
论文题目: 利用 HamNoSys 实现语音到手语翻译

参赛队员姓名: 康哲涵

中学: 中国人民大学附属中学

省份: 北京市

国家/地区: 中国

指导教师姓名: Ali Mirtar



利用 hamnosy 实现语音到手语翻译

康哲涵

摘要：手语是听力障碍人群互相交流思想的一种手势语言，但至今没有成熟的技术实现聋人和健全人之间的交流。本次研究目标立足于已经成熟人工智能领域的语音识别、自然语言处理和三维虚拟人技术，实现可推广应用的语音到自然手语翻译系统。在研究过程中，首先利用 hamnosy、Sigml 和手语字典建立并管理手势库，在此基础上，手语翻译时利用语音识别技术将语音转化为文字，并使用斯坦福研发的自然语言处理工具进行语法分析和自然语言到手语的语法转换，在手势库中找到对应的手势的 Sigml，最后借助三维虚拟人模型合成手语手势，最后通过对简单会话和手语课上在初步测试，基本实现了简单对话的翻译，期望在此基础上进行深入开发，能够真正帮助聋人融入到正常人的社会活动中。

关键字：手语翻译，语音识别，自然语言处理、HamNoSys、Sigml、虚拟人

目录

1 引言.....	5
1.1 现状与问题.....	5
1.2 设计思路.....	6
2. 系统设计.....	6
2.1 系统架构.....	6
2.2 数据结构设计.....	7
2.3 翻译流程设计.....	8
2.4 界面设计.....	8
3. 几个关键技术实现.....	8
3.1 手势库建设与管理.....	10
3.2 语音识别技术选择与实现.....	14
3.3 自然语言与自然手语语法转换.....	15
3.4 手语合成实现.....	16
4. 应用效果测试.....	17
5. 结论.....	17

1 引言

1.1 现状与问题

中国目前患听力障碍的聋人约有 2400 万[3], 他们由于听力言语的障碍, 导致信息缺失、社会沟通功能不畅, 致使不能融入正常的社会生活, 美国著名盲聋作家海伦凯勒说, “盲, 是人和物之间的距离被隔断了, 聋, 是人和人之间的距离被隔断了”。

手语是一种人体运动语言, 通过手的运动和面部表情表达意思, 主要用于聋人之间的相互交流, 我国一直以来重视手语的开发规范, 20 世纪 50 年代开始了中国聋人手语的规范化工作, 到 1994 年, 曾先后出版了《中国手语》(330 个词条)、《中国手语》续集(2 266 个词条) [1]; 2018 年由教育部、国家语言文字工作委员会、中国残疾人联合会联合发布了《国家通用手语常用词表》[4], 从而形成中国手语的“普通话”。很好地保障了聋人之间的交流, 但如何实现聋人与正常人之间的交流, 仍存在障碍。

要实现听障人和健听人自由交流的梦想, 涉及到两方面的技术: 一是手语到语音的翻译, 即听障人打出手语, 计算机通过采集设备获得听障人的手语数据, 通过图像和动作识别出手语含义, 进而翻译成语音, 传达给不懂手语的健听人, 健听人就可以“听懂手语”。二是语音到手语的翻译, 通常也称为手语合成, 计算机将语音翻译为手语并表现出来, 听障人就可以“看懂声音”。手语合成不仅有助于聋人使用各种信息, 而且可以帮助聋人参加各项社会活动, 具有十分重要的应用价值和社会意义。

用计算机合成手语的研究最早提出于 1982 年[5], 山兹(M. Shantz)和波伊兹纳(H. Poizner)描述了一个用计算机合成美国手语的程序。之后, 中国、日本、德国等许多国家都进行了自己国家的手语合成研究[6], 并取得了重要的研究成果。

随着人工智能技术的发展，目前高识别率的语音识别技术已经成为可能，同时，面向中文的自然语言处理也逐渐成熟；基于浏览器的三维显示技术也可广泛应用。这些技术为此次研究提供了前提，使我们可以开发可广泛应用的语音到手语语句级别的翻译系统，本次研究期望利用这些成熟技术实现语音向手语的翻译。

1.2 目标与思路

本次研究基于前人的研究成果，试图将中文语音翻译为自然手语，成果将趋近于谷歌线上翻译模式。本次研究的主要目的为：

- 参考 Google 公司的字典系统，建立可实用的手语字典系统，实现中文手语、美国手语线上字典系统，能够利用虚拟人动画提供直观的手语字典服务。目前网上手语字典都是采用文字描述或录像的方式，这种方式不直观，同时维护性也较差。
- 建立一个可使用的语音手语翻译系统，最终目标是建立一个高正确率、实用型语音手语翻译系统。
- 在此基础上，本次研究试图使翻译系统具有便携性，用户可通过网站、手机、平板电脑等终端访问。可广泛应用到失听人教学、新闻播报等领域。

2.系统设计

2.1 系统架构

本次研究与众多互联网应用系统一样采用三层体系架构。三层体系架构分为展示层，应用层，数据层。

- 展示层：即用户可见的应用界面。展示层允许用户输入数据并查看结果。展示层可应用在不同的设备中，本系统可在笔记本电脑，手机，和平板电脑的 **Chrome**、**Safari** 浏览器等常用浏览器中运行，展示层采用 **Angular** 开发框架，虚拟人模型及驱动类库采用由 **Angila** 大学的开源研究成果，语音识别采用谷歌公司的语音识别 **Javascript** 类库。
- 应用层：三层体系架构中的中间层，主要负责输入输出中的数据传递以及数据处理。应用程序部署在 **Tomcat** 容器中，主要采用 **java** 进行开发，应用了斯坦福大学 **coreNLP** 自然语言处理类库；手势库管理中的 **hamnosys** 和 **sigml** 的相互转换采用了德国汉堡大学的类库。
- 数据层：三层体系架构中的最内层，为系统的正常运转储存数据。本系统采用了 **MySQL** 数据库。

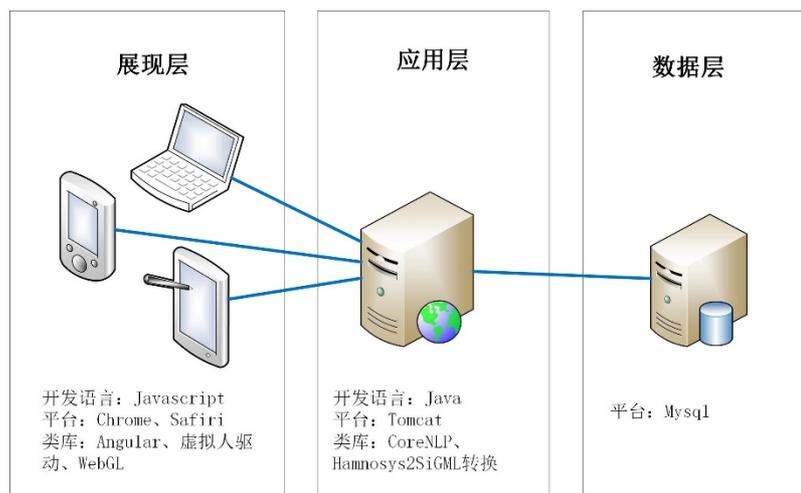


图 1 体系架构图

2.2 数据结构设计

数据库存储了手语翻译系统的相关的基础数据，主要的数据表包括：

- 手势表，主要数据项包括词语，拼音，**Humnosys** 手势库，**SiGML** 手势标记语言，该表为系统的核心表，存储了《国家通用手语常用词表》中的部分中文手语手势和美国手语（**ASL**）的部分手势。

- 同义词表所包含的数据项为词语和同义词，本表的词语与手势表中的词语为外键关联关系。
- 语法表包含自然语言规则与手语语法规则两个数据项。定义了自然语言和自然手语的语法转换规则。

2.3 翻译流程设计

手语到语音的翻译我们的设计思路是首先依据手语字典把各种词的手势分解成基本连续动作和表情等，然后采用标记语言进行表示，这一标记语言能够用三维虚拟人进行动作展示，这个过程我们采用了 **HamNoSys** 来实现。在建立基础手势库的同时，由于手语词汇要远小于自然语言的词汇，因此我们还建立了一个同义词词典，这样就保证了翻译时在手势库中没有对应词的时候能够找到近义词进行替代，然后采用 **Google** 公司和科大讯飞语音识别类库，把语音输入翻译成文字，通过斯坦福大学的自然语言处理工具 **CoreNLP** 句子进行分词、标注和语法转换处理。最后用转换后的文本进行手势动作合成，并驱动三维虚拟人一次播放语法转换后的动作，就达到了从语音到手语的翻译过程。

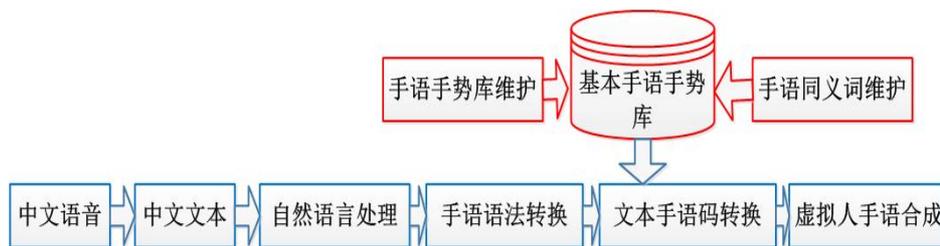


图 2 语音手语翻译转换流程

2.4 界面设计

页面设计主要包括三部分：翻译页面，词典管理页面和语法管理页面。

- 翻译页面包括三大部分：语音录入框，虚拟人和操作菜单。语音录入框右侧有录音按钮，点击录音按钮开始录音，将语音转化为文字显示在录入框中，同时还可以手动输入要翻译的句子。三维虚拟人为页面的主体部分，翻译出的手语通过虚拟人来展示。点击页面左上角的控

制按钮，菜单栏中有设置、管理员登录和语言选择三个选项，在设置中用户可以调节手语展示速度，低速度手语展示适用于想要跟随虚拟人进行手语学习的使用者；管理员输入正确的用户名和密码后可以登录到后台进行词典或者语法库的修改或增加；翻译系统支持中文和英文语音输入，在菜单栏中可以选择语言。

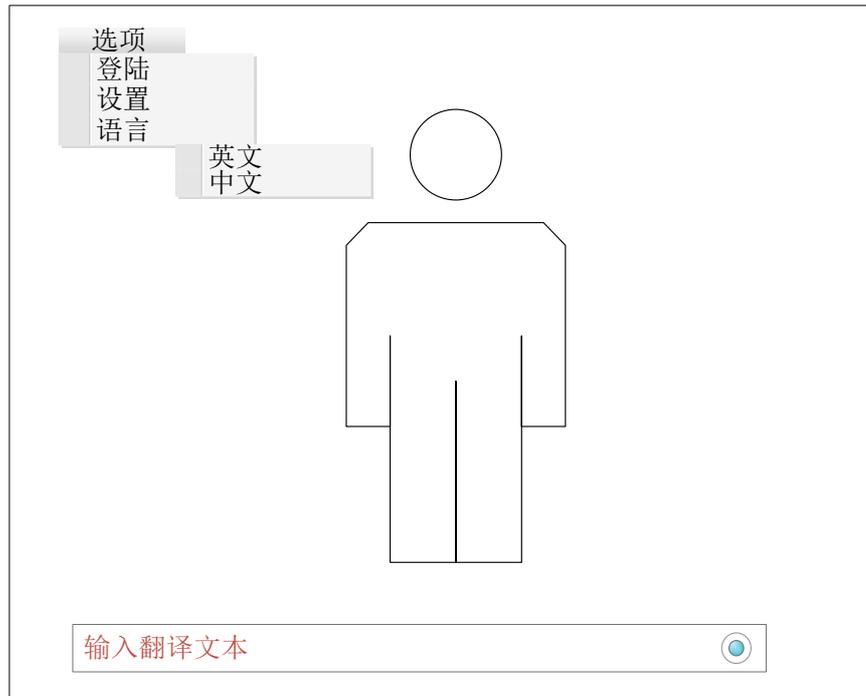


图 3 语音手语翻译页面设计图

- **管理页面**，在左上角可以进行语言的选择，中文或英文。左侧列表为词语分类及词语，中间为录入表单，右侧为虚拟人，用户输入词语及手势后，可在词典中进行手势测试。

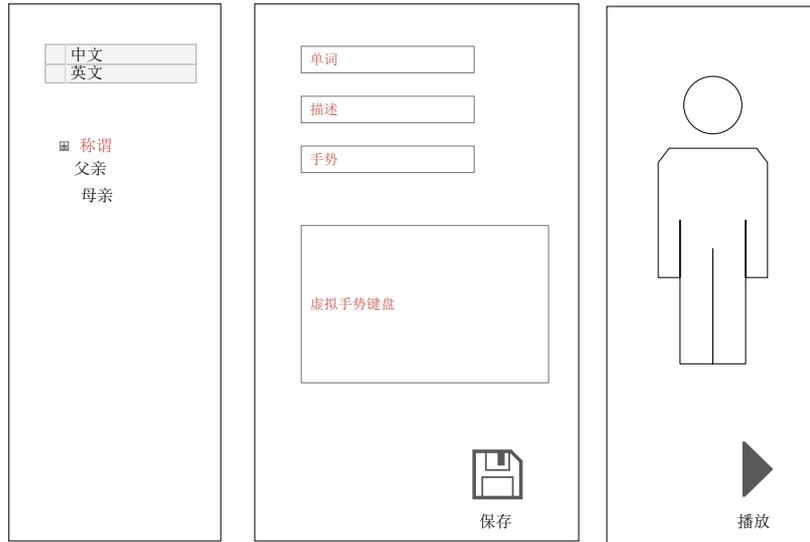


图 4 手语手势库管理页面设计图

- **语法页面**，手语语法管理页面可以让登陆成功的管理员更改或添加语法映射规则。首先进入的是列表页面，其中列包含语言，自然语言与法和所对应的手语语法三列。点击操作按钮可以进行语法规则的删除和更改。在页面中点击添加按钮，可以添加自然语言语法和对应的手语语法规则并进行保存。

3.技术实现

3.1 手语手势编码技术

我们根据身体的不同部位和组合方式把手语词语分为: 手势词语、表情词语、身体词语和口型词语。手势词语指只通过手的性状和运动而表达的词语，是手语的主体。表情词语指在打手势的同时脸上做出配合手势的表情或者是指单独出现的表情, 它们表示各种各样的意思，比如“想”是闭眼凝思。这种表情词语常常是名词和动词的修饰语, 相当于形容词和副词。身体词语指身体在打手势的同时配合手势做出的姿势或者是指单独出现的身体姿势, 它们通常表示角色的

转换, 方位的变化。口型词语指在打手势的同时运用词语的口型来进行辅助表达, 通常为常用词语。

本次研究的核心基于德国汉堡大学开发的汉堡手语编码系统进行对手语的编码。德国汉堡编码系统就如同中文里的拼音, 用标识符来表示手语的形状, 移动, 位置、朝向和动作等信息, 他不依赖与特定国家的手语系统, 所以可以在国际上使用。

手型标识: 手形主要分为拳头, 平手, 分离手指和拇指组合。这四种基本形式与拇指变化和手指弯曲等进行组合, 允许用户为任何给定的手形编写 HamNoSys。一些基本手形如下图 1 所示。

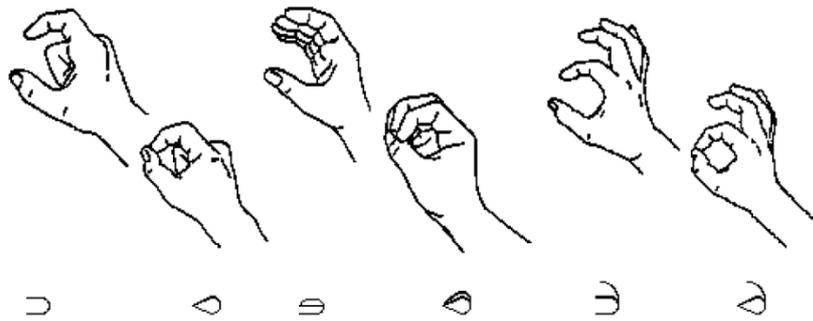


图 5: HamNoSys 基本手型示意图

手势方向标识: HamNoSys 通过组合两个部分来描述给定符号的手的方向: 伸展手指的方向和手掌方向。有三个视角 (手语展示者视图、俯视图和右视图), 用于显示伸展手指相对于手语展示者身体的方向。对于给定的伸展手指方向, 手掌方向也用相同的模型描述。

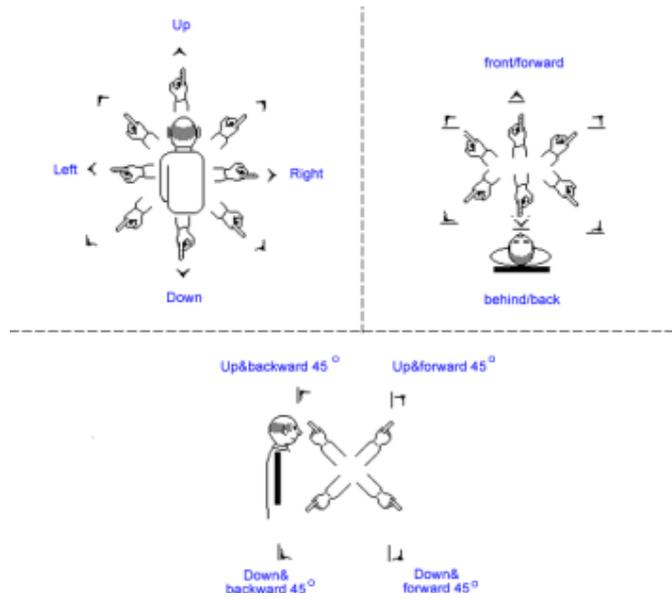


图 6: HamNoSys 手掌方向和手指方向示意图

位置标识: 位置指示符用来表示手势相对于手语展示者的位置, 位置的表示分为两部分: 手相对于身体的位置和手相对于身体的距离。

手的运动标识: 手的移动被分为直线移动, 曲线移动, 环形移动, 折线移动和螺旋形移动。直线移动是平行于身体的移动; 环形移动是手顺时针或者是逆时针的转动。所有形式的手势移动还依据移动的幅度进行划分: 小幅度移动, 大幅度移动等。

非手势标识: 非手势标识包括耸肩、头部运动、面部表情、嘴型变换等, 这些手语中的部分需划分为单独的非手势标识。

SiGML: SiGML 是基于手语手势的以 XML 为标准的标记语言。它把 Hamnosys 手语标识转换为了 XML 格式, 以方便程序的编写和可读性, 根据 hamnosys 标识语言生成的 SiGML 可以驱动虚拟人进行手语手势的展示。

	<pre><?xml version="1.0" encoding="utf-8"?> <sigml> <hns_sign gloss="item_0"> <hamnosys_manual> <hamfinger23 /> <hamextfingerui /> <hampalmu /> </hns_sign> </sigml></pre>
--	--

	<pre> <hamear /> <hamtouch /> <hammoveu /> </hamnosys_manual> </hns_sign> </sigml> </pre>
--	---

图 7: HamNoSys 与 SiGML 对应关系图

3.2 数据库建设与管理

数据库建设与管理主要包含手语字典库管理, 同义词管理, 语法库管理, 用户管理, 下面将详细介绍其实现方法。

- 手语字典管理

手语字典管理目前支持中文手语和美国手语。中文手势库建设依据《中国手语》修订版。中文词语按照词意分类, 如职业, 称谓等, 英文词语分类按照单词首字母进行分类。数据项包含词语、音标、Hamnosys 标识、Hamnosys 虚拟键盘。Hamnosys 键盘可方便管理员对手势进行更改和批量管理。在本次研究中使用的 Hamnosys 键盘管理方便, 在输入标识后即可立即显示手语动画, 所见即所得。

利用 Hamnosys 键盘完成编辑编码后, 在提交保存到数据库之前, 利用程序转换为 SiGML 标识语言, 数据库中同时储存 Hamnosys 编码和 SiGML。完成保存之后, 系统返回 SiGML 以供前台管理界面展示手语动作, 以供调试。

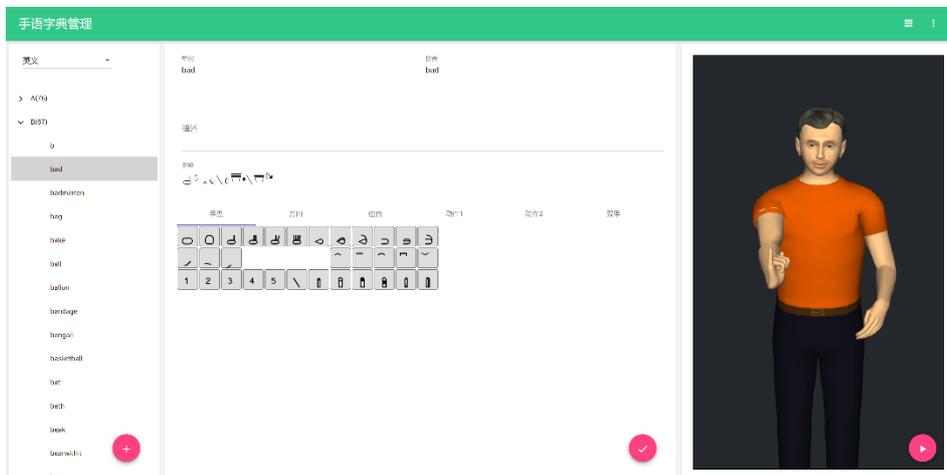


图 8: 手语手势库管理

● 手语字典管理

因手语词汇数远小于中文自然语言词汇数，本次研究的数据库中有同义词库，将每个手语词汇连接多个中文词汇。如当语音输入在手语词汇库中没有的“姥爷”或“外公”一词，系统会自动连接到手语词汇库中的“外祖父”，并找到与其对应的手势。

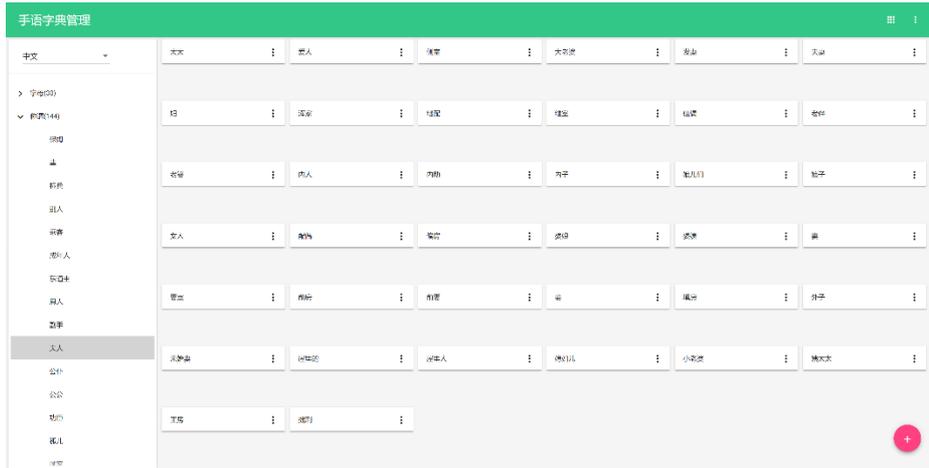


图 8: 手语同义词库管理

● 手语语法管理

语法管理主要定义了自然语言和手语之间的语法转换规则，主要数据项包含语言，自然语言与法和手语语法三个数据项。

● 用户管理

用户管理定义了可以访问管理页面的用户，主要数据项包含用户名、姓名和密码。

3.3 语音识别技术选择与实现

为保证语音识别的准确率，本次研究的语音识别部分采用了现有的成熟技术。通过对百度、科大讯飞和 Google 的语音识别类库进行研究和测试，最后选择 Google 和科大讯飞的语音识别类库，Google 和科大讯飞对于中文和英文的识别率超过 98%，并可以进行智能断句和标点符号预测，由于中国境内不能访问 Google 的服务，因此采用的方式是通过识别客户的浏览器语言类别自动

选择语音识别引擎，中文浏览器缺省使用科大讯飞引擎，英文浏览器缺省使用 Google 语音识别引擎。

3.4 自然语言与自然手语语法转换

手语语法和中文自然语言语法有着较大区别，手语分为文法手语和自然语言手语，文法手语是以汉语语法为基础的手语，与汉语语法相同，但对于汉语基础较差的聋人很难理解。自然语言手语为聋人的母语，语序与汉语有很大的区别，同时，会省略汉语语法中一些虚词。

一般来说，自然语言语法到手语语法的转换主要分为句法和词法转换，常见的词法转换包括：

- 省略规则，如省略虚词，量词等；
- 名词及其修饰语的次序规则：“名词修饰语+名词”转换为“名词+名词修饰语”；
- 专有名词规则：人名和地名等专有名词采用只打声母的简略表达形式

主要语法规则包括：

- 交换位置规则：如在简单句中把主语和宾语调换位置；
- 后置规则，如在疑问句中的能愿动词在手语中应置后，“疑问代词+被疑问对象”转换为被“疑问对象+疑问代词”；

本次研究的语法转换主要依赖于斯坦福大学研发的 **CoreNLP** 自然处理类库进行语法转换。自然语言处理是人工智能和语言学领域的分支学科，主要探讨如何处理及运用自然语言；包括自然语言认知系统、自然语言生成系统、自然语言理解系统几个子领域。斯坦福大学自然语言处理组提供的一系列开源文本分析工具，包括分词、命名实体识别、词性标注和句法分析等，提供了完整的自然语言处理的工具集。词是语言的基本单位，因此将词与词分离则为自然语言处理中的基本步骤。系统首先将语音输入的句子进行分词处理，再进行词性标注和句法分析（确定文本中的主语，谓语，宾语等），在此基础上，利用定义的语法转换规则，如省略、词法转换和句法转换。其次在交流过程中涉及到的

人名以及地名需要通过命名实体识别进行单独处理。手语词汇中除常用词汇外皆使用指拼表达。指拼是指对于专有名词用拼音对应的字母来进行词语的表达

以“青少年应好好学习科学”一句来举例：

进行分词和句子成分标记：

“青少年” “应” “好好” “学习” “科学”
主语 能愿动词 副词 谓语 宾语

根据语法规则：“主语+副词+谓语+宾语→主语+宾语+谓语+副词”和能愿动词置后原则调整词语顺序，变为：

“青少年” + “科学” + “学习” + “努力” + “要”

3.4 手语合成实现

本次研究采用以Html5的WebGL技术为基础的虚拟人进行手语展示，虚拟人手语展示包括虚拟人模型建设、虚拟人运动模型和运动平滑处理等关键技术构成。

为了确保手势的表现，虚拟人关节应具有一定的灵活性。该系统使用三类节点表示一个虚拟人体模型：人体重心，人体关节和人体骨骼段，共有一个人体重心，77个关节和47个骨骼段。各个骨骼段的位置表示了虚拟人的姿态，用户只能看到虚拟人的表面模型，而看不到它的骨架模型。虚拟人表面模型由53块部位组成，每个相互连接的部分之间设有小块公共区域，这样在虚拟人运动时才不会出现“裂缝”。

手语运动模型是由内置在驱动类库中与 Sigml 标签一一对应的数据模型来驱动的。通过对后台传输过来的 SiGML 解析，然后形成驱动虚拟人运动的指令集，驱动虚拟人打出手语。本次研究采用的是基于 Html5 配置的 JASigning 驱动类库来完成的。

每个手语动作间的连接必须进行平滑处理，手势动作中出现卡顿将影响到动作的易懂性，所以虚拟人采用了欧拉插值和球面线性插等算法来确保动作的平滑和连续。

非手势动作包括脸部和唇部的运动，脸部运动用于区分感叹句，疑问句和陈述句。虽然非手势动作不是手语的主体部分，但是缺乏非手势动作会影响手语合成的可懂度。SiGML 支持非手势动作，但目前受限于手势库建模不全，非手势动作将是我们进一步的研究方向。

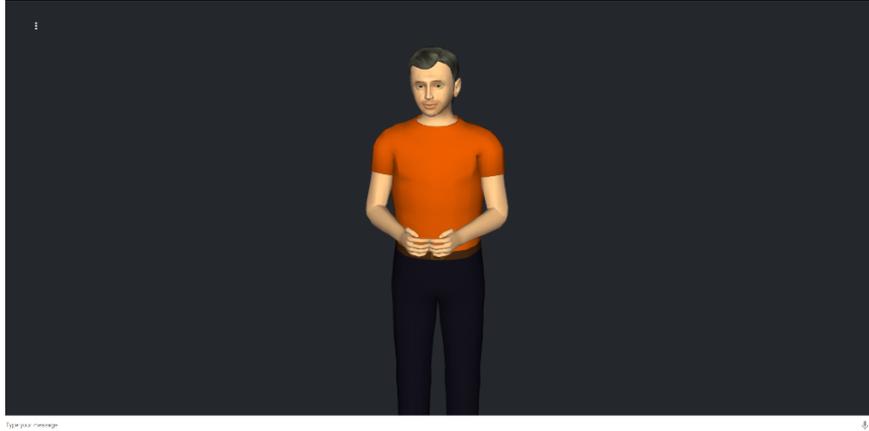


图 9：自然手语翻译界面

4.应用效果测试

由于目前手势库和语法规则库建设还没有完善，应用测试还只能按照特定语句范围进行测试。中文手语测试内容为《中国手语日常会话速成》、手语童话故事《龟兔赛跑》和中国聋人网上的《立春》教学片，测试前对内容涉及的所有词汇进行在手势库进行检查，并对目前已知的语法规则进行确认。

在此基础上每项测试按照正常阅读语速进行 5 次以上重复测试，并对翻译结果与教学片的中手语进行对比，具有一定的准确性。在和中国人民大学星原手语的聋人手语老师确认后，能够基本满足听障人群的理解需求。

5.结论

能够大面积应用的语音手语互译系统具有很大的社会意义，主要体现在消除中国目前患听力障碍人群与健听人的交流障碍，使他们获得社会的信息、融入正常的社会生活，甚至可以承担健听人承担的工作。手语互译系统可以广泛

应用到聋人学校、有聋人的单位或者是残疾人服务部门, 其社会意义无疑是巨大的。

目前本次研究已经完成了语音到自然手语翻译的研究和探索, 在国内率先把 Hamnosys 应用与中国手语翻译, 具有一定的指导意义。为了使该系统能够真正应用, 还有很多需要完善的工作, 主要体现在以下几个方面:

- (1) 建立更加完善的手势库, 录入手语词典中的所有词条。
- (2) 同时与聋人学校的合作, 继续完善语音到手语的语法转换规则, 以便实现更加准确的翻译。
- (3) 对本次采用的开源软件等底层技术进行深入研究, 以实现底层技术开发, 建设能够支持表情的驱动程序等。
- (4) 以网站运营和 APP 下载方式进行免费推广, 期望惠及更广大的失听人群, 建立正向反馈机制, 以便逐步完善。

6. 参考文献

- [1] 中国聋人协会 《中国手语（上下）修订版》 ISBN: 7508030052. 2003
- [2] 中国聋人协会. 中国手语续集. 北京: 华夏出版社, 1988.
- [3] 中国聋人协会. 听力言语残疾人工作知识手册
- [4] 中国残疾人联合会、教育部、国家语言文字工作委员. 国家通用手语常用词表 2018
- [5] 陈益强, 刘军发, 颜庆聪, 何文静. 中国手语合成技术及其应用. 信息技术快报.
- [6] 王兆其, 高文. 基于虚拟人合成技术的中国手语合成方法 软件学报 2002
- [7] Khushdeep Kaur , Parteek Kumar , HamNoSys to SiGML Conversion System for Sign Language Automation, 2016
- [8] Robert Smith, HaNoSys4.0 User guide, 2013
- [10] REPRESENTATION OF THE SIGNS, Extract from Improvement and expansion of a system for translating text to Sign Language.