教育研究前沿进展
ISSN: 2789-5521(print)

雙清學術出版社
Contents lists available at www.qingpress.com
Journal homepage: qingpress.com/zh-cn/journals/1

“网易见外”语音转文字质量评估研究

黄旦华
浙江越秀外国语学院，浙江绍兴，中国
406816705@qq. com

摘要：随着计算机技术和人工智能的发展，语音转文字的识别质量得到极大的提高，也被广泛的应用于商业、教育、和翻译等领域。本文以西班牙学者罗梅罗--弗雷斯科（Pablo Romero-Fresco）等人提出的“NER 模型”为基础，采用“NR 模型”，以 TED 的演讲为素材，使用“网易见外”将其转换为文字，通过详细分析 4 种错误类别，3 个级别的错误，最终计算出语音转写的准确率—这也是评估语音转文字质量的重要标准。结果显示，语音识别虽然识别准确率达到了 98%，但还是存在不同类型的错误，后期需要人工积极干预才能取得满意的结果。研究结果还表明，“网易见外”在断句的准确性、识别单词的准确性方面还有较大的改进空间。

关键字：语音转文字，“NER 模型”，“NR 模型”,准确率

1. 引言

过去的十几年里，以人工智能，大数据，云计算为代表的计算机技术飞速发展，对语言服务行业带来了巨大的影响。让机器能够理解人类的语言是计算机专家和语言学家孜孜不倦的追求。语音识别技术突飞猛进，从萌芽，发展到成熟，已经得到了广泛的运用。国外科技公司亚马逊， IBM， 微软都推出了比较成熟的语音转文字平台，中国的科技公司比如网易、百度、科大讯飞、阿里、腾讯也不甘落后，不断追赶，努力从追随者成为领跑者，他们的产

2789-5521/© Shuangqing Academic Publishing House Limited All rights reserved.
Article history: Received September 28, 2022 Accepted October 7, 2022 Available online October 8, 2022
To cite this paper: 黄旦华(2022). “网易见外”语音转文字质量评估研究. 教育研究前沿进展. 第 2 卷, 第 2 期, 19-29.
Doi: https://doi.org/10.55375/jerp.2022.2.10

作者：黄旦华（1982. 11-），湖北黄冈人，硕士，研究方向：翻译教学，翻译技术。

基金项目：本文系教育部 2021 年第二批产学合作协同育人项目（编号 202102483011），浙江省十三五省级产学合作协同育人项目(浙教办函〔2019〕365)的阶段性成果。

品占据了国内市场较大的份额。“网易见外 – AI 智能语音转写听翻平台”具有音视频翻译及转写功能(见 jianwai.youdao.com)。本文主要研究狭义的语音识别，也就是语音转文字，以罗梅罗--弗雷斯科（Pablo Romero-Fresco）等人提出的研究实时字幕的翻译质量的“NER 模型”为基础[1]，对模型的内容进行了优化，以期对语音转换文本的质量进行科学有效地评估。

2. 文献综述

国外学者研究了如何将语音转文字技术应用于不同的领域的效果。比如，语音转文字技术为改进医疗记录的记录过程、降低记录信息的成本和时间、提高记录质量、提高为患者提供的服务质量以及为医院提供法律保障方面提供了机会[2]。语音转文字技术可运用于英语教学，可以有效提高学生的学习效果和专注力，研究发现大多数的学生对于将语音转文字用于英语学习持积极态度[3]。语音转文字与计算机辅助翻译技术相结合应用于在线跨文化学习活动，检验跨文化交际过程中不同语言的语音转文字和计算机辅助翻译转换文本的准确率，指出了存在的问题及解决方法。

国内学者研究了语音转文字技术应用于大学英语口语训练的效果。比如，采用行动研究法进行研究，结果显示，语音转文字的 APP 练习在提高学生的英语口语能力方面是非常有效的，定性分析数据表明这样的练习可以更好的激发学生的学习热情，让他们更好的参与学习[4]。以“语音转文字”为主题在中国知网进行检索，结果显示，相关研究涉及的中文期刊仅有 10 篇，外文学术期刊有 24 篇；以“语音转文本”为主题在知网进行检索，相关研究涉及的中文学术期刊有 14 篇，外文学术期刊有 2 篇。检索结果表明，国内学者在该领域的研究成果还不多。以“语音识别”为主题在中国知网进行检索，相关研究涉及的中文学术期刊相关的文章是 8746 条，外文的学术期刊相关的文章是 2.06 万条，检索结果反应，国内学者和国外学者在该领域的研究还存在较大差距。

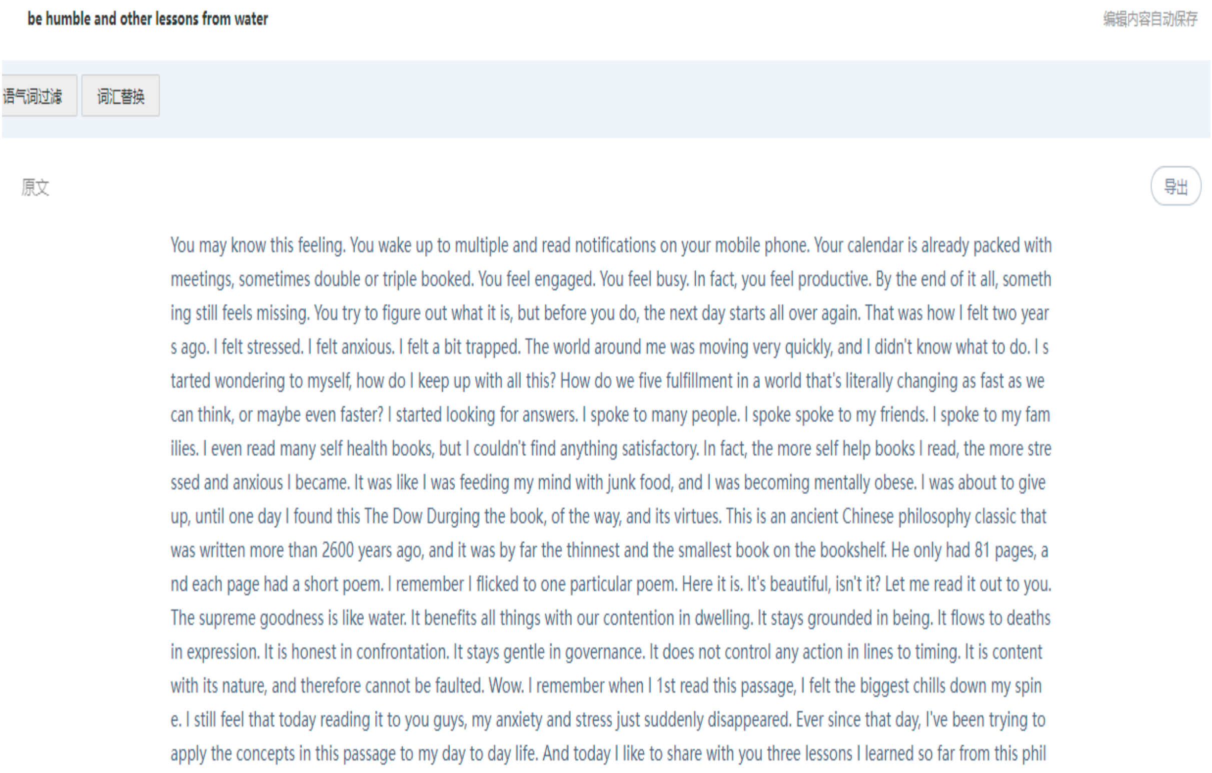
3. 文本质量评估

3.1 研究方法

本研究选取 TED 官网的视频为研究样本。该视频涉及的演讲者为 Raymond Tang。Raymond Tang 具有华裔背景，在海外跨国公司工作，英语发音较为标准。视频中演讲的题目是：保持谦逊——以及其他的“水之道”（Be humble -- and other lessons from the philosophy of water）。整个演讲视频包含片头片尾共 9 分 33 秒。我们通过 TED 官网获得演讲稿，并通过 Office 对数字进行统计。统计结果显示，一共有 1373 个英语单词，最后计算的 WPM(words per minute)四舍五入后为 144 个单词(TED 的总时长是 9 分 33 秒，按照 10 分钟，每分钟 144 个单词,共 1440 个单词。9 分 33 秒，共 1373 个单词，大约就是每分钟 144 个单词)，根据 Virtual Speech 的调查，平均的演讲语速范围为 100-150 WPM，因此本视频语速适中[5]。语言方面，演讲者虽然在中国出生，但是具有丰富的海外学习工作经验，因此英语的语音语调都是比较标准地道，没有明显的口音； 且 TED 演讲的内容都是前期经过精

心准备和彩排过的，在演讲过程中没有不正常的停顿和语气词。内容方面，演讲的内容主要是对中国古典哲学《道德经》的感悟与分享，不涉及专业的领域，几乎没有涉及专业词汇，也没有使用难以理解的修辞，因此素材难度适中。

“网易见外” 语音转写功能支持音频文件上传，且具有语气词过滤和词汇替换功能，等待转换完成，即可获得转换后的文本，可在线编辑文档，也可以导出后进行编辑。（见图1）。笔者首先下载 TED 官网的视频，为了方便直接使用“网易见外”的语音转文字功能（转写功能），下载视频后使用格式转换工具将该视频从 MP4 格式转换为 MP3 格式，然后直接导入网易工作台，自动完成转写，对结果不进行任何修改。作者主要借鉴 Pablo Romero-Fresco Juan Martínez 提出字幕翻译中的 NER 模型，将转换后的文本与官网的文本进行对比分析，评估转换的质量以及错误的类型。



be humble and other lessons from water

编辑内容自动保存

You may know this feeling. You wake up to multiple and read notifications on your mobile phone. Your calendar is already packed with meetings, sometimes double or triple booked. You feel engaged. You feel busy. In fact, you feel productive. By the end of it all, something still feels missing. You try to figure out what it is, but before you do, the next day starts all over again. That was how I felt two years ago. I felt stressed. I felt anxious. I felt a bit trapped. The world around me was moving very quickly, and I didn't know what to do. I started wondering to myself, how do I keep up with all this? How do we five fulfillment in a world that's literally changing as fast as we can think, or maybe even faster? I started looking for answers. I spoke to many people. I spoke spoke to my friends. I spoke to my families. I even read many self health books, but I couldn't find anything satisfactory. In fact, the more self help books I read, the more stressed and anxious I became. It was like I was feeding my mind with junk food, and I was becoming mentally obese. I was about to give up, until one day I found this The Dow Durging the book, of the way, and its virtues. This is an ancient Chinese philosophy classic that was written more than 2600 years ago, and it was by far the thinnest and the smallest book on the bookshelf. He only had 81 pages, and each page had a short poem. I remember I flicked to one particular poem. Here it is. It's beautiful, isn't it? Let me read it out to you. The supreme goodness is like water. It benefits all things with our contention in dwelling. It stays grounded in being. It flows to deaths in expression. It is honest in confrontation. It stays gentle in governance. It does not control any action in lines to timing. It is content with its nature, and therefore cannot be faulted. Wow. I remember when I 1st read this passage, I felt the biggest chills down my spine. I still feel that today reading it to you guys, my anxiety and stress just suddenly disappeared. Ever since that day, I've been trying to apply the concepts in this passage to my day to day life. And today I like to share with you three lessons I learned so far from this phil

图 1： “网易见外”语音转写结果

Figure 1 The screenshot of transcription by Sight Youdao

3.2 NER 模型

NER 模型主要用于评估语内实时字幕，以错误分析为基础，即研究人员统计分析字幕中出现的错误，扣除相应的分数，从而计算字幕的最终得分[6]。NER 模型是以 NERD 模型为基础。[3]NER 模型将字幕翻译的错误分为两类：编辑错误和语音识别错误。且进一步将编辑错误分成增加信息或删减信息，从而导致信息缺失或产生错误的信息，根据错误的严重程度，分别扣除 1 分，0.5 分和 0.25 分。语音识别错误由发音错误/听错或由用于制作字幕的

特定技术引起，分为增加（insertion)、删减(deletion)或替换(substitution)三种错误。同样根据错误的严重的程度，分别扣除 1 分，0.5 分和 0.25 分。 该模型的计算公式为 Accuracy Rate=$\frac{N-E-R}{N}$ ×100。N 表示 Number of words 指的是原文本的单词数量和标点符号数量的总和。E 表示 Edition Errors. R 表示 Recognition Errors.

3.3 NR 模型

由于本研究主要是评估语音转文字的质量，只会对比分析原文本和转换后的文本，不会对文本的内容进行编辑，因此，在借鉴该模型的基础上，去掉了编辑错误这个参数。最终该模型的公式为 Accuracy Rate=$\frac{N-R}{N}$ ×100。N 表示 Number of words 指的是原文本的单词数量和标点符号数量的总和。R 表示 Recognition Errors。错误的类型分为增加（Insertion)，删减(Deletion)，替换(Substitution)。

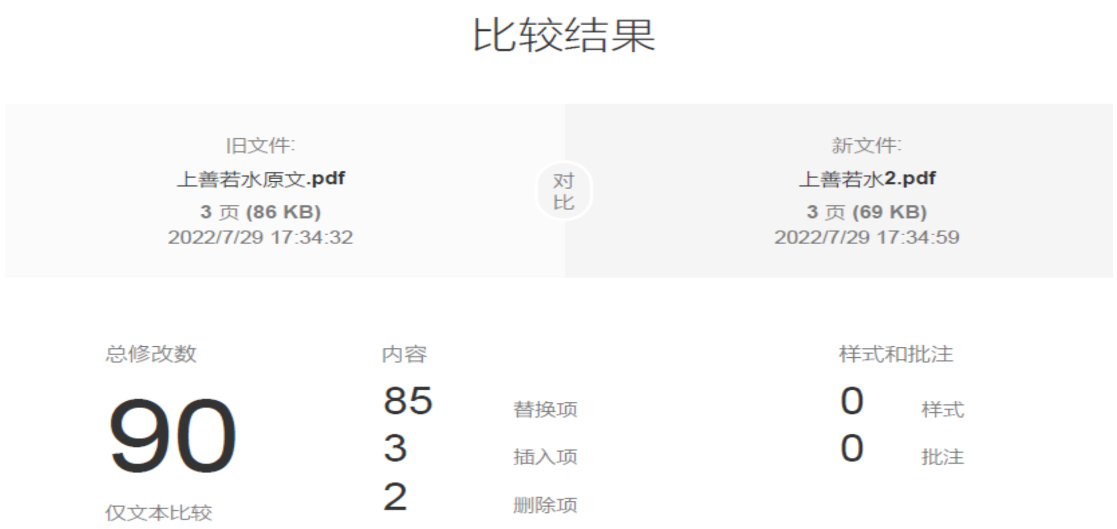
基于 NR 模式，我们可以将原文本和转换后的文本进行对比来识别区分这些错误，错误的类型按照程度可以分为严重（Serious）, 一般(Standard)和轻微(Minor)。每个严重的错误扣 1 分，一般错误扣 0.5 分，轻微错误扣 0.5 分。 Pablo Romeo 在 NER 模型中，列出了三种错误类型，他提出该模型主要是研究实时字幕的翻译质量，在字幕翻译中标点符号往往会以空格的形式取代，所以在识别错误中并没有提出标点符号的错误。但是在文本转换中除了单词的准确影响文本的质量，还有其他的因素也会影响文本的质量。标点符号涉及到断句，其位置也会影响文本的意思。

在上述错误类型的基础上，作者提出了标点符号错误（Punctuations)，将识别错误分成了四类，也就是 Recognition Errors=Insertion + Deletion + Substitution + Punctuation。

3.4 文本转换错误分析

完成对转换后的文本进行错误统计与分类任务时，作者采用了工具和人工相结合的方式。对比文本之前，在不改变文本的内容和标点符号的情况下，将两份文档的格式设置成一致，例如字体、字号和行距都是一致的，然后将其转换成 PDF 文档，并使用 Adobe Acrobat 工具的对比功能将两份文档进行对比。（见图 2）

Adobe Acrobat 对齐两份文档之后显示总修改数是 90，也就是说转换后的文档和原文档有 90 处不一致，即 90 处错误。其中 85 处替换错误，3 处增加错误，2 项是删减错误。该软件能够快速高效的标注两份文件的不同之处，提高了对比的效率，但是该软件的错误分类，与 NR 模型中的错误分类标准并非一致，且该软件没有单独列出标点符号的错误，把该错误列入了替换错误。因此后期需要人工进行干预，将标点符号的错误进行单独分类，避免重叠将标点符号的错误重复计算为替换错误。例如“feeling: you”识别成“feeling. You” 。作为一个标点符号错误，而不是替换错误。



比较结果

旧文件:
上善若水原文.pdf
3 页 (86 KB)
2022/7/29 17:34:32

对比

新文件:
上善若水2.pdf
3 页 (69 KB)
2022/7/29 17:34:59

总修改数	内容		样式和批注	
90	85	替换项	0	样式
	3	插入项	0	批注
仅文本比较	2	删除项		

图 2 Adobe Acrobat 文件对比结果

Figure 2 The comparison result by Adobe Acrobat

通过使用工具以及后期人工的分析对比，最终计算出转换后的文本与原文本共有 96 处不一致的地方。但是经过笔者对比分析，发现有 17 处标点符号的不一致问题，完全可以有不同的标注，且完全不影响意思的理解，例如：原文本“ I felt stressed; I felt anxious. I felt a bit trapped.” 转换后的文本是 “I felt stressed. I felt anxious. I felt a bit trapped.”分号转换成了句号，完全是可行的。又如原文:“The world around me was moving very quickly. And I didn’t know what to do. ”转换后的文本是“The world around me was moving very quickly, and I didn’t know what to do. ”转换前是 独立的句子，转换后是并列句，完全符合语法规范，对意思的理解也没有任何影响。因此这些不一致，不被视为错误。替换导致的错误例如 First 识别成 1st 也不是错误。还有重复的错误不反复统计，例如 Tao Te Ching(道德经)2 次错误拼写。 经过统计最终有 73 处错误。(见表 1)

表 1 错误类型及错误严重程度

Table 1 Category and severity of errors

错误程度 **错误类型**	**严重 Serious**	**一般 Standard**	**轻微 Minor**
增加 Insertion	0	0	2 个 0.50 分
删减 Deletion	0	0	3 个 0.75 分
替换 Substitution	2 个 2 分	19 个 9.5 分	26 个 6.5 分
标点符号 Punctuation	0	8 个 4 分	13 个 3.25 分

根据“NER 模型”中的错误评判标准，严重的错误会改变原文本的意思，在特定语境下可能产生新的意思，且指出严重错误通常是由替换导致的。经过对比发现转换后的文本存在 2 个严重错误。例如“Nor”识别成了“Now”， 导致了句子结构从倒装句变成了疑问句。“It aligns to”识别成了“in lines to”且将原文的疑问句变成了陈述句，这些错误不仅改变了句子的结构，改变了原文的意思，所以算作严重错误。（见表 2）

表 2 严重错误示例

Table 2 Examples of serious errors

原文本	转换后的文本
It doesn’t actually draw any attention to itself, nor does it need any reward or recognition.	It doesn’t actually draw any attention to itself, Now, does there need any reward or recognition?
Does this align to my nature?	This is a line to my nature.

一般错误不会产生新的意思，但是会导致原文本信息的缺失。经过对比分析统计共有 21 个错误，19 个属于替换错误。例如原文本中的“stay behind”识别成了“stake behind”，“unfulfilled”识别成了“unforfill”，“fintech”识别成了“fintexd”，这些错误单词没有产生的新的意思，因此属于一般错误。（见表 3）令人意外的是网易见外识别的单词例如“unforfill”，“fintex”在字典中并不存在，本身拼写都是错误的，也说明“网易见外”识别的算法还是存在改进的空间。

表 3 一般替换错误示例

Table 3 Examples of normal substitution errors

原文本	转换后的文本
Now, we can stay behind closed doors and continue to be paralyzed by our self-limiting beliefs, such as: I will never be able to talk about Chinese philosophy in front of a huge audience.	Now, we can stake behind closed doors and continue to be paralyzed by our self limiting belief, such as, I’ll never be able to talk about Chinese philosophy in front of a huge audience.
whenever I feel stressed, unfulfilled, anxious	whenever I feel stressed, unforfill, anxious
long before the days of bitcoin, fintech and digital technology	long before the days of Bitcoin, fintex and digital technology

一般错误中有 2 个标点符号错误，主要是由于符号位置不对导致断句错误。原本属于一

个句子的成分被分隔到另外一个句子，造成了语义的错误或者逻辑的不通，影响了阅读的体验，但是读者还是可以识别分辨出原有句子的成分，因此该类错误视为一般错误。例如：“And I got more frustrated. By simply shifting my focus…”转换成了 “And I got more frustrated by simply shifting my focus…”。介词短语“by simply shifting my focus”原本是作为第二个句子的状语，转换后却成了第一个句的状语，又如“In our organization, we host a lot of hackathons…”转换成了“We too are expected to constantly reinvent and refresh our skills to stay relevant in our organization. ”同样原文本的介词短语“in our organization ”是作为第二个句子的状语，转换后却成了第一个句的状语。又如原文中“In dwelling , it stays grounded. In being, it flows to depths. In expression, it is honest. In confrontation, it stays gentle. In governance, it does not control. In action, it aligns to timing. ” 居，善地；心，善渊；言，善信；与，善仁；政，善治；动，善时。6 个句子中的结构均相同，都是以一个介词短语开头，加上一个简单句。但是转换后的“It benefits all things with our contention in dwelling. It stays grounded in being. It flows to deaths in expression. It is honest in confrontation. It stays gentle in governance. It does not control any action in lines to timing. ”原本是“In dwelling, it stays grounded. ”In 的前面是一个句号。但是转换后成了 “It benefits all things with our contention in dwelling. ” “in dwelling”与前句连在一起，变成了前句的一个状语。后面的 5 个介词短语的位置全部发生了变化，从原文一个句子的句首转换到另一个句子的句末。（见表 4）这样的错误给理解带来了混乱，增加了理解的难度。

表 4 一般标点符号错误示例

Table 4 Examples of normal punctuation errors

原文本	转换后的文本
And I got more frustrated. By simply shifting my focus from trying to achieve more success to trying to achieve more harmony, I was immediately able to feel calm and focused again.	And I got more frustrated by simply shifting my focus to trying to achieve more success, to trying to achieve more harmony, I was immediately able to feel calm and focused again.
We, too, are expected to constantly reinvent and refresh our skills to stay relevant. In our organization, we host a lot of hackathons, where small groups or individuals come together to solve a business problem in a compressed time frame.	We too are expected to constantly reinvent and refresh our skills to stay relevant in our organization. We host a lot of hackathons, where small groups of individuals come together to solve a business problem in a compressed time frame.
It benefits all things without contention. In dwelling, it stays grounded. In being,	It benefits all things with our contention in dwelling. It stays grounded in being. It

it flows to depths. In expression, it is honest. In confrontation, it stays gentle. In governance, it does not control. In action, it aligns to timing.	flows to deaths in expression. It is honest in confrontation. It stays gentle in governance. It does not control any action in lines to timing.

轻微错误不会影响读者理解原文的意思，典型的轻微错误有大小写错误，省字符和所有格符号的缺失，虚词的增减，不影响理解文本的标点符号错误。读者很容易忽视这些小错误，也有可能发现这些小错误，但它们却不会影响原文的关键要素。整个文本增加类型的错误数量并不多，只有两个轻微的错误。例如原文的“I spoke to my friends. ”转换成了“I spoke spoke to my friends. ”。spoke 重复了两次。又如在原文末尾，演讲者以一句“Thank you!”结束了演讲，但转换后却成了“Thank you. It’s a good thing. ”。增加了“It’s a good thing. ”。但是增加的部分对理解文本没有太大的影响。删减类型的错误数量也较少，同样也只有两个，例如原文的“In fact, it doesn’t feel much at all. ”转写成了“In fact, doesn’t feel much at all. ” 漏掉了一个代词 it. 但漏掉的 it 因为之前的句子已经多次出现，且很容易理解指代的是“水”。又如“towards a rock” 识别成了“towards rock”，“a flower vase” 识别成了“flower vase”。（见表 5）漏掉了不定冠词 a。这些代词和不定冠词的删减，只是语法上存在不规范，本身并不影响对句子意思的理解，因此视作轻微错误。

表 5 轻微删减错误示例

Table 5 Examples of minor deletion errors

原文本	转换后的文本
it doesn’t get agitated. In fact, it doesn’t feel much at all.	It doesn’t get agitated. In fact, doesn’t feel much at all.
it can be a teapot, a cup or a flower vase.	it can be a teapot, a cup or flower vase.
If we think about water flowing towards a rock, it will just flow around it.	If we think about water flowing towards rock, it will just flow around it.

转换后文本也存在较多的轻微替换错误，在所有错误类型中占比超过三分之一。例如“would”识别成了“will”，“teams ”识别成了“team”，“I’d ”识别成了“I’ll”，“it’s”识别成了“is”等。（见表 6）

表 6 轻微替换错误示例

Table 6 Examples of minor substitution errors

原文本	转换后的文本
I would love to hear from you.	I will love to hear from you.

And what’ s interesting to me is that the teams that usually win are not the ones with the most experienced team members,	And what’ s interesting to me is that the team that usually wins are not the one with the most experienced team members
because every day I’ d discover new quirks,	because every day I’ ll discover new quirks,

转换后的文本和原文本标点符号不一致的地方也较多，造成句子的语法错误，影响阅读体验。作为小错误，连字符号错误共有四处，例如原文中“double- or triple-booked”识别成了 “double or triple booked”，“self-helped books” 识别成了“self help books”，“my day-to-day life” 识别成了“my day to day life”，“self-limiting” 识别成了“self limiting”。其他标点符号的错误，例如冒号识别转换成了逗号，原文的双引号识别后被省去了，但是标点符号不规范的问题对句子的理解并没有太大的影响。(见表 7)

表 7 轻微标点符号错误示例

Table 7 Examples of minor punctuation errors

原文本	转换后的文本
Your calendar is already packed with meetings, sometimes double- or triple-booked.	Your calendar is already packed with meetings, sometimes double or triple booked.
In fact, the more self-help books I read, the more stressed and anxious I became.	In fact, the more self help books I read, the more stressed and anxious I became.
Ever since that day, I’ ve been trying to apply the concepts in this passage to my day-to-day life.	Ever since that day, I’ ve been trying to apply the concepts in this passage to my day to day life.
Now, we can stay behind closed doors and continue to be paralyzed by our self-limiting beliefs, such as: “I will never be able to talk about Chinese philosophy in front of a huge audience.”	Now, we can stay behind closed doors and continue to be paralyzed by our self limiting belief, such as, I’ ll never be able to talk about Chinese philosophy in front of a huge audience.
I started wondering to myself: how do I keep up with all this?	I started wondering to myself, how do I keep up with all this?
it’ s perfectly OK to say, “I don’ t know. I want to learn more, and I need your help.”	is perfectly OK to say, I don’ t know, I want to learn more and I need your help.

在分析统计上述各类错误类型及相应的扣分后，采用以下公式计算识别的准确率。

$$\text{Accuracy Rate}=\frac{N-I-D-S-P}{N}\times 100$$

N 的总数为 1570（单词数 1373+标点符号数 197），标点符号包含文中所有的符号，含句号，逗号，冒号，分号，感叹号，问号，双引号。R(IDSP)的总数为 73。

I 增加错误：2 个

D 删减错误：3 个

S 替换错误：2 个严重错误，19 个一般错误，26 个轻微错误

P 标点符号错误：8 个一般错误，13 个轻微错误

$$\text{Accuracy Rate}=\frac{1570-0.5-0.75-18-7.25}{N}\times 100$$

通过计算最终的准确率为 98.3%（保留小数点后一位四舍五入），整体的识别准确率还是比较令人满意。从错误的严重程度来看，严重的错误最少，一般的错误居中，轻微错误最多占所有错误的 60%。从错误类型看，主要的错误集中在替换错误和标点符号错误，占据了所有错误的 90%以上。

4. 结语

经过研究分析，我们认为，“网易见外”语音转文字的准确率达到了 98%以上，虽然增加和删减的错误类型并不多，但是在替换和符号方面，尤其是标点符号位置错误导致断句错误的问题比较突出。还存在识别的单词拼写错误，且识别后的“单词”在字典中并不存在。虽然平台能够快速高效的批量完成转换，改变了传统的转写方式，提高了撰写的效率。转换的结果还远未达到完美的程度，因此现阶段为了确保转换的准确性，还是需要积极发挥人的主观能动性，采取人机耦合的方式对文本进行审校改进完善，才能确保高质量的文本转换。同时在技术方面，相关的企业和研究人员除了研究继续提高单词准确性的基础上，还需要改进语意义切分技术提高断句的准确率，优化算法解决单词拼写错误的问题。

参考文献：

[1] Pablo Romero-Fresco Juan Martínez. “Accuracy Rate in Live Subtitling – the NER Model Audiovisual Translation in a Global Context: Mapping an Ever-changing Landscape” edited by Rocío Baños and Jorge Díaz Cintas, 28-50. London: Palgrave Macmillan, 2015

[2] Ajami S. “Use of speech-to-text technology for documentation by healthcare providers.” The National medical journal of India, 29(3), 148–152, 2016

[3] Shadiev, R., Huang, YM. & Hwang, JP. “ Investigating the effectiveness of speech-to-text recognition applications on learning performance, attention, and meditation.” Education Technology Research and Development, v65, 1239–1261, 2017

[4] Chen, K.T.C. “Speech-to-text recognition in University English as a Foreign Language Learning.” Education and Information Technologies, 27(4), 9857–9875, 2022

[5] Dom Barnard. “virtualspeech.com/blog/average-speaking-rate-words-per-minute.” Virtualspeech, January 20, 2018. Accessed June 15, 2022.

[6] 肖维青,高佳晖.机器翻译字幕质量评估研究——以“网易见外”英译中字幕为例[J].外国语言与文化, 2020, 4(03):95-105

A Case Study of Quality Assessment of Speech-to-Text Recognition by Sight Youdao

Huang Danhua
Zhejiang Yuexiu University, Shaoxing Zhejiang China
Email: 406816705@qq.com

Abstract: With the development of computer technology and artificial intelligence, the quality of Speech-to-Text Recognition（STR）has been dramatically improved, hereafter referred to as STR., and it has also been widely used in business, education, and translation. Based on the NER model proposed by the Spanish scholars Pablo Romero-Fresco et al., this paper adopts the NR model and uses the TED speech as the material. Sight Youdao is used to complete the transcription, and Errors in detail are analyzed in accordance with their categories and severity. Finally, the transcription accuracy of the text, which is an essential criterion for evaluating the quality of STR, is calculated. The results show that although the recognition accuracy rate has reached 98%, there are still different types of errors in the transcribed text. It requires human intervention to achieve satisfactory results. The research results also show that there is still much room for improvement in the accuracy of sentence segmentation and word recognition by Sight Youdao.

Keywords: Speech-to-Text Recognition, the NER model, the NR model, the recognition accuracy rate