

Assistente conversazionale e multimodale per veicoli connessi

Candidato: Eugenio Gallea

Relatori: Fulvio Corno, Luigi De Russis, Alberto Monge Roffarello

INTRODUZIONE

Gli assistenti conversazionali hanno introdotto un nuovo modo di accedere alle funzionalità di un dispositivo. Un assistente è un software che sfrutta algoritmi di intelligenza artificiale per sostenere una conversazione con un utente e capirne eventuali richieste. La possibilità di accedere a funzionalità del dispositivo tramite un'interazione naturale come quella vocale ha favorito l'impiego degli assistenti conversazionali in molti settori. Tra i vari ambiti in cui sono stati impiegati spicca quello automotive. L'introduzione di tali assistenti consente infatti al guidatore di accedere a varie funzionalità del veicolo in maniera immediata. Al momento attuale, nel mercato automobilistico, sono presenti diverse soluzioni: Seat, per esempio, ha integrato Amazon Alexa all'interno del veicolo; Mercedes e BMW hanno creato i loro "Hey Mercedes" e "Hey BMW". A seguito di una valutazione preliminare e un confronto tra queste diverse soluzioni, sono emerse diverse limitazioni di tali assistenti:

1. Essi mancano di integrazione con il veicolo: l'assistente dovrebbe costantemente monitorare la situazione del motore e del veicolo e capire se vi sono presenti anomalie;
2. Sono tutti assistenti reattivi: l'assistente, invece, dovrebbe dimostrarsi proattivo, ossia in caso di presenza di un'anomalia, dovrebbe essere in grado di avvisare l'utente autonomamente, iniziando lui stesso la conversazione;
3. L'assistente non deve distrarre l'utente dalla guida.

OBIETTIVO DELLA TESI

L'obiettivo di questa tesi è la progettazione, lo sviluppo e la valutazione di un assistente conversazionale in auto che permetta al guidatore di accedere e conoscere le funzionalità del veicolo in maniera semplice tramite un'interazione naturale, ma soprattutto senza distrarre dalla guida.

L'assistente deve avere una forte integrazione con il veicolo e monitorarne lo stato al fine di individuare possibili situazioni critiche, in questi casi deve essere in grado di avvisare l'utente iniziando autonomamente una conversazione. Infine deve offrire la possibilità all'utente di richiedere informazioni ed eseguire comandi. In quest'ultimo caso dovrà visualizzare il risultato di tali azioni sul cruscotto del veicolo. Di conseguenza l'interazione è multimodale: vocale e visiva (visualizzazione di dati su schermo). Tale assistente è pensato come impiegabile in un contesto di veicoli connessi, ossia di veicoli in grado di inviare e ricevere dati tramite la rete, di conseguenza punta a sfruttare servizi cloud per svolgere azioni specifiche.

PROGETTAZIONE DEL SISTEMA

Il sistema è stato organizzato in un core centrale residente a bordo del veicolo (Intelligence, evidenziata in arancione nella Figura 1) il quale comunica con diversi servizi cloud (evidenziati in grigio nella Figura 1) al fine di modellare la struttura della conversazione (Conversational Engine) e tradurre da formato testuale a vocale e viceversa le risposte da presentare all'utente o gli input ricevuti dallo stesso (Speech Synthesis & Recognition). Il sistema ha una forte integrazione con l'auto mantenendo una costante comunicazione con essa. Questo consente la lettura dei dati di telemetria da mostrare

all'utente sul cruscotto (velocità, giri al minuto e marcia inserita) e mantiene aggiornata una base dati contenente informazioni sullo stato degli accessori presenti sul veicolo, al fine di potere verificare la presenza di eventuali anomalie. Per emulare una situazione reale è stato utilizzato un simulatore (Car & HMI Simulator, evidenziato in blu in Figura 1).

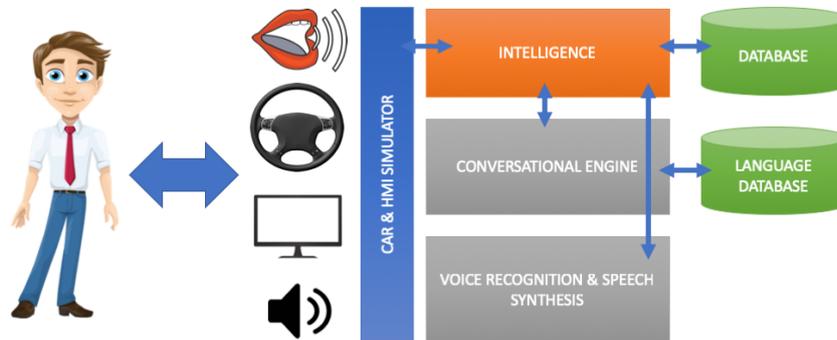


Figura 1. Architettura del sistema.

L'esperienza dell'utente si limita a quest'ultimo modulo. L'interazione con l'utente viene limitata a: interazione vocale, guida e visualizzazione dei dati su schermo. La proprietà di proattività è garantita dal modulo di Intelligence, infatti tramite una verifica sui dati letti dal veicolo (modulo Car & HMI Simulator), tale modulo è in grado di decidere se sia necessario notificare o meno il guidatore di eventuali anomalie o problemi al motore. Il modulo Intelligence, inoltre, è responsabile di gestire le conversazioni e di conseguenza mantenere il contesto (azioni richieste, oggetti di interesse e posizione degli oggetti a cui ci si sta riferendo) dell'ultima conversazione intrapresa dall'utente al fine di potere sostenere un discorso fluido. Il contesto è ricavato grazie ad una comunicazione con il modulo Conversational Engine.

IMPLEMENTAZIONE

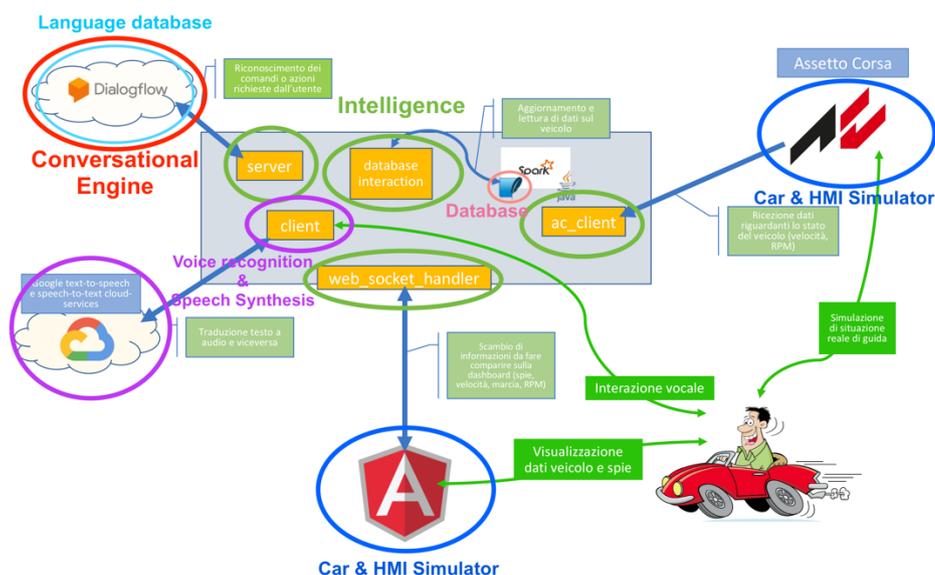


Figura 2. Architettura del sistema con tecnologie impiegate

La Figura 2 mostra l'architettura del sistema con le tecnologie utilizzate per l'implementazione. Il modulo Intelligence è stato sviluppato in Java. Come base dati per lo storage di informazioni riguardanti gli accessori presenti sul veicolo è stato utilizzato SQLite. Il modulo Conversational Engine dell'assistente è DialogFlow, un servizio cloud che permette la modellazione di conversazioni. Esso comunica con l'Intelligence centrale attraverso richieste HTTP POST a un endpoint esposto dalla stessa. È stata creata di conseguenza una REST API utilizzando il framework Spark. Il modulo Car & HMI Simulator è costituito da due parti: Assetto Corsa e una web application sviluppata in Angular. Assetto Corsa fornisce i dati di telemetria, utilizzati, per esempio, per ricavare informazioni derivate, mentre la web application è utilizzata per visualizzare i dati forniti dal simulatore e offrire la possibilità di cambiare lo stato di alcune spie (utile al fine di testare il comportamento dell'assistente in casi particolari). Infine, i servizi cloud utilizzati per la traduzione da formato testuale a formato vocale e viceversa sono offerti da Google: Google Cloud text-to-speech e speech-to-text. Questi ultimi compongono il modulo di Voice Recognition & Speech Synthesis.

RISULTATI E LAVORI FUTURI

Il corretto comportamento del sistema è stato verificato tramite alcuni casi d'uso. Essi possono essere suddivisi in due macro-categorie definite a seconda di chi inizia la conversazione:

1. Guida → Assistente: fanno parte di questa categoria quei casi in cui il guidatore richiede l'esecuzione di comandi oppure richiede informazioni su alcuni accessori;
2. Assistente → Guida: in questa categoria rientrano quei casi d'uso che riguardano avvisi (notifiche di problemi non critici nel veicolo) e alert critici (notifiche da parte dell'assistente di guasti o problemi critici nel veicolo).



Figura 3. Il sistema sviluppato.

L'assistente ha dimostrato di portare a termine la conversazione correttamente in tutti i casi d'uso definiti, inoltre è in grado di interagire correttamente se una nuova conversazione è iniziata in mezzo a un discorso già avviato. Ad esempio: il guidatore richiede l'accensione dei fendinebbia (Guida → Assistente). L'assistente chiede quali, dato che sono presenti sia anteriormente che posteriormente, ma prima che il guidatore possa rispondere si accende la spia della pressione degli pneumatici. A questo punto l'assistente avvia una conversazione notificando l'utente di questo evento (Assistente → Guida) e di fatto iniziando una nuova conversazione. L'utente chiede il significato di tale spia e l'assistente risponde con il significato dell'indicatore.

Nella soluzione sviluppata sono presenti ancora alcuni limiti:

- gestione dei contesti: se il contesto della conversazione viene cambiato a causa di un evento all'interno del veicolo (quindi l'assistente avvia una nuova conversazione), il contesto precedente viene rimosso con impossibilità di recuperare il discorso precedente se l'utente intende ignorare la conversazione avviata dall'assistente;
- le risposte fornite dall'assistente sono hard coded, ossia se la stessa informazione è richiesta più volte, l'assistente risponderà sempre allo stesso modo.

Nei lavori futuri oltre a migliorare i limiti presentati sarà importante effettuare uno studio utente per verificare l'effettiva efficacia dell'interazione con il veicolo, come proposto in questa tesi.