

The Development Trend of the Smart Car Voice Interaction System

iFLYTEK, Zhan Shi & Feifei Tian 30.06.2022

Current status and future development trend

Effect indicators of voice interaction

Covering the full scene of intelligent connected vehicles



Entertainment



Navigation



LBS



Parking



Refueling



Washing

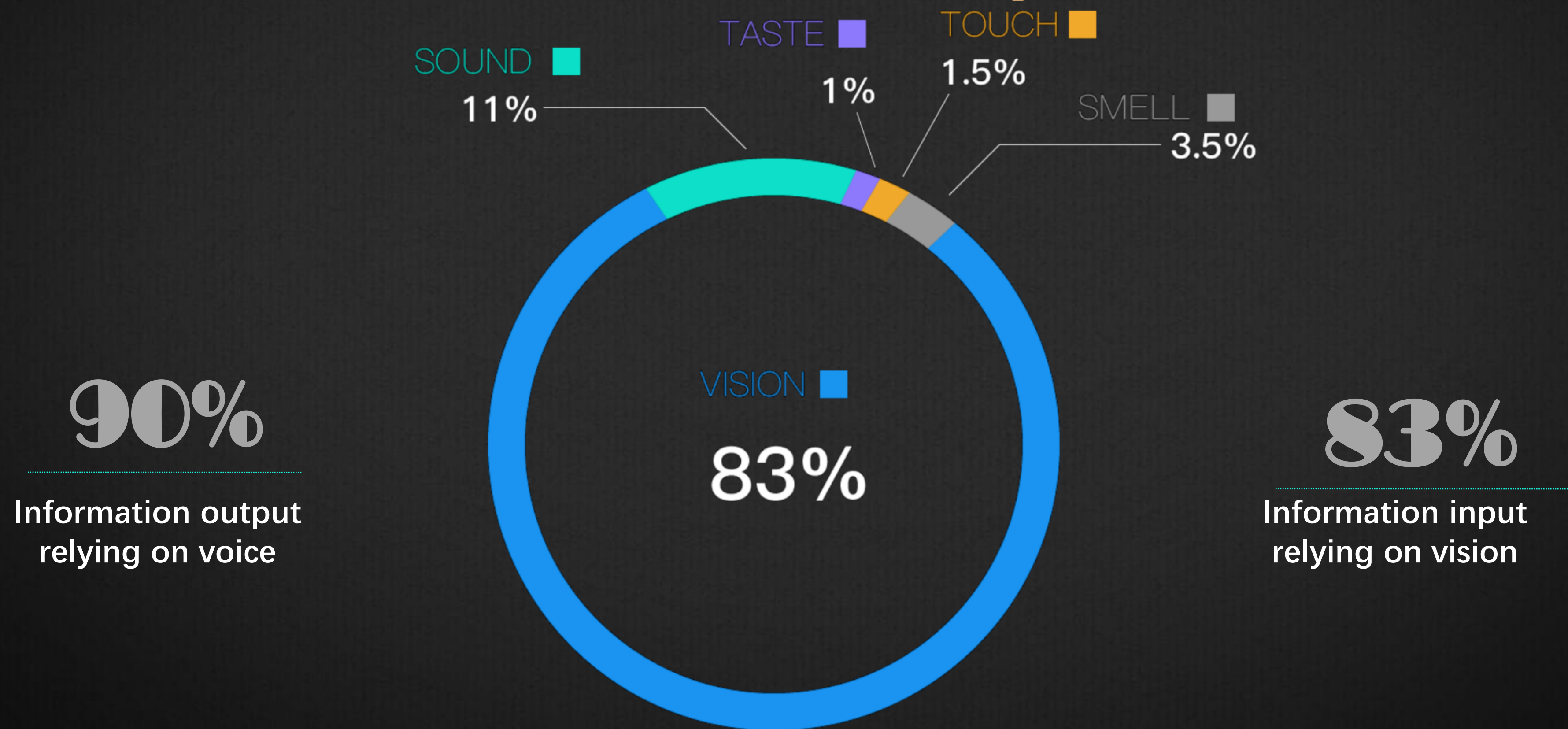


Maintenance



Insurance

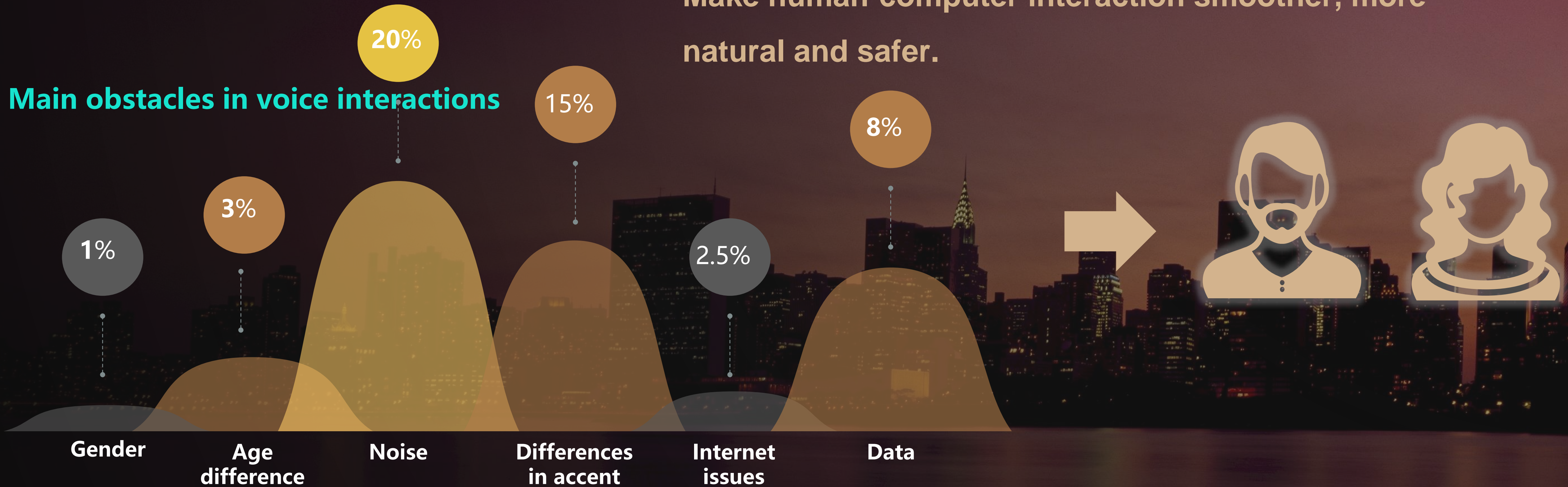
According to the analysis by researchers from Harvard Business School, human brains receive external information from the five senses in the following proportions: taste 1%, touch 1.5%, smell 3.5%, hearing 11% and vision 83%.



Make technology provide the best human-computer interaction tool for user scenario service.

Make human-computer interaction smoother, more natural and safer.

Main obstacles in voice interactions



Reflection on the interactive experience required in the vehicle scene

EASY
Hit the mark
with a single
sentence

SMART
Service care
with better
understanding

INTERACTION

SAFE
Quick
response
and hands-
free

HUMANISTIC
More natural
and smoother
interaction

Meet considerations on simple and natural voice interaction design

Key / voice
wake-up

Hello, FEIYU.

I'm here!

I'd like to go to Sheraton in Jiading.

One-shot

Hello FEIYU, I'd like to go to Sheraton in Jiading.

Sheraton in Jiading.

On-screen texts
as commands

Sheraton in Jiading.

Play *One World, One Dream*

Wake-up
free

Meet considerations on intelligent and humanized voice interaction design

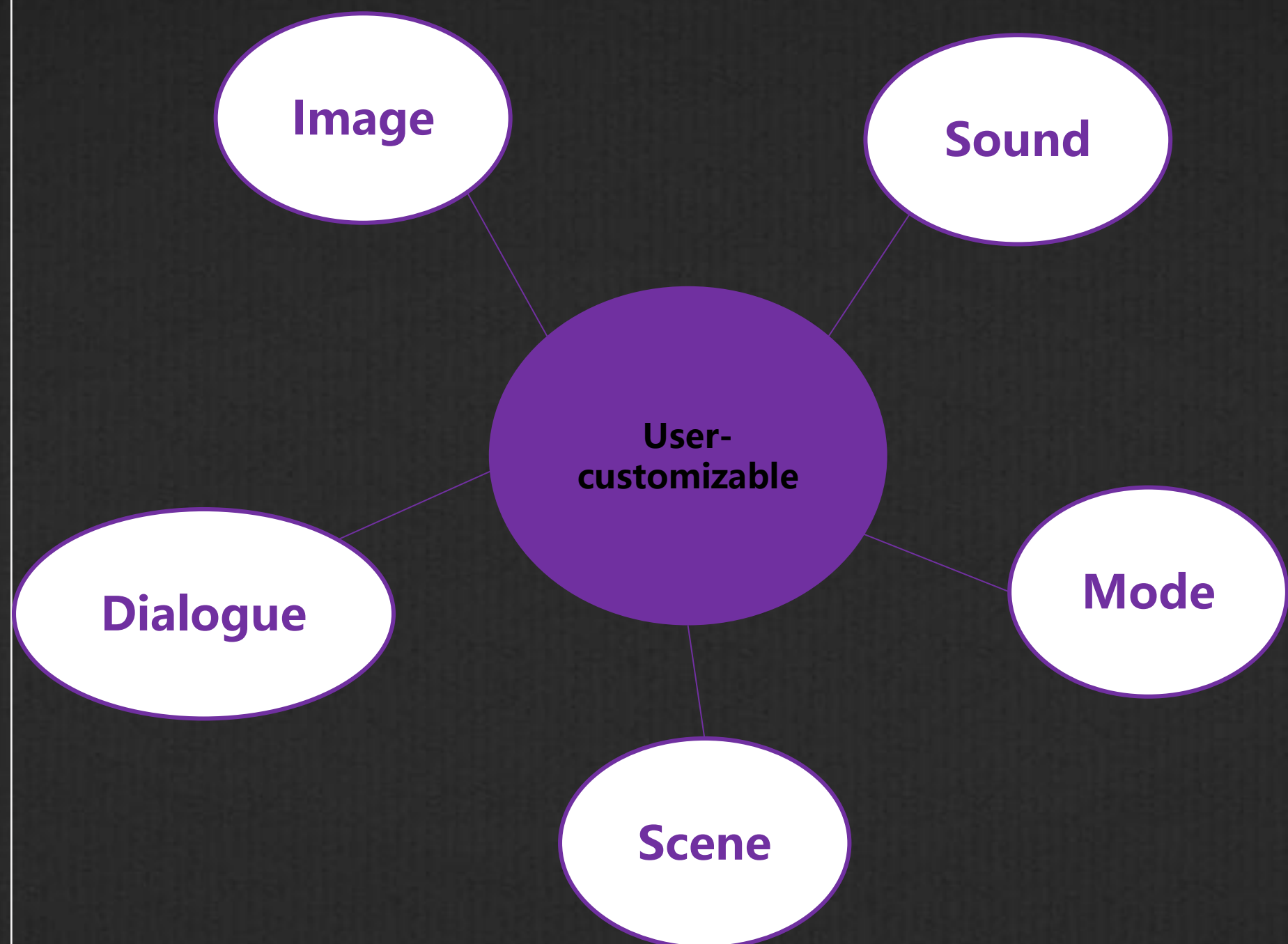
When you tell FEIYU:

- Go to work
- +Add statement / key

the system will initiate:

- Greeting : Hard worker, today's goal is half a ton of bricks.
- Music : Play Beyond's *No more hesitation*
- Navigation : Navigate to company
- AC : External circulation, 25 degrees Fan level 2
- Driving mode : Sport mode
- +Add service

customizable



Supports custom voice commands

Supports multi-task operations

Supports self-triggering under given circumstances

Supports multi-task operations

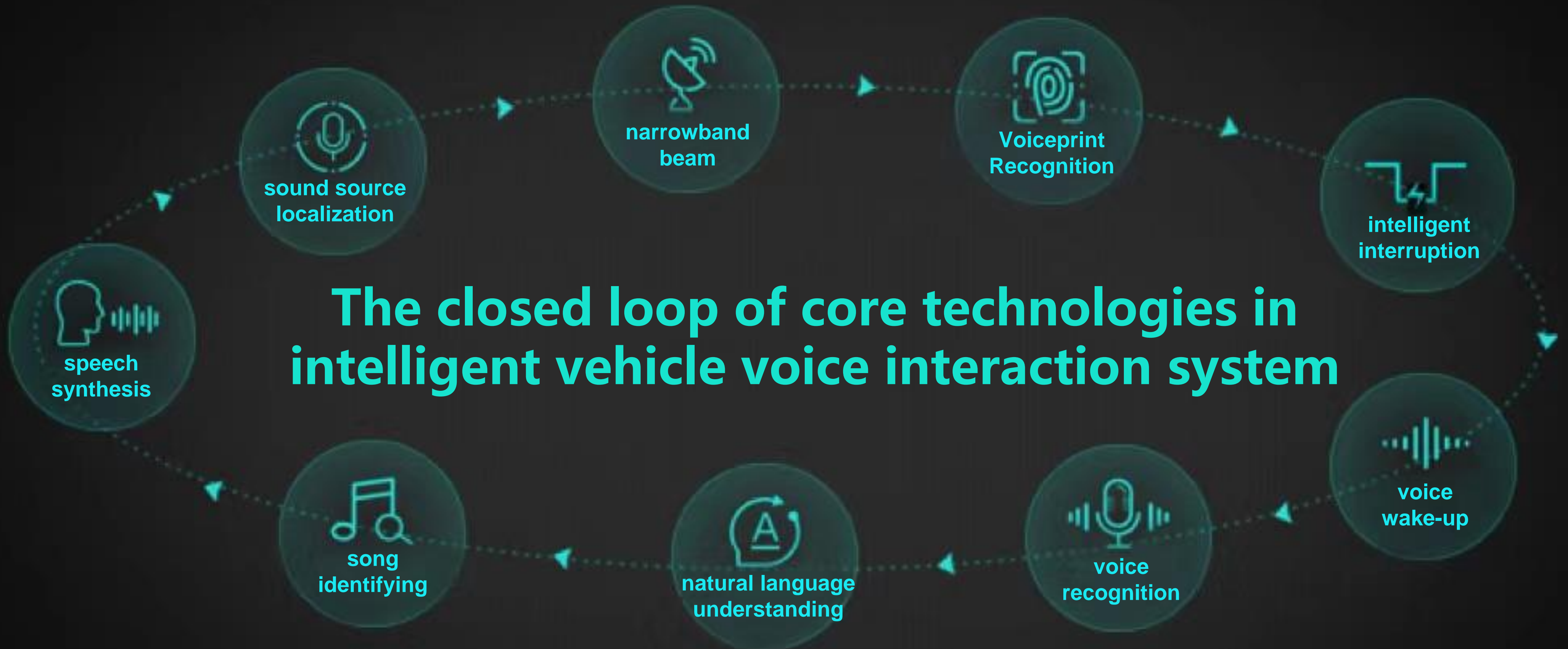
Under the following circumstances:

- Location : within 200 meters of workplace
- Time : Monday - Friday, 7 - 8 p.
- +Add condition

the vehicle will:

- Operation condition : ignition
- Music : Play favourite songs
- Navigation : Navigate home
- Lights : Turn on automatic headlights
- +Add service

self-triggering



The closed loop of core technologies in intelligent vehicle voice interaction system

Future development trend

Intelligent and personalized mobile space

Vehicles have a large number of advanced sensors, they are the best application stage for the AI technology. Vehicles are also a platform for a large amount of multi-dimensional data (human-car-scene) that accompanies users in their travels.

Continuous
breakthroughs
in
AI
technology

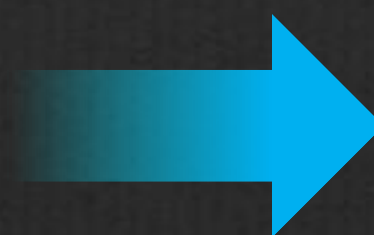
speech recognition

speech synthesis

image recognition

voice interaction

...



Smart and
intimate

Fashionable
and
comfortable

Safe and
reliable



Improvement in user service and user experience
Brand and business model expansion

**Technology Convergence in high
perception of multimodal**



**Data collaboration
of multidimensional cross-scene**



perception

cognition

perception

Listening

Viewing

Multimodal fusion technology based on Speaking, Listening, Viewing, and Representing provides a human-computer interaction AI base for the intelligent cockpit

hearing

interaction management

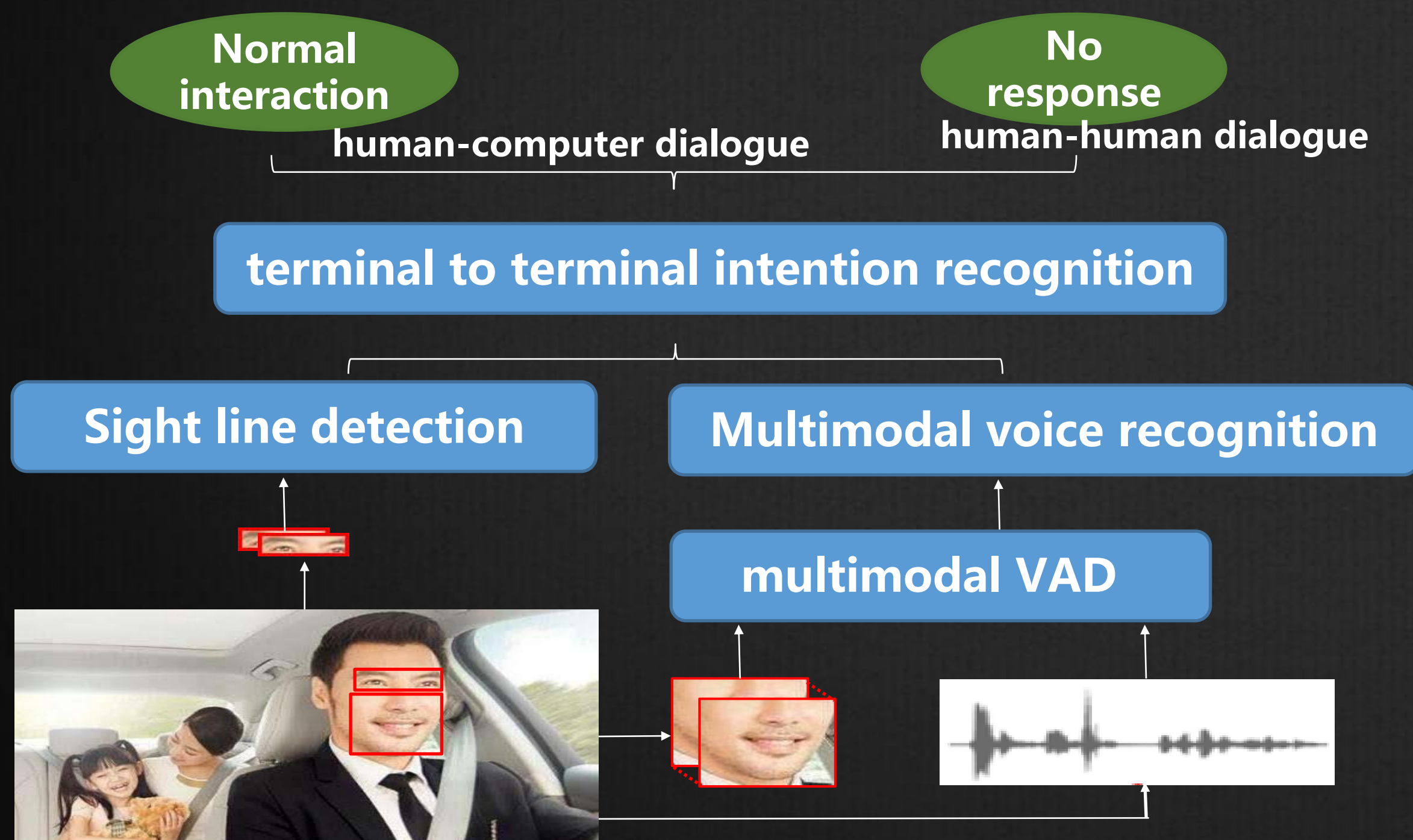
vision

Speaking

Representing

Multimodal wake-up free——iFLYTEK inputs exploration and application facing vehicle multimodal

Interaction scheme of multimodal wake-up free



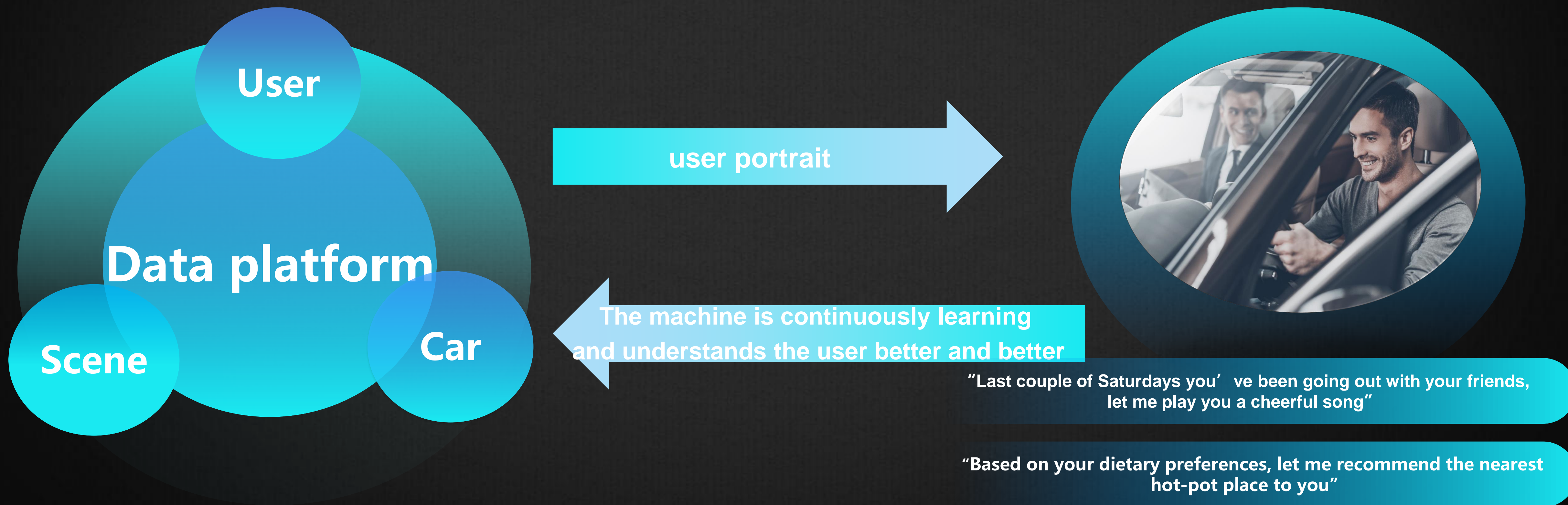
Different signal inputs cause different functions to be implemented

input signal	interaction function implemented
voice recognition	voice interaction based on wake-up
+multimodal VAD	Wake-up free interaction for restricted skills
+ID	improve personalized recognition and interaction effect
+visual attention	improve the effect of specific control commands
+character relationship	multi person wake-up free interaction



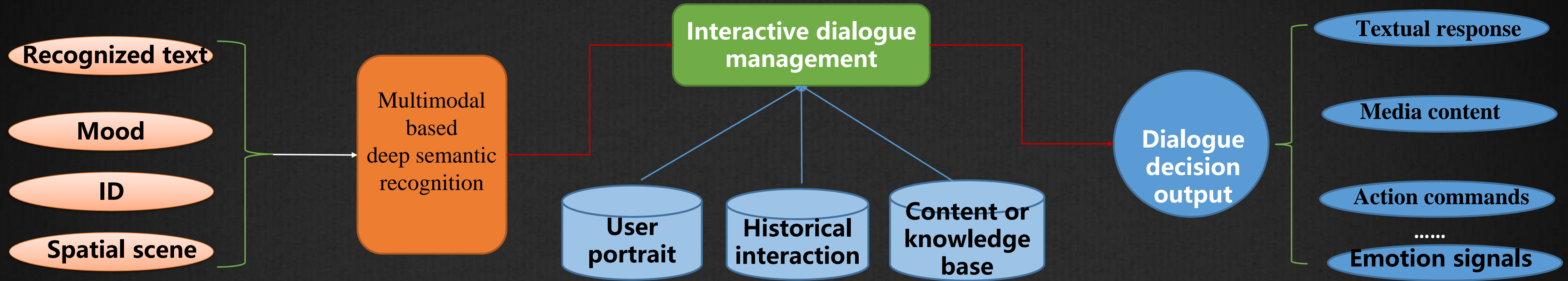
Intelligent and personalized mobile space——“Smart and Intimate”

Work together to create a user personalized data analysis and intelligent recommendation platform to provide users with more active and more convenient services



Scene engine based intelligent interaction

—iFLYTEK is on the path of exploring intelligent interaction technologies in vehicle scene



Serial number	Multimodal input signal	Traditional task-based interaction	New generation active interaction
1	recognized text	realization of voice interaction and operationing commands	active interaction、 active data gathering、 topic recommendation
2	+mood	passive response, quick/single-rounded emotional response	response in accordance with the user's mood (e.g. when the user feels down, the response is sympathetic and comforting, while recommending uplifting content)
3	+ID	unavailable	creating a user portrait with long term memory, so as to achieve personalized interaction
4	+spatial scene	unavailable	designing interaction content based on the surroundings and the setting
5	+terminal platform	unavailable	implementing changes according to the terminal's features (e.g. screen/no screen)

	BASIC AI Assistant	INTERMEDIATE AI Assistant	ADVANCED AI Assistant
Operating features	Visual interaction (2D)	Visual interaction (3D)	Visual interaction (3D)
	User-customizable (sound)	User-customizable (scene, image, sound)	User-customizable (scene, image, sound)
Basic features			See and control
	On-screen texts as commands (HU)	On-screen texts as commands (HU)	On-screen texts as commands (HU&IP)
	Audio assistant functions	Multimodal audios (wake-up free)	Multimodal interaction (audio+vision+emotion)

	Basic	Intermediate	Advanced
AI assistant product levels	Visual interaction (2D)	Visual interaction (3D)	Visual interaction (3D)
	User-customizable (sound)	User-customizable (scene、image、sound)	User-customizable (scene、image、sound)
	Noise Reduction	Noise Reduction	Noise Reduction
	Wake-up	Wake-up	Wake-up
	Recognition	Recognition	Recognition
	Semantic VAD	Semantic VAD	Semantic VAD
	Semantic	Semantic	Semantic
	On-screen texts as conmmands	On-screen texts as conmmands	On-screen texts as conmmands
	-	Lip movement perception	Lip movement perception
	-	Gesture perception	Gesture perception
	-	-	Mood perception
	-	-	Fatigue detection
	-	-	Active interaction of line of sight
	-	-	Shutdown on silence

Current status and future development trend

Effect indicators of voice interaction

Effect indicators of voice interaction

Voice interaction	Functions	Unit	Minimum target	Benchmark vehicle model target
Voice cold start	Voice response	s	≤12	≤2
Voice response speed (high load)	make a phone call	s	≤3	≤2
	Search a song	s	≤4	≤3
	Search a radio station	s	≤4	≤3
	Open the sunroof	%	≤3	≤2
Offline/online response speed	Voice control response speed in navigation	s	≤4	≤3
	Voice control response speed in music	s	≤4	≤3
	Voice control response speed in telephone	s	≤3	≤2
Speech recognition rate	static (parked, windows closed)	%	≥95%	≥95%
	dynamic (60KM/h, uniform speed, asphalt road, windows closed)	s	≥90%	≥90%
	Voice response	s	≤2	≤1
Wake-up rate (static)	Success rate of continuity wake-up for main wake-up words	s	≥95%	≥95%
	Success rate of intermittent wake-up for main wake-up words	%	≥95%	≥95%
	Success rate of wake-up free in navigation	%	≥95%	≥95%
	Success rate of wake-up free in music	%	≥95%	≥95%
	Success rate of wake-up free in telephone	%	≥95%	≥95%



The comparison of measured values and target values for the voice interaction system

Serial number	Category	Test items		Measured value	Target		
1	Speech recognition rate	Dynamic Speech	Scene 1	speed 40km/h; AC off; doors and windows closed	93.83%	92.96%	90%
2			Scene 2	speed 60km/h; AC off; doors closed; front-left window half open	92.83%		
3			Scene 3	speed 60km/h; AC off; doors closed; front-left window fully open	92.33%		
4			Scene 4	speed 60km/h; AC on、 blow face、 air volume 3、 vents up; doors and windows closed	93.00%		
5			Scene 5	speed 80km/h; AC on、 blow face、 air volume 3、 vents up; doors and windows closed	92.80%		
6		Static Speech	Scene 1	doors and windows closed; AC off; engine idle speed	98.00%	96.95%	95%
7			Scene 2	doors and windows closed; AC face blowing mode、 max air volume、 vents down; engine idle speed	96.00%		
8			Scene 3	doors and windows closed; AC face blowing mode、 medium air volume、 vents down; engine idle speed	96.50%		
9			Scene 4	doors and windows closed; AC face blowing mode、 air volume 1、 vents down; engine idle speed	97.30%		
10	Wake-up rate	Static Speech	continuity test of main wake-up words	---	96.60%	95%	
11			intermittent test of main wake-up words	---	96.33%	95%	
12		Dynamic Speech	continuity test of main wake-up words	speed 40km/h; AC off; doors and windows closed	93.00%	90%	
13				speed 60km/h; AC off; doors closed; front-left window half open	91.33%	90%	
14				speed 60km/h; AC on、 blow face、 air volume 3、 vents up; doors and windows closed	91.17%	90%	
15			intermittent test of main wake-up words	speed 40km/h; AC off; doors and windows closed	93.00%	90%	
16				speed 60km/h; AC off; doors closed; front-left window half open	91.00%	90%	
17				speed 60km/h; AC on、 blow face、 air volume 3、 vents up; doors and windows closed	91.37%	90%	
18		Rate of customized wake-up words		continuity test of customized wake-up words	95.50%	95%	
19				intermittent test of customized wake-up words	95.17%	95%	
20		Rate test of scenario wake-up		Success rate of wake-up free in navigation	96.00%	95%	
21				Success rate of wake-up free in music	95.70%	95%	
22			Success rate of wake-up free in telephone	96.67%	95%		
23	Basic voice function	False wake-up rate	False wake-up rate in mute scenario		8 hours 0 time	8 hours 1 time	
24			False wake-up rate in music playing scenario		4 hours 1 time	4 hours 1 time	
25			False wake-up rate in dynamic vehicle scenario		4 hours 1 time	4 hours 1 time	



**iFLYTEK holds the technology of A.I., joining hands with industry partners,
are creating an ultimate driving experience of smart car.**