DEVELOPMENT AND IMPLEMENTATION OF IN-CAR VOICE RECOGNITION

ITU- FG-VM-WG3

WORKSHOP, 30-JUNE-2022

Srinivasagan Ayyappan
Aston Marin Lagonda, United Kingdom

Purpose of this proposal for WG3

This proposal is to add more content, modify Section 7.2 HMI, 7.2.1 Voice Recognition Control of FGVM-03 report

PURPOSE OF VOICE RECOGNITION SYSTEM IN THE CAR

Safety:

Eliminate driver distraction while driving. Use cases such as using Navigation app, dictating hands-free text message, requesting weather update or text message readout.

Convenience:

Less time fiddling with buttons and touchscreens

VR DEVELOPMENT STAGES

Stage 1: Voice command and Control capability for Audio, Media, Navigation and climate control

Stage 2: Interaction with mobile devices to make call, text and control music from USB connected devices

Stage 3: Apple CarPlay- EyesFree-Siri, Amazon Alexa and similar technologies around the globe- Embedded Car Voice technology Concepts

Stage 4: Artificial Intelligence(AI) Algorithms to analyse and enhance Audio and Video Inputs, filter background noise interference for Driver Monitoring Systems, Driver Video recording, Guardian mode implementation for young persons

VR CATEGORIES

- VR categories based on VR functionality required features. Typical selection of voice-controlled features is grouped into three categories: Basic, Intermediate and Advanced.

1. The **basic** voice-activated provision is centred around the car's media and entertainment system. Drivers can use their voice to switch stations, adjust volume, skip tracks, and so on.
2. **Intermediate** systems allow the driver to, on top of the basic functionalities, make and receive telephone calls, program the GPS, and adjust the air-conditioning.
3. More **advanced** technology incorporates an internet connection, which facilitates spoken web browsing and the use of apps.
4. The **ultimate** form will be achieved with autonomous cars.

VR IMPLEMENTATION

- VR can be implemented in three options based on multiple factors such as customer preference, OEM & Supplier requirements/capability

1. Onboard/Local VR – Database and processing locally in the vehicle. Local ASR/NLU.

2. Offboard /Cloud VR – Remote database and processing. Server/Cloud ASR/NLU.

3. Hybrid VR – mixed solution of Local and Cloud database and processing. Rule based+Statistical+Neural

VOICE RECOGNITION FEATURES IMPLEMENTATION

Features	
Automatic Speech Recognition (ASR)	Word segmentation – Forward/Backward maximal matching Recurrent Neural Networks (RNN) Feature Extraction – Begin and end-of-sentence markers, text features Semantic Annotator – Static, dynamic, string lists, etc Rule based classifiers -such as models with expected utterances, phone book access Neural classifier – Convolutional Neural Network (CNN, Recurrent Neural Network (RNN)
Natural Language Understanding (NLU)	Recognizes natural human speech, eliminating restriction to predefined commands. Convenient for driver not to remember 'predefined' commands but recognition rate will be lower
Wake-up by Voice/ Wake-Up-Word (WuW)	Always listening mode with key-word activation removes the need for a "push to talk" button. e.g.1. iPhone already has this feature. VR is activated by "Hey Siri" e.g.2. Wake-up command can be customizable for customer flavor
Combined List Search	Enables recognition from combined large lists recognizing only valid combinations e.g.1. Street + City + State or State + City + Street e.g.2. Song name + Artist for all song list e.g.3. McDonald nearby Destination
Language Recognition (including Dialects)	After speech recognition recognizes the words spoken, detect spoken language automatically and semantic classification extracts the meaning. Various dialects are also to be detected automatically.
Barge-In	Allows user to speak over spoken dialog prompts and be recognized. Customer does not have to wait until dialog prompt finished.

VOICE RECOGNITION FEATURES IMPLEMENTATION

Features	Descriptions
Multi-lingual & Partial Search	Recognizes multi-lingual or partial name and search database such like phonebook, POI category, POI names, Song list and so on. e.g.1. Srini in Phonebook; Voice search by 'Srini' or '西尼' e.g.2. 'I Love Youから始めよう' in song list; Voice search by 'Love' or '始めよう'
1-Shot Voice Destination Entry	Recognizes full sentence for destination entry search without multiple steps. e.g.1. "Banbury Rd, Gaydon, Lighthorne Heath, Warwick CV35 0DB" in 1-shot e.g.2. "பான்பரி சாலை, கெய்டன், லைட்ஹார்ன் ஹீத், வார்விக் **CV35 0DB**" in 1-shot
All Inclusive Main Menu	Enables all commands to be spoken in a single utterance on Home Screen. All predefined commands including domain-specific commands.
Voice Biometric	Recognizes and analyze Voice for Biometric authentication. This cannot be a replacement of car key (or remote key) due to security reason, but can be used for comfortable features such like - Seat/mirror/steering wheel variation (memory) - Recommend favorite radio stations or song list - Bluetooth pairing (when multiple devices connected)

TEXT-TO-SPEECH FEATURES IMPLEMENTATION

Features	Descriptions
Multi-lingual Support	Accurate language identification and high-quality acoustic extensions provide unparalleled foreign language readout (particularly useful for SMS readout)
Naturalness	More natural human-like speech output with less resource consumption (less memory, less size of database, less consumption of processor)
Flexible Speech Generation	Volume and speaking rate can be adjustable at run time for more dynamic and lively effects
Direct phonetic input & Seamless prompt insertion	Allows and offers optimal and seamless read out of off-line phonetic databases with pre-recorded voice. e.g.1. Voice guidance in navigation route: 'in 100 meters, turn left to 'unione internazionale delle telecomunicazioni, ginevra' Blue is pre-recorded voice, Red is TTS from map database
User Dictionaries	Application specific lexica can be phonetically optimized for accurate readout of exceptional pronunciations.
SSML Support	Speech Synthesis Markup Language (SSML) allows for TTS vendor-independent markup which is correctly interpreted by TTS engine. This can be utilized for add-on features of readout weather forecast, favorite stocks, sports news for example.

THANK YOU