

ICS 13.310
CCS A 92

GA

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

GA/T 1587-2019

Test specifications for automatic speaker recognition system

声纹自动识别系统测试规范

(English Translation)

Issued on: September 11, 2019

Implemented on: November 1, 2019

Issued by: Ministry of Public Security

Table of Contents

- Foreword
- 1 Scope
- 2 Normative references
- 3 Terms and definitions
- 4 Test specifications
 - 4.1 Registered Speaker Voice
 - 4.2 Testing Speaker Voice
 - 4.3 Test requirements
 - 4.4 Test indicators
 - 4.5 Test report

Foreword

This standard was drafted in accordance with the rules given in GB/T 1.1-2009.

Please note that some elements of this document may involve patents. The issuer of this document is not responsible for identifying these patents.

This standard is under the jurisdiction of the National Criminal Technology Standardization Technical Committee (SAC/TC 179).

This standard was drafted. Material Evidence Identification Center of the Ministry of Public Security, the Third Research Institute of the Ministry of Public Security, the Expert Appraisal Committee of the Ministry of Security, Guangdong Provincial Public Security Department, Shanghai Public Security Bureau, Anhui Provincial Public Security Department, Hefei Public Security Bureau, Jiangsu Public Security Department, Fujian Public Security Department, Shenzhen Public Security Bureau, Beijing Judicial Identification Center of National Security Bureau, Judicial Identification Center of National Security Department of Guangxi Zhuang Autonomous Region, Beijing Police Academy, Tsinghua University, HKUST Xunfei Co., Ltd.

1 Scope

This standard specifies the test requirements, indicators and reports for the automatic voiceprint identification system.

This standard applies to the testing of voiceprint automatic identification system in the field of public safety, and it can be used in other fields.

2 Normative references

The following documents are essential for the application of this document. For dated references, only the dated version applies to this article Pieces. For undated references, the latest version (including all amendments) applies to this document.

GB/T 35048-2018 Forensic Science Speech and Audio Inspection Terminology

3 Terms and definitions

The terms and definitions defined in GB/T 35048-2018 and the following terms apply to this document.

Voiceprint retrieval A speaker's voice is put into a voiceprint database for retrieval and comparison, and the process of matching with the speakers in the database is obtained.

Voice recognition

The process of putting a speaker's voice in a voiceprint database for retrieval and comparison, and inferring the voice corresponds to the identity of the speaker.

Automatic speaker recognition system A computer system that can automatically complete voiceprint retrieval and comparison or voiceprint recognition tasks.

Valid speech duration the total duration of voice clips that meet the requirements of relevant specifications is generally calculated in seconds.

Registered speaker

A speaker with a known (or presumed known) identity in the system.

Enrollment speech

In the system, it is used to extract the voice of the registered voiceprint information of the speaker.

Test speaker

During the system test, the speaker corresponding to the voice is tested.

Test speech

During system testing, the speaker's unknown voice.

Language

Register or test the language corresponding to the voice.

3.10

Test task trial

Determine whether there is a task of the target speaker in the test voice.

3.11

Target test task

In the test speech, there is a test task of the target speaker.

3.12

Imposter trial

In the test voice, there is no test task for the target speaker.

3.13

Recall ratio

The ratio of the number of verified target test tasks determined by the system to the number of target test tasks in the corresponding answer file.

3.14

Precision ratio

The ratio of the number of target test tasks determined by the system to the correct number of target test tasks determined by the system.

3.15

F1 score

An indicator used in statistics to measure the accuracy of a binary classification model.

$F1 = 2 \times \text{recall rate} \times \text{precision rate} / (\text{recall rate} + \text{precision rate})$

3.16 False alarm probability
the ratio of the number of masquerade test tasks verified by the system to the number of masquerade test tasks in the corresponding answer file.

3.17

Missing probability

The ratio of the number of target test tasks verified by the system to the number of target test tasks in the corresponding answer file.

3.18

Equal error rate

By adjusting the decision threshold, the value when the false alarm rate and the false alarm rate are equal.

3.19

Minimum detection cost function According to the false alarm rate weight and false alarm rate weight, the minimum relevant detection loss cost is calculated by adjusting the judgment threshold.

Minimum detection cost function = false alarm rate weight x false alarm rate underreport rate weight x underreport rate 3.20 Real-time factor for enrollment on the test machine, the ratio of the time required for the system to complete the registration task to the total duration of the registered voice.

3.21

Real-time factor for test on the test machine, the ratio of the time required for the system to complete the recognition task to the total duration of the test voice.

3.22

Registration database Contains a collection of registered speakers and their speech.

3.23

Test database

Contains a collection of test speech.

4.1 Registered Speaker Voice

4.1.1

The registered speaker's voice scale is not less than 100,000.

4.1.2

Voice format. windows pcm wav format, mono, sampling rate can be 8KHz ~ 48KHz, sampling bit number is 16bit.

4.1.3

Each registered voice file contains 1 registered voice.

4.1.3

Each registered voice corresponds to only one speaker.

4.1.4

The cumulative duration of the effective voice of each registered speaker is not less than 20s.

4.2 Testing Speaker Voice

4.2.1

The number of test speakers is not less than 500.