

ICS 13.310
CCS A 91



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

GB/T 47478.1-2026

**Public security - Voiceprint recognition applications - Part 1:
Technical requirements for capture, recognition and database
construction**

公共安全 声纹识别应用 第1部分：采集识别建库技术要求

Issued on: April 30, 2026

Implemented on: November 1, 2026

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

Table of Contents

- 1 Scope
- 4 Source of voiceprint audio files
- 4.2 Requirements for Voiceprint Speech Sample Collection
- 5 Voiceprint data quality inspection and quality grading
- 5.1 General Requirements for Quality Inspection
- 5.3 Voiceprint Data Quality Classification
- 5.3.2 Level
- 5.3.3 Level
- 5.3.4 Level
- 6 Requirements for building a voiceprint audio file library
- 6.2 General Requirements for Database Construction
- 6.3 Voiceprint Algorithm Test Library
- 6.4 Voiceprint Identity Database
- 6.5 Voiceprint-Specific Application Library

Foreword

GB/T 47478.1-2026 | Public security - Voiceprint recognition applications - Part 1: Technical requirements for capture recognition and database construction

GB/T 47478.1-2026 English version. Public security - Voiceprint recognition applications - Part

1 Scope

GB/T 47478.1-2026 is the Chinese national standard covering voiceprint recognition in policing - how a voice sample is captured and what quality it must have, how the comparison is made and scored, and how the reference database is built and maintained. Part 1 of the series, first edition, in force since 1 November 2026, under the Ministry of Public Security. It was issued on 30 April 2026 and takes effect on 1 November 2026, as a first edition. This page is published from the official record of the 2026 edition; the clause text of a standard this recent is not yet in circulation, and the figures, limits and tables it contains are those of the document itself, delivered in full with the English translation.

91 National Standards of the People's Republic of China Public Security Voiceprint Recognition Application Part

1.Data Collection Identify database construction technical requirements Published on 2026-04-

30 Implemented on 2026-11-

01 State Administration for Market Regulation The State Administration for Standardization issued a statement.

1.Scope This document specifies the sources of voiceprint voice files, voiceprint data quality detection and quality classification, and requirements for voiceprint voice file database construction. This document applies to the acquisition of voiceprint files, voiceprint data quality inspection, construction of voiceprint file databases, and voiceprint recognition. application.

4 Source of voiceprint audio files

4.1 Source Classification Depending on how the voiceprint voice file is obtained, the sources of voiceprint voice files are generally divided into two categories.

- a) Direct acquisition. After collecting voiceprint speech samples on-site or remotely using voice acquisition equipment, the voiceprint speech files are saved;
- b) Indirect acquisition. After reading existing electronic data files and performing non-tampering operations such as truncation and merging, the voiceprint is re-saved to obtain the voiceprint. document.

4.2 Requirements for Voiceprint Speech Sample Collection

4.2.1 Factors to be considered during data collection When collecting voiceprint samples, the following factors should be considered and specific requirements should be put forward according to the needs of actual applications.

- a) Data collection environment. environmental reverberation conditions, ambient noise levels, presence of other human voices, etc.
- b) Acquisition equipment. factors such as frequency response, sensitivity, waveform distortion, directivity, and whether a microphone array is used;
- c) Acquisition parameters. sampling rate, quantization accuracy, number of channels, etc.
- d) Data collection constraints. Whether there is supervision at the data collection site, whether the data subject cooperates, and the data subject's speaking style (e.g., reading aloud, speaking casually, etc.). Factors such as dialogue;
- e) Content to be collected. Factors such as whether prompts or specified voice content are used, and the required effective voice duration to be collected;
- f) Other specific factors required by the application. gender, age, language or dialect,

accent, mood, speaking speed, volume, etc. of the person being collected.

4.2.2 Requirements for Voiceprint Identity Database Collection The voiceprint files for the voiceprint identity database should be obtained directly. Guidance should be provided to the person being collected before voiceprint sample collection. Understand the precautions for voiceprint and speech sample collection. The following should be considered regarding the collection site environment, personnel deployment, collection method, and collection equipment. Require.

- a) Verify the identity information of the person being collected before data collection;
- b) The data collection site was quiet with no significant noise;
- c) The voice acquisition device has no obvious echo, electrical noise, or howling sound;
- d) Use a sampling frequency of 8000Hz or 16000Hz, with a quantization precision of 16 bits;
- e) Point the effective acquisition direction of the voice acquisition device towards the person being acquired, and maintain an appropriate distance between the voice acquisition device and the person being acquired. When stationary, disable the automatic noise reduction function of the acquisition device and software;
- f) Do not collect speech when the subject is in a state of obvious physical discomfort;

5.1 General Requirements for Quality Inspection

5.1.1 Based on the acquisition parameters (including sampling rate, quantization accuracy, number of channels, etc.) during the acquisition of voiceprint speech samples, the voiceprint speech can be directly analyzed. Speech quality is tested using audio samples.

5.1.2 For voiceprint-based speech files, the audio data should be read according to the corresponding audio file format, and then the audio data should be processed for voice quality control. Quantity detection.

5.1.3 During the voice quality detection process, the voiceprint data being detected should not be altered.

5.2 Voiceprint Speech Quality Testing Content Voiceprint speech quality testing should include at least the following.

- a) Upper/lower cutoff frequencies. Calculate the upper/lower cutoff frequencies of the speech data based on the sampling rate and sampled values. Note

1. The upper/lower cutoff frequencies are the high/low frequency values located 3dB below the average energy of the passband.

- b) Amplitude and waveform distortion. Determine whether there are amplitude and waveform distortion problems based on the changes in the sampled values.

c) Spraying phenomenon.

Note 2. "Pop-up" refers to the noise phenomenon caused by the airflow being too strong and directly hitting the voice acquisition device during the voice acquisition process.

d) Speech mixing phenomenon. Note

3. Speech mixing refers to the phenomenon of multiple voices in a single audio clip. For example, multiple people speaking at different times, or multiple people speaking simultaneously at the same time.

e) Sudden noise phenomenon. Note

4. Sudden noise refers to noise that is not controlled by the speaker, and that occurs suddenly from the acquisition equipment or acquisition environment, has a short duration, and a high intensity.

f) Signal-to-noise ratio, tested according to formula

5.3 Voiceprint Data Quality Classification

5.3.1 Quality Classification Based on the occurrence of issues such as amplitude truncation, waveform distortion, speech interference, sudden noise, and obvious popping sounds, the voiceprint data quality can be classified from low to high. The high-level classification is divided into three levels. Level 1, Level 2, and Level 3, as shown in Table 1, which are applicable to different applications.

5.3.2 Level

1 The upper cutoff frequency of the voiceprint data is greater than or equal to 3000Hz and the lower cutoff frequency is less than or equal to 500Hz, but does not meet the requirements of... Quality requirements for Level 2.

Note. This level of voiceprint data is suitable for voiceprint recognition applications that do not require high recognition accuracy, such as using voiceprint recognition to achieve personalized services for voice assistants.

5.3.3 Level

2 The upper cutoff frequency of voiceprint data is greater than or equal to 3400Hz and the lower cutoff frequency is less than or equal to 300Hz, indicating the signal strength of the speech segment. Audio with a noise ratio greater than or equal to 25dB, exhibiting issues such as amplitude clipping, waveform distortion, speech interference, sudden noise, and noticeable popping, but still possessing quality problems. The total duration of each segment is less than or equal to 10% of the total audio duration; or the signal-to-noise ratio of the speech segment is less than or equal to 25 dB and greater than or equal to 15 dB. Furthermore, it does not suffer from issues such as amplitude clipping, waveform distortion,