

Interazioni uomo-macchina via voce

Stato dell'arte, applicazioni e tendenze delle tecnologie vocali in ambito business

di Giancarlo Magnaghi Maggio 2010



Le applicazioni vocali, che appartengono alla grande famiglia delle tecnologie di trattamento e riconoscimento automatico dei dati e dei segnali (Automatic Identification and Data Capture - AIDC) hanno l'obiettivo di semplificare le interazioni uomo-macchina.

Le tecnologie relative al trattamento della voce (voice processing) sono diffuse in vari settori di applicazione, come servizi CRM (call center, help desk, operatori telefonici automatici), marketing e vendite (telemarketing, teleselling e indagini di mercato), phone banking (quotazioni, brokeraggio e home banking) e servizi pubblici come il televoto o il pagamento elettronico.

Altre aree di interesse crescente sono il voice picking nei magazzini e le altre attività lavorative in cui è utile operare a mani libere (affiancando le operazioni che vengono compiute con i terminali portatili), le attività di conversione automatica da testo a voce (text-to-speech) e di riconoscimento vocale (speech recognition) sia all'interno delle aziende che sul campo (Field Force Automation - FFA). Queste tecnologie trovano applicazione anche nel campo delle telecomunicazioni (menu e portali vocali, lettura di fax e e-mail, caselle vocali, messaggistica unificata, comando vocale di telefonini e console giochi), nei trasporti (sistemi di comando vocali interattivi negli autoveicoli, sistemi per annunci automatici nei mezzi pubblici, negli aeroporti e nelle stazioni) e negli uffici (dettatura automatica, sistemi di lettura vocale dei documenti). Esistono anche nicchie applicative particolarmente significative, come le interfacce vocali per non-vedenti e ipo-vedenti e le applicazioni in campo medico (dettatura automatica e gestione di prescrizioni mediche, referti, ecc).

Le tecnologie di base

Le tecnologie di voice processing non sono da confondere con la semplice registrazione e riproduzione digitale di parlato e di suoni, come nel caso dei registratori digitali, dei riproduttori MP3, delle audio guide e degli audiolibri. Infatti, il cuore di queste tecnologie sono sofisticati algoritmi matematici di riconoscimento vocale (Automatic Speech Recognition - ASR), verifica sintattica del linguaggio (con utilizzo di grammatiche), comprensione semantica, sintesi del linguaggio partendo dal testo o selezione e composizione intelligente di testi pre-registrati. La base del riconoscimento vocale è il "fonema", ovvero il suono singolo che si ottiene dalla scomposizione delle parole. Ogni lingua ha il proprio set di fonemi e la tecnologia di base del riconoscimento vocale riguarda il riconoscimento dei fonemi. Il sistema software che riconosce i fonemi è detto "motore vocale" (vocal engine). I requisiti di potenza e di memoria delle macchine hardware utilizzate dai motori vocali sono diversi se si devono riconoscere

poche parole o un intero vocabolario, poiché il tempo di elaborazione varia sensibilmente in funzione della lista di parole (vocabolario) che il sistema deve riconoscere. I motori vocali più completi sono quelli utilizzati per le applicazioni di riconoscimento del parlato continuo (continuous speech recognition): i cosiddetti programmi di dettatura o "dittafoni". Per il riconoscimento vocale di poche parole, tipico delle interfacce vocali dei sistemi operativi e delle applicazioni di voice picking, sono sufficienti versioni semplificate dei motori vocali, con vocabolari di poche decine di parole. I motori vocali possono essere "speaker dependent" o "speaker independent". Se si utilizza la modalità speaker dependent, si deve effettuare una profilazione vocale degli utenti, che richiede una fase di apprendimento da parte del sistema. Nella modalità speaker independent, tale profilazione non è necessaria. La profilazione vocale è necessaria soprattutto nei contesti difficili e rumorosi, dove le parole da utilizzare sono poche, e dove pochi minuti di training vengono compensati da migliaia di ore di lavoro senza intoppi. Diversi linguaggi per applicazioni vocali sono stati sviluppati a partire dal 1995 presso AT&T e altre aziende come IBM, Lucent e Motorola, che si riunirono nel 1998 nel VoiceXML Forum (www.voicexml.org) e svilupparono VXML (Voice eXtensible Markup Language): si tratta del linguaggio standard per la creazione di servizi vocali nel mondo web che rappresenta un elemento fondamentale per favorire lo sviluppo di servizi interattivi sulla sintesi e il riconoscimento vocale (dialogo sintetizzato, audio digitalizzato, portali vocali) ed è divenuto il linguaggio comune agli sviluppatori dei servizi, ai creatori degli strumenti di sviluppo e ai produttori di piattaforme e portali vocali. VXML, derivato da XML e simile agli altri linguaggi web, fu riconosciuto nel 2000 dal consorzio W3C (www.w3.org/TR/voicexml) e poi dalla maggioranza delle imprese operanti nel settore ICT (le aziende aderenti al Voice XML Forum sono oggi oltre 350).

Picking vocale

Uno dei principali settori d'applicazione verticale dei sistemi di riconoscimento automatico della voce riguarda le già citate applicazioni di magazzino, dove viene richiesto l'intervento dell'operatore "a mani libere" in alternativa o in affiancamento all'impiego di lettori di codici a barre / RFID e terminali palmari industriali. Nelle applicazioni di prelievo (picking) vocale, gli operatori che lavorano con i sistemi vocali non hanno una lista stampata degli articoli da prelevare a magazzino ma utilizzano il riconoscimento e la sintesi della voce come mezzo di trasmissione delle informazioni. L'operatore è dotato di un terminale dotato di radio, che tiene agganciato alla cintura o in tasca, di una cuffia e di un microfono. L'obiettivo di base del picking vocale è quello di usare la voce come mezzo di comunicazione bidirezionale fra la procedura di gestione operativa del magazzino (Warehouse Management System - WMS) e l'operatore. I messaggi che l'applicazione invia all'operatore (speech out), sostituiscono o affiancano le istruzioni che compaiono sullo schermo del terminale, mentre delle parole pronunciate dall'operatore (speech in) sostituiscono tastiera e scanner. In questo modo, l'operatore può concentrarsi maggiormente sul proprio compito e lavorare in modo più comodo, poiché può usare entrambe le mani e lavorare con i guanti, cosa nient'affatto secondaria in ambienti freddi, come i magazzini di surgelati; inoltre sbaglia molto meno, perché il sistema vocale lo guida passo-passo. Infine, viene ridotta la documentazione cartacea relativa a liste di picking e ad altri moduli. Il settore di applicazione della tecnologia vocale nella logistica non si limita al picking e alla preparazione degli ordini, ma si può estendere anche alle operazioni di inventario o di stock replenishment. I sistemi vocali per la logistica si basano su due filoni ben distinti: terminali standard con opzione voce o terminali dedicati alla sola applicazione vocale. Diversi come prestazioni e come caratteristiche hardware, si devono scegliere in

base alle effettive esigenze. Scegliendo i terminali multimodali si evita l'acquisto di prodotti dedicati e si può ammortizzare più rapidamente il costo di acquisizione dei terminali ripartendolo su differenti applicazioni vocali e non vocali. I terminali dedicati sono generalmente superiori in termini di caratteristiche hardware, come la robustezza complessiva e la qualità delle cuffie/auricolari e del segnale audio. L'attuale tendenza del mercato è verso i terminali di nuova generazione che permettono un funzionamento multimodale: per la voce o per altre modalità di acquisizione dei dati, tanto che si parla di Voice-Assisted Work e la voce è intesa come una modalità di raccolta dati e di controllo delle operazioni che può costituire un valido supporto per molte applicazioni. Per quanto riguarda il software e la connettività alla rete, il traffico generato dai sistemi vocali è molto limitato, poiché viaggiano via radio e in rete solo i caratteri relativi alle transazioni effettuate, mentre tutta l'attività di riconoscimento vocale e di signal processing è fatta sul terminale. Per ottenere prestazioni soddisfacenti, è però indispensabile calibrare il sistema nel suo complesso, poiché il sistema vocale deve essere "sincrono" e deve rispondere in tempo reale, altrimenti diventa un collo di bottiglia. Anche la copertura Wi-Fi per la connessione in rete deve essere di buona qualità. Una rete Wi-Fi sufficiente per un sistema basato su liste di picking trasmesse in modalità batch, può rivelarsi insufficiente per un sistema vocale, che richiede un funzionamento sincro, anziché asincro. I sistemi di picking vocale sono ormai maturi sul piano tecnologico, ma la loro implementazione richiede una revisione dei processi di picking per conseguire l'operatività ottimale. I miglioramenti si vedono in due aree: un netto miglioramento della produttività intesa come quantità di prodotti prelevati in un dato intervallo di tempo (oltre il 15%), e una notevole riduzione dei prelievi errati (i sistemi di riconoscimento vocale permettono di garantire una precisione dei prelievi tendente al 99,99% ,rispetto al 98% dei sistemi tradizionali). Bisogna inoltre considerare la riduzione dei costi di stampa dei documenti cartacei di picking, la possibilità per l'operatore, grazie alle mani libere, di concentrarsi di più sui propri compiti; il feed-back in tempo reale, che permette una gestione proattiva, l'aggiornamento degli stock in tempo reale e la riduzione dei tempi di formazione.

Il futuro

In futuro, le tecnologie vocali saranno sempre più presenti in tutte le interfacce utente sia dei dispositivi locali (comando vocale di elettrodomestici, telefoni e altri terminali, sistemi di domotica) sia delle reti telefoniche e Internet. Nei computer tablet e negli altri terminali evoluti, in futuro saranno disponibili tre tipi di interfacce naturali: touch screen, interfaccia vocale e riconoscimento della scrittura manuale (handwriting), e si diffonderanno le "applicazioni multimodali", in grado di gestire più modalità di input (touch screen, voce, tastiera, mouse, keypad, penna) e di output (parlato sintetizzato, testo, immagini, video), che permetteranno di utilizzare a piacere la voce, il tatto o la vista per interagire con l'ecosistema tecnologico circostante.

Alcune aziende attive nel campo delle tecnologie vocali

Questa tabella, che non ha alcuna pretesa di completezza, elenca in ordine alfabetico alcune aziende attive nel campo delle tecnologie vocali. Non sono citate le società che producono sistemi telefonici e di networking, che utilizzano tutte le tecnologie vocali nei loro sistemi IVR, ACD, caselle vocali, ecc.

Apple (www.apple.com)	Apple (www.apple.com) I sistemi operativi per Mac sono dotati di programmi di gestione dei comandi vocali. Apple ha incluso alcune funzioni di ricerca vocale in iPhone e iPod.
Aton (www.aton.eu)	Aton produce la soluzione onVoice, che permette una semplice integrazione con ERP o WMS, grazie all'architettura modulare della piattaforma M3, Aton Mobility Management, su cui è basato onVoice. Applicazioni nei magazzini del settore food.
Di.Tech (www.ditechonline.it)	Di.Tech, da oltre vent'anni fra i leader in Italia nei progetti logistici per la grande distribuzione, è attiva nelle soluzioni vocali dal 2005 e produce il sistema Di.Vo. (Di.Tech Voice).
Google (www.google.com)	Google Voice è un servizio di telefonia VoIP offerto da Google (disponibile per smartphone Android e su altre piattaforme) che permette di trasferire le chiamate e i messaggi di testo da un numero virtuale a diversi apparecchi telefonici, fissi o mobili, accedere online a un'unica casella vocale, ottenere via sms la trascrizione della chiamata. L'estensione vocale di Google Chrome, permette agli utenti di chiamare i numeri di telefono direttamente dal proprio browser.
Hi Pro Solutions (www.hipro.it)	Hi Pro Solutions sviluppa soluzioni applicative utilizzando tecnologie di identificazione automatica come barcode, RFID e voce. Opera in tre settori convergenti e complementari: raccolta dei dati, sicurezza, automazione della forza lavoro.
IBM (www.ibm.com)	Ibm è una veterana nel campo vocale e ha sviluppato la tecnologia ViaVoice. Alcuni esempi di prodotti: Ibm WebSphere Voice Server; Ibm DirectTalk Speech Recognition per Aix permette di gestire le applicazioni tramite comandi vocali; Ibm DirectTalk Text-to-Speech per Aix consente di trasformare il testo scritto in discorso. Ibm DirectTalk Beans per Java permette di creare applicazioni call; Eclipse Voice Tools è uno strumento di sviluppo di applicazioni Vocali open source basate sullo standard VoiceXML.
Lexter (www.lexter.it)	System Integrator di sistemi logistici. Realizza varie soluzioni basate su terminali indossabili per la lettura di codici a barre e vocali.
Loquendo (www.loquendo.com)	Società del gruppo Telecom Italia, Loquendo è uno dei leader internazionali delle tecnologie vocali, grazie alle tecnologie di sintesi vocale (Loquendo TTS), riconoscimento automatico del parlato (Loquendo ASR), verifica e identificazione del parlatore (Loquendo SV), disponibili in 25 lingue, e alle piattaforme Speech Server e VoxNauta.
Microsoft (www.microsoft.com)	I sistemi operativi Windows sono dotati di programmi di gestione dei comandi vocali. La tecnologia Microsoft Speech Recognition contenuta nel frame work .Net permette di aggiungere le funzioni di riconoscimento vocale a un sito web. Il Microsoft Office Communications Server gestisce voce e unified messaging
Nuance (www.nuance.com)	Nuance è la più nota società che opera nel campo delle tecnologie vocali. Produce software vocali per tutte le applicazioni: le soluzioni TTS (Text To Speech) Vocalizer, disponibili per 50 lingue, i prodotti di riconoscimento vocale ASR (Automatic Speech Recognizer), le soluzioni professionali e per il magazzino (motore vocale VoCon 3200), applicazioni in campo medico e automobilistico, i programmi più diffusi di dettatura automatica Dragon Naturally Speaking per Windows e MacSpeech Dictate per Mac.
Psion Teklogix (www.psionteklogix.com)	Psion Teklogix offre soluzioni vocali complete, basate su terminali vocali indossabili, auricolari, software, integrazione e servizi professionali che consentono l'attivazione vocale di sistemi di gestione del magazzino e l'accesso in tempo reale alle applicazioni.
Vocollect (www.vocollect.com)	Società specializzata nella produzione di sistemi logistici (hardware e software) con tecnologie vocali. Vocollect Voice è un vocal engine disponibile su hardware di diversi produttori (Intermec, LXE, Motorola, Psion Teklogix). VoiceArtisan è uno strumento per la produzione di soluzioni software che integrano Vocollect Voice con diversi sistemi host.
Zetes (www.zetes.it)	Zetes è un integratore paneuropeo di soluzioni di raccolta dati, basate su varie tecnologie nell'ambito dell'architettura 3iV Crystal.