Assistenti vocali:

l'Intelligenza Artificiale a portata di voce

Paolo **Casagranda**, Francesco **Russo**, Raffaele **Teraoni Prioletti**
Rai - Centro Ricerche, Innovazione Tecnologica e Sperimentazione

In questo articolo viene introdotto il concetto di assistente digitale a controllo vocale, chiarendo la situazione del mercato ed i principi generali di funzionamento.

Un assistente vocale è un programma che, opportunamente addestrato, può dialogare con interlocutori umani grazie alla capacità di riconoscere, sintetizzare ed elaborare il linguaggio naturale dei comandi vocali. Gli assistenti vocali intelligenti sono ormai pervasivi ed estremamente rilevanti per la radio, e permettono un utilizzo immediato dell'Intelligenza Artificiale sotto diversi aspetti: riconoscimento e sintesi vocale, riconoscimento della richiesta, attuazione della risposta.

Per meglio comprenderne le applicazioni, al CRITS è stato costruito un prototipo per servizi radiofonici evoluti che è stato valutato anche da un gruppo di utenti. Lo studio ha confermato la rilevanza degli assistenti vocali per la radio, trovandone applicazioni possibili e accennando alcuni fondamentali requisiti sui contenuti.

Gli *assistenti vocali* sono tra i protagonisti della recente evoluzione tecnologica e stanno raggiungendo rapidamente molti ambiti della nostra vita. Un assistente vocale è un programma che, opportunamente addestrato, può dialogare con interlocutori umani grazie alla capacità di riconoscere, sintetizzare ed elaborare il linguaggio naturale dei comandi vocali. Spesso l'assistente vocale è identificato con lo *smart speaker*, il dispositivo fisico che più comunemente lo ospita. Grazie al supporto dell'interfaccia vocale, il mercato degli smart speaker è in rapida crescita e anche le più recenti analisi di mercato condotte negli USA all'inizio del 2020 confermano questo trend [1]. Basti pensare che negli USA il 24% degli americani dai 18 anni in su (60 milioni circa) possiede già uno smart speaker e il 54% ha usato qualche tipo di tecnologia a comando vocale, come gli assistenti vocali su smartphone, smart speaker o altri dispositivi. A partire dal 2014, anno in cui **Amazon** ha introdotto il primo dispositivo, molte altre aziende sono entrate nel mercato con i propri smart speaker e il tasso di adozione di questa tecnologia è risultato più veloce rispetto a qualsiasi altro dispositivo di consumo [2].

Al di là dell'utilizzo degli smart speaker come semplici altoparlanti, la funzionalità principale che questi dispositivi forniscono è l'*Intelligenza Artificiale* (IA). È infatti grazie alle tecnologie di IA che l'assistente vocale costruisce prima un modello per la comprensione del linguaggio parlato e dopo elabora una risposta adeguata, rendendo possibile il dialogo tra utente e macchina.

201_026_3.00

È proprio per evidenziare questa capacità che gli assistenti vocali vengono anche chiamati *assistenti vocali intelligenti*. I dati mostrano, inoltre, che assistenti vocali e smart speaker stanno diventando un veicolo importante per la fruizione della radio e dell'audio in genere, in aggiunta e spesso in sostituzione alle vecchie abitudini di ascolto. Questo scenario li rende estremamente rilevanti per la radiofonia creando sia opportunità che sfide per il mondo dei broadcaster radiofonici.

L'articolo è organizzato nel modo seguente: ad una panoramica dell'attuale offerta di mercato e dei principali utilizzi degli assistenti vocali, seguirà un'introduzione tecnica al loro funzionamento. L'analisi si concluderà con la descrizione delle funzionalità più rilevanti di un prototipo implementato al **CRITS Rai** per valutare le potenzialità di questa tecnologia in ambito radiofonico.

Panoramica dell'offerta di mercato

Comunemente l'assistente vocale è integrato in dispositivi che possono essere prodotti anche da aziende diverse da quelle dello stesso assistente vocale, come smart speaker, telefoni, tablet, PC, TV nonché negli abitacoli di alcune automobili.

Le maggiori aziende che hanno implementato un assistente vocale sono **Amazon** con *Alexa*, **Google** con *Google Assistant*, **Baidu** con *Duer*, **Alibaba** con *Genie*, **Xiaomi** con *Xiao*, **Apple** con *Siri*, **Samsung** con *Bixby*, **Microsoft** con *Cortana* e **Huawei** con *Celia*.

Oltre ai prodotti commerciali sono disponibili progetti open source per l'implementazione degli assistenti vocali. Le tecnologie open source offrono opzioni flessibili a startup e sviluppatori per sperimentare e costruire prodotti orientati al proprio settore di interesse. *Mycroft*, *Kalliope*, *Open Assistant*, *Jasper* e *Leon* sono esempi di assistenti vocali open source.

L'aumento delle vendite di smart speaker degli ultimi anni è da attribuire principalmente al successo degli assistenti vocali, di cui gli smart speaker rappresentano un comodo strumento di accesso. Da alcuni mesi è iniziata anche la vendita di dispositivi con display, i cosiddetti *smart display*, che rendono la risposta dell'assistente vocale più completa grazie al supporto dello schermo (es. *Amazon Echo Show* e *Google Nest Hub* (Fig. 1)) e, in alcuni casi, permettono di effettuare videochiamate.

Col passare del tempo sempre più brand ed aziende hanno lanciato la propria proposta di dispositivi smart con integrazione di uno degli assistenti vocali esistenti, contribuendo a diversificare il mercato. Attualmente, aziende come **JBL** e **Sony** hanno indirizzato la propria scelta verso *Google Assistant* mentre **Yamaha**, **Ultimate Ears** e **Harmand Kardon** hanno optato per *Alexa*. **Marshall**, **Bose** e **Polk** hanno scelto di produrre dispositivi con *Google Assistant* o con *Alexa* mentre **Sonos** offre la possibilità di utilizzare più assistenti vocali nello stesso dispositivo.

Secondo il report di **Canalys** del febbraio 2020 [3], il mercato globale ha avuto una crescita del *52%* nel Q4 del 2019 con *49,2 milioni* di unità vendute. I cinque maggiori produttori di assistenti vocali digitali mondiali, ordinati per numero di unità vendute nel Q4 del 2019, sono stati **Amazon** con *15,6 milioni* di vendite, **Google** con *12,5*, **Baidu** con *5,7*, **Alibaba** con *5,6* e **Xiaomi** con *4,6*. **Apple** si è fermata su un numero di vendite inferiore, in parte dovuto alla differente clientela di riferimento a cui punta l'azienda.

Fig. 1 – *Rai Radio 2* su *Google Nest Hub*

Fig. 2 – Vendite globali di smart speaker e crescita annuale nel 2019 secondo *Canalys* *(fonte [3])*

Spedizioni e crescita annuale degli smart speaker in tutto il mondo
Impulso del mercato degli smart speaker per Canalys: 2019

Venditore	Spedizioni 2019 (milioni)	Quota mercato 2019	Spedizioni 2018 (milioni)	Quota mercato 2018	Crescita annuale
Amazon	37.3	29.9%	24.2	31.1%	+54%
Google	23.8	19.1%	23.4	30.0%	+2%
Baidu	17.3	13.9%	3.6	4.6%	+384%
Alibaba	16.8	13.5%	8.9	11.4%	+89%
Xiaomi	14.1	11.3%	7.1	9.1%	+97%
Altri	15.4	12.3%	10.8	13.8%	+43%
Totale	124.6	100.0%	78.0	100.0%	+60%

Nota: le percentuali potrebbero non arrivare al 100% a causa dell'arrotondamento
Fonte: Analisi degli smart speaker di Canalys (spedizioni), Febbraio 2020

Per l'intero anno 2019 sono stati venduti un totale di circa *125 milioni* di smart speaker con un aumento del *60%* rispetto al 2018 (Fig. 2). Le vendite in Cina sono più che raddoppiate in un anno grazie a **Baidu**, **Alibaba** e **Xiaomi**. In particolare, **Baidu** ha ottenuto un'impressionante crescita nelle vendite passando da *3,6 milioni* nel 2018 a *17,3 milioni* nel 2019.

Alcune aziende hanno adattato gli assistenti vocali per l'utilizzo in contesti specifici, ad esempio alberghi, settore ospedaliero e automotive. Nel settore ricettivo, ad esempio, si stanno implementando attività per migliorare i servizi di concierge, facilitare la riproduzione di musica, il controllo della temperatura o l'illuminazione della camera, la ricerca di servizi locali e persino il check-out (ad es. **Amazon** e **Alibaba**) [4]. Un altro settore da tenere in considerazione è quello ospedaliero dove gli assistenti vocali possono essere utili a pazienti, infermieri e medici. Un rapporto di **IHS** (*Information Handling Services*) descrive l'uso di smart speaker negli ospedali: i comandi vocali possono essere impiegati per controllare televisori e apparati posti nelle stanze dei pazienti, nonché per inoltrare richieste verso i dispositivi mobili utilizzati da medici e infermieri. Un esempio di tale uso è una piattaforma basata su *Alexa* che da febbraio 2019 è impiegata in un progetto pilota presso l'ospedale **Cedars-Sinai Medical Center** di West Hollywood, in California [4]. Inoltre, si ritiene che gli smart speaker possano essere di grande aiuto a persone ipovedenti (almeno 2,2 miliardi di individui a livello globale secondo l'**Organizzazione Mondiale della Sanità**) e agli anziani, rendendo molto più semplice l'interazione con la tecnologia.

L'ultimo settore che citiamo è l'automotive, che sta manifestando un grandissimo interesse nell'implementazione dell'assistente vocale all'interno dei veicoli. *Alexa*, l'assistente vocale di **Amazon**, è già presente nell'abitacolo di alcuni modelli di auto, tra cui Toyota, Audi, Ford [5]. General Motors introdurrà *Android Automotive* di **Google** nei suoi veicoli dal 2021 come già fatto dall'alleanza Renault-Nissan-Mitsubishi, da Fiat Chrysler Automobiles e da Volvo [6]. Attualmente i principali produttori di automobili offrono modelli che già supportano *Apple CarPlay* o hanno in programma di introdurlo [7]. **Alibaba** ha annunciato di aver siglato un accordo, per i veicoli destinati al mercato cinese, con Audi, Renault, Volkswagen e Honda [8]. **Baidu** ha firmato un accordo con Ford, BMW e Volkswagen che entrano a far parte di un consorzio di più di 130 aziende globali per collaborare ad *Apollo*, la piattaforma di tecnologie di *guida autonoma open source* [9]. La piattaforma *Houndify*, fornita da **SoundHound**, è stata utilizzata per lo sviluppo di un proprio assistente vocale da Mercedes, Kia, Honda, Hunday e PSA Group [10]. Infine, altre case automobilistiche come General Motors, FCA, Toyota, Ford, Audi e BMW utilizzeranno, su alcuni veicoli, la tecnologia di **Cerence** che ha da poco presentato il *Cognitive Arbitrator*, un sistema sviluppato per consentire l'utilizzo di più assistenti vocali all'interno dell'auto [11].

Come funziona un assistente vocale

La grande sfida dell'assistenza vocale è l'utilizzo del linguaggio naturale per l'accesso a differenti servizi, tra cui l'ascolto della radio in streaming. L'elaborazione di una richiesta utente prevede l'esecuzione di due fasi logicamente distinte: la *gestione dell'input dell'utente* e la *produzione di un risultato* da restituire come risposta. Di seguito descriveremo l'architettura di funzionamento generale degli assistenti vocali facendo riferimento alle due piattaforme utilizzate per i test: *Google* e *Amazon*.

Il successo di un'applicazione, qualunque sia il dispositivo su cui deve funzionare, è fortemente influenzato dalla sua facilità di utilizzo. Risulta fondamentale, quindi, riuscire a creare un'interfaccia vocale che sia semplice e intuitiva. I comandi vocali, prima di poter essere eseguiti, devono essere compresi e interpretati. La gestione dell'input dell'utente è sicuramente uno dei processi più complessi a causa della varietà e imprevedibilità con cui è possibile fare le richieste. L'uso dell'Intelligenza Artificiale per assolvere a tale scopo risulta fondamentale. I produttori di assistenti vocali mettono a disposizione degli sviluppatori gli strumenti utili per creare applicazioni specifiche di terze parti per il proprio assistente vocale (es. *Google Dialogflow* e *Amazon Alexa Skill Kit*). Tutta la logica usata per l'elaborazione delle richieste si trova, in genere, nel *cloud* proprietario.

In **Google** il fulcro di un'applicazione (*Action*) è la costruzione del *Dialogflow agent*, un agente virtuale incaricato di gestire le conversazioni con l'utente finale e responsabile del riconoscimento della richiesta dell'utente. Per espletare la sua funzione, l'agente virtuale utilizza un modello di comprensione del linguaggio naturale costruito con l'ausilio di avanzate tecniche di *Machine Learning* (**ML**) e *Natural Language Understanding* (**NLU**). La prima operazione che viene eseguita su una nuova richiesta è la trascrizione del parlato in testo attraverso tecniche di *riconoscimento vocale automatico* (**ASR**). L'agente virtuale applica, quindi, il modello costruito al testo della richiesta con lo scopo di comprendere le espressioni dell'utente, associarle agli *intent* (azioni che soddisfano la richiesta dell'utente), estrarre i parametri utili. L'addestramento dell'agente virtuale è fondamentale nella definizione di una *Action*. L'addestramento avviene attraverso un insieme di frasi di training scelte dal progettista che devono essere di qualità e quantità sufficienti da permettere la costruzione di un modello efficace per il riconoscimento dell'intent. Quando un'espressione dell'utente assomiglia ad una delle frasi di training, il corrispondente intent viene innescato. Non è necessario definire ogni possibile esempio poiché le tecniche di machine learning si occuperanno di espandere il modello con frasi simili a quelle di training. Il flusso della conversazione deve essere progettato in modo che la Action sia sempre in grado di interpretare la richiesta.

Anche **Amazon**, l'altro produttore che abbiamo utilizzato per i nostri test, per gestire l'input dell'utente opera con un'architettura del tutto simile a quella appena descritta per Google e che ragionevolmente accomuna anche agli altri principali produttori di assistenti vocali (Fig. 3).

Come già anticipato, ogni volta che l'utente inoltra una richiesta all'assistente virtuale viene attivato il corrispondente *intent* per l'elaborazione della risposta. Nel caso più semplice la risposta che viene restituita all'utente è predefinita e statica ma, più frequentemente, è frutto di un ragionamento più complesso.

La Radio e gli Assistenti Vocali

L'impiego di moderne tecniche di IA permette di effettuare complesse analisi sui dati per rendere le risposte dell'assistente vocale accattivanti e utili per l'utente. Al **CRITS Rai** è stato progettato ed implementato un prototipo che permette un accesso evoluto ai contenuti di *Radio Rai* attraverso gli assistenti vocali. Le piattaforme utilizzate nel corso della sperimentazione sono state *Google* e *Amazon* e, attraverso una user evaluation condotta su più di 130 volontari, sono emersi chiari apprezzamenti verso alcuni dei servizi esposti con il prototipo, che brevemente descriviamo qui di seguito [12].

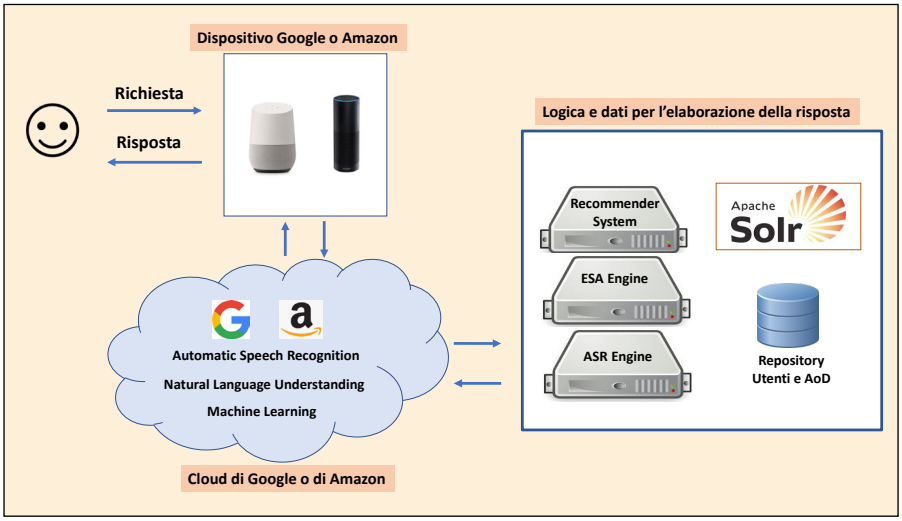

Fig. 3 – Schema di funzionamento del prototipo del CRITS Rai

Ricerca con parola chiave

Questa funzionalità permette la ricerca di parole all'interno del contenuto audio (nel seguito **AoD**, Audio on Demand*). I contenuti trovati vengono ordinati per score decrescente e proposti, uno alla volta, all'ascolto dell'utente che può, a propria discrezione, saltare da un contenuto all'altro. Alla base di questa funzionalità c'è* Apache Solr*, una potente piattaforma open source di ricerca e indicizzazione del testo.*

Ricerca per similarità semantica

Per ogni AoD che l'utente ascolta, c'è la possibilità di chiedere all'assistente vocale una lista di contenuti simili ad esso. Se l'AoD proposto è apprezzato dall'ascoltatore, risulta immediato poterne chiedere altri che potrebbero incontrare il suo interesse.

Contenuti raccomandati

Permette di iniziare l'ascolto di AoD raccomandati. L'utilizzo di questa funzionalità presuppone la presenza di un sistema di raccomandazione che crea una lista di AoD suggeriti e ordinati per score decrescente.

Le funzionalità appena descritte sono state progettate con l'idea di avvicinare il modello di radio tradizionale a forme di fruizione più recenti e di successo, che fanno della flessibilità e della personalizzazione caratteristiche imprescindibili. Anche l'uso di *contenuti atomizzati*, cioè di contenuti suddivisi in atomi semanticamente coerenti, ha contribuito fortemente a migliorare la resa del prototipo grazie alla possibilità di raggiungere direttamente i segmenti di audio rilevanti per l'utente.

Alla base dei servizi esposti dal prototipo c'è l'impiego di differenti tecniche di IA. Solo grazie alla conoscenza profonda degli AoD e delle preferenze dell'utente è possibile creare associazioni rilevanti tra AoD e utente. Nel prototipo, la similarità tra contenuti è calcolata attraverso una comparazione semantica realizzata con tecniche di *Explicit Semantic Analisys* (**ESA**) che utilizzano l'intelligenza artificiale per costruire un interprete semantico che mappa un testo in una sequenza ponderata di concetti [12].

Conclusioni

In questo articolo abbiamo introdotto i concetti di *assistente digitale a controllo vocale* e di *smart speaker*, uno dei dispositivi che più comunemente lo integrano. Gli assistenti intelligenti sono ormai pervasivi ed estremamente rilevanti per la radio, e permettono un utilizzo immediato dell'IA sotto diversi aspetti: riconoscimento e sintesi vocale, riconoscimento della richiesta, attuazione della risposta. Per meglio comprenderne le applicazioni, è stato costruito un prototipo per servizi radiofonici evoluti, valutato da un gruppo di utenti. Lo studio ha confermato la rilevanza degli assistenti vocali per la radio, trovandone applicazioni possibili e individuando alcuni fondamentali requisiti sui contenuti.

BIBLIOGRAFIA

[1] NPR e Edison Research, *The Smart Audio Report*, Edison Research (web), 30/04/2020, https://www.edison-research.com/the-smart-audio-report-2020-from-npr-and-edison-research/ (ultimo accesso 08/10/2020)

[2] *Activate Technology & Media Outlook 2020*, Activate Consulting (web), https://activate.com/outlook/2020/ (ultimo accesso 08/10/2020)

[3] *Canalys: Global smart speaker market to grow 13% in 2020 despite coronavirus disruption*, Canalys (web), 27/02/2020, https://www.canalys.com/static/press_release/2020/pr20200227.pdf (ultimo accesso 08/10/2020)

[4] P. Bajpai, *An Overview Of The Smart Speaker Market*, Nasdaq (web), 20/12/2019, https://www.nasdaq.com/articles/an-overview-of-the-smart-speaker-market-2019-12-20 (ultimo accesso 08/10/2020)

[5] *Vehicles with Alexa*, Amazon (web), https://www.amazon.com/b?node=17744356011&ref=ALEXA_AUTO_VEHICLES (ultimo accesso 08/10/2020)

[6] U. Lawrence, *Android Automotive OS provides the smarts for new Polestar 2 electric sedan*, IEEE Spectrum (web), 04/03/2020, https://spectrum.ieee.org/cars-that-think/transportation/advanced-cars/android-automotive-os-news-polestar-2-electric-sedan (ultimo accesso 08/10/2020)

[7] *CarPlay - Available Models*, Apple (web), https://www.apple.com/ios/carplay/available-models/ (ultimo accesso 08/10/2020)

[8] B. Kinsella, *Alibaba extends Tmall Genie with a new in-car smart speaker pertnership with automakers Audi, Honda, and Renault*, Voicebot.AI (web), 16/06/2019, https://voicebot.ai/2019/06/16/alibaba-extends-tmall-genie-with-a-new-in-car-smart-speaker-partnership-with-automakers-audi-honda-and-renault/ (ultimo accesso 08/10/2020)

[9] Reuters Staff, *VW taps Baidu's Apollo platform to develop self-driving cars in China*, Reuters (web), 02/11/2018, https://www.reuters.com/article/us-volkswagen-autonomous/vw-taps-baidus-apollo-platform-to-develop-self-driving-cars-in-china-idUSKCN1N71J1 (ultimo accesso 08/10/2020)

[10] *Powered By Houndify*, Houndify (web), https://www.houndify.com/powered-by-houndify (ultimo accesso 08/10/2020)

[11] E. H. Schwartz, *FCA will use Cerence voice recognition tech in all vehicles*, Voicebot.AI (web), 20/03/2020. https://voicebot.ai/2020/03/19/fiat-chrysler-will-use-cerence-voice-recognition-tech-in-all-vehicles/ (ultimo accesso 08/10/2020)

[12] P. Casagranda, F. Russo, R. Teraoni Prioletti, *Evolution of Radio Services in the era of Voice-Controlled Digital Assistants*, documento presentato a "IBC 2019 Conference", settembre 2019, Amsterdam, https://www.researchgate.net/publication/335928463_Evolution_of_Radio_Services_in_the_Era_of_Voice-Controlled_Digital_Assistants