

# PARLARE CON LE MACCHINE

Feedback: redazione.cbritaly@matedizioni.it

DI GIANCARLO MAGNAGHI

Nuance Communications è uno dei principali fornitori di soluzioni vocali e di imaging per i mercati business e consumer di tutto il mondo. L'azienda, che ha la sede principale a Burlington (Massachusetts), conta circa 6000 dipendenti a livello mondiale, di cui 800 in Europa, e uffici nelle maggiori città Europee di cui uno in Italia, vicino a Milano. Il fatturato a livello mondiale nell'anno fiscale 2009 è stato di un miliardo di dollari, con una crescita del 25% rispetto all'anno precedente.

Nuance opera soprattutto nelle forniture Business-to-Business, ha grandi clienti come gestori di telecomunicazioni, produttori di apparati telefonici, di telefonini e di automobili. I principali prodotti per il mercato consumer sono quelli per la dettatura (Dragon) e per il riconoscimento di testi (Omnipage).

Per conoscere meglio questa azienda importante in un settore in rapida crescita, abbiamo intervistato **ALAN RANGER, DIRETTORE MARKETING EMEA** (Europa, Medio Oriente e Africa).

*Qual è la missione di Nuance e in quali mercati opera?*

La nostra missione è fornire interfacce naturali tra persone e macchine/sistemi di comunicazione, attraverso il trattamento automatico della voce – l'interfaccia più naturale – e dei testi. La maggior parte dei

call center che utilizzano il riconoscimento vocale a livello evoluto lo fanno con i nostri prodotti. Per esempio, le sole società Fortune 500 hanno più di 4.000 call center in tutto il mondo basati sulle nostre soluzioni. Quasi tutti i principali produttori di automobili utilizzano le nostre tecnologie per i sistemi di bordo (navigazione, viva-voce, comandi vocali, intrattenimento). Questo sta diventando un mercato molto grande, grazie alla diffusione dei navigatori e dei sistemi cellulari di bordo, che non devono distrarre il conducente dalla guida. I comandi vocali aumentano la sicurezza poiché consentono di non staccare gli occhi dalla strada e le mani dal volante, inoltre le nostre tecnologie TTS (*Text-to-Speech*) consentono di leggere i messaggi di posta elettronica e gli SMS, mentre con le funzioni ASR (*Automatic Speech Recognition*) si possono dettare email e SMS. I nostri prodotti software per il riconoscimento vocale e per la digitazione predittiva T9 (che permette di aumentare la velocità di digitazione grazie ai dizionari disponibili per varie decine di lingue), sono quelli più diffusi a livello mondiale, poiché sono presenti in oltre 4 miliardi di telefoni cellulari.

Negli smart phone, come iPhone Apple, oltre a digitare le mail utilizzando il touchscreen, è possibile dettarle grazie alle funzioni di riconoscimento vocale e ascoltare i messaggi in arrivo tramite le funzioni

**Intervista  
ad Alan Ranger,  
Direttore Marketing  
EMEA di Nuance**

TTS. Una particolare nicchia applicativa è la dettatura medica, supportata dal prodotto Dragon Medical, utilizzato per esempio dai radiologi per dettare i referti relativi a radiografie, TAC e altri strumenti di *medical imaging*. Queste applicazioni sono già diffuse negli ospedali degli Stati Uniti e iniziano a diffondersi anche in Europa, in particolare in Inghilterra. Altre aree in cui operiamo sono il riconoscimento delle immagini (*image recognition*) e i programmi OCR (*Optical Character Reading*).

### **Quali sono le principali tecnologie ASR?**

Ci sono due diversi tipi di prodotti per il riconoscimento del discorso. Quelli dotati di un vocabolario limitato (come nei menu vocali dei call center) che non richiedono apprendimento, e i prodotti che devono riconoscere il parlato continuo, con dizionari di migliaia di termini, che richiedono una fase di apprendimento per migliorare la precisione. Il sistema comunque apprende con l'uso, impara a conoscere lo speaker e diventa sempre più preciso. Per esempio, i sistemi voice-to-text utilizzati dagli operatori di telefonia mobile e nelle applicazioni di unified messaging per convertire i messaggi vocali in messaggi testuali, memorizzano e affinano automaticamente i profili vocali degli utenti e funzionano sempre meglio.

### **Quale hardware è necessario per supportare le applicazioni vocali?**

Dal punto di vista tecnico, le funzioni vocali più semplici sono gestite dai microprocessori e dai DSP (*Digital Signal Processor*) dei terminali telefonici e delle schede audio dei PC, mentre per le funzioni ASR più complesse svolte dalle compagnie telefoniche e dai call center si

utilizzano server dedicati alle funzioni di voice processing (*voice server*). Per esempio, nella dettatura su iPhone, la voce viene digitalizzata e pacchettizzata su iPhone e inviata a un server che provvede al riconoscimento e alla conversione in testo.

### **È possibile mischiare lingue diverse nello stesso discorso e avere un'interpretazione semantica dei discorsi?**

Sicuramente. Per esempio in Svizzera, dove ci sono tre lingue ufficiali, abbiamo sistemi che permettono di mischiare in tutti i modi queste tre lingue, capiscono in quale lingua sono poste domande e rispondono nella medesima lingua. Nei call center, i nostri agenti virtuali sono in grado di comprendere domande poste in formato libero dagli utenti, come "vorrei parlare con il signor Bianchi", evitando di percorrere menu vocali con decine di scelte e di livelli, pongono a loro volta domande se sono necessarie precisazioni e rispondono a voce, esattamente come un operatore umano.

### **Com'è organizzata la ricerca in questo campo così specialistico?**

La maggior parte della ricerca viene svolta al nostro interno. Nuance impiega oltre mille scienziati specializzati nei vari aspetti delle tecnologie vocali, e abbiamo alcuni centri di ricerca nelle città universitarie per facilitare l'acquisizione di giovani talenti. Per esempio, abbiamo un team ad Aachen (Aquisgrana), in Germania, che lavora con i principali produttori di automobili per ottimizzare il filtraggio acustico dei sistemi di bordo, e abbiamo una sede anche a Cambridge.



Alan Ranger, Nuance

Altri team sono dedicati al riconoscimento automatico del discorso e delle immagini.

### **Quali sono i principali vantaggi ottenibili con le tecnologie e le applicazioni vocali?**

I vantaggi ottenibili utilizzando un'interfaccia vocale sono in termini di produttività (per esempio la velocità di dettatura è tripla di quella di scrittura), facilità d'uso, diminuzione degli errori, miglioramento della user experience e della sicurezza sul lavoro. L'utilizzo della voce in automobile aumenta notevolmente la sicurezza della guida. Nelle applicazioni al servizio dei consumatori, un call center con cui è possibile parlare in modo naturale riduce fino al 40% il numero di chiamate in strada- te in modo non corretto.

### **Quali sono le principali tendenze e gli sviluppi previsti?**

Ci sono molte aree di sviluppo e

## TESTIMONIANZE

stanno nascendo molte nuove applicazioni. Uno dei problemi più complessi è il filtraggio per eliminare i rumori di fondo dalla voce. Stiamo lavorando con i principali produttori di veicoli per migliorare le tecnologie di filtraggio acustico dei sistemi di bordo.

Intensi sviluppi anche nelle applicazioni mobili, come la compilazione di form con la voce e l'integrazione della voce con le interfacce touch screen di tablet, PDA e smart phone. In espansione anche le applicazioni biometriche che si basano sul riconoscimento dell'impronta vocale (*voice print*) dello speaker, come *voice security* (l'auto che "ubbidisce" solo alla voce del proprietario, o la verifica dell'identità del chiamante al telefono) e *voice verification* (riconoscimento dello speaker attraverso la pronuncia di una parola o frase stabilita), spesso in accoppiamento con password e codici segreti.

I sistemi di *natural language call steering* consentono un dialogo telefonico automatico completo e interattivo tra l'interlocutore al telefono e il sistema, rendendo possibili applicazioni specializzate come l'assistenza telefonica automatizzata ai clienti (*over the phone customer care*). Per esempio, Vodafone Spagna ha implementato un sistema che, quando un cliente chiama, invia un menu di scelta direttamente sul telefonino, che il cliente utilizza per navigare nell'applicazione e trovare la soluzione al suo problema. Stiamo andando verso un blend di modalità di comunicazione tramite smart phone: voce, touch-screen, applicazioni interattive intelligenti. Un ulteriore trend è quello della personalizzazione della voce dei sistemi text-to-speech (*voice parametrics*), già diffusa negli Stati Uniti (ad esempio Bank of Ameri-



---

***La missione di Nuance è fornire interfacce naturali tra persone e macchine/sistemi di comunicazione, attraverso il trattamento automatico della voce e dei testi***

---

ca), che sta iniziando a diffondersi anche sul mercato europeo.

Poiché per l'immagine aziendale è importante il modo di parlare dei sistemi vocali automatici, Nuance lavora con alcune tra le maggiori aziende per creare voci personalizzate che vengono utilizzate nei sistemi vocali automatizzati. Le tecnologie vocali di Nuance supportano più di 50 lingue e accenti regionali. La ricerca "The voice for your Brand", recentemente commissionata da Nuance e svolta su un campione italiano di 1000 utenti, ha evidenziato l'importanza della voce per l'immagine aziendale, dove l'utilizzo della voce rappresenta valore aggiunto dei call center e identifica l'azienda negli spot pubblicitari televisivi o radiofonici, nonché l'impatto che la voce può avere sulla percezione del brand da parte dei clienti, poiché è

nella natura umana reagire in maniera differente a diversi tipi di voci, in base al timbro, al tono e all'accento. Dalla ricerca sono emerse alcune caratteristiche che rendono certe voci più gradevoli di altre. Per esempio gli accenti più graditi sono quello toscano (42%) e romagnolo (32%), che risultano piacevoli ma anche maggiormente comprensibili. Si nota una leggera preferenza verso una voce femminile quando si parla con un call center, mentre una voce maschile è più convincente in occasione di un reclamo. Quasi metà del campione (43%) si dichiara più disposto ad ascoltare un'offerta se posta con voce suadente. Nuance è in grado di individuare, insieme ai clienti, un adeguato profilo vocale del proprio brand e di realizzare una voce sintetica rispondente alle caratteristiche e alla personalità dell'azienda. **B**