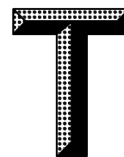


ICS 35.240
CCS L 67



团 体 标 准

T/TAC 11—2024

机器同声传译系统基本要求与测试方法

Basic requirements and test methods for machine simultaneous translation system

2024-03-30 发布

2024-04-30 实施

中国翻译协会	发 布
中国标准出版社	出 版

目 次

前言 III

引言 IV

1 范围 1

2 规范性引用文件 1

3 术语和定义、缩略语..... 1

4 机器同声传译系统参考框架 2

5 基本要求 3

6 测试方法 7

附录 A（规范性） 平均擦除率的计算 12

参考文献 13

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专利的责任。

本文件由中国翻译协会提出并归口。

本文件起草单位：科大讯飞(上海)科技有限公司、中国外文局翻译院、鹏城实验室、北京百度网讯科技有限公司、阿里巴巴(中国)网络技术有限公司、华为终端有限公司、小米通讯技术有限公司、安徽听见科技有限公司、中译语通科技股份有限公司、合肥智能语音创新发展有限公司、科大讯飞股份有限公司、中国科学院自动化研究所、上海交通大学、上海华东电信研究院、北京外国语大学、上海外国语大学、广东外语外贸大学、西安外国语大学。

本文件主要起草人：张为泰、刘俊华、刘聪、胡国平、刘颖、王飞、邢玉堂、杜楠、梁倩如、相洋、孙萌、杨宝嵩、张穗云、李响、王玮、宗浩、吴江照、张家俊、陈谐、彭莉、王华树、张爱玲、戴光荣、曹达钦。

引 言

随着跨语言国际会议、文化传播和沟通交流日益增多,机器同声传译的需求与日俱增。机器同声传译是一种通过机器实现的实时翻译技术,能够在不打断说话人的情况下对说话内容进行翻译。这项技术是我国人工智能产业中的重要技术布局方向,具备客观信息的识别与记忆完整度高、术语与词汇库大、可低成本长时间高效工作等优势,因此,利用机器同声传译辅助或部分场景代替人工翻译已成为一种趋势。

然而,在机器同声传译技术的发展过程中,存在语音识别难度高、口语化问题严重、延迟与效果难以平衡、上下文难以实现一致性与连贯性等问题。同时,在机器同声传译产品的创新过程中,产品定位、功能、效果等缺少统一标准,导致需求方面临选择困扰。

为了满足中国国际化交流的需求,确保机器同声传译系统的质量,亟须制定机器同声传译系统基本要求与测试方法标准。本文件旨在为评价机器同声传译系统的优劣提供统一的评价标准,从而推动机器同声传译技术、产品以及行业的规范发展。

机器同声传译系统基本要求与测试方法

1 范围

本文件规定了机器同声传译系统的系统参考框架、基本要求和测试方法。

本文件适用于机器同声传译系统的设计、开发、应用和维护,也可用于指导第三方测评机构对机器同声传译系统进行功能评估和系统验收。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义、缩略语

3.1 术语和定义

下列术语和定义适用于本文件。

3.1.1

语音识别 automatic speech recognition

将人类的声音信号转化为文字或指令的过程。

[来源:GB/T 21023—2007,3.1]

3.1.2

语音合成 text to speech

通过机械的、电子的方法合成人类语音的过程。

注:该过程所产生的语音称为合成语音,和人的发音器产生自然语音相区别,有时也叫人工语音(artificial speech)。

[来源:GB/T 21024—2007,3.1]

3.1.3

机器翻译 machine translation

使用计算机应用程序将一种自然语言自动翻译成另一种自然语言。

[来源:GB/T 40036—2021,3.1.1,有修改]

3.1.4

机器同声传译系统 machine simultaneous translation system

利用语音识别(3.1.1)、语音合成(3.1.2)、机器翻译(3.1.3)等技术,实现从一种自然语言到另一种自然语言的同声传译过程的自动化系统。

3.1.5

虚拟人合成 virtual human synthesis

通过数字图像处理、语音合成技术生成的虚拟人物形象、表情、动作、语音的音视频数据过程。

3.1.6

热词 hot word

命名实体、专业术语、新词以及在语音识别或机器翻译过程中实现难度大的词汇。