

团 体 标 准

T/xxx —2023

语音转写准确率检测方法

The method for detecting the accuracy of speech transcription

（征求意见稿）

2023 - XX - XX 发布

2023 - XX - XX 实施

广东省质量检验协会 发布

目 次

1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 检测过程和检测分析内容.....	3
5 测试流程.....	4
6 标注文本标记方法.....	4
7 转写准确率计算.....	5
参 考 文 献.....	6

前 言

本文件按照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》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这些专利的责任。

本文件由中移互联网有限公司提出。

本文件由广东省质量检验协会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布

语音转写准确率检测方法

1 范围

本文件规定了视频通话语音转写业务端到端业务质量指标，供业务试点及商用阶段进行语音转写业务端到端质量分析和管理使用。

本文件规定了面向视频通话语音转写业务规模应用的测试方案，适用于通信公司语音转写业务规模试验阶段的使用。

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。

SF/Z JD0403003-2015 计算机系统用户操作行为检验规范

SF/Z JD0403004-2018 软件功能鉴定技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

视频通话 video call

视频通话语音转写被定义为在视频通话场景下进行的语音转写过程。该过程涉及将视频通话中的口头语言或语音信息转换为书面文本形式。视频通话语音转写要求考虑特定于视频通话的环境和要求，确保在这种情境下实现准确的语音转写，适应各种语音质量和网络状况，以及满足特定领域的需求。

3.2

语音转写 phonetic transcription

语音转写是将口头语言或语音信息转换为书面文本形式的过程。该过程通常涉及使用自动语音识别(ASR)技术，通过将语音信号转换为文本，实现对语音内容的准确记录。语音转写在不同场景中有广泛的应用，包括但不限于会议记录、语音搜索、语音助手等。

3.3

业务质量指标 business quality metrics

业务质量指标被定义为用于衡量视频通话语音转写业务性能和效果的一组标准化指标。这些指标旨在评估语音转写系统的准确性、时效性、稳定性以及对视频通话场景的适应性。业务质量指标可能包括但不限于词错误率(WER)、字符错误率(CER)、句子错误率(SER)等，以综合评估语音转写在业务中的质量水平。

3.3.1 词错误率 word error rate

- WER是用于衡量语音转写准确性的指标之一。
- WER表示在转写结果中与参考文本不匹配的词的百分比。

- 计算方法包括插入（系统多转写了什么不在参考中的词）、删除（系统漏掉了参考文本中的哪些词）、替换（系统替换了参考文本中的哪些词）的操作数。

3.3.2

字符错误率 character error rate

- CER是另一种用于评估语音转写准确性的指标。
- CER表示在转写结果中与参考文本不匹配的字符的百分比。
- 与词错误率类似，计算方法包括插入、删除和替换的字符操作数。

3.3.3

句子错误率 sentence error rate

- SER是用于评估整体语音转写准确性的指标。
- SER表示在转写结果中与参考文本不匹配的句子百分比。
- 这是一个更高层次的评估，考虑了完整句子的正确性。

3.4

字准确率 word accuracy

视频通话过程中，语音转化出来的文字中，转写正确的字数占所有字数的百分比。

3.5

语料 corpus

为5G新通话业务提供语音物料，具体为一定时长的语音文件。

3.6

原始文本 original text

指与语料内容相匹配的文字文本。

3.7

总字数 total word count

特指原始文本文字总长度。

3.8

转写文本 transliterated text

由5G新通话提供的能力对语料转写得出的字幕输出的文本文件。

3.9

标注文本 annotated text

基于原始文本，将转写文本与原始文本的区别（包括多转写字词和漏转写字词）标注出来的文本文件。

3.10

主叫 caller

5G新通话业务进行中，发起视频通话的一方终端。

3.11

被叫 callee

5G新通话业务进行中，接受视频通话的一方终端。

3.12

测试终端 test terminal

为测试5G新通话业务提供测试载体的手机。

3.13

安静环境 quiet environment

此处特指噪音音量不高于40分贝的环境下。

3.14

字词 word

最小表义单位，可以是单字、双字词或多字词。

3.15

实义词 notional word

包括名词、动词、形容词、数词、量词、副词。

4 检测过程和检测分析内容

4.1 检测要素

4.1.1 所需硬件

所需硬件应包含：

- a) 音频播放设备；
- b) 支持5G网络（VoNR）和视频通话功能的安卓手机；
- c) 已开通转写业务的测试号卡。

4.1.2 所需软件

所需软件应包含：

- a) 具备图像帧数截取功能的视频播放软件；
- b) 手机端录屏软件，可以同时录制视频和音频。

4.1.3 所需素材

所需素材应包含：

- a) 语种为普通话的测试语料音频文件，要求普通话等级为甲等一级，语料播音清晰，无明显杂音，语料长度不少于1000字，语速均衡，语料内容应无自创词，语义应可被大众理解。
- b) 上述测试语料的文字原始文本，要求文本字句与语料一一对应，且无明显语法错误，字句词语没有明显歧义。

4.2 检测前准备工作

检测前准备工作应包含：

- a) 确保所处环境相对安静，需要在没有明显杂音和人声干扰的安静环境下，噪音音量不大于40分贝；
- b) 所处环境网络信号强度稳定在-80dBm以上；
- c) 确保测试号卡已开通转写业务且测试终端能够正常收音；
- d) 音频播放设备可正常放音，音量不低于50分贝。

4.3 检测环境的搭建

检测环境的搭建应包含：

- a) 当前处于安静的环境中，区分测试终端A、B分别为主叫和被叫，且A所在空间声音无法传至B处；
- b) 测试终端A作为主叫，对测试终端B发起视频呼叫，测试终端B应答呼叫；
- c) 连接视频通话后，测试终端B打开屏幕录制工具，进行录制；
- d) 音频播放设备对测试终端A播放测试语料，查看测试终端B确认转写功能开启；
- e) 结束通话与视频录制，确认视频画面与声音可正常播放。

5 测试流程

测试流程宜满足 SF/Z JD0403004-2018 的要求，包括但不限于以下要求：

- a) 处于安静环境下，音频播放设备准备好待播放的语料，确认播放位置处于语料起点；
- b) 测试终端 A 作为主叫，对测试终端 B 发起视频呼叫，测试终端 B 应答呼叫并开始视频录制；
- c) 测试终端 B 能正常接收并外放语料播音内容，且屏幕下方的字幕框中持续出现由语料转写出的字幕，等待语料播放结束；
- d) 语料播放结束，挂断电话并停止录屏；
- e) 输出该录屏文件，使用视频播放软件导入该文件，查看转写后的文本，并将写文本转录至 word 文档，形成转写文本；
- f) 检测记录，与鉴定活动有关的情况及时、客观、全面地记录，保证测试过程和结果的可追溯性，包括检测时间、检测人员以及设备信息等。

6 标注文本的标记方法

6.1 文本比对方法

6.1.1 比较原则

以短句为单位将原始文本与转写文本进行比对。由句号、逗号、问号、感叹号、分号、冒号、存折号分隔的文本均可认为是短句，另外，如转写文本中出现了空格或换行，那么由空格和换行分隔的文本也可认为是短句。

6.1.2 短句文本对比

短句文本对比应包含以下内容：

- a) 在转写文件中查找原始文本的第一个字词：如找到，则确认起始位置，继续查找第二个字词；如在转写文件中未能找到原始文本的第一个字词，则查找第二个字词，如未能找到则继续查找，直至本句原始文本结束；
- b) 找到第一个字词后，从转写文本第一个字词处开始，按查找第一个字词方法，查找原始文本第二个字词，直到查找完成所有原始文本；
- c) 将查找到的转写文本字词与原始文本进行对照，如转写文本覆盖原始文本字词数量不少于原始文本的 35%，且存在实义词，则认为短句对比有效，对转写文本进行标注，标注方法参考本章后续内容；如不满足上述条件，则认为短句对比无效，整句标注为多转写。

6.1.3 短句查找

短句查找应包含以下内容：

- a) 将原始文件中的第一句与转写文件第一句进行对比：如短句对比有效，则依次进行第二句的对比；如对比无效，记录对比无效的短句位置；
- b) 当原始文件中的第一句与转写文件第一句对比无效时，将原始文件中的第一句与转写文本的第二句进行对比：如对比有效，则依次将原始文本与转写文本的下一句进行对比；如对比无效，则使用原始文本的第二句与转写文本第二句进行对比；
- c) 在 b) 情况下，如原始文本的第二句与转写文本第二句对比有效，则将原始文本第一句整句记为漏转写，如对比无效，则将原始文本第一句与转写文本第三句进行对比，依此类推

- d) 将上述方法整理如下,设原始文本第 x 句 O_x 与转写文本第 y 句 T_y 对比有效,如尚未开始对比,则 x 与 y 取 0, 设原始文本共有 m 个短句, 转写文本共有 n 个短句, 则使 O_{x+1} 与 T_{y+1} 进行对比, 如对比有效, 则令 $x=x+1$, $y=y+1$, 如对比无效, 则使 T_{y+z} 与 O_{x+1} 至 O_{x+z} 的阶乘进行对比, 其中 $z>2$, 每对比一轮且对比结果均为无效则 z 加 1, 如对比有效, 则新一轮对比的 x 与 y 分别取对比有效的原始文本与转写文本的位置。中间跳过的原始文本均记录漏转写, 跳过的转写文本均记为多转写。

6.2 标注转写文本中的错误转写字词

将原始文本原样保留在标注文本中, 其中未能转写正确的字词, 为转写错误字词, 在标注文本中标记为红色。

6.3 标注转写文本中的多转写字词

转写文本两个转写正确字词之前的文本为多转写字词, 按转写文本中的位置放置在标注文本中, 并标记为蓝色。

6.4 标注转写文本中的漏转写字词

两个转写正确字词之间, 仅在原始文本有字词而在转写文本中为空的字词为漏转写字词, 按原始文本中的位置放置在标注文本中, 并标记为绿色。

7 转写准确率计算

转写字幕准确率计算, 行业通用的字准确率和行准确率两个指标。本文件选择字准确率, 指标计算公式如下:

$$x = \frac{A}{B} \times 100\%$$

式中:

x ——字准确率;

A ——正确识别的单词数量;

B ——总单词数量。

注: 假设有100个单词, 语音识别系统输出的识别结果有90个单词是正确的。那么字准确率可以计算如下:

$$\text{字准确率} = (90 / 100) \times 100\% = 90\%$$

参考文献

- [1] 《中国移动5G新通话业务总体技术要求》
 - [2] 《中国移动5G新通话业务测试规范》
 - [3] 《中国移动5G新通话业务质量指标体系规范》
 - [4] 《中国移动VoNR+能力网元北向接口规范》
 - [5] 《中国移动5G VoNR+总体技术要求》
 - [6] 《中国联通LTE网络业务测试规范语音业务手册》
-