

Input vocale: un esempio di interfaccia

di Raffaello De Masi

Parlare di interfaccia vocale fa prevedere immediatamente grossi problemi. Delle difficoltà di una macchina nel gestire input vocali e, più genericamente, sonori, abbiamo parlato a più riprese nelle pagine di MCmicrocomputer a proposito di intelligenza artificiale. Vedemmo, allora, come alla facile e relativamente ben realizzabile fase di output sonoro, corrispondesse il controaltare dell'input che si presentava complesso e ben difficile da realizzare in maniera valida ed efficiente

Che l'input vocale potesse essere la strada maestra per la gestione di handicap legati a disfunzioni motorie e addirittura della stessa parola apparve subito chiaro. Una persona dalle ridotte possibilità manipolative che, anche attraverso device dedicati, come tastiere a sfioramento, a pozzetto, a soffio o a membrana, mostrava ancora gravi problemi di adattamento all'uso di una macchina di comunicazione, nella fattispecie un computer, sarebbe divenuta, attraverso tecnologie di input vocale, dal parlato al sillabato fino al grunt converting, estremamente più efficiente nella gestione dei contatti e delle comunicazioni con l'esterno.

Così, al di fuori della pura ricerca nel campo della A.I., ricerca peraltro indispensabile per la messa a punto, almeno, degli algoritmi di gestione dell'input sonoro stesso, ricercatori «commerciali» cominciarono ad esplorare nuove strade per la gestione corretta della ca-

tena suono-calcolatore-effetto. Mentre l'interfacciamento tra la seconda e la terza fase era piuttosto semplice, restava quello della prima, che per una serie di motivi, permaneva molto complessa.

Riducendo il problema all'osso (chi ha interesse allo studio specifico del problema può far riferimento ai passati articoli nella rubrica di A.I. ed alla relativa bibliografia che, all'epoca, pubblicammo), potremmo dire che esistono tre difficoltà principali alla gestione dell'input vocale stesso: la generalizzazione del timbro vocale, la gestione delle frasi e la gestione del rumore di fondo.

Vediamo i problemi un poco più da vicino. Il primo è forse il problema di più complessa soluzione; sebbene la stessa frase, detta da Pino Locchi e da Tina Lattanzi (per citare due voci famosissime e immediatamente riconoscibili, appartenenti a due doppiatori che hanno prestato la voce, rispettivamente, ad HAL, il computer di 2001 Odissea nello spa-



Voice Navigator II

Costruttore:

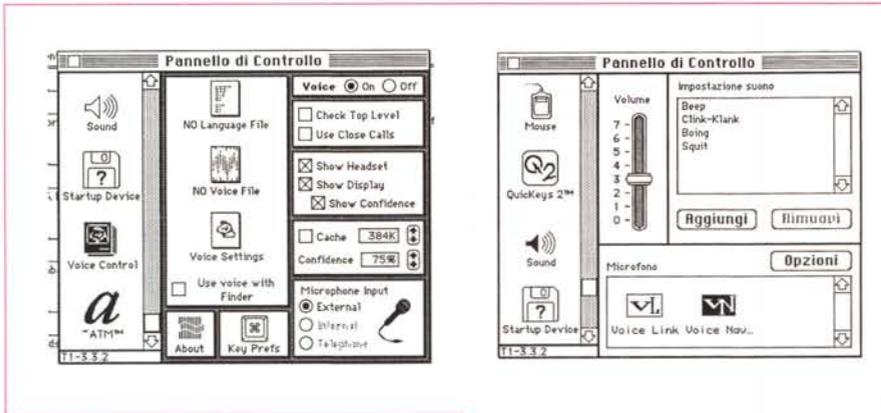
Articulate Systems Inc.
Orriston, Mass. USA.

Distributore:

Modo srl
Via Masaccio, 11 - 42100 Reggio Emilia
Tel.: 0522/512828

Prezzo:

L. 1.450.000



L'uso dei cdev-INIT per il setup del package; interessante la combinazione qualità di suono-rapporto di compressione, che permette di inserire input ben articolati senza occupare eccessiva memoria.

zio ed a personaggi come Greta Garbo e Katharine Hepburn) significhi la medesima cosa, per una macchina, almeno all'attuale stato della conoscenza, è esattamente il contrario. Non esisterebbe cioè implementatore di software dedicato al riconoscimento capace di costruire un programma che permetta alla macchina di identificare le due voci.

Secondo problema, non meno trascurabile, è la gestione delle frasi: pochi di noi si rendono conto che, nel parlare, le parole non sono staccate l'una dall'altra. Quando chiediamo «Metti lo zucchero nel caffè», indipendentemente dal significato della frase stessa, generalmente diciamo «Mettilozuccheronelcaffè» senza alcuna separazione tra le parole. Allo stato attuale della tecnica (e probabilmente chissà per quanto tempo ancora) non esiste macchina capace di eseguire la scansione della frase per separare i suoni di senso finito, il tutto in tempo ragionevolmente breve. Così la comprensione del parlato si può ridurre a quella di singole parole, o, al più, alla analisi di una frase, pronunciata in maniera irrealistica (e non sempre facile) scandendone le parti una per una.

Terzo problema, il rumore. Facciamo un esempio, forse banale, ma che si adatta bene al nostro caso. Ascoltando musica d'orchestra, è possibile, con un poco d'attenzione, separare e «ascoltare» solo un certo tipo di strumento, ad esempio, i violini o il pianoforte. La stessa cosa avviene quando parliamo in un ambiente gremito di persone; anch'esse discutenti tra di loro. Se ci concentriamo in maniera attenta sul nostro interlocutore e su quanto egli dice, il nostro cervello ha l'incommensurabile potenza di «separare» il discorso del no-



stro partner da quello dal rumore di fondo. Probabilmente perderemo qualche parola, ma qui l'intuito permetterà di ricostruire i particolari mancanti ed è probabile che la comprensione sia se non totale, almeno adeguata.

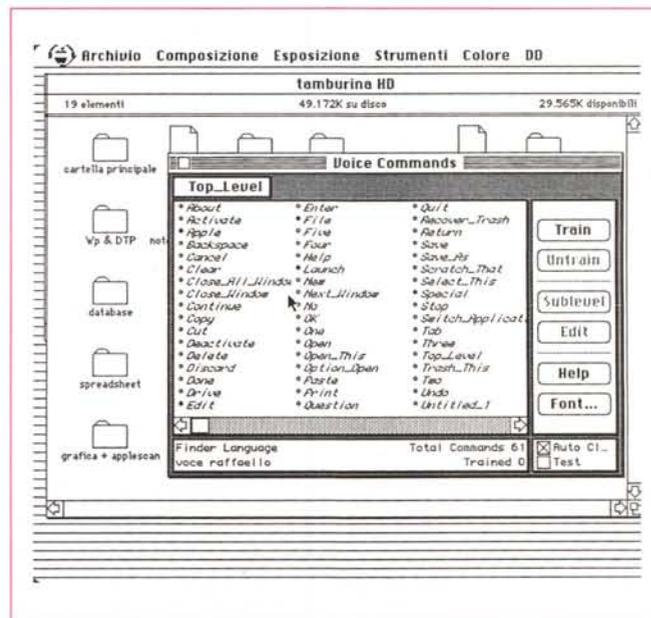
Per una macchina, quello che giunge alla periferica di input (verosimilmente un microfono) è tutto «materiale utile», essa non possiede metro né tecniche

per discernere «l'interlocutore» e così non può fare altro che analizzare tutto quanto, in termini di suono, giunge ad essa. E così, i primi esperimenti di conversazione tra macchina e persona avvenivano sempre con lo stesso sperimentatore, usando una tecnica di conversazione un tantino innaturale, il tutto in una camera silente e anecoica.

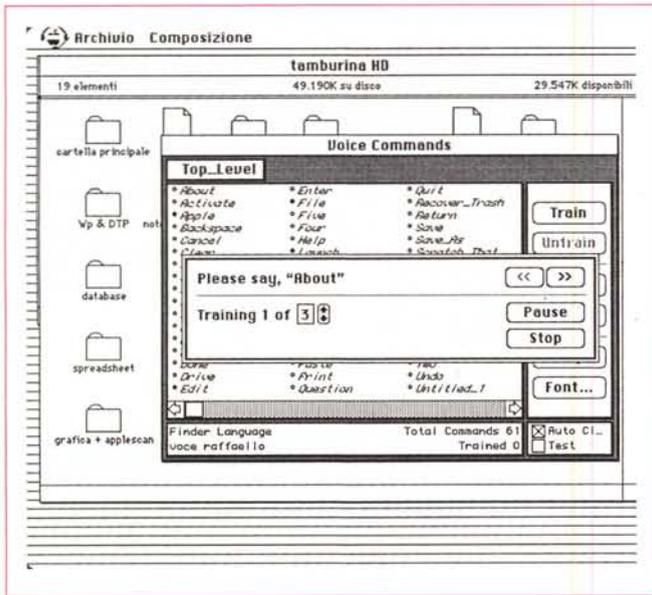
Questa tecnica che, sebbene primordiale, risale a non più di qualche anno fa, è stata poi perfezionata e raffinata per renderla più elastica ed efficiente. Lo sviluppo è stato articolato in due direzioni. Da una parte, la creazione di software di gestione sempre più raffinato ed efficiente, dall'altro la specializzazione di hardware dedicato per sopperire, in maniera adeguata, alle problematiche di base appena esposte. L'attrezzatura della quale vi parliamo in questa puntata è una adeguata ed efficiente proposta di soluzione che abbraccia l'una e l'altra tematica.

Il Voice Navigator II

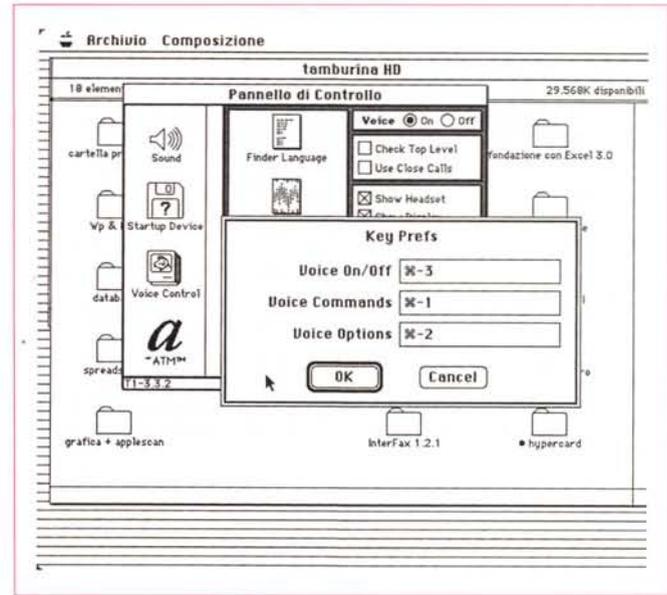
Qualche mese fa, proprio in questa rubrica, ho avuto modo di parlare di un progetto di informatica specializzata, dedicata al recupero di portatori di handicap, e realizzata nell'ambito della provincia di Salerno. La struttura fu voluta dall'ALIAS, che, su mia consulenza, commissionò e realizzò un'aula sperimentale, a Salerno, destinata alla riabilitazione e all'inserimento professionale di suoi pazienti, e alla sperimentazione, appunto, di attrezzature per l' almeno parziale recupero di portatori gravi di handicap, in cui le tecniche di riabilitazione tradi-



La fase di istruzione dei comandi vocali. Nell'esempio abbiamo tutte le chiamate del Finder, ma è possibile aggiungere comandi personalizzati o addirittura macro.



La vera e propria fase di input vocale. Attraverso una semplice finestra è possibile assegnare messaggi a un ordine, in un numero praticamente illimitato di forme.



Gli shortcut per la chiamata dei menu principali di VN. È possibile comunque avere lo stesso risultato con comandi diretti, a voce.

zionali avessero dato risultati modesti o addirittura nulli.

L'aula progettata e realizzata tutta su macchine Macintosh, si basava su 16 posti di lavoro, collegati in rete AppleTalk, e gestiti da un host che faceva capo al docente. La realizzazione del progetto di base fu, ovviamente, semplice; i problemi cominciarono a presentarsi quando si dovette affrontare la gestione, appunto, degli handicap più particolari.

Alcuni contatti con riviste americane mi permisero di intravedere una possibilità di soluzione al problema; così sono venuto in contatto con una serie di istituzioni, peraltro tecnologicamente estremamente avanzate, che mi hanno messo a disposizione materiale e risultati sperimentali di notevole valore. Ebbi così la possibilità di accedere a diverse periferiche e a programmi specializzati che mi permettono oggi, di poter affrontare con risultati affidabili la maggior parte degli handicap con cui, nella mia attività di consulente, mi trovo quotidianamente a contatto.

E così ritorniamo all'argomento di questa puntata. Un gruppo numeroso di disabili aveva problemi motori, di coordinamento muscolare e, comunque, di accuratezza di manipolazione tali da impedire l'utilizzazione di periferiche di input (tastiera, mouse) originali o opportunamente modificati. Inoltre l'accesso al mondo esterno, e alle sue attrezzature, dal telefono al televisore alla semplice apertura della finestra erano virtualmente preclusi, salvo la costante

presenza di un assistente. Ero addirittura in presenza di disabili che, accanto al limitato o assente uso degli arti aveva anche gravi problemi di fonazione, tanto da poter emettere solo una serie di suoni gutturali, ancorché articolati e differenziati.

Così, nella mia ricerca e nei numerosi contatti avuti con produttori e rivenditori degli Stati Uniti ho potuto acquistare e sperimentare la macchina oggetto di questa prova (regolarmente distribuita in Italia anche dalla Modo di Reggio Emilia), che ha il vantaggio e la caratteristica di poter essere adattato, proprio perché si basa su tecniche di input assai poco specializzate, a un numero abbastanza esteso di problemi.

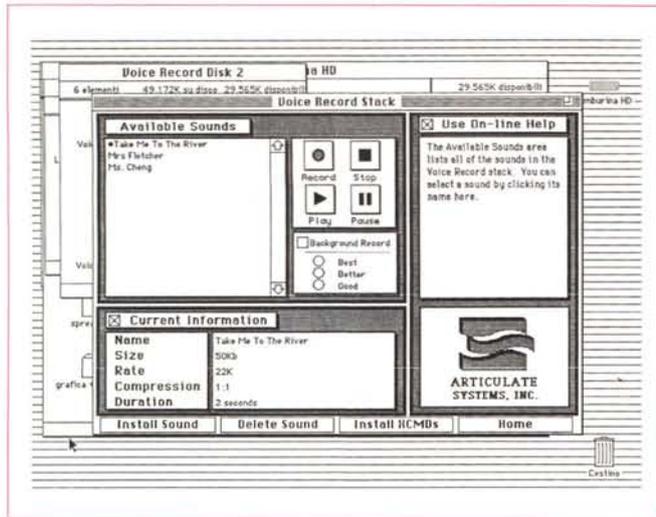
Dopo questa prolusione tanto lunga, è ovvio che Voice Navigator non poteva essere che un tool di input vocale. Si tratta, infatti, di una attrezzatura rappresentata da un campionatore, corredato di microfono, e da un software specializzato destinato alla gestione dell'interfaccia. Il package è rappresentato da una parte hardware e una nutrita quantità di software.

L'hardware, custodito in una grossa scatola di cartone e polistirolo espanso contiene il campionatore vero e proprio (il Voice Navigator II, tanto per intenderci esiste anche il Voice Navigator, tout court, pacchetto precedente a questo, di più ridotte prestazioni, ma che è ancora in produzione e vendita), di un microfono dedicato (il DeskTop Mike, ma può essere utilizzato qualunque microfono direzionale a basso fruscio, anche

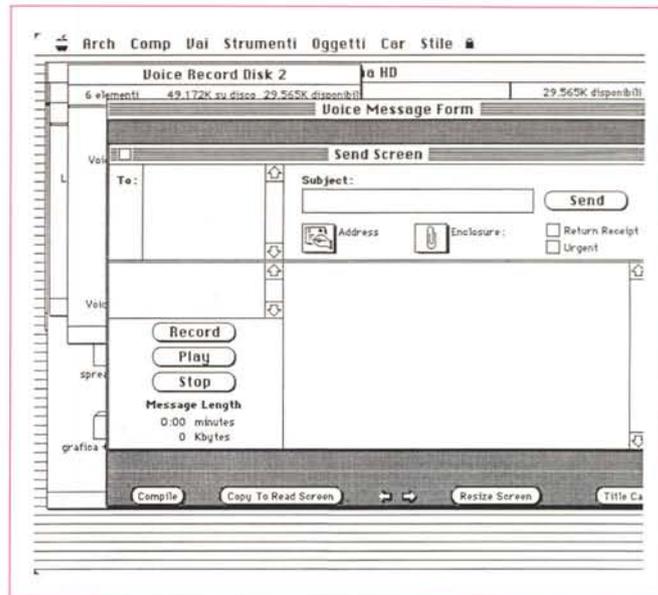
se la cosa può ritenersi per lo meno superflua), di una serie di cavi di connessione e di basette per l'hardware, basi che permettono di disporre le attrezzature secondo le specifiche esigenze, e di alcuni grossi manuali accompagnati da ben sei dischetti di software. Come al solito non mancano una miriade di fogli volanti aggiuntivi, i read me dell'ultima ora, le schede di registrazione hardware e software e una serie di cartoncini, etichette, adesivi da attaccare un po' dappertutto, visto che probabilmente Articulate Systems, la casa produttrice dell'apparecchiatura, è quella che a mio avviso fornisce un supporto all'utente dei più raffinati efficienti e specializzati (non a caso la garanzia su tutto il fornito, l'assistenza e gli upgrade sono assicurati a vita).

Il Voice Navigator vero e proprio è rappresentato da un parallelepipedo di plastica nera, dal design elegante e raffinato, dotato di una serie di tasti e selettori rossi, delle dimensioni approssimative di una piccola agenda, ma sorprendentemente pesante. Ad esso si abbina il Mike, piccolo (ma estremamente selettivo) microfono montato su una lunga e sottile base snodabile e supportato direttamente dal Navigator o da una basetta cilindrica di ghisa.

Il software fornito in corredo è strutturato in due blocchi diversi rispettivamente in due e quattro dischetti. I primi due si riferiscono alla sezione Voice Record, gli altri quattro al Voice Control (e contengono il vero e proprio programma Voice Navigator). Vediamo come at-



Gli stack Hypercard forniti di supporto, contenenti una serie di utility destinate a rendere più semplice la fase di input.



traverso di essi si passa alla gestione tramite voce del computer.

Le due parti del pacchetto

L'uso di Voice Navigator è basato su due fasi, peraltro abbastanza ovvie: il Voice Record e il Voice Control. Ambedue sono applicazioni software destinate a gestire l'interfaccia e a pilotare l'hardware dedicato.

Il primo è un blocco di software che permette di registrare, editare, testare, cancellare e conservare suoni, rumori, frasi, messaggi o tutto quanto altro udibile da orecchio umano. Per essere precisi tutte le operazioni descritte sono riferibili a file Audio-IFF, Audio-IFF Compressi, e FSSD; inoltre Voice Record può accedere a risorse di tipo 'snd' (ricordiamo che una risorsa 'snd' è semplicemente un metodo di rappresentare un suono all'interno di un altro file senza modificare il contenuto del file stesso). L'uso dei due dischi è regolato da un installer, che gestisce il Voice Record Control Panel, un cdev di scrivania abbinato a un DA, e il software di compressione (MACE- Macintosh Audio Compression-Expansion). Il secondo disco contiene SoundWave, applicazione ben nota per registrare e manipolare suoni, due mail form dedicate rispettivamente a Microsoft Mail e a CE Quick-Mail, e il Voice Record Help, dall'ovvio uso e significato.

L'uso e lo scopo di questa parte di software è abbastanza ovvia. È il blocco che permette di registrare i suoni in arrivo, di campionarli secondo la frequenza desiderata, di eventualmente modifi-

carli. Si tratta, manco a dirlo, di un package estremamente sofisticato, che addirittura si autoconfigura a seconda del System presente; esso è praticamente tutto rappresentato da cdev e INIT, per cui è assolutamente trasparente all'utente.

Nella versione 6.07 o 7 il software è pienamente compatibile, anche se non supporta le utility proprie di questa versione (in ogni caso Articulate System ha già promesso un upgrade, ovviamente gratuito!) il cdev di autoinserisce nell'INIT SOUND del pannello di controllo. Schiacciando questo compare la finestra di configurazione che permette di scegliere l'hardware di input (Voice Navigator o Voice Link) la qualità di registrazione (che è, ovviamente proporzionale al consumo di memoria), il rapporto di compressione, e il guadagno (a proposito di qualità di registrazione, che è direttamente legata alla frequenza di campionamento, abbiamo notato che, anche scegliendo quella più bassa, ben difficilmente si sono verificati errori di interpretazione del messaggio; probabilmente la qualità maggiore è riservata all'uso del Voice Navigator come grunt-converter, dove occorre un più preciso controllo dell'input vocale da analizzare).

Eseguito il setup tutto è pronto per la registrazione dei messaggi che, come abbiamo già detto, possono essere analizzati poi attraverso la finestra propria di Sound (praticamente simile a quella che pilota il microfonino delle macchine LC e SI più recenti) o, in maniera più raffinata (e più complessa) attraverso l'applicazione SoundWave, ben nota nel mondo Macintosh.

Macchina-mondo esterno-messaggio vocale

Eseguita l'installazione del Voice Record, l'applicazione per l'input nudo e crudo, passiamo adesso all'operazione più raffinata, pieghiamo cioè la macchina ai nostri voleri «parlati». Premettiamo che la cosa può essere fatta sia attraverso Voice Navigator II (che è l'hardware che proviamo) sia attraverso il modello I, sia ancora utilizzando Voice Link (tutte e tre le attrezzature, con prestazioni via via decrescenti, sono prodotte dalla Articulate System). C'è da dire, ancora, che la fase di input vocale vera e propria può avvenire anche attraverso il microfono di serie negli LC-SI, ma le prove da me effettuate hanno dato risultati non sempre buoni; inoltre il microfono fornito da Articulate System è eccezionalmente direzionale, di elevata qualità e molto più facile da posizionare e maneggiare (Articulate Systems fornisce, addirittura, in alternativa, una cuffia e un microfono trasmittente a spilla, per chi desidera utilizzare la tecnica anche allontanandosi dal posto di lavoro).

Anche qui la maggior parte del software, di eccezionale complessità e qualità, è quasi tutto guidato da cdev. La prima cosa da fare è quella, ovviamente, della istruzione. Sebbene, infatti il pacchetto utilizzi tecniche proprie di MacinTalk, il package vocale proprietà Apple, l'interfaccia vocale stessa è assolutamente svincolata dalla pronuncia specifica della lingua inglese, cosa che ha reso virtualmente inutile MacinTalk in Italia. In altri termini, al contrario di

alcuni altri pacchetti che adottano una pronuncia inglese per la gestione dei comandi, e che quindi divengono virtualmente inutili in altri idiomi, Navigator adotta suoni, frasi, rumori «imparati» direttamente dall'utente. Questo, abbinato alla elevata definizione di campionamento dell'input stesso, permette di personalizzare i comandi in funzione del singolo utente. Così diversi operatori allo stesso computer adotteranno librerie di comandi diversi, personalizzati, e, per questo più facili da tenere a mente.

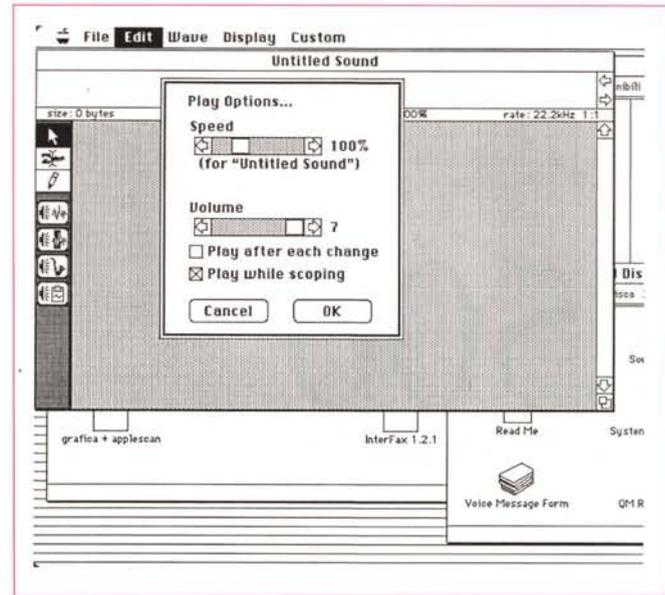
Adottiamo, quindi, la soluzione più sofisticata e vediamo come la macchina diviene docile ai comandi attraverso Voice Navigator. La prima cosa da fare è l'installazione di tutto il software, che consta di ben 7 cdev-INIT, di un file di help estremamente efficiente, di una applicazione, e di alcuni stack Hypercard, con comandi già predefiniti.

La cosa più importante è rappresentata dalle librerie: la libreria è uno dei due pilastri fondamentali su cui si basa l'uso del pacchetto; l'altra è la collezione di suoni-frasi-messaggi alla libreria collegati. Una piccola pausa per capire come funziona il tutto.

Basiamoci sulla finestra del Finder, all'accensione: il menu del finder ha una serie di comandi, ad esempio Taglia, Copia, Incolla, Archivio, Nuova Cartella, Colore, Esposizione, e così via, comandi che possono essere principali o gerarchici. Bene, una libreria comandi è tutto l'elenco dei comandi eseguibili raggruppati insieme.

Invochiamo Voice Navigator (la prima volta, alla installazione, attraverso una shortcut; dalla volta successiva sarà sufficiente dare un ordine attraverso il microfono; in default, comunque, Voice è sempre attivo). Selezioniamo la libreria dei comandi del Finder (solo la prima volta, le successive l'apertura avvorrà automaticamente); ci verranno mostrati tutti i comandi (la libreria mostra i comandi del System USA, tipo ShutDown o NewFolder, ma si possono agevolmente cambiare da finestra tutti gli ordini) e clicchiamo uno di essi; si apre una finestra che chiede di pronunciare la frase-parola-messaggio corrispondente all'ordine; in default si possono abbinare tre input allo stesso ordine, ma il numero può essere modificato. Perché più input? Semplice, per dare ordini con tono di voce diverso o per dare diversi messaggi (ad esempio Taglia e Cut) allo stesso ordine. Il software, con molta discrezione ci avvisa se non ha capito bene l'ordine e ci chiede di ripetere. Ovviamente tutti gli ordini possono essere modificati, sostituiti o riascoltati (nel caso, probabile le prime volte, in cui non ci ricordiamo più come abbiamo dato

Una fase di utilizzo dell'applicazione Sound-Wave, fornito a complemento, e destinata alla manipolazione e alla modifica dei messaggi stessi.



l'ordine). Tutti i comandi vanno inseriti in una libreria personale, vale a dire in un file che contiene la «voce» del singolo utente. Appare così chiara la funzione delle librerie (ne vengono fornite una cinquantina, relative alle più diffuse applicazioni, da MacWrite, a Word, a File-Maker, a FreeHand e così via). La cosa interessante è che una serie di comandi, comuni a tutte le applicazioni, come Stampa, Taglia, Chiudi, Salva e così via, possono essere inseriti in un file «globale», comune a diverse (o a tutte) le applicazioni.

Non solo, ma si possono aggiungere, a qualunque libreria, comandi specializzati, personali, se ne possono escludere altri, si può addirittura riscrivere una libreria, si può abbinare ad un comando la scrittura di una frase, una macro, una sequenza preregistrata attraverso MacroMaker o QuickKeys 2, l'adozione di uno stile o di una gerarchia, la creazione di un grafico, la stessa definizione di un disegno, addirittura lo scrolling controllato o il drag sullo schermo. È possibile, anche, per un istruttore con un poco di buona volontà, accedere alla programmazione aggiuntiva stessa del pacchetto che, basata essenzialmente sulla adozione e sull'aggiunta di risorse, permette di costruire applicazioni specifiche per il singolo handicap. L'unico limite alle opzioni è veramente la operatività e la fantasia dell'operatore.

Conclusioni

Il package ha bisogno di una fase di training e di istruzione tanto più lunga quanto più sofisticazione si chiede all'u-

so dell'interfaccia. Fortunatamente questa specializzazione può essere raggiunta per gradi successivi, e la estrema (praticamente illimitata) possibilità di personalizzazione permettono di costruire interfacce vocali estremamente efficienti e self-tailored. Si tenga conto della adattabilità del pacchetto come «grunt converter», nel cui campo di applicazione Voice Navigator oscura tutti i concorrenti. E si tenga a mente la possibilità di far «rispondere» a voce il Macintosh ai nostri comandi.

Tanto per spingere le cose più lontano, e per risolvere un problema di un paziente AIAS affetto da quadriplegia, ho interfacciato il Mac, così comandato, a una serie di servizi dell'abitazione del paziente stesso. Ho costruito così una applicazione (utilizzando lo ZBasic Zedcor integrato con Resource Maker, fornito con lo stesso pacchetto, che mi ha risolto la creazione delle risorse stesse) che permetteva al paziente (che circolava per casa, essendo collegato alla macchina da un microfono trasmettitore indipendente) di eseguire chiamate telefoniche (attraverso HyperCard e un modem Interfax, o, ancora meglio utilizzando l'eccellente HyperTel provato sulle pagine Macintosh quattro mesi fa) di aprire la porta e le finestre, di accendere il televisore, e così via. L'applicazione, che ho realizzato in meno di tre mesi, ha permesso di risolvere problemi logistici eccezionali, e può essere considerata la punta di diamante di infiniti altri campi di applicazione, ben più semplici da realizzare e alla portata di qualunque assistente o terapeuta.

MB

475.000
workstation e server SUN installati

3.500
applicazioni disponibili

1
posizione di SUN nelle classifiche
di settore (IDC)

9.800.000
prezzo d'ingresso della famiglia SPARC

34
produttori di sistemi SPARC

TUTTI I NUMERI DI UN SUCCESSO

Numeri, non parole. Numeri che ha solo SUN. Perché solo SUN ha scelto di dedicarsi completamente alle workstation e ai server UNIX/RISC. Scegliere però non basta. Bisogna avere i numeri, per conquistare il successo. E il numero di computer SUN installati - o, perché no, quello delle applicazioni disponibili - dimostra che SUN ha fatto centro... e con lei chi l'ha scelta.

039.60551 SUN Italia: ancora un numero per darvi tutte le informazioni che volete.



sun
microsystems

SUN MICROSYSTEMS ITALIA SpA - Centro Colleoni Andromeda 1
via Paracelso, 16 - 20041 Agrate Brianza (MI)

MILANO

PADOVA

ROMA

TORINO