

TU PARLI il pc esegue

I sistemi di riconoscimento vocale sono sempre più evoluti e precisi. E presto potrebbero sostituire mouse e tastiera, ma anche il telecomando tv

DI FEDERICO GUERRINI

L'esempio più evidente è Siri, l'assistente vocale di Apple. Tu parli allo smart phone, chiedendo indicazioni sulle condizioni meteorologiche o sui film in programmazione, e lui ti risponde, azzeccando in molti casi anche le sfumature e il contesto.

È una delle funzioni più apprezzate dell'iPhone 4S, quella che ha mandato in visibilibio all'inizio di quest'anno i fan della Mela nel corso della presentazione al Ces, la fiera dell'elettronica di Las Vegas. Ma Siri, che ha avuto il merito di sdoganare e rendere familiare al vasto pubblico la tecnologia del riconoscimento vocale, non è che la punta dell'iceberg di una rivoluzione, quella del riconoscimento vocale, destinata a modificare a breve termine il modo con cui interagiamo con gli oggetti.

«Siri ha avuto un impatto enorme per il cognome che si portava dietro, quello di Apple», dice Giuseppe Riccardi, a capo del Signals and Interactive Systems Lab dell'Università di Trento, «ma in realtà non è un'invenzione nuova. Molti non sanno che già da cinque o sei anni è possibile parlare col proprio telefonino e chiedere ad esempio le indicazioni per raggiungere un certo locale o negozio».

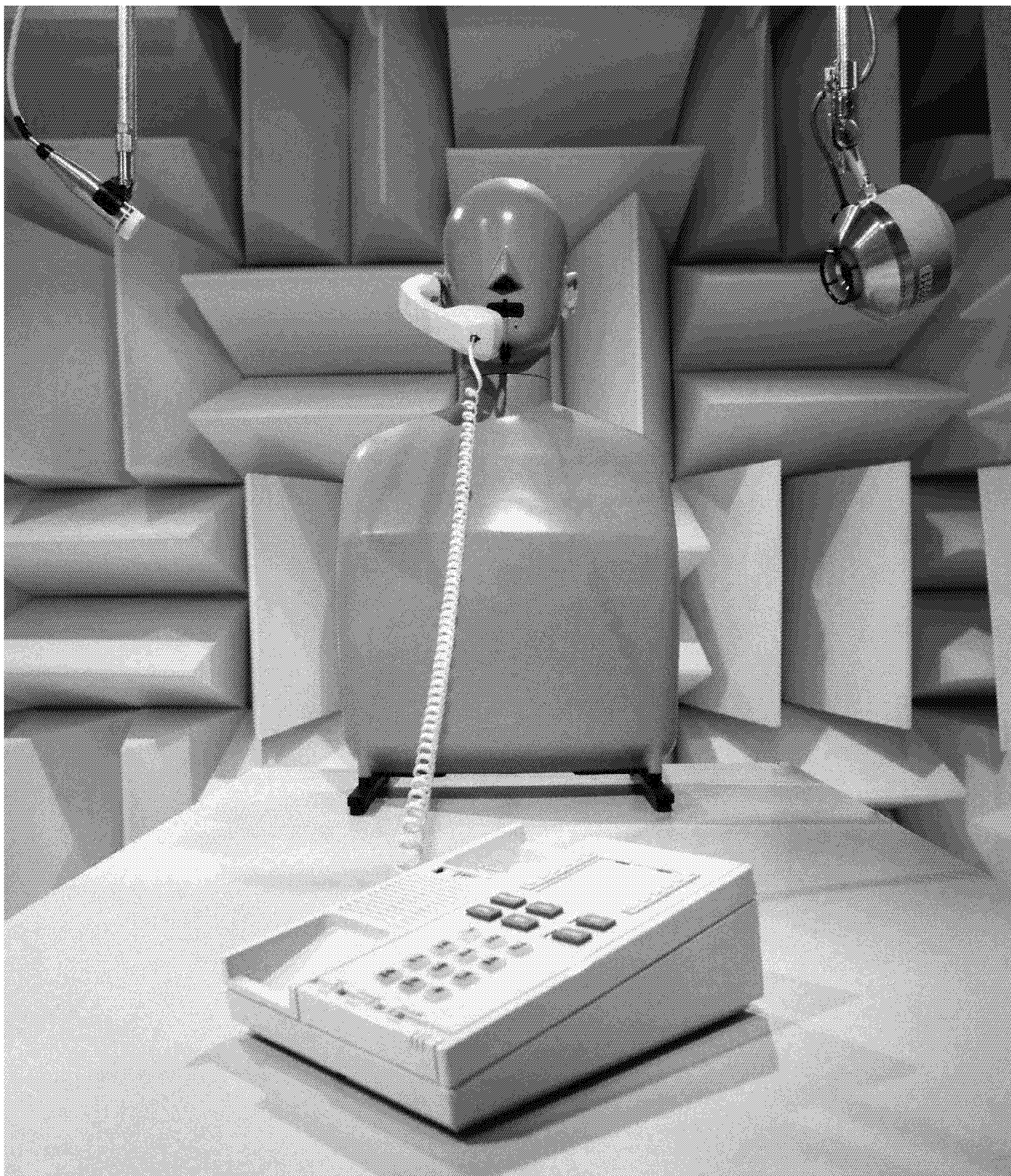
Non solo Apple, ma anche Google che integra la ricerca vocale e le cosiddette "azioni vocali" negli smartphone con sistema operativo Android e Microsoft che include la propria tecnologia Tell Me nei Windows Phones, effettuano da tempo ricerche in questo campo. E in realtà il mondo delle tecnologie di interazione verbale ha radici antiche: già nel 1962, alla fiera mondiale di Seattle gli ingegneri di Ibm mostrarono al pubblico Shoebox, un rudimentale elaboratore in grado di effettuare calcoli matematici sulla base di input vocali. Negli anni Novanta furono sviluppati diversi sistemi per aiutare persone con difficoltà motorie a comandare un Pc attraverso un microfono. Le prime applicazioni commerciali di questo tipo di tecnologia iniziarono verso la metà di quel decennio, con alcuni esperimenti nei call center. Oggi la compagnia aerea American Airlines adopera Wally, un assi-

I risponditori automatici non ci chiederanno più di premere un tasto, ma di spiegargli il problema

stente vocale per gestire le chiamate dei clienti. Lanciato a luglio 2011, Wally nei primi cinque mesi di utilizzo ha ridotto i tempi di risposta ed è apprezzato dai clienti per la capacità di fornire sua sponte informazioni aggiuntive, sulla base dei dati custoditi nel proprio database. Di avvisare, per esempio, il cliente se sta raggiungendo un certo saldo punti sulla tessera di "frequent flyer" o di proporre una migliore combinazione di voli, se chi chiama percorre spesso lo stesso tragitto.

Ed è solo l'inizio. Uno dei settori su cui stanno lavorando i ricercatori è quello dell'utilizzo della tecnologia di riconoscimento del parlato per analizzare le relazioni fra le persone. «Le macchine», spiega Riccardi, «non dovranno per forza giocare un ruolo attivo; potranno restare in secondo piano, osservare l'interazione in corso e se necessario prendere provvedimenti sulla base delle sfumature emozionali captate». Nel caso di un incendio, mettiamo, comprendere il grado di panico o sicurezza delle persone coinvolte ed emettere messaggi rassicuranti oppure, tornando ai call center, individuare sfumature di irritazione o rabbia nella voce di un cliente e allertare un responsabile in carne e ossa affinché prenda in mano la situazione. ▶

Foto: L. Pishoyev - Corbis



Per il mondo a caccia di accenti

Il lavoro di Linne Ha - percorrere il mondo alla ricerca di campioni esotici da portare a casa e catalogare - ricorda un po' quello di un entomologo. Solo che invece di farfalle, Linne (che lavora per Google) raccoglie voci. Destinazione dei reperti: il programma Voice Search dell'azienda di Mountain View, con cui è possibile interrogare il motore di ricerca senza bisogno di tastiera. La tecnologia di riconoscimento vocale infatti si basa su un sistema statistico e, per capire le parole, il sistema mette a confronto l'input parlato con un modello statistico specifico per la lingua selezionata. Di solito per costruire il modello si usano campioni sonori presi da aziende specializzate, che però sono privi di accenti. Di qui l'idea di registrare le voci di gruppi di utenti locali. Loro parlano, poi Ha e il suo team raccolgono gli audio. Per questo li chiamano "voice hunters", cacciatori di voci.

Wally e Siri hanno qualcosa in comune: la tecnologia che li rende così efficienti è opera di Nuance, l'azienda del Massachusetts leader mondiale nel campo del riconoscimento vocale, nota principalmente per il software di dettatura Dragon Naturally Speaking, ma le cui ricerche spaziano nei campi più svariati, dalla sanità all'intrattenimento. In collaborazione con Ibm, ad esempio, Nuance

sta lavorando a un progetto per sfruttare il supercomputer Watson per fornire diagnosi mediche accurate per i pazienti sulla base dei dati forniti a voce dai medici. Il progetto, per cui sono stati stanziati 25 milioni di dollari, è ancora agli inizi, ma già oggi l'azienda commercializza un'applicazione in grado di trascrivere le note dettate a voce da specialisti e quindi interfacciarsi con le cartel-

le cliniche elettroniche dei pazienti e individuare eventuali criticità: per esempio dei farmaci che, combinati insieme, possono provocare effetti collaterali indesiderati. Nuance inoltre commercializza prodotti come Dragon Tv, una tivù interattiva in cui i canali si cambiano a voce e Dragon Go!, un'app per iPhone e Android che consente di effettuare operazioni di vario tipo: da prenotare un ta-

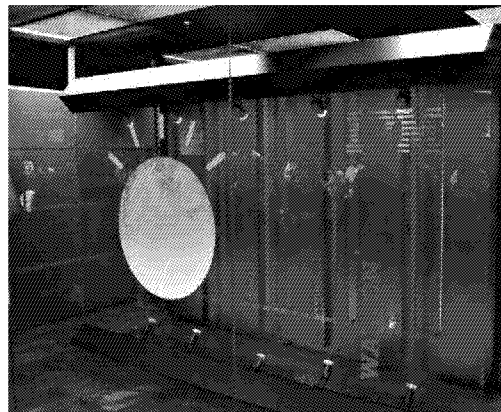


Foto: G. Halli - Corbis

volo a un ristorante o un biglietto per un concerto, solo col pronunciare un semplice comando.

Non c'è settore o quasi in cui le tecnologie sviluppate dall'azienda americana o dai suoi concorrenti non abbiano ipotizzato di sostituire, almeno parzialmente, i dispositivi di input manuale con il parlato: dalle macchine per caffè alle Smart Tv (produttori come Lg, Samsung e Panasonic hanno già presentato i loro "schermi intelligenti"), dalle auto, dove la voce serve ad attivare funzioni di infotainment (accendere lo stereo, controllare il navigatore) ai frigoriferi con cui dialogare per regolare la temperatura o ordinare la spesa via Internet, alla casa in cui basta pronunciare un comando per alzare o abbassare le tapparelle, accendere la luce e far scorrere l'acqua del rubinetto.

Tuttavia, anche il riconoscimento vocale ha i suoi limiti. «Ci sono molte situazioni», dice a "l'Espresso" il pro-

fessor Jim Glass, direttore del Spoken Language System Groups del Massachusetts Institute of Technology, «in cui il parlato può essere la modalità adeguata, quella più naturale, che ti consente di compiere determinate azioni lasciando liberi gli

occhi e le mani, ma anche molte in cui non è ottimale: non inserirei la mia password dicendola a voce alta, né farei uso del parlato per comandare il pc nel corso di lezioni in una classe affollata; quella vocale dovrà rimanere sempre una modalità opzionale, da associare a quella manuale».

Ad ogni modo, si va verso un'epoca popolata da "badanti" robotici in grado di ascoltare e comprendere qualunque nostra conversazione e agire di conseguenza per soddisfare le nostre esigenze. Ma, la privacy? «Finché i dati ven-



UN APPARECCHIO DELLA AMERICAN AIRLINES. NELL'ALTRA PAGINA: IL SUPERCOMPUTER WATSON

gono immagazzinati in locale», continua il professor Glass del Mit, «i potenziali rischi sono minori. Altrimenti dipende dalla consapevolezza e dall'atteggiamento dell'utente. Già oggi, con Siri molti non sono consapevoli del fatto che ogni parola pronunciata viene trasferita e conservata nella nuvola. O se se ne accorgono, ciò non gli impedisce di continuare a usare il dispositivo. Ma, a mano a mano che crescerà attorno a noi il numero di dispositivi in ascolto, è un problema di cui la società dovrà di certo tenere conto». ■