

ICS
CCS

DA

中华人民共和国档案行业标准

DA/T XXXXX—XXXX

录音录像档案语音识别转写工作规范

Specification for speech recognition of audio and audio-visual records

(征求意见稿)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

国家档案局 发布

目 次

前 言	1
引 言	2
1 范围	3
2 规范性引用文件	3
3 术语和定义	3
4 总体要求	4
5 组织与管理	5
5.1 工作机构	5
5.2 工作人员	5
5.3 工作场地	5
5.4 管理制度	5
5.5 工作流程	5
5.6 工作文件	5
5.7 元数据	5
6 前期准备	5
6.1 制定工作方案	5
6.2 制定技术方案	6
7 语音识别转写的实施	6
7.1 评估与导入	6
7.2 识别转写	6
7.3 成果修改校正	7
7.4 质量要求	7
7.5 成果整理输出	8
8 语音识别转写成果管理与应用	9
8.1 成果管理	9
8.2 成果应用	9
参 考 文 献	11

前 言

本文件按照 GB/T1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由国家档案局提出并归口。

本文件起草单位：安徽省档案馆、讯飞智元信息科技有限公司、中国人民大学信息资源管理学院。

本文件主要起草人：黄玉明、张海剑、朱霖露、邓晓文、徐拥军、牛力、陈姗姗、李衍、高峰。

引 言

录音录像档案语音识别转写事关档案资源建设和档案工作数字化转型。利用新一代信息技术对录音录像档案进行数字化处理，实现语音识别和转写，已成为保护和利用档案资源的重要途径。本文件旨在明确录音录像档案语音识别转写工作的总体要求和实施规范，为各级档案部门和相关工作人员提供科学、规范的操作指南，推动档案信息资源的深度开发和利用。

录音录像档案语音识别转写工作规范

1 范围

本文件规定了录音录像档案语音识别转写工作的组织、实施、技术和管理要求。

本文件适用于各级各类档案馆、机关、团体、企事业单位和其他社会组织形成的以汉语、常见少数民族语言及国际通用语种录制的录音录像档案语音识别转写工作及相关服务平台建设应用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件。不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 3947-1996 声学名词术语

GB/T 21023-2007 中文语音识别系统通用技术规范

GB/T 34083-2017 中文语音识别互联网服务接口规范

GB/T 41807-2022 信息安全技术 声纹识别数据安全要求

DA/T 62-2017 录音录像档案数字化规范

DA/T 63-2017 录音录像类电子档案元数据方案

DA/T 68.2-2020 档案服务外包工作规范 第2部分：档案数字化服务

DA/T 78-2019 录音录像档案管理规范

3 术语和定义

DA/T 1-2000、DA/T 58-2014 界定的以及下列术语和定义适用于本文件。

3.1

录音录像档案 audio and audio-visual record

具有凭证、查考和保存价值并归档保存的录音录像文件。

[来源：DA/T 78—2019, 3.3]

3.2

语音识别转写 speech recognition

将人类的声音信号转化为文字或指令，同时进行特征要素提取的过程。

注：改写 GB/T 21023—2007, 定义 3.1

3.3

语音识别转写结果 speech recognition result

语音信号经语音识别转写后得到的文字或声纹。

注：改写 GB/T 34083—2017, 定义 3.5

3.4

识别速度 recognition speed

单位时长录音录像档案通过语音识别转写技术获取结果所需要的时间。

3.5

识别准确率 recognition accuracy

通过语音识别转写技术能够正确识别出的文字与实际语音内容之间相符合的程度。

[来源：GB/T 21023-2007, 5.2.1]

3.6

音素 speed sound

最小语音单位。

[来源：GB/T 3947-1996, 10.3]

3.7

声学模型 acoustic model

一种将输入的语音信号转换为对应的音素或音节序列的模型。

3.8

语言模型 language model

一种将声学模型输出的音素或音节序列转换为自然语言文本的模型。

3.9

声纹 voiceprint

人的语音中所蕴含的、能用以表征和标识数据主体的生物学特性和行为特性的总称。

[来源：GB/T 41807—2022, 3.3]

3.10

热词集 hot words

由用户或应用程序定义的结构化关键词集合。

[来源：GB/T 34083-2017, 3.6]

3.11

时间码 time code

一种对语音信号中出现的单词、短语或句子等单元进行时间标记的编码方式。

4 总体要求

4.1 语音识别转写工作应纳入数字档案馆（室）或电子档案管理信息系统建设范畴，统筹规划、有序推进，逐步实现常态化。语音识别转写工作可与录音录像类档案数字化工作同步开展。

4.2 语音识别转写工作应科学实施，成果应与录音录像档案建立准确、可靠的关联关系。语音识别转写工作实行服务外包的，应按照 DA/T 68.2-2020 的要求执行。

4.3 语音识别转写工作应加强过程管理和质量把控，捕获相关元数据，确保工作过程规范、成果可靠、

数据安全。

4.4 语音识别转写工作应以有利于档案信息检索、智能编目、数据挖掘、开发利用为原则。

4.5 语音识别转写工作应符合涉密档案管理、知识产权保护、个人信息保护等要求。

5 组织与管理

5.1 工作机构

应建立语音识别转写工作组织，对组织实施、协调管理、安全保障、技术支持、监督检查、成果验收等进行统筹规划，确保语音识别转写工作顺利开展。

5.2 工作人员

应配备熟悉档案业务且具备语音识别转写工作基础知识的工作人员，可根据需要聘请外部有关技术专家或顾问。

5.3 工作场地

应设置专门语音识别转写工作场地，按 DA/T 68.2-2020 的要求，结合转写工作需要，合理布局档案评估与导入、识别转写、成果修改校正、成果整理输出等工作区域。

5.4 管理制度

应制定并实施科学、规范的语音识别转写工作管理制度，包含岗位管理、人员管理、场地管理、设备管理、数据管理、质量管理、进度管理、安全管理等方面，有效保障档案安全和成果质量。

5.5 工作流程

语音识别转写工作流程包括评估与导入、识别转写、成果修改校正、质量要求、成果整理输出五个业务环节。

5.6 工作文件

应加强对语音识别转写过程中形成的工作文件管理，明确工作文件的整理、归档、移交等管理要求。工作文件主要包括工作方案、技术方案、审批书、流程单、验收报告、成果移交清单等，采取服务外包的应包括项目招标文件、投标文件、中标通知书、项目合同、保密协议等。

5.7 元数据

语音识别转写过程中应完整捕获元数据，元数据项的设置参照 DA/T 63-2017，应包括但不限于转写时长、转写环境、转写引擎、转写时间、成果文件名称、成果文件大小、成果文件格式等元数据。

6 前期准备

6.1 制定工作方案

6.1.1 应依据语音识别转写工作的计划、目标等有关要求，在充分调研的基础上制定工作方案。

6.1.2 语音识别转写工作方案内容包括语音识别转写对象、工作目标、工作内容、经费核算、工作流程、技术方法、人员安排、验收依据、安全管理措施等内容。

6.2 制定技术方案

6.2.1 应根据语音识别转写工作方案制定语音识别转写过程中各系统技术方案，包含语音识别转写引擎、语音识别转写软件、基础设施与网络等，所涉及的系统应满足如下条件：

- 与数字档案馆（室）或电子档案管理信息系统相同的安全保护级别；
- 与相关体系内其他系统接口、数据格式及技术服务框架具有可兼容性；
- 对相关体系内其他系统、网络的功能和性能无明显影响；
- 具备高安全性、可靠性，可在国产软硬件环境下稳定运行。

6.2.2 装备语音识别转写引擎，对录音录像档案进行高速度和高识别准确率的语音识别转写。应满足如下条件：

- 支持语音识别转写文本、文本翻译、文本摘要及声纹提取；
- 具备对噪音的强抵抗能力，能够有效屏蔽较大程度的噪音干扰；
- 具备对出现的异常语音方式（口吃、哼唱、口语化、复述等）处理能力；
- 能准确判别干扰词（停顿词、语气词、重复词等），提高识别准确率。

6.2.3 装备语音识别转写软件，可高速、准确输出识别成果。确认以下各项功能可正确实现：

- 对录音录像档案进行导入及预处理，录像档案自动提取音频；
- 识别语种预测及语音识别转写引擎手动选择；
- 自动语音识别转写；
- 语音识别转写文本后处理及自动校验；
- 输出语音识别转写成果，文本成果支持人工校正，支持热词集管理；
- 数据阅览及导出，文本成果及翻译成果支持与档案同步对应阅览。

6.2.4 基础设施与网络，包括引擎运行服务器、软件应用服务器、工作计算机、有线网络等，应支持系统正常运行。

7 语音识别转写的实施

7.1 评估与导入

7.1.1 语音识别转写实施前，应对开展语音识别转写的录音录像档案质量进行评估，评估内容一般包括语言信息、格式、采样率、信噪比、音量、背景音、时长及文件大小等，应满足以下各项要求：

- 语言信息：以汉语、常见少数民族语言及国际通用语种录制；
- 录音档案格式：WAV、MP3 等格式；
- 录像档案格式：AVI、MPEG、MP4、FLV 等格式；
- 音频采样率：不低于 16kHz，宜为 44.1kHz；
- 音频信噪比：不低于 30db；
- 音量：适中且稳定，避免过小或过大，以减少处理难度和提高识别准确率；
- 背景音：避免有复杂的背景音干扰语音识别转写效果；
- 时长：单个文件应不超过 5 小时；
- 文件大小：单个文件应不超过 10GB。

7.1.2 评估通过后，需将符合语音识别转写要求的录音录像档案导入到语音识别转写软件。对评估不满足要求的应按照 DA/T 62-2017 等相关标准要求重新处理。

7.2 识别转写

7.2.1 音频预处理

7.2.1.1 在识别处理前，应根据音频混响的特征，对输入的语音波形信号进行信号增强和噪声抑制处理。

7.2.1.2 应对音频流分析检测，划分音频信号中语音信号和非语音信号（无声段或背景噪声），确定有效语音起止点，提取语音信号部分。

7.2.1.3 应对声学特征分析提取，自动检测语音识别转写有效的、噪声鲁棒性且区分性好的声学特征，经标准化优化处理后提取，以提升语音识别的准确性和鲁棒性。

7.2.2 音频识别

7.2.2.1 识别时，应自动检测待识别音频的语种，确定语音识别转写引擎语种类型，并支持手动选择。对需翻译为简体中文的应同步设置。

7.2.2.2 应通过声学模型和语言模型的联合建模完成音频识别。

7.2.2.3 声学模型应对语音信号自动抽取特征信息，并映射到语音的音素或词汇上，识别为初步文本。

7.2.2.4 语言模型应对语音信号自动语义分析，结合上下文信息推测出最可能的文本结果。应支持通过系统自动学习或人工自定义方式补充热词集，优化识别结果，以提升文本的生成准确率。

7.2.2.5 宜支持通过语言模型对语音信号中不同语音片段的说话人声纹识别，并注册说话人声纹。声纹识别注册方法参见 SJ / T 11380-2008 的规定。

7.2.3 文本处理

7.2.3.1 应对转写文本进行自动语法修正或手动处理为符合阅读习惯的格式。

7.2.3.2 分句与分段。应对转写文本按语义自动切分成若干个子句与段落，并添加标点。

7.2.3.3 文本顺滑。应对转写文本自动过滤无意义的字或词，如停顿词、语气词、非强调性重复词等。

7.2.3.4 数字转换。应对转写文本自动按照常用书面表达方式转换，设置数字、字母、日期、货币、号码等不同场景的数字表达方式。

7.2.3.5 文本标记。宜对转写文本自动按照语音片段标记对应的说话人。

7.2.3.6 文本翻译。宜对少数民族语言及国际通用语种转写文本自动进行简体中文翻译及规整。

7.2.3.7 文本摘要。宜对转写文本自动提取关键信息，归纳为简短的内容概述。

7.3 成果修改校正

7.3.1 应支持对转写文本进行自动语义识别和校正，通过热词集对识别后文本中的特殊词汇、语句自动分析更正，如人名、地名、机构名、专有名词、多音字、同音不同字等。热词集应具备手动更新、自动更新和自动学习功能。

7.3.2 应支持对转写文本与音频、视频同步阅览，支持以人工方式进行对照、修正等校正功能，以满足更高识别准确率的要求。

7.3.3 宜支持对语音识别转写成果按说话人分离、聚类同步阅览，以满足快速校正的要求。

7.3.4 宜支持语音识别转写成果文本摘要内容的手动调节及重新自动生成，以提高摘要的准确性。

7.4 质量要求

7.4.1 识别准确率

7.4.1.1 语音识别转写对以汉语、常见少数民族语言及国际通用语种录制，声音质量好、说话人发音标准的识别准确率应在 95%以上。以标准汉语普通话录制，其识别准确率应在 98%以上。

7.4.1.2 语音识别转写对声音质量差、杂音多、讲话人发音不标准的识别准确率应在 80%以上。

7.4.1.3 语音识别转写的识别准确率应以 WCR 度量其指标。识别结果中的错误分为插入错误、删除错误与替换错误，相应的错误率为字错误率（Word Error Rate:WER），其与字准确率（Word Correct Rate:WCR）加和为 100%，其计算公式如下：

$$\text{字准确率: WCR} = 100\% - ((S+I+D)/N) \times 100\%$$

式中：

N——正确文本字数；

D——删除错误字数；

I——插入错误字数；

S——替换错误字数；

7.4.2 识别速度

7.4.2.1 识别速度指标与识别准确率指标应同时适用。

7.4.2.2 语音识别转写速度 1 小时音频不超过 10 分钟完成识别转写，1 小时视频不超过 30 分钟完成识别转写。

7.5 成果整理输出

7.5.1 成果整理

7.5.1.1 语音识别转写成果主要包括纯文本文件、声纹文件、字幕文件等。

7.5.1.2 语音识别转写成果应记录不同说话人时间码，应与档案原句中说话人起止点一致，并支持与档案原件同步浏览。录像档案语音识别转写成果，还应支持转写文本以字幕形式同步浏览，成果含翻译文本的应支持双语字幕。

7.5.1.3 应支持对转写成果的调用、编辑、导出、搜索。

7.5.2 成果输出

7.5.2.1 录音档案语音识别转写成果应保存为纯文本文件、声纹音频文件。录像档案语音识别转写成果应保存为纯文本文件、视频字幕文件、声纹音频文件。

7.5.2.2 应以录音录像档案的件为单位输出，单独保存语音识别转写成果为附件。语音识别转写成果保存规则参见表 1：

表 1 语音识别转写成果保存规则

档案类型	录音录像档案语音识别转写成果	成果保存格式
录音档案	识别文本、翻译文本、摘要文本	TXT 文件
	说话人声纹	PCM 文件
录像档案	识别文本、翻译文本、摘要文本	TXT 文件
	说话人声纹	PCM 文件
	识别文本、翻译文本	SRT 文件

7.5.2.3 应以档号为基础对语音识别转写成果命名，应确保语音识别转写成果命名的唯一性。一件档案保存为多个语音识别转写成果时，应按档号结合语音识别转写成果顺序流水号为语音识别转写成果命名。

示例 1：档号为 X001-LY·2025-D30-0001 的单人录音档案，对应的语音识别转写成果文件名为 X001-LY·2025-D30-0001_0101.txt 和 X001-LY·2025-D30-0001_0201.pcm。

示例 2：档号为 X001-LY·2025-D30-0001 的两人录音档案，对应的语音识别转写成果文件名为 X001-LY·2025-D30-0001_0101.txt、X001-LY·2025-D30-0001_0201.pcm、X001-LY·2025-D30-0001_0202.pcm。

示例 3：档号为 X001-LX·2025·Y-0001 的单人录像档案，对应的语音识别转写成果文件名为 X001-LX·2025·Y-0001_0101.txt、X001-LX·2025·Y-0001_0201.pcm、X001-LX·2025·Y-0001_0301.srt。

示例 4：档号为 X001-LX·2025·Y-0001 的两人录像档案，对应的语音识别转写成果文件名为 X001-LX·2025·Y-0001_0101.txt、X001-LX·2025·Y-0001_0201.pcm、X001-LX·2025·Y-0001_0202.pcm、X001-LX·2025·Y-0001_0301.srt。

7.5.2.4 应保存档案原件与语音识别转写成果时间码对应关系，方便对档案的快速检索定位阅览。

7.5.2.5 应支持按照档案著录规则和元数据规范，自动保存相关元数据。相关元数据应纳入数字档案馆（室）或电子档案管理信息系统中统一管理。

7.5.3 成果验收

7.5.3.1 应采用计算机自动检验与人工检验相结合的方式对语音识别转写成果进行验收检验。

7.5.3.2 验收检验内容包括语音识别转写文本、翻译文本、摘要文本、语音文字同步阅览情况、说话人标记情况、元数据、工作文件和存储载体等。

7.5.3.3 能够采用更换语音识别转写引擎或比较先后两次语音识别转写结果的计算机自动检验方式进行 100%检验，对于无法用计算机自动检验的，应对语音识别转写成果进行人工抽检验收，被抽检录音录像档案个数及时长总和均不应低于该批次的录音录像档案的 5%。

8 语音识别转写成果管理与应用

8.1 成果管理

8.1.1 应保持语音识别转写成果与对应档案原件、目录、元数据之间的逻辑层次和关联关系。

8.1.2 语音识别转写成果输出文件应与档案原件统一保存。

8.1.3 语音识别转写成果中，声纹音频文件的存储、管理、传输应符合 GB/T 41807-2022 的规定。

8.1.4 应建立语音识别转写成果权限管理机制，与录音录像档案原件相同。

8.1.5 语音识别转写成果应与录音录像档案原件同步开展数据备份、长期保存等工作。

8.2 成果应用

8.2.1 应建立语音转写成果索引库并与录音录像档案原件对应，通过数字档案馆（室）或电子档案管理信息系统实现全文检索，支持档案原文溯源查询及文本摘要信息的快速阅览。

8.2.2 可利用语音识别转写成果，结合数据挖掘技术，开展关键实体要素标签库建设，辅助档案自动著录。

8.2.3 可利用语音识别转写成果，结合声纹识别技术，构建说话人人物声纹库，通过相关信息系统实现说话人确认、说话人辨认、说话人检出及说话人追踪等应用。

8.2.4 可利用语音识别转写成果，构建录音录像档案编研数据库，辅助开展档案编研开发、专题展陈。

8.2.5 可利用语音识别转写成果，结合人工智能技术，开展知识关联管理、内容推荐、开放审核等工作。

参 考 文 献

- [1] GB/T 1.1-2020 标准化工作导则 第 1 部分：标准化文件的结构和起草规则
 - [2] GB/T 18894-2016 电子文件归档与电子档案管理规范
 - [3] GB/T 20000 标准化工作指南
 - [4] GB/T 20001 标准编写规则
 - [5] GB/T 31497-2024 信息技术 安全技术 信息安全管理监视、测量、分析和评价
 - [6] GB/T 36464.1-2020 信息技术 智能语音交互系统 第 1 部分：通用规范
 - [7] GB/T 37036.5-2023 信息技术 移动设备生物特征识别 第 5 部分：声纹
 - [8] GB/T 41813.1-2022 信息技术 智能语音交互测试方法 第 1 部分：语音识别
 - [9] DA/T 31-2017 纸质档案数字化规范
 - [10] DA/T 77-2019 纸质档案数字复制件光学字符识别（OCR）工作规范
 - [11] 档案数字化外包安全管理规范
-