

ICS 35.240.01
CCS L77



**NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA**

GB/T 44089-2024

**Information technology - General technical requirements of full
duplex speech interaction system**

信息技术 全双工语音交互系统通用技术要求

Issued on: May 28, 2024

Implemented on: May 28, 2024

**Issued by: State Administration for Market Regulation;
Standardization Administration of the PRC**

Table of Contents

1 Scope	1
2 Normative references	1
3 Terms and Definitions	1
4 Abbreviations	2
5 System reference functional framework and interaction process	2
5.1 System reference functional framework	2
5.2 System Interaction Process	4
6 Functional Requirements	5
6.1 Core Requirements	5
6.2 Acoustic Treatment Layer	6
6.3 Speech Recognition Layer	6
6.4 Dialogue Processing Layer	6
6.5 Speech Synthesis Layer	6
7 Performance Requirements	6
7.1 Speech Recognition Layer	6
7.2 Dialogue Processing Layer	7
7.3 Speech Synthesis Layer	7
10 Reference	11

Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for standardization work Part

1. Structure and drafting rules for standardization documents" Drafting. This document was proposed and coordinated by the National Technical Committee for Information Technology Standardization (SAC/TC28). This document was drafted by: China Electronics Technology Standardization Institute, iFLYTEK Co., Ltd., Midea Group (Shanghai) Co., Ltd., Shenzhen Tencent Computer Systems Co., Ltd., Beijing Baidu Netcom Technology Co., Ltd., China Telecom Group Co., Ltd., Xiaomi Communications Technology Co., Ltd., China Mobile (Hangzhou) Information Technology Co., Ltd., Qingdao Haier Technology Co., Ltd., Fuzhou Data Technology Research Institute Co., Ltd., Shenzhen Intellifusion Technology Co., Ltd., Beijing Telecom Planning and Design Institute Co., Ltd., AISpeech Technology Co., Ltd., Hangzhou Fangde Intelligent Technology Co., Ltd., Antelope Industrial Internet Co., Ltd.,

Hefei Intelligent Voice Innovation Development Co., Ltd., Shenzhen Sihe Technology Co., Ltd. Company, Shanghai Intelligent Manufacturing Functional Platform Co., Ltd., Beijing Jitong Huasheng Technology Co., Ltd., and Mashang Consumer Finance Co., Ltd. The main drafters of this document are: Dong Jian, Xu Yang, Jia Yijun, Liu Ying, Song Wenlin, He Yongchun, Yu Lei, Sudan, Yuan Jie, E Lei, Cai Yasen, Mei Linhai, Zhao Pei, Liu Cong, Yang Zhen, Lei Zong, Gong Sheng, Fan Shuai, Hong Pengda, Huang Chao, Li Linlu, Fang Bin, Chen Ming, Hu Guoping, Yang Yifan, Liu Zhiqiang, Bi Shengnan, Ding Qiang, Gao Yu, and Li Xu. Information Technology Full-duplex Voice Interaction System General technical requirements

1 Scope

GB/T 44089-2024 sets the general technical requirements for full duplex speech interaction systems, the voice assistants that listen while they speak rather than waiting for the user to finish. Full duplex is what makes a spoken interface feel like a conversation, because the user can interrupt, correct and continue without a wake word each time, and it is technically much harder than half duplex: the system must cancel its own output from the microphone signal, decide continuously whether incoming speech is addressed to it, and manage a dialogue whose turns overlap. The standard sets the system architecture and its functional components and then the requirements: the audio front end including acoustic echo cancellation, noise suppression and the wake and interruption detection, the speech recognition and endpointing under barge-in, the dialogue management and turn taking, the response latency, the multi-round and context handling, the rejection of speech not addressed to the system, and the requirements on performance testing, reliability, privacy and security. It took effect on 28 May 2024.

This document specifies the reference functional framework, interaction process, functional requirements and performance requirements of a full-duplex voice interaction system. This document is applicable to the design, development, application, testing and maintenance of full-duplex voice interaction systems.

2 Normative references

This document has no normative references.

3 Terms and definitions

The following terms and definitions apply to this document.

3.1 Full-duplex A communication method capable of transferring data in both directions simultaneously.

3.2 functional unit A hardware or software entity that is capable of accomplishing a specific goal.

3.3 Speech recognition The process of converting human voice signals into text or instructions. [Source: GB/T 21023-2007, 3.1]

3.4 Enable the functional unit to understand the intention of human speech. [Source: GB/T 36464.1-2020, 3.11]

3.5 speech synthesis The process of synthesizing human language through mechanical and electronic methods. [Source: GB/T 21024-2007, 3.1]

3.6 Telephony The dialogue text content with certain logic used in the interaction process.

3.7 dialogue management Follow up the current conversation status and context input, update the conversation status, and generate the required actions based on the conversation processing logic. dialogue action.