dei documenti* (o messaggi) sono incluse le annotazioni e ridistribuzione di documenti (o messaggi) ricevuti in visione.

La *analisi* di dati o di situazioni va intesa come finalizzata alla presa di decisioni mentre la *lettura* ha scopi più generali di arricchimento delle conoscenze professionali.

Infine, delle *attività non direttamente produttive* fanno parte la attesa di un evento, l'organizzazione del lavoro proprio o dei collaboratori, la tempificazione delle attività, la ricerca di informazioni o di persone (diretta o per telefono), la archiviazione, le trascrizioni, etc.

Tenuto conto della popolazione media nei tre ruoli, è interessante rilevare, come fa la fig. 18, quanto la valutazione oggettiva della ripartizione del tempo nelle cinque attività si discosti dalla stima che soggettivamente viene fatta dai rappresentanti degli stessi ruoli.

Si può, in particolare, constatare come la attività di comunicazioni venga notevolmente sottodimensionata dagli interessati e venga invece sovradimensionata quella di analisi e la voce relativa ad attività non direttamente produttive.

Lo scostamento nel tempo di comunicazione si giustifica con la inclinazione naturale a considerare il processo di scambio delle informazioni più efficiente di quanto in realtà non lo sia.

Per quanto riguarda la analisi, è evidente il tentativo di dare maggior peso a una attività che si ritiene qualificante della mansione, mentre per i compiti non produttivi è altrettanto immediato constatare che la loro sopravalutazione deriva dal giudizio negativo che su di essi, magari inconsciamente, viene formulato in quanto ritenuti di carattere operativo e non confacenti alla immagine del ruolo.

In ogni caso, tornando alla ripartizione effettiva, si rileva immediatamente come più del 70% del tempo lavorativo dei ruoli che stiamo esaminando vada ad attività di comunicazioni e a quelle non direttamente produttive.

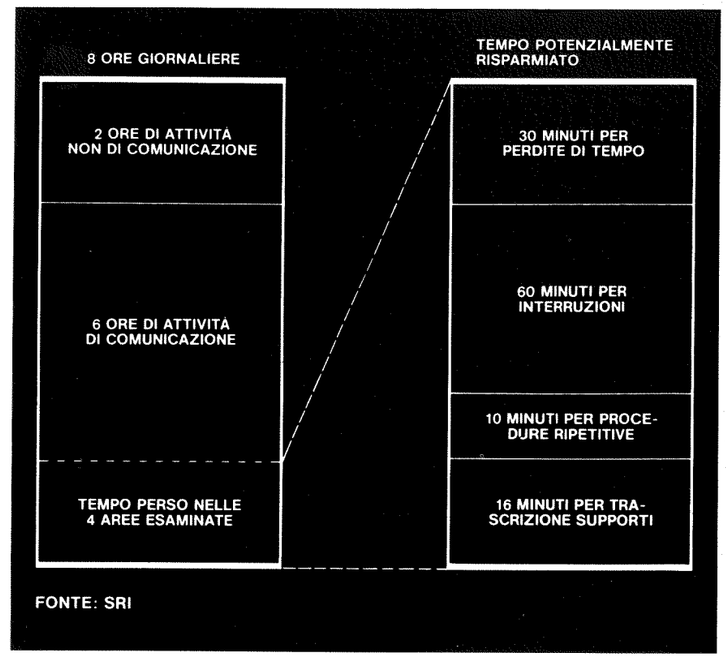
Questa costatazione è sufficiente per formulare una prima e semplice strategia di intervento che, rispetto ad altre, ha almeno il pregio della chiarezza di obiettivi.

Precisamente, iniziando dalle attività non direttamente produttive e analizzandone la struttura, ci si rende subito conto di come esse possano convenientemente venir in gran parte delegate al personale di supporto, o aumentandone l'organico o, assai preferibilmente dal punto di vista imprenditoriale, dotandolo di strumenti che gli consentano di operare con maggiore produttività.

Questo primo intervento, se attuato correttamente (cioè con l'opportuna gradualità e gli opportuni investimenti in formazione, oltre che in automazione) è potenzialmente in grado di apportare due significativi vantaggi sulla organizzazione del lavoro; il primo, consistente nell'arricchimento delle mansioni del personale di supporto e il secondo nell'alleggerimento del carico di lavoro specialistico e direttivo.

Quest'ultimo vantaggio si traduce in una maggiore quantità di tempo che, a parità di impegno, potrà essere dedicata ai compiti istituzionali del ruolo.

Fig. 20 - Posta elettronica risparmio nel lavoro direttivo



Dal punto di vista della motivazione in ambedue i casi ci si muove nella direzione di un miglioramento della professionalitá.

Per quanto poi riguarda la attività di comunicazione, la fig. 19 stabilisce un confronto tra la comunicazione telefonica e quella attuabile tramite la cosiddetta "posta elettronica„ che consiste nel trasmettere a uno o più destinatari un messaggio (o un documento) senza che si richieda la presenza contemporanea degli interlocutori ai terminali (il sistema di posta elettronica e in grado di memorizzare le informazioni e renderle disponibili al destinatario quando quest'ultimo ne faccia richiesta).

Come si può constatare, la posta elettronica presenta un insieme di vantaggi (essenzialmente riconducibili alla asincronia dello scambio dovuta alla capacità di archiviazione) che, per tutta una vasta gamma di comunicazioni di ufficio (di carattere operativo e tecnico) la rendono di gran lunga preferibile al telefono.

Una analisi approfondita condotta presso lo Stanford Research Institute [8] porta addirittura a prevedere che il solo impiego sistematico della posta elettronica nella attività direttiva potrebbe dar luogo a un risparmio di tempo dell'ordine del 20% (vedi fig. 20).

In ogni caso, attraverso la posta elettronica, si razionalizza il flusso delle comunicazioni che può così disporre di 3 "canali di trasporto„ diversi a secondo del contenuto delle comunicazioni stesse.

Gli argomenti di carattere "politico„ continueranno ad essere trattati in riunioni interpersonali; quelli che richiedono decisioni urgenti, attraverso il telefono, mentre tutti gli altri potranno avvalersi del nuovo canale con grande vantaggio nella pianificazione del lavoro (vedi ancora la fig. 20 dove le interruzioni e le perdite di tempo nelle comunicazioni tradizionali arrivano a pesare per circa 1 ora e mezzo sulle 8 lavorative).

La fig. 21 fornisce una stima cautelativa del fattore di miglioramento della "produttività„ che la adozione della strategia indicata può arrecare ai ruoli specialistici e direttivi.

Nella stessa figura sono anche messi in evidenza gli strumenti applicativi che tale strategia supportano tenendo conto dell'attuale stato dell'arte tecnologica, strumenti che in parte sono già stati discussi nelle considerazioni precedenti.

Precisamente il *trasferimento informazioni* riguarda specificamente l'uso della posta elettronica; la *memorizzazione e reperimento informazioni* la possibilità che lo stesso sistema ha di immagazzinare e rendere disponibili a chi ne è autorizzato le informazioni; la *gestione attività* rappresenta quanto è stato compreso nelle "attività non direttamente produttive„ e che è coperto dai servizi cosiddetti di "agenda elettronica„, di scadenziario, etc.

Con la *elaborazione personale* si intende la possibilità di dare all'utente (specialista, quadro intermedio o direttivo) una autonoma capacità di creare semplici sequenze di calcolo sui dati di suo specifico interesse, qualora, a suo giudizio, la ripetitività di tali sequenze lo giustifichi (è opportuno a questo proposito ricordare le considerazioni sui linguaggi "end user„ fatte al punto 10). Infine con il termine *"teleconferenza„* si vuole accennare a un ulteriore mezzo di comunicazione affiancabile con profitto, sul medio-lungo termine, all'incontro interpersonale, al telefono e alla posta elettronica.

Si tratta della possibilità di tenere

STRUMENTO	RISPARMIO TEMPO POTENZIALE (IN %)	% DEL RISPARMIO TOTALE DI TEMPO
TELECONFERENZA	0.5	2.7
TRASFERIMENTO INFORMAZIONI	4.4	29.9
MEMORIZZAZIONE E REPERIMENTO INFORMAZIONI	4.2	28.6
ELABORAZIONE PERSONALE	4.0	27.2
GESTIONE ATTIVITÀ	1.7	11.6
	14.8	100

FONTE: BOOZ. ALLEN

Fig. 21 - Strumenti di miglioramento della prestazione

riunioni tra due o più gruppi di persone geograficamente distanti utilizzando canali di comunicazione audio, o anche video (in questo caso con frequenze di due ordini di grandezza superiori a quelle foniche e quindi a costi oggi ancora proibitivi); possibilità sperimentata da anni e con successo nella NASA (l'Ente USA che si occupa del settore spaziale) ma che non è ancora diffusa nella realtà aziendale europea e che, anche al di là del costo, richiederà certamente un notevole tempo di adeguamento; di qui l'esiguo contributo che si ritiene oggi possa dare al risparmio di tempo potenziale.

Una osservazione conclusiva sullo strumento *gestione attività* costituito da un supporto automatico di gestione degli impegni, dello scadenziario, della pianificazione nell'impiego di risorse comuni, etc.

Il contributo che tale strumento dà al miglioramento del lavoro specialistico/direttivo è relativamente modesto perchè valutato nella ipotesi che la gran parte delle attività non direttamente produttive (vedi fig. 17) vengano, secondo la strategia suggerita, allocate al personale di supporto a sua volta dotato di quanto necessario per migliorare la produttività e quindi messo in grado di assorbirle.

13. Modelli di valutazione dell'investimento (*)

Determinati così gli obiettivi conseguibili con una opportuna strategia di intervento, resta aperto il problema di valutare la entità dell'investimento che tali obiettivi finanziariamente giustificano.

La fig. 22 visualizza una previsione corrente di crescita dell'investimento per addetto degli attuali 2.5÷3 K$ (vedi punto 6 precedente) ai 16 K$ del 1990 ma non specifica come tale crescita debba avvenire e a fronte di quali obiettivi.

È chiaro che ciascuna azienda sceglierà il criterio che meglio si confà al suo stile operativo e a quello della sua direzione, tenuto conto del mercato in cui opera: così una azienda innovativa che opera in un settore dinamico tenderà ad anticipare i vantaggi della automazione con un alto livello di investimenti iniziali, mentre una azienda che si può permettere di assumere un atteggiamento cauto tenderà ad assumere iniziative, e quindi ad investire, soltanto dopo che siano consolidate le esperienze.

(*) Per questa parte l'Autore desidera ringraziare il dr. Enrico Parazzini e la dott.ssa Daniela Nan che gli hanno assicurato la necessaria consulenza e che hanno sviluppato i modelli di cui qui si riportano i risultati.

In tutti i casi, è fuori di dubbio la necessità di disporre di un modello finanziario che sappia correlare i livelli di investimento con i miglioramenti conseguibili.

La metodologia da seguire è quella classica che giustifica l'investimento quando il valore attuale dell'incremento di produttività (ritenuto per definizione pari al tempo risparmiato) sia uguale al valore attuale del capitale investito.

L'incremento di "produttività„ può essere espresso in termini di risparmio sul costo del personale ma anche l'investimento per posto di lavoro può venire misurato in percentuale della stessa voce.

In base al rapporto si viene così a determinare quanta parte del costo annuo di personale (impiegato esecutivo, specialista, quadro intermedio o superiore) possa essere investita in automazione ufficio al manifestarsi di un dato incremento di produttività.

Se si opera nel settore industriale, tenendo conto del tasso d'incremento dei salari prevedibile sulla base 1981 e del costo del capitale alla data (o del ritorno di investimento atteso), si perviene, applicando il modello descritto, alla tabella di fig. 23 che stabilisce, in funzione del miglioramento di produttività,

Fig. 22 - Dinamica di crescita degli investimenti

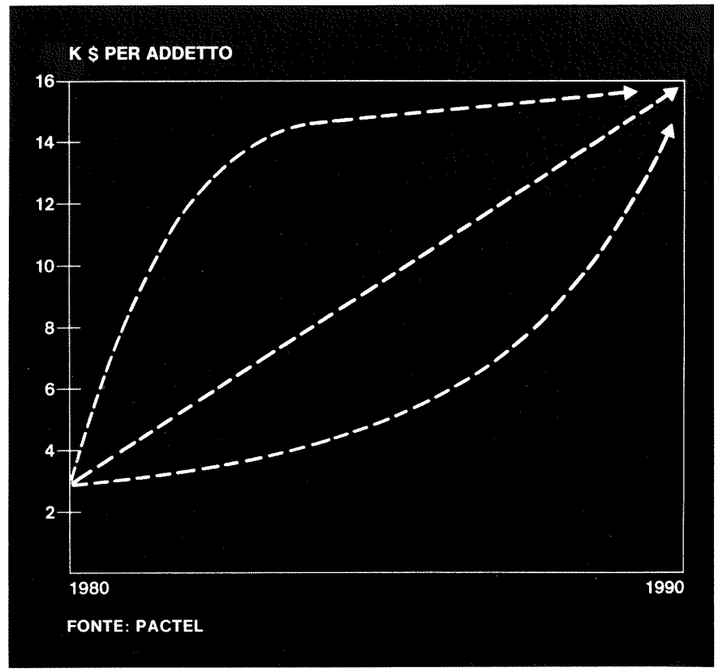


Fig. 23 - Investimento per posto di lavoro nelle aziende manifatturiere

POSTO DI LAVORO \ INVESTIMENTO IN KLIRE	Δ MIGLIORAMENTO 10%	15%	20%
5° LIVELLO	11.760	17.850	23.730
6° LIVELLO	14.560	22.100	29.380
7° LIVELLO	19.600	29.750	39.550
DIRIGENTE	34.720	52.700	70.000
INVESTIMENTO MEDIO PER POSTO DI LAVORO (IN KLIRE)	15.400	23.375	31.075

Fig. 24 - Investimento per posto di lavoro nelle aziende finanziarie

POSTO DI LAVORO \ INVESTIMENTO IN KLIRE	Δ MIGLIORAMENTO 10%	15%	20%
IMPIEGATO	16.700	24.900	33.400
FUNZIONARIO	31.500	47.000	63.000
DIRIGENTE	73.100	109.000	146.000
INVESTIMENTO MEDIO PER POSTO DI LAVORO (IN KLIRE)	18.900	28.200	37.800

Fig. 25 - Miglioramento della prestazione nel lavoro esecutivo

TIPO DI DATO	MIGLIORAMENTO
• PREVISIONALE	20%
• CONSUNTIVO SU ESPERIENZE A REGIME	25-30%

Fig. 26 - Miglioramento della prestazione per "knowledge worker"

ANNO	INVESTIMENTI (IN $)	MIGLIORAMENTO
1	2300	--
2	2500	10%
3	2500	18%

quanto sia ragionevole spendere in automazione del posto di lavoro dell'impiegato appartenente alle varie categorie del contratto metalmeccanici o del dirigente, in un conto economico che partendo dal 1981 si estenda sino al 1985.

Se si opera nel settore finanziario il modello rimane lo stesso ma, anzichè il costo del capitale, come indicatore di reddività si può assumere il tasso di rendimento sugli impieghi.

La tabella di fig. 24 dà i risultati cui tale applicazione conduce.

Come si può agevolmente constatare, anche per incrementi di produttività che, alla luce delle considerazioni del punto precedente, appaiono largamente conseguibili, le cifre che possono essere investite in automazione del posto di lavoro sono di rilevante entità e, in generale, superiori a quelle necessarie per l'acquisto di apparecchiature stante lo stato attuale della tecnologia.

Una volta di più si dimostra come l'investimento in informatica, se correttamente indirizzato e attuato, non tema confronti quanto a redditività.

14. Il risultato di alcune esperienze

Dopo aver valutato i miglioramenti conseguibili e gli investimenti che ne vengono giustificati, corre l'obbligo di verificare se, nelle esperienze sinora attuate nel campo della automazione ufficio, tali miglioramenti siano stati effettivamente raggiunti.

Ci si riferirà per questo ad esperienze essenzialmente statunitensi per le maggiori quantità di applicazioni cui si può attingere.

La fig. 25 fornisce un consuntivo riguardante il recupero di efficienza nelle attività del personale d'ordine che, ai nostri fini, è interessante non tanto di per sè quanto perchè dimostra la effettiva capacità di assorbire gran parte delle attività non direttamente produttive dei ruoli cui è di supporto, a organico sostanzialmente inalterato.

La fig. 26 riguarda invece in modo diretto i ruoli specialistici e direttivi (cioè gli "knowledge worker„ di fig. 16). Si può costatare come, anche con investimenti inferiori a quelli giustificabili secondo il modello esaminato precedentemente, gli incrementi di produttività conseguiti (come miglioramento della prestazione, fatta pari al tempo guadagnato) possano essere assai superiori a quelli ipotizzati.

Queste risultanze confermano la validità della strategia di automazione suggerita a quella del modello finanziario che ne sta a fondamento.

15. Conclusione

Al termine di questa analisi delle prospettive della automazione di ufficio o, meglio, della razionalizzazione e automazione delle attività che vi si svolgono facendo uso della tecnologia informatica, conviene ripercorrere il cammino compiuto per trarre alcune considerazioni conclusive.

Si è dimostrata l'indilazionabilità della opzione informatica nel campo delle attività di trattamento informazioni delle società industriali avanzate; in questo ambito generale si sono discusse differenze e aspetti di continuità della automazione ufficio con le applicazioni classiche della elaborazione dati; infine, sono state proposte delle strategie di automazione di cui ci si è preoccupati di fornire una ragionevole giustificazione finanziaria.

Non si è volutamente affrontato il tema della modellizzazione delle attività di ufficio, sia per il carattere introduttivo dell'articolo, sia perchè si ritiene oggi prematuro af-

Fig. 27 - Pianificazione di un intervento di OA

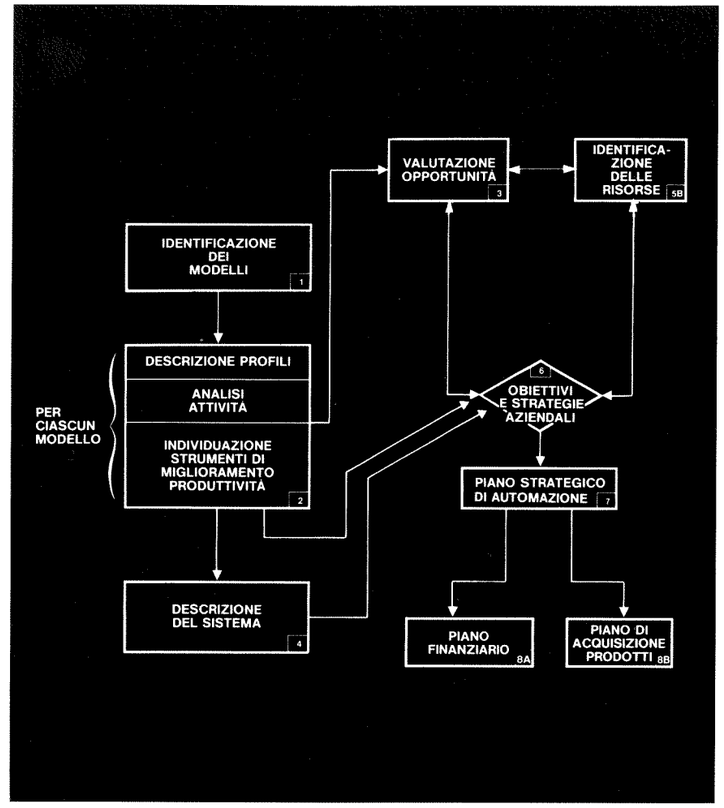


Fig. 28 - Responsabilità di attuazione

LOCALIZZAZIONE RESPONSABILITÀ	%
MIS	50
COMITATO DIRETTIVO	20
AMMINISTRAZIONE	30

FONTE: CW

Fig. 29 - Stato della utenza USA nelle applicazioni OA

FASE DI AVANZAMENTO DELLA AUTOMAZIONE UFFICIO	%
GIÀ AVVIATA	72
IN AVVIAMENTO	5
IN VALUTAZIONE	20
NESSUN PROGRAMMA	3

FONTE: CW

frontare il problema in questi termini mancando ancora una visione organica dell'argomento; la strategia di intervento suggerita prescinde infatti da questo aspetto introducendo come semplice punto di avvio una analisi dei ruoli destinatari della automazione.

Su queste basi il processo di pianificazione di un intervento nell'ufficio può essere schematizzato nelle 8 fasi della figura 27.

Il modello (o i modelli) di partenza è (sono) costituito(i) dall'universo degli utenti potenziali del sistema di automazione ufficio; di tali utenti vengono definiti i profili di attività con l'aiuto di questionari ma non solo di essi, stanti le osservazioni del punto 12.

Vengono successivamente individuati gli strumenti potenzialmente in grado di migliorare la prestazione nel senso più volte discusso e inquadrati (punto di fondamentale importanza) in una opportuna visione sistemica che deve in prospettiva conciliarsi con quella informatica.

Data la pervasità dell'intervento e l'impatto sulla organizzazione del lavoro il progetto deve far parte integrante degli obiettivi e delle strategie della azienda e, assai piu che nel caso della informatica classica, deve essere indirizzato e seguito nel suo sviluppo dalla direzione; ciò in particolare, significa che opportunità e risorse vanno conciliate nell'ambito della pianificazione aziendale di medio-lungo termine (3-5 anni di orizzonte temporale).

Il piano finanziario e quello di acquisizione prodotti discendono a questo punto dal piano strategico di automazione che la azienda ha elaborato.

Ma a chi deve, organizzativamente, venire assegnata la responsabilità di attuazione di un tale piano? La tabella di fig. 28 dà una risposta a questa domanda quale risulta da una inchiesta condotta a fine 1981 dal quindicinale Computer world su circa 400 aziende abbonate.

Come si può vedere, il 50% delle aziende assegna la responsabilità al settore dell'informatica aziendale giudicando che la competenza specifica sia un prerequisito essenziale anche per l'automazione ufficio; a favore di questa tesi stanno le necessità di integrazione tra i due mondi già discusse al punto 11.

Il 20% costituisce un apposito comitato di direzione in cui sono rappresentati, oltre al settore informatico, tutti i settori interessati al progetto con lo scopo di facilitare il pieno coinvolgimento della azienda nel progetto stesso.

In quest'ambito è lecito pensare, anche se non specificato dall'inchiesta, che il compito propositivo e quello più specificatamente operativo siano di pertinenza del settore informatico.

Un 30% infine giudica il problema dell'ufficio un problema di costi e quindi di pertinenza della amministrazione; va però ricordato che, anche negli USA, spesso il settore informatico è inquadrato nella Direzione Amministrativa.

Sempre con riferimento alla stessa inchiesta é anche interessante rilevare la diffusione delle applicazioni di automazione ufficio nel campione di aziende scelto.

La tabella di fig. 29 mostra, a questo proposito, come solo una esigua minoranza (3%) non abbia ancora, a fine 1981, preso in esame il problema; pressoché la totalità delle aziende o ha già messo in atto dei progetti, o lo sta facendo, o quanto meno, sta valutandone l'opportunità.

Torna qui utile ripetere l'osservazione già fatta al punto 11 sulla maggior cultura informatica su cui si può contare negli USA; tuttavia, proprio per le ragioni che sono state illustrate all'inizio di questo articolo, é sul terreno del trattamento delle informazioni che verrà misurata nel prossimo futuro la competitività del sistema economico e quindi é su questo terreno che l'Europa dovrà gareggiare con gli Stati Uniti (e, non dimentichiamolo, con il Giappone) pena il retrocedere a posizioni subalterne.

Le decisioni vanno quindi prese con la massima tempestività per recuperare il ritardo e mettersi quanto prima in condizioni di trarre il massimo vantaggio dalle esperienze già maturate.

16. Bibliografia

[1] P. SYLOS LABINI, *Saggio sulle classi sociali,* Laterza, Bari, 1978.

[2] S. NORA, A. MINC, *L'informatisation de la societé,* Annexe tome 1: Nouvelle informatique et nouvelle croissance, La Documentation Française, Paris, 1978.

[3] G. DORGET, *Le coût du système d'information de l'enterprise,* Istitut d'Informatique IBM, Paris, 1974.

[4] B. SPATARO, *L'elaborazione elettronica nell'industria manifatturiera italiana,* Quaderni di Informatica, 2, 1978.

[5] G. OCCHINI, *Redditività economica e aspetti organizzativi nelle applicazioni informatiche,* Quaderni di Informatica, 2, 1978.

[6] H. ZIMMERMANN, *OSI Reference Model - The ISO Model of architecture for open systems interconnection,* IEE Trans. on Commun., April 1980.

[7] *Issues in implementing office automation,* International Data Corporation, Framingham, Mass., 1981.

[8] J.M. BAIR, *Communication in office of the future: where the real payoff may be,* Stanford Research Institute International, 1978.

Il riconoscimento automatico della voce

TOM EDMAN

HONEYWELL
Technology Strategy Center
Minneapolis, USA

1. Introduzione

Agli inizi degli anni '70, nell'ambiente della ricerca si diffuse la convinzione che fosse ormai maturo il momento per risolvere il problema del riconoscimento automatico della voce.

In accordo a tale convinzione, la Defense Advanced Research Projects Agency (DARPA) distribuì per cinque anni dei fondi mirando a realizzare un salto in avanti nelle tecniche di riconoscimento del parlato. Quindici milioni di dollari furono dati a otto laboratori di ricerca, sia industriali che universitari, e alla fine dei cinque anni, si disse che gli obiettivi finali erano stati raggiunti (confr. tabella di fig. 1).

Tuttavia, a giudizio anche degli enti partecipanti, la conclusione generale fu che un sistema flessibile, affidabile e commercialmente valido di riconoscimento della voce era ancora lontano da venire.

La sofisticazione dei dispositivi di riconoscimento del parlato attualmente disponibili è modesta. Essi non sono realmente capaci di "capire il parlato", ma solo di riconoscere un numero limitato di parole isolate. Tuttavia, a parte l'impiego in giocattoli o giochi, essi cominciano a trovare applicazione in prodotti industriali, militari e di automazione dell'ufficio. Tutti gli studi di mercato dicono che l'uso della voce per comunicare con le macchine crescerà rapidamente e che l'Input/Output vocale sarà una cosa normale negli anni '90.

La Honeywell è impegnata da parecchi anni in questo settore della tecnologia. I gruppi di ricerca e sviluppo più attivi sono il "Technology Strategy Center", il "Systems and Research Center," il "Corporate Technology Center", la "Action Communication Systems," la "Residential Division", mentre un'altra dozzina di centri o divisioni ha interessi nell'argomento. In effetti, la Honeywell ha già diversi prodotti che usano la voce. Per esempio, nel "Roadrunner Network Management System" esiste un sistema di riconoscimento del parlato, per per-

Fig. 1 - Questa tabella confronta gli obiettivi posti dalla Defence Advanced Research Projects Agency (DARPA) con i risultati ottenuti da HARPY, un dispositivo di riconoscimento della voce progettato alla Carnegie-Mellon University. L'apparente successo del progetto DARPA è in effetti illusorio, e non si è comunque tradotto in un successo commerciale.
NOTA: "Branching factor" è un indice di complessità sintattica e si riferisce al numero di parole che, in media, seguono una data parola in una frase.

OBIETTIVI DARPA (NOV. 1971)	RISULTATI HARPY (NOV. 1976)
• Accettare parole non isolate	• Si
• provenienti da molti	• 5 (3 maschi, 2 femmine)
• parlatori interagenti	• Si
• in una stanza tranquilla	• Stanza terminali
• usando un buon microfono	• Microfono a basso rumore
• con modesto adattamento sul singolo utente	• 20 frasi per l'addestramento del sistema sul singolo utente
• accettando 1.000 parole	• 1.011
• usando una sintassi artificiale	• Branching factor medio = 33
• in un campo limitato	• Ricerca documentaria
• con errore semantico < 10%	• 5%
• in poche volte il tempo reale	• 80 volte il tempo reale
• su una macchina da 100 MIPS	• su una macchina da .4 MIPS, usando 256K di memoria, al costo di $ 5 per frase elaborata

mettere l'uso di carte di credito nelle telefonate a lunga distanza, il "Datanetwork" offre risposte a voce, e la Residential Division sta realizzando un centro di controllo per edifici che include messaggi e allarmi vocali.

Anche se in questa introduzione si è fatto cenno al problema generale della comunicazione vocale, ossia alla sintesi della voce e al suo riconoscimento, il presente articolo tratta specificatamente il tema del riconoscimento, il meno sviluppato, il più difficile e, sotto diversi aspetti, il più interessante dal punto di vista concettuale e tecnologico.

2. Il problema

Che cosa rende il riconoscimento automatico del parlato così difficile?

Alcuni problemi sono evidenti. La voce varia con l'età, il sesso e l'origine geografica del parlatore. A questi fattori che generano differenze tra diverse persone occorre aggiungere i fattori che introducono delle variabilità nella voce di una stessa persona. Stress, fatica, salute ed emozioni influenzano la voce di un individuo. Infine, le parole sono moltissime, e molte di esse hanno un suono simile. Una persona colta può avere un vocabolario di circa 100.000 parole conosciute, ed un vocabolario parlato di 3.000 - 8.000 parole.

Oltre a ciò, il numero di frasi differenti che possono essere formate con tale vocabolario è, ovviamente, enorme.

Il termine "varianza" è normalmente usato per descrivere alcune di queste intuitive caratteristiche della voce. Che il parlato sia altamente variabile significa che esistono poche relazioni fisse tra le unità del messaggio (parole, morfemi, fonemi, sillabe) e le proprietà del segnale acustico. Risulta in effetti molto difficile identificare caratteristiche invarianti del parlato che consentano di riconoscere automaticamente le parole.

Il segnale fonico ha origine nel tratto vocale, un buon modello del quale è un tubo chiuso ad un estremo da uno stantuffo. (fig. 2). L'energia è introdotta in questo tubo sotto forma di impulsi periodici, che corrispondono a regolari movimenti vibratori delle corde vocali, o come turbolenza dell'aria, che viene a crearsi quando questa è forzata a passare attraverso strettoie formate da palato, lingua o denti.

Più importante della sorgente del suono è, tuttavia, il modo con cui esso viene modificato dalla configurazione del tratto vocale al momento della comunicazione. Il tratto vocale agisce come un filtro e la sua configurazione (ossia posizione e forma della lingua, delle labbra e della mascella) ad ogni dato momento determina quali frequenze vengono accentuate e quali diminuite e pertanto quali sono le frequenze dominanti della forma d'onda.

La rappresentazione nel dominio del tempo della forma d'onda di pressione rivela poche strutture significative. La rappresentazione spettrale, ossia nel dominio delle

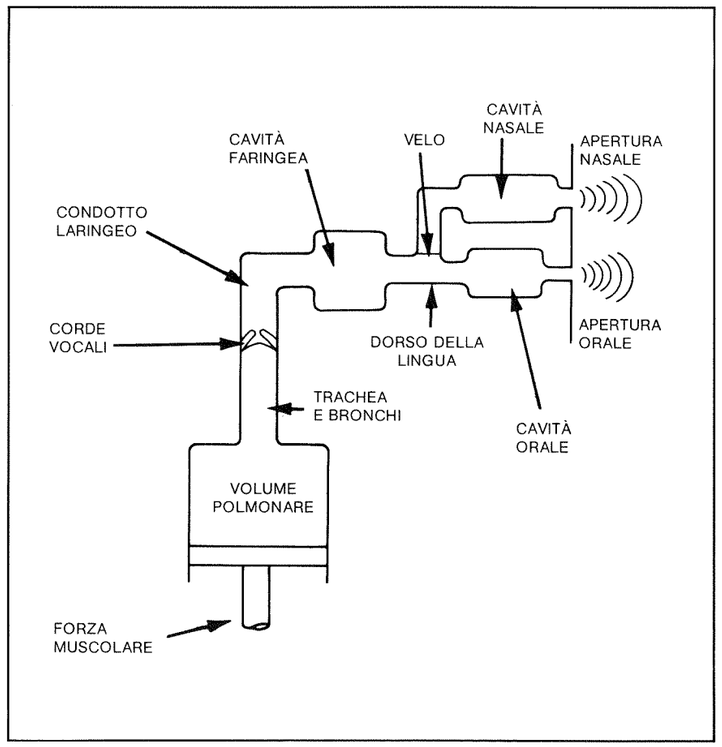
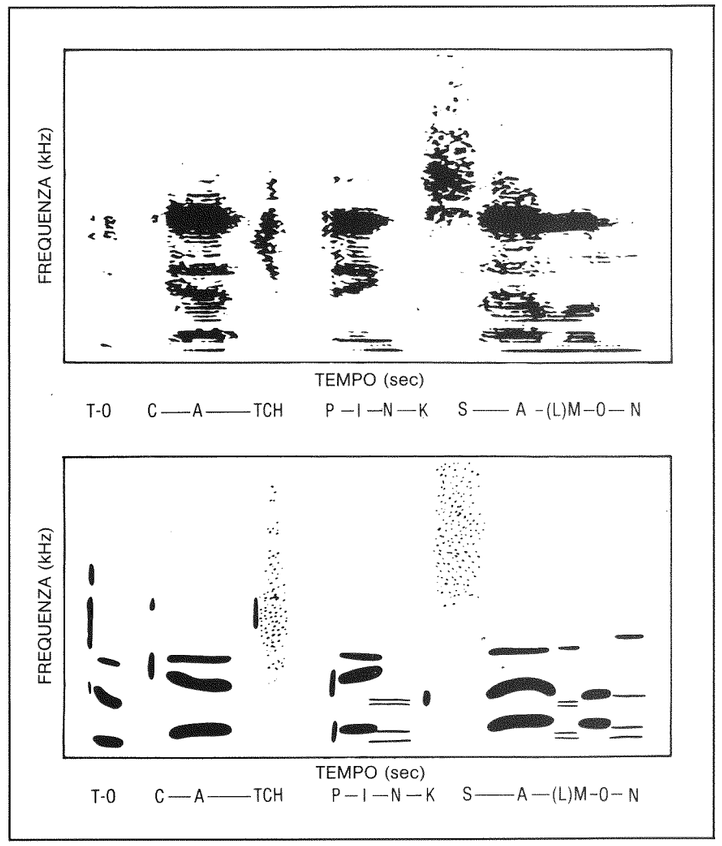


Fig. 2 - Modello del sistema vocale umano. Esso consente di studiare e di riprodurre il meccanismo di formazione della voce.

Fig. 3 - Spettrogrammi di una frase. L'asse x corrisponde al tempo, l'asse y alla frequenza e l'annerimento del tratto all'energia relativa.
La figura in alto rappresenta lo spettrogramma reale, quella in basso è una schematizzazione del medesimo, che ne mette in evidenza le parti più significative.



frequenze, fornisce invece maggiori informazioni ed è pertanto preferita nell'analisi del parlato. Lo spettrogramma del parlato (fig. 3) consente la rappresentazione di tre parametri: tempo, sulle ascisse, frequenza, sulle ordinate, ed energia relativa come scala di grigi.

La prima caratteristica di uno spettrogramma che colpisce è la presenza di regioni o bande di frequenza di concentrazione dell'energia, che si muovono secondo pattern apparentemente intricati. Queste aree di concentrazione dell'energia, che corrispondono alle risonanze naturali di una particolare configurazione del tratto vocale, sono chiamate *formanti*; la concentrazione di energia in corrispondenza della frequenza più bassa è chiamata prima formante (F1), la seconda è indicata come seconda formante (F2), e così via. La voce umana presenta formanti fino a F5 o F6, ma poche informazioni intelligibili sono presenti oltre F3. Infatti, una voce sintetica contenente due formanti è del tutto comprensibile, anche se suona in modo innaturale.

Le informazioni che sono richieste per riconoscere il parlato risiedono nei movimenti delle formanti, sebbene le relazioni tra questi movimenti e le unità di parlato non siano nè semplici nè ovvie. Anche il più esperto ricercatore del settore può vantare solo una modesta capacità nella lettura di uno spettrogramma, ossia nel dire quale frase è rappresentata da un dato spettrogramma. Uno studioso del MIT, dopo anni di pratica, ha sviluppato una eccellente abilità nella lettura di spettrogrammi e afferma di poterla insegnare. La sua abilità è fuori dubbio, ma sembra molto difficile che possa essere facilmente trasferita ad altri.

Come regola generale, le consonanti sono accompagnate da movimenti rapidi in F2, mentre le vocali sono associate con porzioni statiche dello spettrogramma. È tuttavia impossibile dire qualcosa di più preciso su queste relazioni. Consideriamo il semplice caso delle vocali. Esse non si spostano nello spazio delle frequenze e possono perciò essere caratterizzate dalle frequenze centrali di F1 e F2. Tuttavia, se si tracciano le frequenze centrali di vocali pronunciate più volte da una varietà di parlatori, il grafico rivela una spiacevole dispersione (fig. 4). Vocali differenti presentano aree in comune; in altri termini, una data vocale non è identificata univocamente da singoli punti nello spazio F1/F2.

Le consonanti non sono da meglio, anzi sono complessivamente peggiori. Anche se esse sono caratterizzate da movimenti della seconda formante, la direzione e la velocità dello spostamento sono determinate non dalla consonante, ma dalle vocali che la precedono e la seguono. Prendiamo, per esempio, la pro-

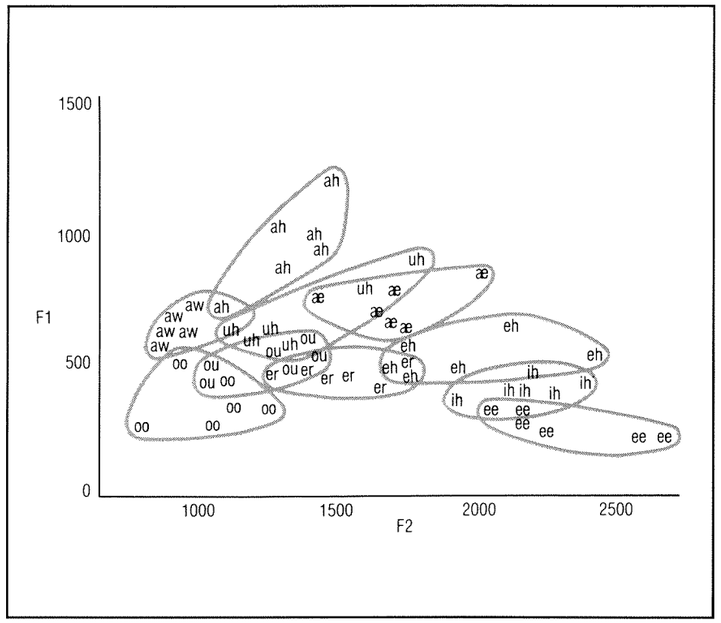


Fig. 4 - Grafico rappresentativo di vocali pronunciate da una varietà di persone.
Come si vede, c'è una larga dispersione delle frequenze centrali delle formanti F1 e F2 che identificano ciascuna vocale. È questa la ragione del fallimento di molti tentativi di sviluppare sistemi di riconoscimento della voce indipendenti dal parlatore.

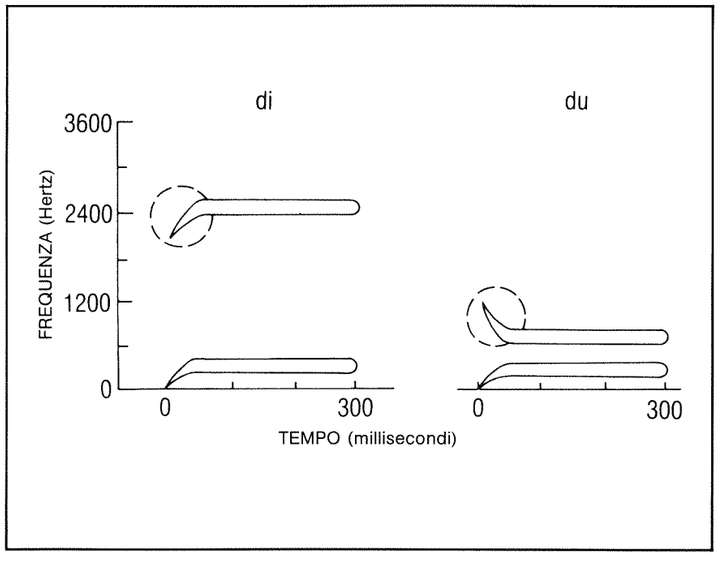
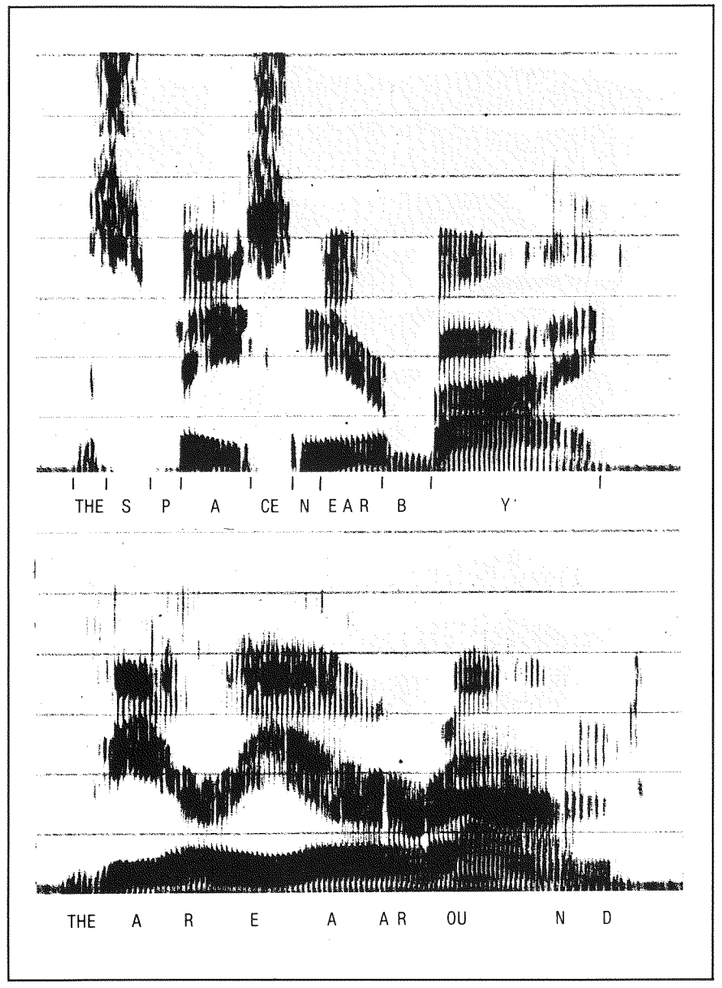


Fig. 5 - Questi spettrogrammi sintetici illustrano il fenomeno della coarticolazione, che rende difficile il riconoscimento.
Come si vede, la lettera "d" ha componenti diverse a seconda della vocale che la segue.

nuncia della lettera "d". Come è illustrato in fig. 5, la porzione della seconda formante associata con "d" è competamente diversa a seconda che la "d" è seguita da una "u" o da una "i".
Questo tipo di problemi derivano dal fatto che il tratto vocale è in movimento continuo mentre si parla e perciò la posizione della mascella, dei denti e delle labbra sono influenzate dalla loro posizione precedente e da quella seguente. Questo fenomeno è chiamato *coarticolazione* ed è una delle caratteristiche del parlato che più contribuisce alla sua varianza. Per esempio, quando la "t" di "cat" si combina con la "y" di "your", ciò da alla frase "you gave the cat your dinner" un suono simile a "you gave the catcher dinner". In questo contesto "you" suona molto diverso dallo "you" pronunciato isolatamente.
Questo esempio indica pure che spesso non vi sono confini tra le varie parole pronunciate, ma che una parola sconfina nell'altra. Come mostrato in fig. 6, mentre vi possono essere separazioni tra le parole in frasi dove si alternano fonemi dissimili ("the space nearby"), tale separazione può mancare quando vi è

Fig. 6 - Una delle ragioni per cui esistono pochi sistemi che riconoscono il parlato connesso è la difficoltà di segmentare una frase nelle parole che la costituiscono.
Come queste figure illustrano, i suoni sono più separati in frasi in cui si alternano vocali e consonanti ("the space nearby"), che in quelle in cui vi sono sequenze di vocali ("the area around").



una sequenza di vocali ("the area around"). In definitiva, può essere veramente difficile definire i confini delle parole.

Questa discussione su varianza e coarticolazione non esaurisce il problema del riconoscimento del parlato, ma da un'idea della sua difficoltà e del perchè non vi furono soddisfacenti risposte alla gara indetta dalla DARPA. Tuttavia, il programma DARPA e le conseguenti ricerche in campo industriale ed accademico hanno fatto imparare molte cose, la più importante delle quali è che *il discorso non può essere riconosciuto semplicemente analizzando il segnale relativo.* È spesso necessario ricorrere alla *sintassi*, ossia alle regole che definiscono l'ordine grammaticalmente ammissibile delle parole, e alla *semantica,* un termine che include sia il significato delle parole che la conoscenza dell'argomento del discorso, per essere in grado di distinguere parole che le varianze e le coarticolazioni del parlatore hanno rese confuse.

Un'altra lezione imparata dalle ricerche in questione è che alcuni dei problemi possono essere evitati ponendo accettabili limitazioni alla capacità dei dispositivi di riconoscimento. In effetti, alcuni dispositivi attualmente sul mercato, che si basano soltanto sull'analisi del segnale, hanno raggiunto prestazioni accettabili e stanno cominciando ad avere parecchie applicazioni.

3. Come risolvere il problema

Le riviste specializzate sono piene di soluzioni enormemente sofisticate, complesse e intelligenti del problema del riconoscimento del parlato. I sistemi commerciali, invece, non sono nè sofisticati nè complessi. Essi funzionano perchè impongono all'utente certe restrizioni che eliminano alcuni degli aspetti più onerosi del problema. I sistemi commerciali non tentano di aggiustare automaticamente le differenze tra i parlatori; ogni utente deve addestrare il sistema a riconoscere la sua voce, pronunciando tutte le parole del vocabolario fissato. È difficile realizzare un sistema che possa estrarre le singole parole da un discorso; perciò i sistemi

FABBRICANTE	TIPO DI DISPOSITIVO	DIMENSIONI DEL VOCABOLARIO	RIPETIZIONI DI UNA PAROLA PER L'ADDESTRAMENTO	MINIMA PAUSA TRA PAROLE (m sec)	ACCURATEZZA DEL RICONOSCIMENTO (%)	PREZZO DELL'UNITÀ ($)
• INTERSTATE ELECTRONICS	D,I	100	6-10	160	99	2.225
• THRESHOLD TECHNOLOGY	D,I	60	1-4	25	99	13.000
• AURICLE	D,I	80	3	300	99	2.500
• VOTAN	D,I	160	1	200	99	4.500
• NEC	D,C	120	1	0	99	50.000

Fig. 7 - Questa tabella riporta le caratteristiche di dispositivi commerciali per il riconoscimento della voce, fornite dai fabbricanti. Un'altra valutazione è data nella tabella di fig. 11.
Tutti i sistemi sono dipendenti dal parlatore (D) e riconoscono solo parole isolate (I), salvo l'ultimo che riconosce parole connesse (C).

commerciali lasciano all'utente il compito di segmentare la frase in parole, e gli richiedono di far precedere e seguire ogni parola da una adeguata pausa di silenzio. Infine, i sistemi commerciali rinunciano a grandi vocabolari; la maggior parte hanno vocabolari di non più di 100 parole e il funzionamento ottimale si ha con un quinto di tale dimensione.

In conclusione, la maggioranza dei sistemi commerciali sono *dipendenti dal parlatore,* richiedono *parole isolate* ed hanno *vocabolari modesti.* Attualmente, una mezza dozzina di aziende producono dispositivi del tipo indicato (vedi tabella di fig. 7). La maggior parte dei sistemi oggi disponibili sono sul mercato da parecchi anni ed hanno registrato una scarsa evoluzione in tale periodo. Tutti questi sistemi funzionano sul principio del confronto; ossia, il riconoscimento avviene trovando il miglior accoppiamento tra il pattern acustico della parola in esame ed un membro di un gruppo di parole di riferimento ("sagome"), il cui spettro è registrato nella memoria della macchina.

Sistemi di questo tipo ignorano evidentemente di che cosa si parla. Infatti essi funzionano ugualmente bene con ogni tipo di segnale acustico di breve durata, purchè esso abbia una larghezza di banda simile a quella del parlato. Essi possono, per esempio, essere installati nel cofano di un'automobile per ascoltare dei rumori sintomatici di problemi del motore o della trasmissione. Sistemi "ignoranti" del tipo descritto possono essere molto accurati, però sono inevitabilmente limitati. È praticamente impossibile realizzare il riconoscimento di un discorso vero e proprio senza introdurre qualche conoscenza della sintassi.

Un tipico sistema di riconoscimento, con dipendenza dal parlatore e per parole isolate, consiste di una base di dati contenente le sagome delle parole e tre blocchi funzionali: rivelatore di inizio/fine della parola, analizzatore dello spettro e confrontatore dei pattern (fig. 8).

4. Rivelazione di inizio/fine e analisi spettrale

La prima cosa che un sistema di riconoscimento deve fare è di catturare una parola, discriminando il discorso dal rumore di fondo e distinguendo colpi di tosse, sospiri ecc., dalle parole. La maggior parte dei sistemi sono progettati per essere usati con microfoni per distanza ravvicinata e a cancellazione di rumore, così che certi problemi di rumore sono minimizzati in partenza. Tuttavia, poichè il suono delle parole può variare largamente in ampiezza e durata, riconoscere l'inizio e la fine di una parola non è un compito banale. Un approccio normalmente seguito è di misurare diversi parametri in modo da aumentare la affidabilità del riconoscimento dell'inizio/fine della parola; ad esempio, si possono usare delle statistiche relative a ingressi a microfono aperto, combinazioni di livelli energetici di bande di frequenza chiave, livelli di potenza, tempi di salita e durata dei segnali.
Il secondo passo è l'analisi spettrale.

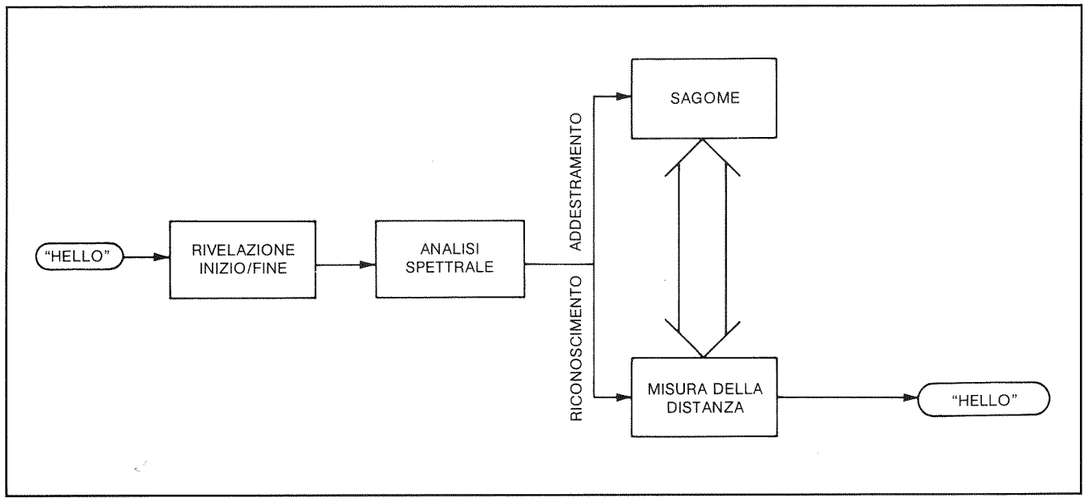


Fig. 8 - Schema a blocchi di un sistema commerciale di riconoscimento della voce. Durante una sessione di addestramento, l'utente ripete più volte ogni parola del vocabolario; gli spettri relativi sono registrati, in forma compressa, come "sagome" di riferimento. Durante il riconoscimento, il sistema analizza lo spettro della parola pronunciata e lo confronta con le "sagome". La parola viene identificata con la sagoma ad essa più vicina. Il sistema include un meccanismo per riconoscere l'inizio e la fine di una parola.

Nei sistemi commerciali sono stati realizzati diversi approcci al problema. Il sistema più semplice è quello di contare il numero di volte che la forma d'onda di tensione attraversa un livello di riferimento in un dato intervallo di tempo. Questo metodo, detto "zero-crossing" (ZCR), è chiaramente piuttosto rozzo. Se il segnale è una sinusoide, il sistema misura la frequenza, ma se il segnale è di forma complicata come il parlato, non si sa bene cosa si misura col sistema ZCR. Questo approccio rimane tuttavia interessante poichè funziona, almeno per piccoli vocabolari, senza richiedere la digitalizzazione del segnale e consentendo una analisi in tempo reale del segnale stesso.

Un sistema un po' più sofisticato è l'analisi dello spettro mediante un banco di filtri. Questo è l'approccio usato nei dispositivi commerciali attualmente più venduti. Un sistema tipico ha 16 filtri che coprono una banda da 100 a 5.000 Hz, anche se la distribuzione dei filtri lungo tale banda può essere abbastanza arbitraria. Alcuni banchi di filtri cercano di imitare l'orecchio umano, che presenta una risposta non lineare nel suo campo di funzionamento, mentre altri sistemi dividono semplicemente la banda passante totale in intervalli uguali. Come il metodo ZCR, i banchi di filtri forniscono di una parola un profilo spettrale a bassa risoluzione, i cui dettagli sono limitati dal numero dei filtri del banco come pure dalla selettività di ciascun filtro. Come lo ZCR, i banchi di filtri si sono dimostrati un metodo pratico per il riconoscimento della voce ed hanno il merito di fornire una valutazione dello spettro in tempo reale e a basso costo.
Tecniche di analisi spettrale con più alta risoluzione sono state usate in un paio di dispositivi di riconoscimento di tipo avanzato. La tecnica più comune è l'uso della "trasformata rapida di Fourier ("FFT = Fast Fourier Transform). Versioni della FFT adatte ad un efficiente impiego dell'elaboratore sono state sviluppate negli anni '60 e consistono in una generalizzazione della serie di Fourier che, come noto, riporta una qualsiasi forma d'onda periodica alla somma di una serie infinita di onde sinusoidali. Prima di cominciare l'elaborazione con la FFT occorre, però, digitalizzare e memorizzare la forma d'onda. Pochi fabbricanti hanno finora adottato la FFT; la ragione è essenzialmente economica, in quanto il costo combinato di memoria, potenza di elaborazione e convertitori analogico/digitali richiesti da luogo ad un prodotto non competitivo in termini commerciali. Oltre a ciò, si ritiene comunemente che il riconoscimento della voce debba essere fatto in tempo reale o in un tempo molto vicino ad esso.

Per contro, anche le più veloci FFT, finchè non siano implementate direttamente con hardware specializzato, sono probabilmente troppo lente per applicazioni commerciali.

Una ulteriore tecnica di analisi spettrale, che ha in comune alcuni problemi con la FFT, è la "codifica predittiva lineare" (LPC = Linear Predictive Coding). Questo algoritmo, molto in voga attualmente, è stato in origine sviluppato per la sintesi della voce. L'assunto che sta dietro il suo uso nel riconoscimento della voce è che le stesse informazioni richieste per la sintesi di una parola sono valide anche per il suo riconoscimento.

Il sistema LPC prende le mosse dal fatto che le strutture del tratto vocale sono relativamente inerti e si muovono lentamente. Ciò significa che una data configurazione del tratto vocale, e per conseguenza la corrispondente gamma d'onda di pressione, può essere predetta in base alla configurazione immediatamente precedente del tratto vocale stesso. La stima dell'ampiezza della forma d'onda ad un dato istante viene effettuata, pertanto, usando una combinazione lineare delle ampiezze della forma d'onda in alcuni istanti precedenti (circa una dozzina). Alcune note trasformazioni matematiche (Laplace, Z) permettono di portare la rappresentazione LPC del tratto vocale dal dominio del tempo a quello delle frequenze. Attualmente è disponibile in hardware la sintesi LPC della voce, ma altrettanto non esiste per l'analisi spettrale; di conseguenza, il riconoscimento col sistema LPC, come col FFT, comporta degli oneri computazionali che ne limitano l'impiego.

5. Formazione delle "sagome"

Lo stadio di elaborazione successivo all'analisi spettrale è il "pattern matching", cioè il riconoscimento della forma in esame mediante confronto con sagome di riferimento. I sistemi di riconoscimento di tipo commerciale richiedono la registrazione delle sagome durante una sessione di addestramento della macchina prima del suo uso. È in questo senso che i dispositivi attuali sono dipendenti dal parlatore; esiste un'alta probabilità di effettuare un corretto riconoscimento solo per quegli utenti che hanno addestrato il sistema e non per chiunque altro.

Il problema, nella formazione delle sagome, è di riuscire a rappresentare le variazioni normali della pronuncia di una parola pur mantenendo dei caratteri distintivi ben definiti delle sagome. Per consentire variazioni della pronuncia, si effettua la media di parecchi campioni della stessa parola, o si fanno altri trattamenti matematici di tali campioni. Alcuni sistemi di riconoscimento richiedono uno o due campioni da ciascun utente, ma altri insistono nel richiederne otto o dieci. Alcuni dispositivi sono abbastanza sofisticati da raccogliere statistiche sulla varianza di ciascuna parola, ma essi rappresentano l'eccezione piuttosto che la regola.

La forma in cui le sagome vengono immagazzinate dipende largamente dallo schema usato per l'analisi spettrale. In ogni caso, anche gli schemi più semplici richiedono che i dati forniti dall'analizzatore di spettro vengano semplificati prima dell'immagazzinamento. Un sistema tipico di riduzione dei dati è quello usato dalla Interstate Electronics (fig. 9). In tale sistema, l'uscita di ciascun filtro passabanda viene rettificata, amplificata logaritmicamente e campionata ogni 10 millisecondi. Una parola viene quindi rappresentata da una matrice di numeri con 16 colonne (1 colonna per filtro) e, per una durata media della parola (da 250 a 350 millisecondi), con un numero di righe da 25 a 35. Le entrate di questa matrice sono originariamente interi positivi ma sono subito ridotti drasticamente. Le entrate in ciascuna colonna sono infatti riscritte come 0 oppure 1, a seconda la loro relazione con un valore di riferimento colonnare; in tal modo, ad ogni "fetta" di 10 millisecondi corrisponde una parola di 16 bit.

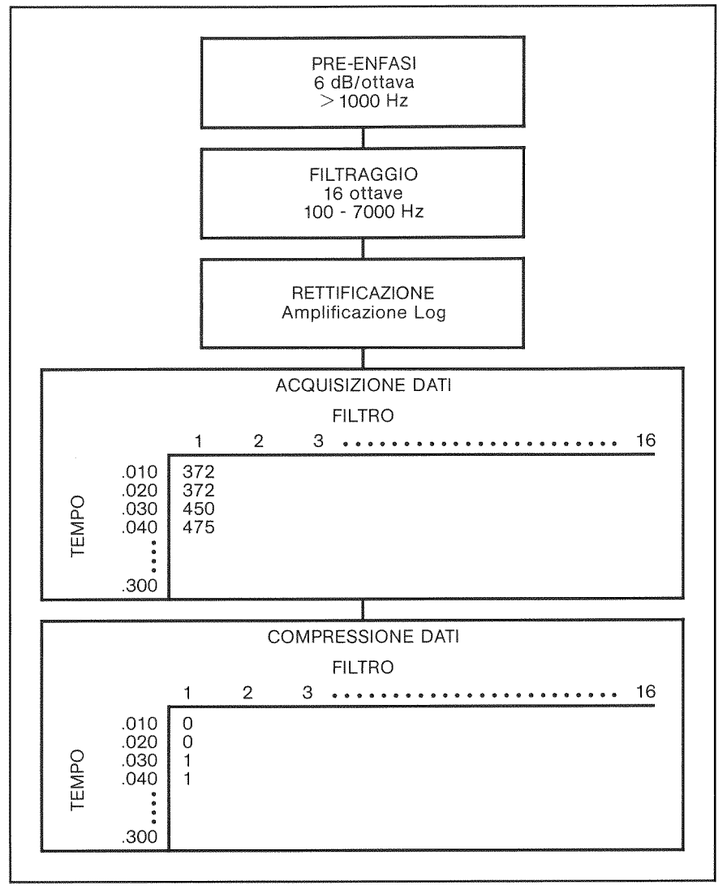
Schemi di riduzione dei dati come questo rendono la memorizzazione delle sagome economicamente ragionevole e fanno si che le dimensioni del vocabolario non siano limitate dalla capacità della memoria. La limitazione viene, invece, imposta dal processo finale di riconoscimento, cioè dal confronto della parola in esame con le sagome. Il problema risiede nel numero di possibili combinazioni e nel tempo di elaborazione. Quanto maggiore è il numero di parole accettate, tanto più numerosi sono i confronti da fare e quindi il tempo necessario. Oltre a ciò, all'aumentare del numero delle parole diventa più difficile avere parole con marcate differenze spettrali e quindi aumenta la possibilità di errori. Per tutto ciò, la maggior parte dei dispositivi pone forti limiti al vocabolario, sia per risparmiare tempo che per migliorare l'affidabilità.

6. Pattern matching

Il pattern matching consta di due operazioni separate: l'allineamento temporale e l'effettivo processo di confronto. Anzitutto, occorre normalizzare sia la parola in esame che le parole che funzionano da sagoma, in modo che abbiano lunghezza identica. Ciò è reso necessario dal fatto che la durata di una parola può variare sostanzialmente da una enunciazione all'altra. Per questa ragione, occorre realizzare un pattern matching che sia sensibile alla sequenza delle caratteristiche di una parola ma non alla loro durata.

I sistemi di normalizzazione temporale possono essere costosi o a buon mercato e operano conseguentemente in modo differente. I sistemi di basso costo hanno un approccio piuttosto grezzo al problema. Infatti, durante la formazione delle sagome e il pattern matching, ogni parola è forzata a restare entro una singola lunghezza standard. Una parola che è più corta dello standard viene allungata ripetendo quante volte necessario le sue fette temporali, una parola più lunga viene invece ridotta eliminando un certo numero di tali fette. I sistemi più costosi e sofisticati trattano invece le parole con maggior rispetto. La sagoma di ogni parola viene immagazzinata con la sua lunghezza naturale. Poi, durante il processo di

Fig. 9 - La maggior parte dei dispositivi commerciali impiega uno schema di compressione dei dati che consente di registrare le "sagome" delle parole in memorie di dimensioni ragionevoli.
La figura illustra lo schema di compressione della Interstate Electronics, descritto nel testo. In sostanza, la parola viene campionata ogni 10 millisecondi e ogni campionatura dà luogo ad una parola di 16 bit.



confronto, porzioni della sagoma vengono allungate o compresse affinchè ogni sagoma sia il più simile possibile alla parola in esame. Ciò è realizzato mediante una routine di software chiamata programmazione dinamica, e la sua applicazione rappresenta un significativo progresso nel riconoscimento della voce.

Le attuali routine di pattern matching sono, per il resto, simili nei sistemi sia di basso che di alto costo. Un modo di vedere il processo di riconoscimento è di immaginare ogni sagoma come un insieme discreto di punti in uno spazio multidimensionale astratto, i cui assi corrispondono a importanti caratteristiche del parlato. Una routine di pattern matching colloca ogni parola in esame dentro tale spazio e riconosce una parola identificando il suo più prossimo vicino. Identificare significa in questo caso calcolare la distanza tra la parola in esame e tutte le sagome e determinare la distanza minima. Vi sono molti modi per definire e valutare le distanze in uno spazio multidimensionale. Il modo più ovvio per definire la distanza è una generalizzazione del concetto euclideo, ma si usano anche altri metodi, come la misura della distanza spettrale. La scelta della metrica è considerata un argomento di originalità da parte dei costruttori, ma una attenta lettura della documentazione suggerisce che vi è in effetti un ampio uso di tecniche topologiche e statistiche standard.

Molti dei sistemi disponibili includono un algoritmo di rigetto che entra in gioco a questo punto, e ha lo scopo di respingere parole non comprese nel vocabolario. Come regola generale, se una parola non si accorda con una sagoma entro certe tolleranze, oppure se si accorda colla sagoma precedente e con quella successiva circa altrettanto bene, la parola viene rigettata. Sfortunatamente, alte frequenze di rigetto non si traducono direttamente in un basso tasso di errore. Vi è uno scotto da pagare. Tipicamente si producono da due a quattro rigetti per ogni errore che viene eliminato. Pertanto vengono spesso tollerate probabilità fino al 50% di riconoscere erroneamente parole spurie. I valori di accuratezza forniti dai fabbricanti tipicamente non conteggiano come errori i falsi riconoscimenti; si fa cioè l'assunzione che gli input illegali devono essere controllati dall'utente.

7. Il riconoscimento della voce nel prossimo futuro

I fabbricanti di dispositivi per il riconoscimento della voce hanno tratto vantaggio dai recenti pro-

		OGGI	FRA 2 ANNI	FRA 5 ANNI
HARDWARE	UNITÀ DI ELABORAZIONE	CPU 8 bit	CPU 16 bit	CPU 16-32 bit
	PRE-ELABORAZIONE ACUSTICA	Filtro passabanda Zero-crossing	Chips per analisi spettrale	Chips per analisi spettrale
SOFTWARE	ELABORAZIONE DEL SEGNALE	Fast Fourier Transform Linear Predictive Coding Dynamic Programming	Clustering Discriminant Analysis	?
	LINGUISTICA	Nessuna	Prosodia Fonologia	Sintassi
	ALGORITMI DI RICONOSCIMENTO	Euristici	Basati sulla conoscenza	Adattivi

Fig. 10 - Nei prossimi 5 anni si può prevedere l'applicazione al riconoscimento della voce di nuove tecnologie provenienti da diverse discipline scientifiche e tecniche. Potenti tecniche statistiche, quali clustering e discriminant analysis, saranno impiegate per risolvere il problema della variabilità del parlato. Studi linguistici sulle regole che governano i pattern sonori di un dato linguaggio (prosodia e fonologia) diventeranno importanti, così come lo studio delle sequenze ammissibili di parole (sintassi). Infine, mentre la maggior parte degli algoritmi odierni è scritta su base euristica, ossia partendo dall'esperienza e dall'intuizione del progettista, si ritiene che abbastanza presto verranno usate nel riconoscimento della voce le tecniche della Intelligenza Artificiale per rappresentare conoscenze complesse, come prosodia e sintassi, e per realizzare sistemi adattivi.

gressi della tecnologia dei circuiti integrati, ed esistono ormai dei riconoscitori elementari realizzati su un singolo chip. Ci sono stati invece modesti progressi per quanto concerne le tecniche di riconoscimento in se. I laboratori accademici sono rimasti piuttosto tranquilli. Dopo il progetto DARPA, l'attenzione si è spostata dal riconoscimento del parlato allo studio di come esso viene generato o a più ampi settori dell'intelligenza artificiale. Anche in campo commerciale, annunci di una certa rilevanza sono stati piuttosto rari. Qualche nuovo sistema (dipendente dal parlatore, per parole discrete) è stato messo sul mercato, mentre vecchi dispositivi sono stati ringiovaniti, riducendone il costo e le dimensioni. Ma tutti i dispositivi commerciali sono concettualmente sempre simili a quelli della prima generazione introdotta circa dieci anni fa.

La tecnologia del riconoscimento della voce non è più moribonda. Una intensa attività sta prendendo piede dietro la cortina di riservatezza dei grandi laboratori industriali e dovremmo presto avere manifestazioni tangibili di quanto sta avvenendo. Un segno di queste attività è fornito dalla migrazione di ricercatori dall'ambiente accademico a quello industriale. Alcune persone di alta qualificazione, tra le migliori nel mondo, non si trovano oggi più al MIT o Carnegie-Mellon ma piuttosto nelle aziende tecnologiche e nei gruppi di imprenditori scientifici della California.

Come conseguenza, si può predire che vedremo tra non molto significativi sviluppi della tecnologia di riconoscimento del parlato, un cambiamento non tanto nelle prestazioni dei dispositivi commerciali, quanto nell'approccio ai problemi del riconoscimento. Diventa infatti sempre più chiaro che il riconoscimento del parlato è un problema troppo difficile per risolverlo semplicemente mediante l'analisi del segnale, e che i dispositivi di riconoscimento debbono conoscere qualcosa sul linguaggio. Per esempio, essi devono sapere che la vocale iniziale della parola "one" è fortemente influenzata dalla successiva nasale "n", e che in inglese un insieme di tre consonanti all'inizio di una parola comincia con una "s", è seguito da "p" o "k" e finisce con una "r" o una "l".

I nuovi dispositivi verranno progettati da gruppi di lavoro interdisciplinari. Psicologi contribuiranno con informazioni sulla percezione umana della parola, ingegneri porteranno nel team le nuove tecniche digitali per l'elaborazione dei segnali, linguisti introdurranno analisi fonologiche e sintattiche, ricercatori di intelligenza artificiale contribuiranno con rappresentazioni adattive e flessibili dei dati (fig. 10). Tutto ciò può sembrare forzoso e può anche suonare come una ripetizione del progetto DARPA. Tuttavia, per il fatto che sta avvenendo nell'ambiente industriale anzichè in quello accademico, il fenomeno sarà pilotato dall'obiettivo di offrire un prodotto e di avere un ritorno dell'investimento, e fornirà perciò probabilmente risultati meno concettuali ma più pratici.

Nel frattempo, dovremmo vedere dispositivi che riconoscono un discorso connesso, anche se rimarranno dipendenti dal parlatore e potranno essere progettati ad hoc per specifiche applicazioni. Appariranno anche sistemi indipendenti dal parlatore; essi avranno un vocabolario ridotto, fissato dal produttore, ma aumenteranno drammaticamente l'interesse per il riconoscimento della voce. Infine, sembra inevitabile l'avvento di qualche tipo di dispositivi ibridi. I Bell Labs, per esempio, stanno sperimentando un riconoscitore di numeri connessi, indipendente dal parlatore, e un'altro laboratorio sta lavorando su un "segnalatore di parole", cioè su un dispositivo in grado di riconoscere solo alcune parole, ma di estrarle da una normale conversazione.

Infine, nel breve termine, continueremo a vedere rifacimenti di vecchi prodotti. I fabbricanti continueranno a ritoccare i loro dispositivi per avvicinare sempre più le prestazioni di field con quelle di laboratorio, per integrarli meglio nei sistemi utente e per ridurre il prezzo e l'ingombro. Diventerà però sempre più difficile collocare sul mercato oggetti di questo tipo.

8. Il riconoscimento della voce attualmente

Finora i dispositivi commerciali di riconoscimento della voce non hanno rimunerato finanziariamente i loro sviluppatori o produttori. Nessun fabbricante di questi dispositivi ha finora tratto profitti e la maggior parte di essi sono ancora nel settore solo perchè hanno alle spalle una grande e diversificata azienda multinazionale. All'inizio del 1982, infatti, la più vecchia azienda produttrice di riconoscitori della voce, la Heuristics Inc., ha dichiarato fallimento, mentre la casa madre dell'attuale leader del mercato ha dichiarato che il 1982 deve portare profitti, altrimenti...

È difficile dire quali sono le ragioni esatte per cui questi dispositivi non hanno mantenuto le loro promesse. Il prezzo relativamente alto e interfacce difficili hanno certo deluso gli utenti coraggiosi e scoraggiato quelli timidi. Ma la ragione più probabile è che non sono state trovate abbastanza applicazioni in cui la voce presenta persuasivi vantaggi sui sistemi competitivi. Come molti altri nuovi prodotti, i dispositivi per il riconoscimento della voce sono una "soluzione in cerca di un problema".

9. Un prodotto in cerca di applicazioni

L'interesse generale dei dispositivi per il riconoscimento automatico della voce è raramente messo in discussione. Ci sono delle applicazioni in cui la voce è solo un mezzo più comodo, altre in cui diventa il solo mezzo a disposizione.

Si ha questo secondo caso quando gli occhi e le mani non bastano; quando cioè si devono svolgere controlli e ricerche visive simultaneamente con manipolazioni di oggetti e comandi e, tuttavia, è richiesto di fare qualcos'altro. La voce può essere usata per fare "qualcos'altro". Il pilota di un areo da combattimento si trova in una situazione di questo tipo e perciò non deve sorprendere se il Department of Defense è la sorgente primaria di contratti per ricerche nel campo del riconoscimento della voce. Altri esempi di compiti in cui gli occhi e le mani sono già impegnati si trovano in molte attività di produzione, selezione e montaggio.

10. L'addestramento del sistema

Un'area di cui il potenziale utente di dispositivi di riconoscimento della voce può preoccuparsi è l'addestramento del sistema. La domanda più critica è "Come fare la sessione di addestramento per garantire la creazione di sagome realmente rappresentative". In base all'esperienza, occorre anzitutto che l'ambiente dove si effettua l'addestramento sia il più possibile vicino a quello reale di utilizzo. Questo può sembrare ovvio, ma, per esempio, durante i primi test di un dispositivo per la cabina di pilotaggio di aerei l'addestramento venne fatto in un ambiente tranquillo e controllato. La cabina di un aereo è invece un ambiente acusticamente molto sporco, e perciò i risultati di field furono insoddisfacenti. Era successo semplicemente che le parole di ingresso comprendevano l'elevato livello di rumore della cabina, che non era invece incluso nelle sagome realizzate altrove. Una volta effettuato l'addestramento nella cabina dell'areo, si ottennero subito ottimi risultati, malgrado la presenza di un livello di rumore superiore a 110 dB.

L'addestramento non riguarda solo il dispositivo di riconoscimento, ma anche l'utente. Questi deve famigliarizzarsi col dispositivo e superare la "soggezione del microfono" per ottenere le migliori prestazioni di cui è capace. Se durante l'addestramento egli parla in modo innaturale, le sagome non si adatteranno alle parole pronunciate quando egli parla in modo rilassato. In generale, è preferibile che l'utente parli in modo naturale quando lavora con dispositivi di riconoscimento piuttosto che sforzarsi di parlare chiaro o attentamente, semplicemente perchè un modo marcatamente artificiale di parlare altera i tempi. Infine, si possono usare alcuni trucchi per migliorare la qualità delle sagome. Per esempio, quando l'utente registra diverse sagome di ciascuna parola del vocabolario, conviene che ripeta le parole in ordine sparso, anzichè tutte in fila.

Se si prendono questi ed altri accorgimenti per garantire un eloquio naturale durante l'addestramento, anche un nuovo utente può ottenere

FABBRICANTE	MODELLO	INDIPENDENZA DAL PARLATORE	PAROLE CONNESSE	NUMERO DI SOSTITUZIONI	ERRORI PER VOCI MASCHILI	ERRORI PER VOCI FEMMINILI
• VERBEX	1800	SI*	SI*	10 (0,2%)	2	8
• NIPPON ELECTRIC	DP-100	NO	SI*	60 (1,2%)	1,4%	1,0%
• THRESHOLD TECHNOLOGY	T-500	NO	NO	73 (1,4%)	1,2%	1,7%
• INTERSTATE ELECTRONICS	VRM	NO	NO	147 (2,9%)	2,0%	3,7%
• HEURISTICS	7000	NO	NO	300 (5,9%)	4,4%	7,3%
• CENTIGRAM	MIKE 4725	NO	NO	366 (7,1%)	6,3%	8,0%
• SCOTT INSTRUMENT	VET/1	NO	NO	646 (12,6%)	11,2%	14,0%

Fig. 11 - La tabella mostra i risultati di una attività indipendente di valutazione di dispositivi commerciali per il riconoscimento della voce [5]. Lo studio è il più rigoroso tra quelli pubblicati; si riferisce però a dispositivi disponibili nel 1980 e 1981. Attualmente molti dei fabbricanti dichiarano prestazioni migliori.

NOTA. L'asterisco indica che non è stato effettuato il test relativo. Tutte le prove sono state fatte addestrando il sistema sull'utente e usando parole separate.

buoni risultati pressochè immediatamente. Tuttavia, mentre vari studi mostrano che un funzionamento soddisfacente può essere ottenuto con solo tre ore di pratica, per ottenere le migliori prestazioni può essere necessario un periodo di due o più settimane. Una ragione di questo graduale miglioramento è che ogni corretto riconoscimento costituisce un rinforzo, in quanto sia l'utente che il dispositivo tendono a condizionarsi sottilmente l'un l'altro.

Infine mentre la maggior parte degli utenti riescono ad ottenere buoni risultati, l'esperienza mostra che pochi di essi risultano incapaci di far riconoscere la propria voce, anche se sono ovviamente motivati a farlo. Ancora non è chiaro perchè ciò accada.

11. La velocità

Una delle prime domande che gli utenti potenziali fanno sui dispositivi di riconoscimento della voce (o meglio, la domanda che fanno subito dopo aver chiesto: "Cosa succede se sono raffreddato?") è: "Quanto veloce è l'ingresso vocale?". Generalmente con ciò essi intendono confrontare l'ingresso a voce con quello da tastiera.

Un abile dattilografo arriva a 60-80 parole al minuto. La velocità di una normale conversazione è di circa 150 parole al minuto; la voce sembra essere quindi decisamente più conveniente. Però i dispositivi di riconoscimento commerciali non riescono a seguire il ritmo di una normale conversazione connessa; occorre fornire loro parole separate. Facendo delle pause tra le parole, la velocità di ingresso si abbassa a 50-70 parole al minuto. Sulla base dei dati disponibili, risulta che le velocità di ingresso vocale sono solo leggermente maggiori di quelle che si hanno usando un qualche tipo di tastiera, per compiti semplici e di breve durata. Se il compito dura a lungo, la velocità di ingresso vocale scende a 25-35 parole al minuto, anche con un utente esperto.

D'altra parte, l'ingresso vocale comincia ad eclissare le altre forme di ingresso quando aumenta la complessità del compito o se si debbono svolgere più compiti contemporaneamente. Infatti, l'ingresso vocale aggiunge un'altro canale di elaborazione. Nei casi in cui la voce è usata per un compito e la tastiera per un'altro, la velocità di ingresso vocale rimane la stessa che per un solo

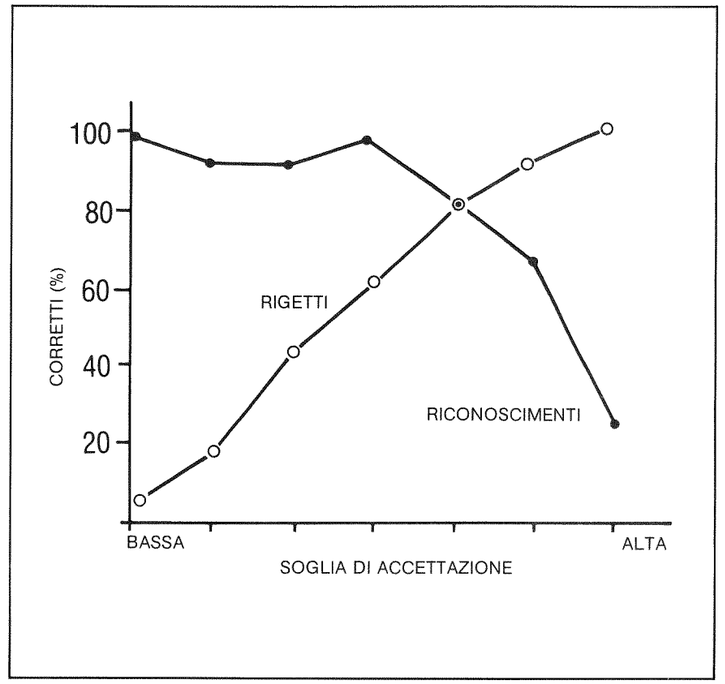


Fig. 12 - La figura mostra la relazione tra il tasso di riconoscimento e quello di rigetto, sulla base di prove effettuate in Honeywell su una unità commerciale. Un aspetto critico messo in luce da questo grafico è che il livello di accettazione deve essere posizionato relativamente in basso se si vogliono ottenere elevati tassi di riconoscimento. Normalmente, i fabbricanti trascurano il problema delle false accettazioni di ingressi illegali poichè assumono che l'utente abbia il completo controllo dell'ingresso. Ciò, tuttavia, può non verificarsi in certe applicazioni. Questo è un esempio di argomenti cui non si è forse data finora sufficiente attenzione e di cui il potenziale utente deve essere conscio.

compito vocale (25-35 parole al minuto), e così pure il tasso di errori, mentre sia la velocità che l'accuratezza degradano se i due differenti compiti sono eseguiti manualmente.

12. Misura delle prestazioni

Non siano in grado di dare una risposta soddisfacente a parecchie domande relative alla accuratezza di questi dispositivi. Esiste un solo studio pubblicato in cui si confrontano in modo rigoroso le prestazioni di alcuni di essi [5]. Se si confrontano i dati di targa pubblicizzati, si corre il rischio di prendere grossi abbagli a causa delle differenti condizioni in cui i dispositivi sono stati provati o delle differenti definizioni di errore usate dai fabbricanti.

Il modo standard di misurare le prestazioni di questi dispositivi è di calcolare la loro propensione a fare "errori di sostituzione". Tale propensione può essere formalmente definita come il rapporto tra il numero di falsi riconoscimenti e quello degli input legali (cioè delle parole la cui sagoma è registrata). Spesso, anzichè agli errori di sostituzione, ci si riferisce al suo complemento statistico, cioè l'accuratezza.

La tabella di fig. 11 fornisce una valutazione indipendente di sette dispositivi commerciali. Da questi dati, l'accuratezza risulterebbe veramente elevata; per il dispositivo medio si avrebbe un valore superiore al 97%. Le prove di field non hanno, però, in generale confermato questi dati che, in ogni caso, non caratterizzano completamente le prestazioni di un dispositivo per il riconoscimento della voce.

Da un lato, infatti, è sviante parlare solo degli errori di sostituzione. Un dispositivo per il riconoscimento della voce può fornire infatti cinque tipi di risposta:

per parole legali
- accettazione corretta
- accettazione errata
- rigetto errato

per parole illegali
- accettazione errata
- rigetto corretto

Di queste risposte, tre sono erronee. Di conseguenza, la valutazione di un dispositivo può variare notevolmente a seconda che vengano o no controllati tutti i tipi di errore.

Consideriamo, per esempio, la relazione tra falsi rigetti e accuratezza. Se al progettista del dispositivo è consentito di variare il tasso di falsi rigetti, si possono ottenere valori di accuratezza molto elevati; si possono infatti avere molti rigetti, ma pochi falsi riconoscimenti. Per fortuna, anche i fabbricanti di dispositivi ammettono che permettere un alto tasso di falsi rigetti è un modo scorretto di aumentare l'accuratezza e cercano di mantenere il tasso di rigetti tra il 2 e il 4%.

L'altro tipo di errore, la falsa accettazione, non è trattata in modo altrettanto scrupoloso. In parte, ciò sembra ragionevole. Le false accettazioni sono infatti definite come il tasso di parole riconosciute a fronte di ingressi illegali, e il sottomettere ingressi illegali sembrerebbe completamente sotto il controllo dello utente, ma se si consente al progettista del dispositivo un livello arbitrario di false accettazioni, ciò significa permettergli di ottenere fraudolentemente alti valori di accuratezza, esattamente come se gli si concedessero alti tassi di rigetto. Se il vocabolario del dispositivo consiste di due sole parole, "yes" e "no",

e l'utente dice "not", "now", "no" o "go", e tutte queste parole sono riconosciute come "no" ma le false accettazioni non vengono contate, il valore di accuratezza che ne risulta vi dice realmente poco sulla accuratezza con cui il dispositivo è capace di discriminare differenti parole.

La Honeywell ha raccolto dei dati sulle relazioni tra false accettazioni e accuratezza (fig. 12). La relazione tra rigetti corretti e riconoscimenti corretti è quella che ci si poteva aspettare. Alzando la soglia di decisione aumenta la percentuale di riconoscimenti corretti, al costo però di una diminuzione della percentuale di rigetti corretti. L'aspetto critico di questo grafico è la posizione del punto di incrocio. Un 80% di riconoscimenti corretti non è probabilmente accettabile per la maggior parte delle applicazioni, così come può essere inaccettabile la elevata percentuale di false accettazioni che si accompagna a più alti tassi di riconoscimenti corretti.

13. Bibliografia

[1] KLATT D.H., *Trends in Speech Recognition,* Prentice-Hall, 1980

[2] GLUCKSBERG S., DANKS J.H., *Experimental Psycholinguistics,* Erlbaum Ass., 1975

[3] SKINNER T.E., *Speaker Invariant Characterization of Vowel, Liquids and Glides Using Relative Formant Frequencies,* Proceed. of the 94 th Meeting of the Acoustic Society of America, Dec. 1977

[4] LEVINSON E., LIBERMAN M.Y., *Speech Recognition by Computer,* Scientific American, vol. 244, n.4, April 1981

[5] DODDINGTON G.R., SCHALK T.B., *Speech Recognition: turning Theory into Practice",* IEEE Spectrum, Sept. 1981

[6] RABINER L.R., SCHAFER R.W., *Digital Processing of Speech Signals,* Prentice-Hall, 1979

[7] DIXON N.R., MARTIN T.B., *Automatic Speech and Speaker Recognition,* IEEE Press, 1979.

NOTA - Il materiale qui presentato è tratto dall'articolo "Speech Recognition" dello stesso Autore, pubblicato su *"Scientific Honeyweller",* vol. 3, n. 3, Sept. 1982.

L'Intelligenza Artificiale ed il gioco degli scacchi

BARBARA PERNICI

Progetto di Intelligenza Artificiale
Politecnico di Milano

1. Introduzione

L'Intelligenza Artificiale è lo studio delle idee che rendono i calcolatori capaci di fare le cose che fanno sembrare intelligenti le persone.

Gli scopi centrali di tale disciplina sono, come definiti in [Wins 77], quello di rendere i calcolatori più utili e quello di capire i principi che rendono possibile l'intelligenza.

Lo studio dell'intelligenza con il calcolatore è infatti un metodo per studiare l'intelligenza in generale.

I calcolatori offrono dei vantaggi particolari in questo studio:

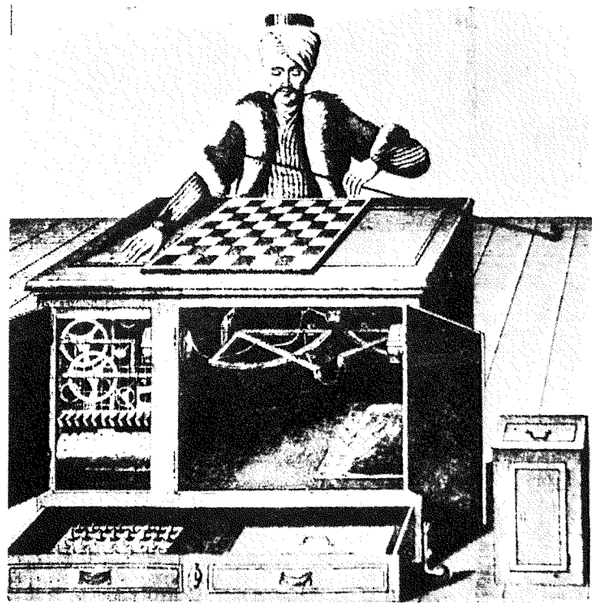
- i calcolatori sono soggetti ideali per effettuare esperimenti e per determinare all'interno di un esperimento quanto è importante un singolo frammento di conoscenza:
- il lavoro con i calcolatori ha portato ad un nuovo modo di espressione che rende più potente il ragionare sul pensiero.
- i modelli su calcolatore sono precisi e consentono di scoprire errori, concetti e sviste in essi contenuti;
- si può individuare un limite superiore alla quantità di informazione da elaborare per svolgere un certo compito.

Caratteristica dell'Intelligenza Artificiale è che non si propone di simulare l'intelligenza umana. Cerca di individuare i principi che si applicano a tutti i processori intelligenti, non privilegiando quindi, ma neppure escludendo, i metodi propri dell'intelligenza umana. È un punto di vista che porta ad una nuova metodologia ed a nuove teorie.

Un risultato potrebbe essere l'acquisizione di nuove idee per migliorare l'intelligenza umana.

Tecniche dell'Intelligenza Artificiale portano a:

- riconoscere analogie e semplici concetti, esaminare lo stesso problema da diversi punti di vista per ottenere una buona descrizione del problema stesso;
- sfruttare i vincoli naturali nei problemi;



Pseudo automa giocatore di scacchi del '700

- esplorare alternative;
- utilizzare strutture di controllo nella risoluzione dei problemi;

Le tecniche dell'Intelligenza Artificiale vengono applicate a numerosi campi: alla linguistica, per il riconoscimento dell'inglese parlato e scritto; al riconoscimento di espressioni, ad esempio nella risoluzione di alcuni test di intelligenza; alla dimostrazione automatica di teoremi e alla risoluzione automatica di problemi;
alla risoluzione di problemi specialistici, come calcoli matematici, integrazioni non numeriche, diagnosi cliniche, analisi di spettrogrammi di massa in chimica; all'apprendimento autonomo; alla visione; alla robotica, per compiere automaticamente quel lavoro industriale di tipo meccanico, che essendo sporco, pericoloso, noioso e poco remunerativo viene rifiutato dagli uomini; ai modelli dei processi psicologici, per capire il pensiero umano dal punto di vista dell'elaborazione dell'informazione.

I risultati ottenuti in ciascuno di questi campi forniscono nuove tecniche e nuovi spunti che possono essere applicati in campi apparentemente molto dissimili.

Una caratteristica particolare dell'Intelligenza Artificiale è l'interdisciplinarietà.

Il problema di giocare a scacchi con il calcolatore è stato scelto come oggetto di ricerca sin dalla creazione dei primi calcolatori.

Per questo problema sono state inventate tecniche particolari, alcune poi trasferite nel campo della risoluzione automatica dei problemi, ed anche sono state sperimentate tecniche di Intelligenza Artificiale già esistenti [Bowd 53, Feig 63, Newe 72].

Tra tutti i giochi studiati nel mondo dell'Intelligenza Artificiale, gli scacchi sono risultati i più interessanti per alcune loro caratteristiche.

Innanzi tutto gli scacchi sono un gioco a *informazione completa.* Entrambi i giocatori sono in grado di vedere tutta la scacchiera e niente è nascosto, come avviene in altri giochi, come ad esempio nel mastermind, dove il tentativo di identificare i colori che sono stati nascosti dall'avversario è l'essenza stessa del gioco.

Inoltre ciascun giocatore sa sempre a chi tocca muovere e quali sono le mosse possibili (legali) in ciascuna posizione.

Una seconda caratteristica degli scacchi è che *non vi è alcun elemento casuale,* come avviene nei giochi di carte o con i dadi. Terzo, una partita di scacchi termina sempre con un *risultato* ben definito: vittoria, parità o sconfitta.

Queste caratteristiche sono comuni a molti giochi e a molti problemi, ma gli scacchi risultano i più interessanti per la loro complessità. Anche se teoricamente sarebbe possibile, è in pratica impossibile calcolare quale sarebbe il risultato di una partita nel caso in cui entrambi i giocatori giocassero sempre la mossa migliore, detta anche il valore del gioco.

Good ha calcolato che dalla posizione iniziale sulla scacchiera possono derivare, effettuando sempre mosse legali, circa 10^{46} posizioni diverse [Good 68]; è chiaramente impossibile esaminare tutte queste posizioni per calcolare il valore del gioco. Di conseguenza è necessario cercare di risolvere il problema con tecniche di intelligenza artificiale [Nils 71, Nils 80, Wins 77]; le persone non esaminano tutte le possibili varianti a partire da una data posizione, ma considerano solo un numero di mosse ridotto, a seconda della loro abilità e seguendo i principi basilari del gioco. Per risolvere questo problema in Intelligenza Artificiale si tratta innanzitutto di rappresentare bene il problema stesso e poi di applicare a ciascuna rappresentazione i relativi metodi di soluzione. Diverse tecniche sono state sperimentate per tentare di risolvere questo problema.

Le principali sono:

- spazio degli stati
- sistemi di produzione
- tavole di decisione
- reti semantiche
- riduzione a sottoproblemi.

Segue una rassegna storica dell'impiego di queste tecniche negli scacchi ed una loro breve descrizione [Pern 82]. Si dovrebbe fare in realtà una distinzione fra programmi in grado di giocare un'intera partita e programmi che considerano solo un aspetto del gioco (di solito combinazioni di matto oppure un certo tipo di finale).

Da un punto di vista metodologico, però, entrambi i tipi di programmi sono interessanti e quindi verranno qui considerati [Clar 77, Clar 79].

2. Spazio degli stati

L'articolo nel quale sono contenuti i principi basilari del gioco degli scacchi con il calcolatore fu pubblicato già nel 1950 da Shannon [Newb 75].

Nel suo articolo Shannon definisce gli stati e gli operatori degli scacchi.

Gli *stati Pi* sono le posizioni sulla scacchiera e i relativi *operatori* Mi sono le mosse possibili e legali per il giocatore con il tratto in una data posizione Pi.

Applicando a ciascuna posizione Pi tutti i relativi operatori Mi si ottiene una struttura ad albero del tipo di Fig. 1. PO è detta radice dell'albero e le posizioni ottenute nella parte terminale dell'albero sono dette foglie.

Nella posizione data PO una mossa *Mi* può essere scelta *a caso,* oppure si può assegnare un *valore* ad ogni posizione risultante da PO costruendo l'albero del gioco come in Fig. 1 e scegliere la posizione migliore.

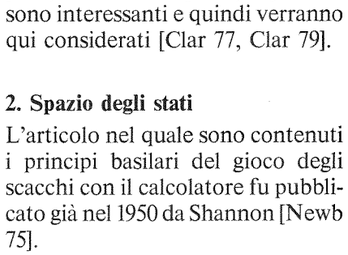


Fig. 1 - Un esempio di albero del gioco.

La definizione dei criteri che portino a corrette valutazioni delle posizioni è molto difficile.

Shannon propose una *funzione di valutazione* come somma pesata dei seguenti elementi:

- bilancio del materiale (valore complessivo dei pezzi per ciascuno dei due giocatori; al Re che deve essere sempre presente sulla scacchiera, viene dato un valore superiore di uno o due ordini di grandezza rispetto agli altri pezzi);
- struttura dei pedoni (ad esempio si penalizza il fatto che due pedoni dello stesso colore vengano a trovarsi sulla stessa colonna);
- mobilità dei pezzi (in quante posizioni diverse può recarsi ciascun pezzo dalla casa in cui è collocato).

Supponiamo che il giocatore che ha il tratto in PO debba massimizzare questa funzione; in questo modo sceglierà la mossa che porta allo stato in cui la funzione assume il valore maggiore (Fig. 2).

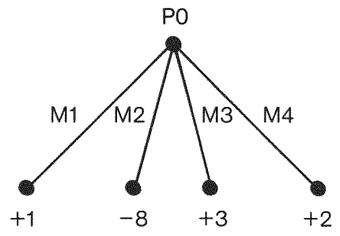


Fig. 2 - La mossa migliore.

Data la difficoltà di valutare con precisione una posizione guardando una sola mossa in avanti, allo stesso modo si può analizzare un'ulteriore mossa in avanti per valutare meglio la posizione (Fig. 3), creando un ulteriore livello dell'albero del gioco, tramite gli operatori (mosse) applicabili rispettivamente in P1, P2, P3, P4.

Fig. 3 - Albero a due livelli.

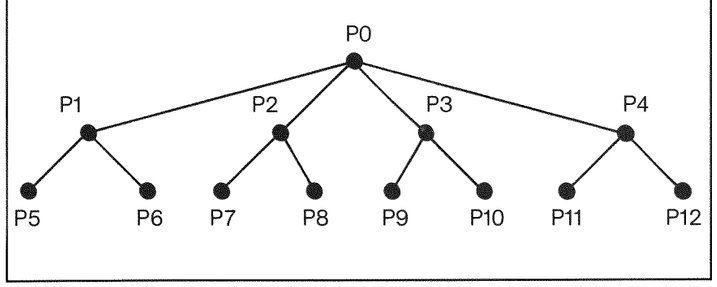
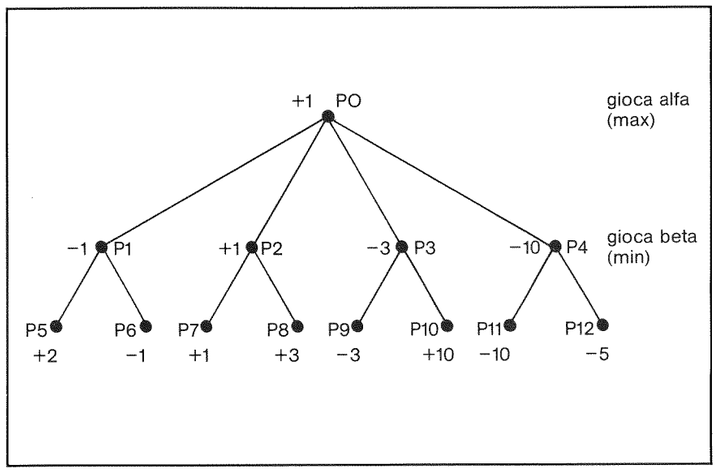


Fig. 4 - Minimax.



Tutte le posizioni finali sono valutate ed un valore adeguato è riportato alla posizione iniziale PO.

Questo è fatto con la procedura di *minimax.* Il primo giocatore, come detto prima, sceglie come mossa migliore la mossa che lo porta ad uno stato in cui la funzione di valutazione ha il suo valore più alto. L'altro giocatore farà la stessa cosa al contrario: sceglierà quella mossa che lo porta ad una posizione che sia il meno possibile favorevole al suo avversario, scegliendo la posizione in cui la funzione di valutazione assume il valore minore.

Nella Fig. 4, anche se la posizione migliore per alfa dopo due mosse è P10, non conviene giocare la mossa che porta a P3, perchè beta da P2 sceglierà invece la mossa che porta a P9, che è per lui migliore.

In questo approccio agli scacchi con il calcolatore si presentano alcuni problemi:

1) La dimensione dell'albero

Il fattore di ramificazione (numero di mosse possibili a partire da ciascuna posizione) è molto alto, in media 35, cosicchè anche se si considerano nella valutazione soltanto due mosse in avanti, si devono valutare circa $35^2 = 1225$ posizioni. Il numero di posizioni da valutare cresce esponenzialmente con il numero di livelli dell'albero (se si considerano n livelli, si devono valutare circa 35^n posizioni).

2) Effetto orizzonte

Se la profondità dell'albero (cioè il numero di mosse da considerare) è fissata e se, per esempio, l'ultima mossa effettuata a partire da una posizione in cui vi è parità di materiale è stata la cattura di un pezzo di alfa, la posizione finale è valutata con un pezzo in meno per alfa. In realtà due cose possono essere successe: o alfa ha effettivamente perso il pezzo, o l'ultima mossa è l'inizio di uno scambio di pezzi, che porta poi nuovamente a una situazione di parità di materiale (Fig. 5).

Si potrebbe pensare di estendere l'albero (Shannon) finchè non siano state esaminate tutte le catture, ma questo non è sufficiente.

Nella figura 6 l'alfiere a4 è perso qualunque mossa venga giocata dal bianco, ma anche estendendo l'albero il programma non se ne renderà conto e continuerà a giocare altre mosse in altri settori della scacchiera per evitare la sequenza di mosse

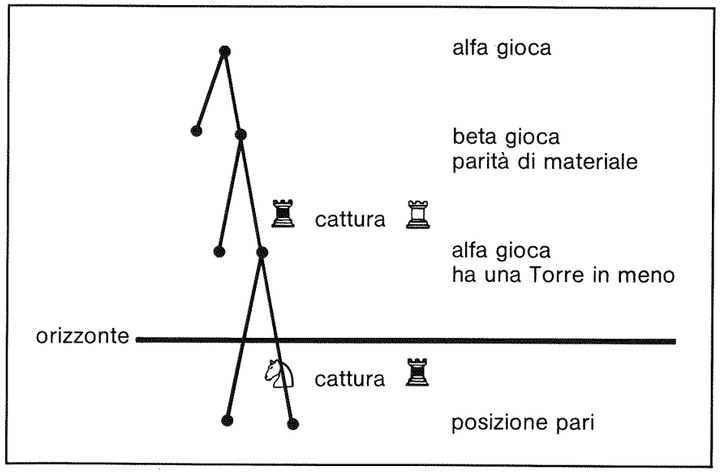


Fig. 5 - L'effetto orizzonte in uno scambio di Torri.

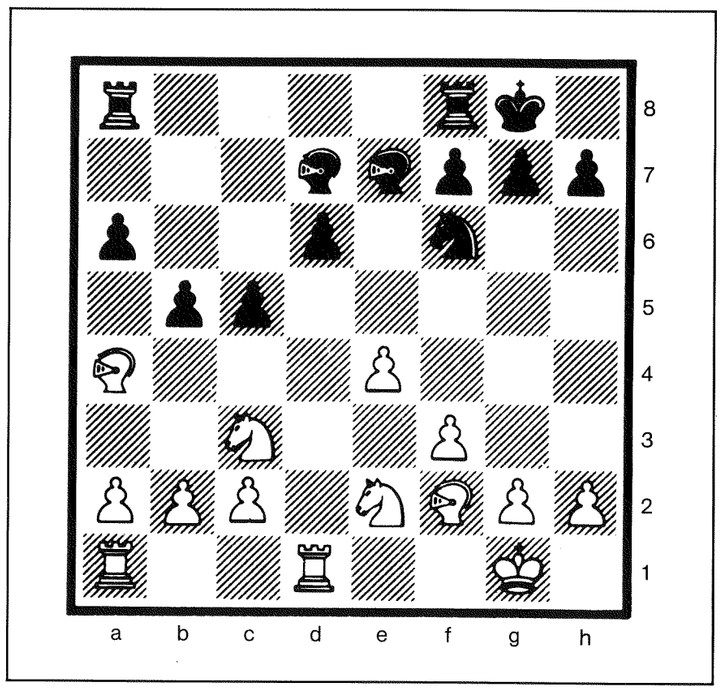


Fig. 6 - Un esempio di effetto orizzonte.

1. Ab3 (che salva momentaneamente il pezzo), c4; in realtà la sequenza non è evitata, ma viene rimandata oltre l'orizzonte.

Per risolvere il primo dei due problemi citati una soluzione è quella di evitare di generare tutte le mosse possibili: verrà scelta un'*euristica*, cioè un criterio da seguire nella generazione delle mosse, per evitare di considerare mosse illogiche [Harr 75]. Questa soluzione porta ad una *potatura anticipata* dell'albero, con l'eliminazione di tutti quei rami che derivano da una mossa illogica.

Questo metodo era già contenuto in una proposta B di Shannon (alberi a livelli variabili e potatura anticipata), e fu poi migliorata da Turing nel 1953 [Newb 75], il quale, oltre a migliorare la funzione di valutazione proposta da Shannon, esaminava oltre l'orizzonte (a partire da un certo livello fisso dell'albero in poi) solo le *mosse considerabili:*

1) ricattura di un pezzo;
2) cattura di un pezzo indifeso;
3) cattura di un pezzo da parte di uno di minor valore;
4) una mossa che dà scacco matto.

In questo modo viene introdotto il concetto di *posizione statica,* l'unico tipo di posizione in cui possono essere fatte considerazioni di tipo statico tramite una funzione di valutazione.

In questo modo, seppure in modo limitato, si possono risolvere sia i problemi derivanti dalle dimensioni dell'albero che dall'effetto orizzonte.

I programmi descritti finora non furono eseguiti realmente su calcolatore, ma furono piuttosto proposte teoriche.

Solo Turing simulò un'esecuzione del suo programma, eseguendolo a mano. Ciò nonostante queste proposte contengono molte delle idee con cui sono progettati i calcolatori e microcalcolatori per gli scacchi ancor oggi.

Seguirono dei tentativi di realizzazione pratica di queste proposte negli anni '50, sui lentissimi calcolatori allora a disposizione.

Con il passare del tempo si registrarono miglioramenti dovuti essenzialmente alla maggiore velocità di calcolo dei calcolatori, che permetteva di esaminare un maggior numero di mosse, all'introduzione di nuove euristiche e al miglioramento della funzione di valutazione [Adel 75, Adel 79, Berl 79, Mous 79].

È difficile confrontare le prime realizzazioni fra loro, poichè dipendevano fortemente dal calcolatore che utilizzavano, sia riguardo alla velocità di calcolo che riguardo alla quantità di memoria disponibile. Inoltre i programmi giocarono pochissime partite pubblicate.

Verso la fine degli anni '50 fu introdotto un altro tipo di euristica: il *generatore di mosse.*

Le mosse vengono generate solo in funzione di alcuni *aspetti del gioco* (difesa del re, sviluppo dei pezzi, difesa dei pezzi, attacco dei pezzi), come fece Bernstein nel 1958 [Newb 75, Feig 63], o come funzione di alcuni *obiettivi* (controllo del centro, equilibrio del materiale) per i quali vengono costruiti dei processi separati, costituiti da un generatore di mosse separato per ciascun obiettivo. Ognuno di questi processi ha una funzione di valutazione per la valutazione del raggiungimento del proprio obiettivo ed un generatore di mosse per l'analisi all'interno del processo. Questo tipo di approccio fu realizzato da Newell, Shaw e Simon nel 1958 [Newb 75]. Nel loro programma è molto importante la modularità: ogni obiettivo è separato dagli altri e può venire utilizzato nella fase del gioco in cui è più adatto (gli obiettivi durante diverse fasi di una stessa partita possono essere contrastanti; ad esempio mentre nell'apertura e nel centro di partita il re deve essere protetto e rimanere ai margini della scacchiera, nel finale al contrario deve essere portato verso il centro della scacchiera nella maggior parte dei casi).

Le difficoltà che sorgono nell'utilizzare questo approccio riguardano la valutazione globale della posizione ed il tempo di calcolo necessario, dato il numero molto elevato di obiettivi. D'altra parte, se, per evitare questi problemi, si considera solo un numero limitato di obiettivi, il risultato è un programma che gioca molto male.

L'altra novità contenuta nel programma di Newell et a1. è l'uso dell'*algoritmo alfa-beta,* basato su un semplice principio: se una mossa giocata da un giocatore (alfa) è già stata confutata da una possibile contro-mossa di beta, è del tutto inutile esaminare le altre possibili contro-mosse dell'avversario in quella posizione. Se qualcuna di esse fosse più favorevole ad alfa certamente beta la scarterebbe. Un esempio di potatura dell'albero con l'algoritmo alfa-beta è illustrato in Fig. 7.

Nella Fig. 7, visitando l'albero in profondità si valutano le posizioni P4, P5, P6. Supponendo che beta debba minimizzare e alfa massimizzare la funzione di valutazione, viene riportato il valore (+7) alla posizione P1. Continuando la visita dell'albero in profondità si considerano poi le posizioni P7 (+14) e P8 (+3).

A questo punto sappiamo che se alfa gioca la mossa M1 ottiene una posizione valutata +7, e se gioca M2 raggiungerà una posizione valutata +3 (oppure ancor meno, se beta ha mosse migliori di M8). Se ne conclude che non vale la pena di analizzare le posizioni che possono essere raggiunte dalla posizione P2, e che si può andare direttamente a valutare P10, che deriva da P3. Il valore di P10 è +1, e questo risultato porta immediatamente alla conclusione che nella posizione P0 la mossa migliore per alfa è M1. Utilizzando l'algoritmo alfa-beta in questo caso si sono dovute valutare 6 posizioni anzichè 9; l'algoritmo risulta in genere ancor più conveniente se il fattore di ramificazione è maggiore.

Questo è un esempio di *potatura posticipata* dell'albero: i rami dello albero che è inutile esaminare non vengono considerati.

Il minimax applicato sull'albero del gioco con o senza la potatura dell'albero con l'algoritmo alfa-beta porta in entrambi i casi allo stesso risultato, con un notevole risparmio di tempo di calcolo se si utilizza l'algoritmo alfa-beta.

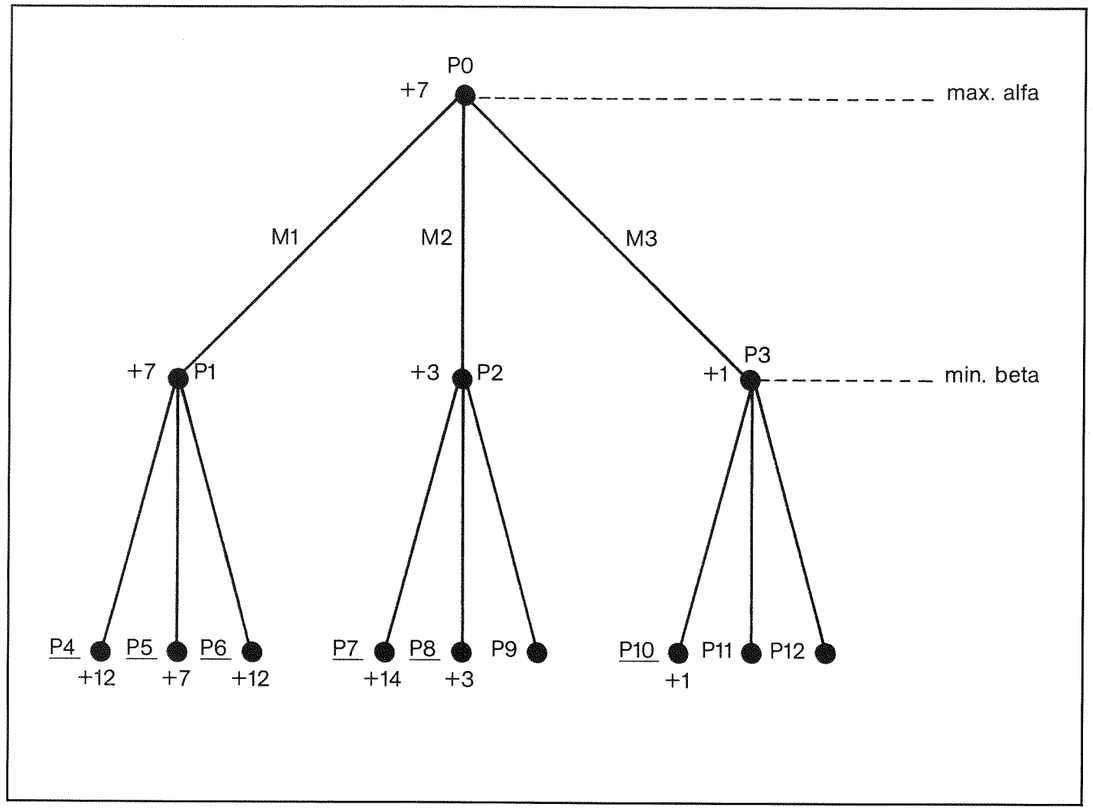


Fig. 7 - Esempio di algoritmo alfa-beta.

Un ulteriore miglioramento nelle prestazioni di un programma che gioca a scacchi può essere ottenuto *ordinando le mosse* nell'albero, in modo tale da cominciare ad esaminare l'albero con l'algoritmo alfa-beta dalla posizione ottenuta con la mossa ritenuta la migliore (Greenblatt Chess Program - 1967) [Newb 75]. Questo accorgimento rende la ricerca della mossa migliore nell'albero molto più rapida.

Il programma citato sopra, che gioca abbastanza bene, è uno dei più importanti nella storia dei programmi per gli scacchi. Esso giocò diverse partite anche in tornei regolari ed ottenne un punteggio nella classifica statunitense che lo poneva al di sopra del livello di un principiante.

Nel 1970 si tenne il primo Campionato per Calcolatori negli Stati Uniti. Nel 1974 ci fu il primo Campionato Mondiale a Stoccolma.

Tutti i programmi che giocano in torneo usano molto la ricerca tramite l'albero del gioco. I due migliori programmi sono stati per anni CHESS, di Atkin e Slate (USA) e KAISSA, di Arlazarov e Donskoy (URSS) [Frey 77].

La ragione per la quale tutti questi programmi fanno uso dell'albero è che il tentativo di fare sofisticate considerazioni sulle posizioni dell'albero per cercare di ridurre le ramificazioni richiede spesso una quantità eccessiva di tempo e talvolta è inutile. Sebbene i giocatori umani non analizzino mai sequenze molto lunghe di mosse, o lo facciano in maniera molto selettiva, cioè esaminando le mosse che ritengono più logiche, questo modo di procedere non può essere realizzato in maniera ragionevole in un calcolatore [Botv 70]; il risultato è che in genere risulta più conveniente fare una ricerca ad albero il più estesa possibile con euristiche e accorgimenti tipo l'algoritmo alfa-beta, poichè questo metodo risulta il più adatto alla velocità di calcolo tipica di un calcolatore. Se avessimo dovuto aspettare di sapere come gli uomini ragionano quando fanno un calcolo i calcolatori non esisterebbero ancora!
Questo però non significa che non si possano trovare altri metodi che, pur non utilizzando il modo di ragionare tipico degli uomini, ne considerino però alcuni aspetti [Berl 73].

È più facile trovare questo approccio non nei programmi che giocano partite complete, ma in quelli che considerano solo un aspetto del gioco [Chur 79].

Ci sono molti programmi "tattici", per un gioco di tipo brillante e combinativo, ad esempio per la risoluzione di problemi di matto in un numero fissato di mosse, che evitano di esaminare tutte le mosse legali, ma seguono dei principi generali.

In questo campo un lavoro notevole è quello realizzato da Berliner nel 1974 come tesi di dottorato [Berl 74]. Pur utilizzando ancora la ricerca ad albero, introduce alcune innovazioni:

- Una *migliore rappresentazione* della scacchiera. Se la scacchiera viene rappresentata bene è più facile applicare considerazioni generali alle posizioni.

  Vengono introdotti alcuni concetti scacchistici, come quello di linea di azione dei pezzi.
- Il *sistema di rivendicazione*. Un *valore presunto* che specifica qual'è il giocatore che sta meglio e di quanto viene assegnato alla posizione da esaminare.

  Se durante la ricerca con l'albero vengono trovati dei valori della funzione di valutazione che si discostano molto dal valore presunto, questo significa che uno dei due giocatori ha giocato una mossa debole e che quindi non vale la pena di proseguire l'analisi delle mosse che derivano da quella posizione.
- La *ricerca della causa*. È questa la novità maggiore. Se una mossa viene confutata, si cerca di individuare se la confutazione è dovuta alla mossa stessa o se esisteva già nella posizione precedente un pericolo per cui la confutazione non è dovuta all'ultima mossa gicata (ad esempio nella Fig. 6 qualunque mossa giochi, il bianco perde l'alfiere; questo fatto non può essere attribuito ad una mossa giocata in questa posizione).
- *Stati del giocatore* (aggressivo, difensivo, ecc.). Questi stati vengono usati per generare diversi tipi di mosse, adatte al tipo di stato considerato.

È stato raggiunto un buon livello di gioco nelle posizioni tattiche.

Per questo motivo la ricerca si rivolge ora all'obiettivo di ottenere un buon *gioco posizionale*. Mentre il gioco di tipo tattico è basato più sul calcolo di un gran numero di mosse che sulle idee, nel gioco posizionale queste sono essenziali.

Consideriamo la posizione di Fig. 8.

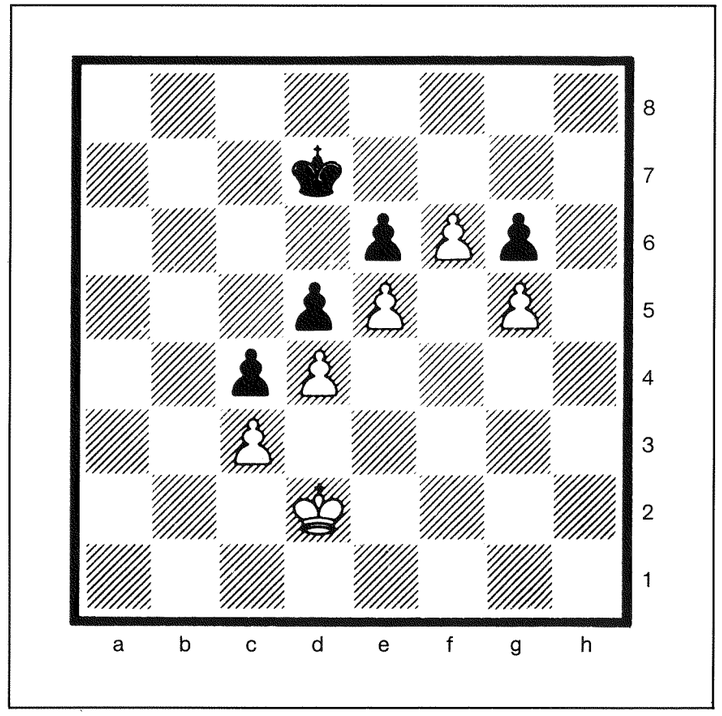
Se si ha per il finale un'euristica che raccomanda di portare il re verso il centro, il bianco non vincerà, mentre vincerebbe facilmente avvicinando il re al proprio pedone in f6 attraverso la linea a. Invece con l'euristica menzionata, che è tipica per i finali, il bianco continuerà a giocare mosse di re nelle case e3-e2-d2 senza trovare il piano giusto per vincere.

D'altra parte non è possibile esaminare tutte le mosse a partire da questa posizione perchè, anche considerando solo le mosse di re, il fattore di ramificazione è nella maggior parte dei casi uguale ad 8 e porta ad un numero eccessivo di posizioni da considerare.

Da queste considerazioni si evidenzia la necessità di utilizzare altri metodi.

Un metodo che può essere associato a quello dello spazio degli stati comporta l'utilizzo di una funzione

Fig. 8 - I piani nel finale.



di valutazione di tipo SNAC, cioè con coefficienti variabili gradualmente e non lineari.

Questo metodo è stato applicato con successo da Samuel alla dama [Samu 63] e da Berliner al backgammon [Berl 79], in questo caso senza ricerca tramite l'albero, a causa dell'eccessivo fattore di ramificazione di questo gioco. I risultati sono ottimi: il programma di Berliner è riuscito persino a battere il campione del mondo in carica di backgammon!

3. Sistemi di produzione e tavole di decisione

Anche se ormai il livello di gioco ottenuto impiegando l'approccio dello spazio degli stati è buono, rimane ugualmente molto lontano da quello dei grandi maestri. Si potrebbe pensare di migliorare ulteriormente la ricerca tramite l'albero, la funzione di valutazione e così via, ma ai programmi mancherebbe ugualmente quello che i giocatori umani usano ancor più dell'esame delle varianti: la *conoscenza scacchistica* [Eise 73]. Il solo tipo di conoscenza contenuta nei tradizionali programmi che impiegano lo spazio degli stati è quella che riguarda le mosse legali in una data posizione e la valutazione statica delle posizioni, fatta essenzialmente confrontando il numero e il tipo di pezzi presenti sulla scacchiera per i due giocatori, dopo che è stato giocato un certo numero di mosse. In questo modo, più mosse si considerano, più precisa risulta la valutazione della posizione.

A partire da questa sezione verranno illustrati i tentativi fatti dai vari ricercatori negli ultimi anni per introdurre la "conoscenza" nei programmi per gli scacchi utilizzando tecniche già utilizzate in altri settori dell'Intelligenza Artificiale.

Un settore che richiede una grande quantità di conoscenza è quello dei *sistemi esperti* [Mich 79].

I sistemi esperti già esistenti sono capaci di dare consigli in tempo reale riguardo argomenti specialistici della medicina e della chimica. Vengono costruiti utilizzando le conoscenze degli specialisti di un dato settore ed hanno ottime prestazioni, a volte persino migliori di quelle degli specialisti stessi, sia pure solo nel loro limitato campo di applicazione.

Anche un programma per giocare a scacchi può essere considerato un sistema esperto [Brat 79, Bram 80a].

Questo tipo di approccio è particolarmente indicato nella fase del finale di partita, ma può essere utilizzato anche per risolvere problemi di tipo tattico [Pitr 77, Wilk 79].

La base di conoscenza di un sistema esperto è costituita da *regole di produzione*. Ogni regola di produzione è composta da un insieme di condizioni ed un insieme di azioni. Quando le condizioni di una o più regole sono verificate in una data situazione, le corrispondenti azioni vengono eseguite.

Ogni regola di produzione è del tipo:

se 'condizione' allora 'azione'

La parte 'azione' può essere costituita da una data azione, un valore di un parametro, una sequenza di azioni, uno schema di azioni, una lista di obiettivi o un'altra struttura di decisione usata per guidare un modulo che genera azioni.

Concatenando un certo numero di regole, si può ottenere un processo di *inferenza deduttiva*. Inoltre un vantaggio della struttura usata in un sistema esperto e che è possibile l'acquisizione di nuove regole anche tramite un processo di apprendimento autonomo. Questo fu scoperto in un programma per la diagnosi di malattie delle piante. Le regole di produzione sintetizzate dalla macchina davano risultati migliori di quelle proposte originariamente da uno specialista in quel campo.

È evidente che un approccio di questo tipo è molto interessante anche per il gioco degli scacchi. I risultati finora ottenuti sono molto promettenti, anche se vi sono dei problemi per un'applicazione in questo campo. Infatti individuare quali condizioni sono verificate è un'operazione molto costosa in termini di tempo di calcolo e può essere utile solo se l'impiego di un sistema di produzione riduce drasticamente le dimensioni dell'albero, che non può mai essere del tutto eliminato nel gioco degli scacchi.

Le regole di produzione negli scacchi vengono impiegate per ottenere dei piani, simili a quelli dei giocatori umani, ma il *numero di piani* in una posizione deve rimanere limitato.

Un altro notevole problema riguarda la dimensione della *base della conoscenza*. I sistemi esperti finora progettati operano in campi molto specializzati. Anche i programmi di scacchi di questo tipo sono sempre limitati ad un particolare aspetto del gioco, ad esempio un certo tipo di finale o combinazione di centro partita. Se si volesse un sistema esperto per gli scacchi in grado di giocare un'intera partita la sua base di conoscenza dovrebbe comprendere migliaia di regole e la verifica delle condizioni di queste regole richiederebbe un'enorme quantità di tempo. Nonostante questo svantaggio questo approccio sembra molto promettente per il futuro.

Una tecnica molto simile alla precedente è quella che implica l'uso di *tabelle di decisione*. Una tabella di decisione è composta da un insieme di possibili condizioni e per ogni combinazione di queste condizioni sono elencate delle azioni da intraprendere (Fig. 9).

La regola R1, ad esempio, consiglia di intraprendere l'azione A1 quando è verificata la condizione C1 e non è verificata la condizione C3.

Questo metodo è stato utilizzato da Michie e Bratko per il finale di re e torre contro re e cavallo [Brat 78, Brat 79, Brat 80].

Con una tabella per questo finale e *l'albero delle varianti forzate,* in cui si considerano solo le mosse migliori per il giocatore in vantaggio e tutte le mosse dell'altro giocatore, i due ricercatori hanno dimostrato che i libri teorici su questo finale non sempre riportavano varianti esatte [Brat 78].

4. Le reti semantiche

Un altro aspetto generale che viene considerato poco nei programmi di scacchi è quello delle relazioni che ci sono tra i diversi pezzi della scacchiera.

Fig. 9 - Un esempio di tabella di decisione.

		R1	R2	R3	regole
condizioni	C1	SI	NO	–	
	C2	–	SI	–	
	C3	NO	SI	SI	
azioni		A1	A2	A2	
			A3	A3	
				A4	

In genere nei programmi esistenti i pezzi sono messi in relazione alla casella della scacchiera sulla quale si trovano, e le relazioni tra di essi passano attraverso le caselle della scacchiera a causa di questa rappresentazione. In questo modo ad esempio non c'è modo di sapere direttamente se un pezzo attacca dei pezzi dell'avversario. Il programma deve esaminare tutte le caselle che il pezzo può raggiungere in una mossa e poi vedere se in queste caselle ci sono pezzi avversari.

Nel '73 è stata fatta da Michie l'interessante proposta di una *rete semantica* che rappresenta le relazioni tra i pezzi (vedi Fig. 10) [Mich 73].

Le reti semantiche sono state usate in problemi come quello della visione per riconoscere gli oggetti.

Potrebbe essere possibile, ad esempio, riconoscere immediatamente se un pezzo è in presa.

Sarebbero però necessari un approfondimento di questo argomento e probabilmente l'introduzione di nuovi tipi di relazioni, dal momento che quelle indicate in Fig. 10 forniscono solo una piccola quantità di informazione sulla posizione, e la stessa rete può rappresentare più di una posizione con caratteristiche differenti [Sali 76], e quindi perde parte della sua validità.

Questo aspetto è stato considerato finora solo tramite alcune strutture di dati ad hoc, ad esempio per determinare se un pezzo è catturabile [Berl 74], ma mai in maniera molto estesa ed approfondita.

5. Riduzione a sottoproblemi

Un problema che si presenta utilizzando l'albero delle varianti è che l'"albero" in realtà non è un albero, ma un grafo. Infatti si può raggiungere la stessa posizione con diverse sequenze di mosse. Si fa uso però dell'albero del gioco, perchè è molto più semplice considerare un albero piuttosto che un grafo.

Un problema in un certo senso simile al precedente è che alcuni aspetti di una posizione in realtà non necessitano di essere valutati nuovamente dopo l'esecuzione di una mossa. Questo problema si presenta anche in robotica quando si cerca di capire che cosa è cambiato nell'ambiente dopo un'azione.

Per evitare la ripetizione della valutazione di parti di una posizione che non sono cambiate per effetto della mossa fatta (vedi ad es. Fig. 6 e Fig. 8), il problema originale di valutare una posizione e scegliere quindi una mossa può essere diviso in sottoproblemi, e la valutazione separata di aspetti della posizione consente di non dover riconsiderare quegli aspetti che sono rimasti invariati.

Questo modo di procedere è molto simile a quello di un giocatore umano che, per varie ragioni (memoria e tempo a disposizione limitati, schemi fissi di gioco, ecc.) tende ad esaminare la scacchiera a settori e non nel suo insieme. D'altra parte questo tipo di approccio, applicato ad un programma, accentuerebbe il difetto che si presenta nel gioco degli uomini, che spesso tendono a considerare solo una parte della scacchiera, trascurando così alcune mosse "strane" o imprevedibili, che però possono cambiare completamente la situazione sulla scacchiera [De Gr 65].

6. Le "macchinette" che giocano a scacchi.

Per concludere, si presenta una breve storia delle "macchinette" che giocano a scacchi, cioè di quei microcalcolatori di modeste dimensioni in grado di giocare a scacchi, che sin dal loro apparire nel non lontano 1976 hanno letteralmente invaso il mercato.

La prima di esse fu prodotta da una piccola ditta, la Fidelity, a partire dall'idea originale di un dipendente della ditta stessa. Quest'idea di costruire una piccola macchina con la

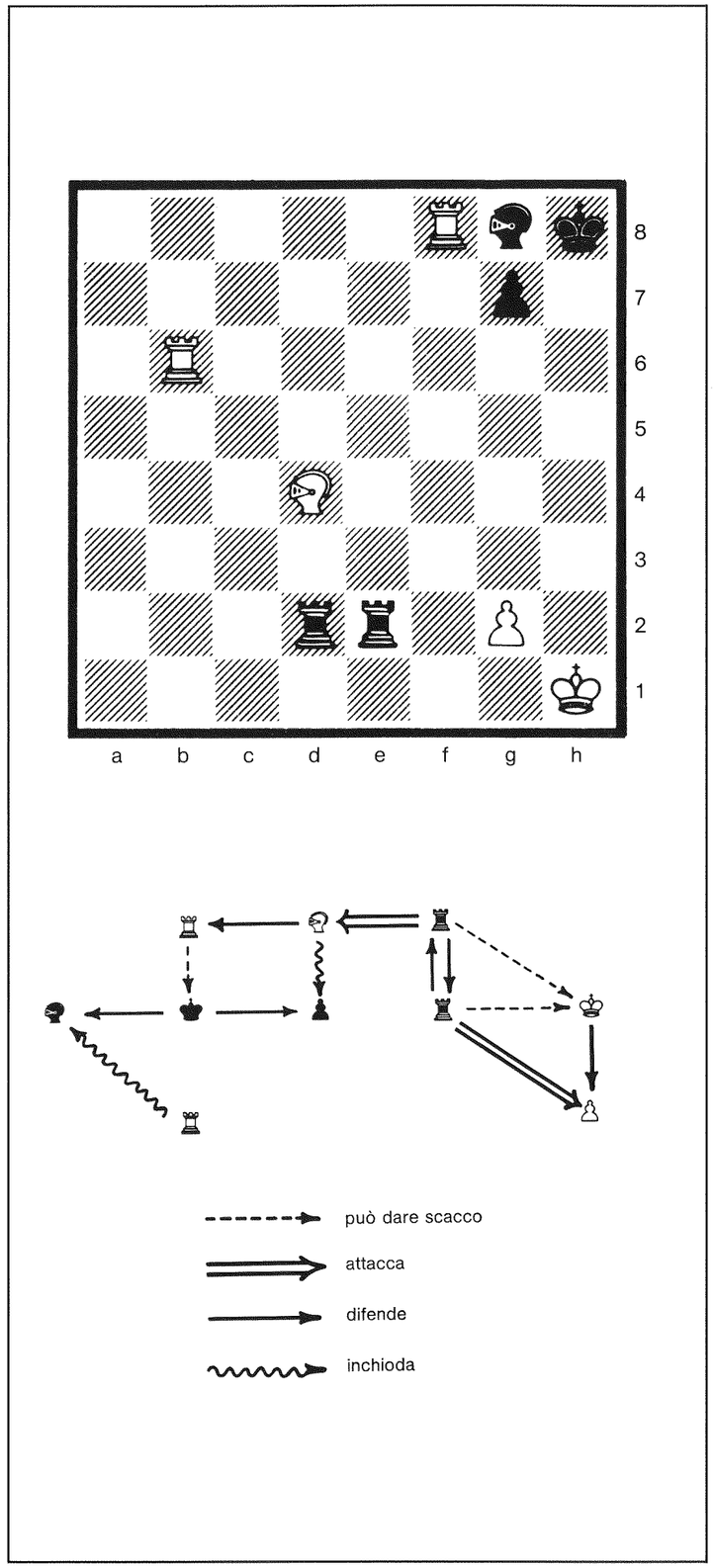


Fig. 10 - Rete semantica.

quale l'uomo si potesse cimentare in una partita a scacchi senza alcun intervento esterno ebbe un grande successo. E questo nonostante la macchinetta non facesse altro che muovere i pezzi, non effettuando neppure, tra l'altro, mosse particolari come l'arrocco, la promozione di un pedone giunto in ottava traversa e così via.

L'anno seguente, nel '77, venne presentata una versione migliorata che effettuava una ricerca ad albero a tre livelli.

Nel 1978 si registrò un notevole miglioramento con l'estensione dell'albero a 10 livelli, riducendo progressivamente il fattore di ramificazione tramite euristiche all'aumentare del livello. Il risultato è una macchina che riesce già a battere una buona parte dei dilettanti.

Fino al 1979 non vi fu alcun contatto tra il mondo della ricerca scientifica ed il mondo della produzione commerciale delle macchinette. In tale anno, però, questo contatto comincia a verificarsi con l'introduzione nelle macchinette dell'algoritmo alfa-beta, nel programma Sargon 2.5 dell'Applied Concepts, che ha avuto un'amplissima diffusione e che rappresenta un vero salto di qualità rispetto alle macchinette precedenti.

Bisogna però aspettare il 1980 perchè venga affrontato anche il problema dell'effetto orizzonte, e la ricerca ad albero venga fatta proseguire fino al raggiungimento del punto di quiescenza.

È in quest'anno che si svolge a Londra il primo Campionato del Mondo per microcalcolatori, che riveste una notevole importanza anche commerciale, dal momento che coinvolge un giro d'affari intorno ai 100 milioni di dollari.

Vince il Champion della Fidelity, un prototipo, secondo si classifica il Morphy, programma in commercio dell'Applied Concepts.

Nei nuovi modelli viene migliorata l'interfaccia con l'utente mettendo dei sensori sulla scacchiera, consentendo di effettuare le mosse in modo naturale muovendo i prezzi su una scacchiera e non tramite tastiera.

Inoltre in alcuni modelli appare la separazione del software dall'hardware, che consente di utilizzare la stessa macchina per moduli di programma diversi, sia per diverse fasi del gioco, che per edizioni successive di programmi per la stessa macchina. Ciò consente soprattutto di ridurre i prezzi, non costringendo il cliente a buttare via tutta la macchina di cui è già in possesso per comprarne una nuova di prestazioni superiori.

Nel 1981 si è svolto il 2° Campionato Mondiale, in Germania ad Amburgo, diviso in due categorie, una riservata ai prototipi, vinta dal Champion Elite della Fidelity ed una riservata alle macchinette di serie, vinta dal Mark V, programmato dal maestro internazionale britannico Levy.

Attualmente tutte le macchinette che giocano ad un buon livello 'conoscono' tutte le regole degli scacchi, sono di più facile impiego rispetto ai primi esemplari, 'pensano' in tempi brevi ed utilizzano le tecniche ereditate dall'Intelligenza Artificiale.

Non sono ancora in grado di battere regolarmente giocatori classificati a livello nazionale, ma i miglioramenti avvengono con una rapidità tale che non è possibile fare previsioni.

Nel campo dei grandi calcolatori, che negli Stati Uniti hanno raggiunto punteggi nella graduatoria nazionale a livello magistrale, è già successo che un calcolatore abbia vinto il Campionato in uno degli stati USA. Il titolo però è stato assegnato al secondo classificato, un uomo!

7. Conclusioni

In questo articolo si è fatta una breve storia delle tecniche usate nei programmi che giocano a scacchi. Anche se i programmi attuali hanno raggiunto un buon livello di gioco, si possono però individuare alcuni notevoli difetti:

- difficoltà (o incapacità) di elaborare piani a lungo termine nel considerare le mosse possibili
- le mosse sono esaminate separatamente l'una dall'altra, senza alcuna concatenazione tra di esse
- mancanza di una visione globale della scacchiera.

La ricerca in questo campo dovrebbe seguire queste direzioni, per trovare una soluzione a questi problemi, anche considerandone per semplicità solo alcuni, almeno per il momento. Un programma che giochi veramente bene a scacchi può essere ottenuto solo considerando il gioco degli scacchi nella sua completezza e complessità, e non come arida sequenza di mosse.

8. Bibliografia

[Adel 75] ADELSON-VELSKY G.M., ARLAZAROV V.L. e DONSKOY M.V., *"Some Methods of Controlling the Tree Search in Chess Programs"*, Artificial Intelligence 6, 1975.
[Adel 79] ADELSON-VELSKY G.M., ARLAZAROV V.L. e DONSKOY M.V., *"Algorithms of Adaptive Search"*, Machine Intelligence 9, 1979.
[Arla 79] ARLAZAROV V.L. e FUTER A.L. *"Computer Analysis of a Rook End-game"*, Machine Intelligence 9, 1979.
[Berl 73] BERLINER H.G., *"Some Necessary Conditions for a Master Chess Programm"*, Proc. 3° IJCAI, 1973.
[Berl 74] BERLINER H.G., *"Chess as Problem Solving: the Development of a Tactics Analyser"*, Test di PhD, Carnegie-Mellon Univ., 1974.
[Berl 78] BERLINER H.J., *"A Chronology of Computer Chess and its Literature"*, Artificial Intelligence 10, 1978.
[Berl 79] BERLINER H.J., *"On the Construction of Evaluation Functions for Large Domains"*, Proc. IJCAI-79, 1979.
[Botv 70] BOTVINNIK M.M., *"Computer, Chess and Long-Range Planning"*, Springer-Verlag, 1970.
[Bowd 53] BOWEDER B.V., *"Faster than Thought"*, Sir Isaac Pitman and Sons, Ltd., 1953.
[Bram 79] BRAMER M.A., *"Testing Correctness of Strategies in Game-playing Programs"*, Proc. IJCAI-79, 1979.
[Bram 80a] BRAMER M.A., *"Pattern-based Representation of Knowledge in the Game of Chess"*, Proc. AISB-80, 1980.
[Bram 80b] BRAMER M.A., *"Correct and Optimal Strategies in Game Playing Programs"*, The Computer Journal, 23(4), 1979.
[Brat 78] BRATKO I. e MICHIE D., *"Advice Table Representation of Chess End-Game Knowledge"*, Proc. AISB/GI Conf. on Artificial Intelligence, 1978.
[Brat 79] BRATKO I., *"Implementing Heuristics Using the ALI Advice-Taking System"*, Proc. IJCAI-79, 1979.
[Brat 80] BRATKO I. e MICHIE D., *"An Advice Program for a Complex Chess Programming Task"*, The Computer Journal 23(4), 1980.
[Chur 79] CURCH K.W., *"Co-ordinate squares: a Solution to many Chess Pawn Endgames"*, Proc. IJCAI-79, 1979.
[Clar 77] CLARKE M.R.B., *"Advances in Computer Chess 1"*, Edinburgh University Press, 1977.
[Clar 79] CLARKE M.R.B., *"Advances in Computer Chess 2"*, Edinburgh University Press, 1979.
[Davi 77] DAVIS R. e BUCHANAN B.G., *"Meta Level Knowledge: Overview and Applications"*, Proc. 5° IJCAI, 1977.

[De Gr 65] DE GROOT A., *"Thought and Choice in Chess"*, Mouton & Co, 1965.
[Eise 73] EISENSTADT M. e KAREEV Y., *"Toward a Model of Human Game Playing"*, Proc. 3° IJCAI, 1973.
[Feig 63] FEIGENBAUM E.A. e FELDMAN J., *"Computers and Thought"*, Mc Graw Hill, 1963.
[Film 79] FILMAN R.E., *"The Interaction of Observation and Inference"* Tesi di PhD, Stanford University, 1979.
[Frey 77] FREY P.J., *"Chess Skill in Man and Machine"*, Springer-Verlag, 1977.
[Gill 72] GILLOGLY J.J., *"The Technology Chess Program"*, Artificial Intelligence 3, 1972.
[Good 68] GOOD I.J., *"A Five-year Plan for Automatic Chess"*, Machine Intelligence 2, Edinburgh, 1968.
[Guid 81] GUIDA G. e SOMALVICO M., *"Multi-Problem-Solving: Knowledge Representation and System Architecture"*, (proposto per la pubblicazione), 1981.
[Harr 75] HARRIS L.R., *"The Heuristic Search and the Game of Chess. A Study of Quiescence, Sacrifices, and Plan Oriented Play"*, Proc. 4° IJCAI, 1975.
[Kozd 73] KOZDROWICKI E.W. e COOPER D.W., *"COKO III: The Cooper-Koz Chess Program"*, Comm. ACM 16(7), 1973.
[Levy 76] LEVY D.N.L., *"Chess and Computers"*, Batsford, 1976.
[Mars 79] MARSLAND T.A., *"A Bibliography on Computer Chess"*, Machine Intelligence 9, 1979.
[Mich 73] Michie D., *"The Path to Championship Chess by Computer"*, Computers and Automation 1, 1973.
[Mich 79] MICHIE D., *"Expert Systems in the Micro-Electronic Age"*, Edinburgh University Press, 1979.
[Mich 81] MICHIE D., *"A Theory of Evaluative Comments in Chess with a Note on Minimaxing"*, The Computer Journal 24 (2), 1981.
[Mous 79] MOUSSOURIS J., HOLLOWAY J. e GREENBLATT R., *"CHEOPS: A Chess-oriented Processing System"*, Machine Intelligence 9, 1979.
[Newb 75] NEWBORN M., *"Computer Chess"*, Academic Press, 1975.
[Newe 72] NEWELL A. e SIMON H.A., *"Human Problem Solving"*, Prentice-Hall, Inc., 1972.
[Nils 71] NILSSON N., *"Problem-solving Methods in Artificial Intelligence"*, Mc Graw-Hill, 1971.
[Nils 80] NILSSON N., *"Principles of Artificial Intelligence"*, Tioga Publishing Company, 1980.
[Pern 81] PERNICI B., *"Il Gioco degli Scacchi e la Risoluzione Automatica dei Problemi"*, Atti Convegno Nazionale sui Giochi Creativi, Siena, 1981.
[Pern 82] PERNICI B., *"Problem Solving in Game Playing: Computer Chess"*, Cybernetics and Systems: An International Journal 13(1), 1982.
[Pitr 77] PITRAT J., *"A Chess Combination Program which Uses Plans"*, Artificial Intelligence 8, 1977.
[Reks 79] Reksten H., *"An Annotated Bibliography on Computer Chess"*, Dept. of Comp. and Inf. Sciences, Univ.of Delaware, Technical Report No. 79/4, 1979.
[Sali 76] SALIBA M.A., *"Aspetti della Simulazione Elettronica del Comportamento Umano nella Scelta di Strategia: il Caso degli Scacchi"*, Tesi di Laurea in Matematica, Pisa, 1976.
[Samu 63] SAMUEL A.L., *"Some Studies in Machine Learning Using the Game of Checkers"*, Computers and Thought, Mc Graw-Hill, 1963.
[Wilk 79] WILKINS D., *"Using Plans in Chess"*, Proc. IJCAI-79, 1979.
[Wins 77] WINSTON P.H., *"Artificial Intelligence"*, Addison-Wesley, 1977.
[Zobr 73] ZOBRIST A.L. e CARLSON F.R. JR., *"Il Calcolatore a Lezione di Scacchi"*, Le Scienze 62, 1973.

"COLLANA DEI QUADERNI DI INFORMATICA": DUE NUOVI TESTI SUI LINGUAGGI DI PROGRAMMAZIONE

***"La programmazione in Pascal"* di Dorothy A. Walsh**
e *"Il linguaggio Fortran"* di Alessandra e Enrico Spoletini

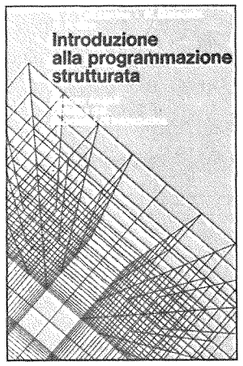
La "Collana dei Quaderni di Informatica" edita dalla Franco Angeli e curata dalla Honeywell Information Systems Italia si è arricchita di due nuovi titoli: "La programmazione in Pascal di *Dorothy A. Walsh* (pp. 364, Lit. 20.000) e "Il linguaggio Fortran" di *Alessandra e Enrico Spoletini* (pp. 315, Lit. 15.000).
Essi fanno seguito, nella Collana, ai due precedenti testi su linguaggi di programmazione: il "Basic" e il "Cobol" di *Enrico Spoletini.*
La "Collana dei Quaderni di Informatica" (otto titoli finora pubblicati, vedine l'elenco in fondo) è nata nel 1976 e rappresenta la proiezione sul piano librario della rivista "Quaderni di Informatica" edita, a partire dal 1974, dalla Honeywell Information Systems Italia.

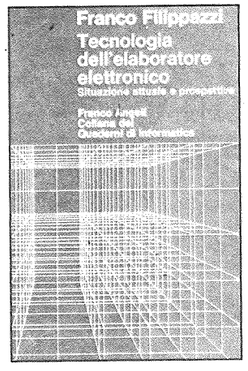
Uno dei più importanti e innovativi sviluppi della recente informatica è la creazione di linguaggi di programmazione che, alle caratteristiche tipiche dei cosidetti linguaggi ad alto livello, associano quella di consentire una facile realizzazione di "buoni programmi" (oggi, nell'età adulta dell'elaborazione dati, ben si comprende come non basti più "programmare", ma occorra ottenere subito programmi di "buona qualità", ossia privi di quegli errori logici che nei primi tempi dell'informatica richiedevano lunghe messe a punto).
È questo il caso di Pascal, un linguaggio realizzato dallo svizzero *Niklaus Wirth* tra la fine degli anni 60 e l'inizio degli anni 70, mentre cioè era in corso quel vasto processo di riflessione sui formalismi dei linguaggi di programmazione e sulla struttura dei programmi che è poi culminato nella formulazione dei moderni concetti della programmazione strutturata. Prodotto di questa fase evoluta dalla storia dell'informatica, il Pascal è un linguaggio di programmazione che per la sua funzionalità, flessibilità, facilità d'uso, leggibilità e per la sua intrinseca attitudine a favorire la redazione ordinata dei programmi, è destinato ad assumere una posizione sempre più rilevante fra i linguaggi di programmazione nei prossimi anni, in particolare fra i nuovi e sempre più larghi strati di utenti raggiunti dalla diffusione dei piccoli e piccolissimi elaboratori (non a caso viene sempre più largamente adottato sui microelaboratori).
In considerazione di quest'ultimo fatto il libro della Walsh non si limita alla pura descrizione dei formalismi del linguaggio, ma intende affrontare il problema della programmazione in generale, anche se con specifico riferimento alla scrittura di programmi in Pascal. Oltre a insegnare il linguaggio cui è intitolato, esso fornisce pertanto una gamma di informazioni e suggerimenti sullo "scrivere programmi" o, più generalmente, sul "produrre software" quale raramente si trova raccolta in un unico volume.
Il libro è infatti ricco di note di programmazione, richiama principi di base degli algoritmi, introduce concetti su metodiche di produzione di software, analizza la problematica dell'affidabilità del software e le modalità di prova dei programmi, non trascurando di illustrare, in una sintetica quanto completa panoramica, i termini e i concetti generali d'uso corrente nel mondo dell'elaborazione dei dati.
Inoltre la ben definita trattazione degli argomenti, il tono didattico ed una certa studiata "ridondanza" di taluni termini e concetti fondamentali, in modo che in qualche misura possa essere letto utilmente anche "a caso", ne fanno un libro di facile consultazione.
Per tutti questi motivi, e non ultimo per il fatto che la descrizione teorica del linguaggio è sempre supportata da esempi e addirittura da interi programmi, questo libro si presta anche ad essere vantaggiosamente impiegato come testo ausiliario in corsi generali e di introduzione alla programmazione.

Dorothy A. Walsh, nata New York, ha mosso i primi passi nell'elaborazione dati nel 1955, in una delle prime soft-

ware house nate negli Stati Uniti. Da allora ha svolto come consulente una serie di attività (dalla realizzazione di compilatori alla conduzione di grossi progetti software all'insegnamento e alla pubblicazione di saggi e libri) maturando in particolare una notevole esperienza nelle metodologie di software.
Attualmente lavora a Roma come consulente indipendente e insegna Computer Science al John Cabol International college, sede distaccata a Roma di un'università americana.

Realizzato dall'americano *John Backus* nel 1956 e definitivamente stabilizzatosi con la versione Fortran IV nel 1962, il Fortran è il più vecchio fra i linguaggi di programmazione oggi utilizzati ed è rimasto il linguaggio più diffuso per la risoluzione di problemi scientifico-matematici, in cui è andato sempre più specializzandosi.
Per questa sua grande diffusione esso è stato oggetto di definizione da parte dell'American National Standard, il cui controllo non ha però impedito il proliferare di numerosissimi "dialetti", ciascuno dei quali offre dei vantaggi settoriali, ma ognuno dei quali presenta lo svantaggio della non trasferibilità dei programmi.
La maggior parte delle forme dialettali, inserite nei diversi Fortran presenti sul mercato, riguardano la possibilità di utilizzare particolari forme organizzative dei dati disponibili sull'elaboratore che si utilizza, e le relative modalità di accesso. Questa possibilità rende compatibili i flussi generati con programmi Fortran e quelli generati con programmi scritti in altri linguaggi di programmazione.
In questo volume sono esposti gli elementi del Fortran IV e sono evidenziati le innovazioni più significative introdotte con la versione Fortran 77. Non viene trascurato inoltre di descrivere alcune utili e diffuse forme dialettali.
La prima parte è dedicata alla terminologia prevista dallo standard; la seconda ad una introduzione globale al linguaggio fatta attraverso semplici programmi; la terza alla presentazione sistematica delle specifiche del Fortran; la quarta infine riporta una serie di programmi relativi a problemi reali di natura diversa.
Le appendici presentano i concetti base relativi all'elaboratore, alcuni elementi classici di programmazione strutturata, una panoramica su una medodologia di programmazione ed informazioni su memorizzazione di dati ed ordini di grandezza.





Alessandra e *Enrico Spoletini* sono entrambi laureati in matematica e sono entrambi docenti universitari. Alessandra, dopo aver insegnato algebra nel corso di laurea in matematica all'Università di Milano, è attualmente docente di geometria presso il Politecnico di Milano.
Enrico, dopo essersi a lungo occupato di didattica della programmazione, linguaggi e metodologie presso la Honeywell Information Systems Italia, è ora docente di analisi matematica e macchine calcolatrici nel corso di laurea in fisica all'Università di Milano. Enrico Spoletini è anche autore di un volume sul Basic e di un volume sul Cobol, entrambi pubblicati nella "Collana dei Quaderni di Informatica".

Volumi finora pubblicati nella "Collana dei Quaderni di Informatica"
1. *Franco Filippazzi* "Tecnologia dell'elaboratore elettronico. Situazione attuale e prospettive".
2. *Enrico Spoletini* "Il Basic: teoria ed esercizi".
3. *Enrico Spoletini* "Il Cobol: teoria ed esercizi".
4. *Gaetano A. Lanzarone, Marco Maiocchi, Roberto Polillo* "Introduzione alla programmazione strutturata".
5. *D. Ferrari, G. Serazzi, A. Zeigner* "Le prestazioni degli elaboratori elettronici. Misura/Valutazione/Ottimizzazione".
6. *Giulio Occhini* "L'informatica nella gestione aziendale. Aspetti e prospettive d'impiego".
7. *Alessandra* ed *Enrico Spoletini* "Il linguaggio Fortran".
8. *Dorothy A. Walsh* "La programmazione in Pascal". ✧