# 一种日语语音识别系统构造方法

## 背景介绍

语音识别是将声音转换成文字的技术。语音识别需要大量标注好的语音数据进行模型训练，否则很难达到实用化的效果。数据的采集需要大量人力物力和时间成本，很难在短期积累大量数据。对中文而言，可以通过向数据公司购买或线上数据外包标注的等法，实现较多的数据积累；然而，当构造其它语言的语音识别系统时，必须重新积累该语言的数据，带来巨大成本开销。**本发明关注如何利用现有的中文数据资源，快速开发一个日语语音识别系统。**

本发明基于如下事实：[汉语](https://zh.wikipedia.org/wiki/%E6%BC%A2%E8%AA%9E)，特别是古汉语自古以来就通过[汉字](https://zh.wikipedia.org/wiki/%E6%BC%A2%E5%AD%97)、[汉语](https://zh.wikipedia.org/wiki/%E6%BC%A2%E5%AD%97%E8%A9%9E#.E6.97.A5.E6.9C.AC.E8.AA.9E.E7.9A.84.E6.BC.A2.E5.AD.97.E8.A9.9E)对日语的表记及词汇、语素上有很强影响。日语音韵中的拗音等就是来自于汉语，此外日语模仿古代汉语书面语的文法、语法的行为，使得人们在日语文法、语法和文体上都能看见汉语对日语的影响。中文和日文这种相似性可以作为先验知识，利用汉语的数据和模型资源，构造日语识别系统。

本发明的基本思路是，**基于中文和日文在发音、表义等方面的相似性，利用现有较大量的汉语数据或已训练模型，实现对日语识别系统的有效训练**，如下图所示。

**中文数据**

日文数据

日语语音识别系统训练

中文

模型

日语模型

**图1. 基于汉语知识的日语语音识别系统构造方法**

## 发明内容和思路

为构造一个日语语音识别系统需要大量的日语语音和文本资源。获取这些资源通常非常困难。

**本发明的目的是提出一种利用少量日语数据即可构造出一个有效的日语语音识别系统的方法。**

**本发明所提出的方法是利用汉语和日的相似性，充分利用汉语的数据和已有模型，将其中包含的知识用于日语识别系统的模型训练过程中，实现日语语音识别系统的快速有效构造。**

本发明所提出的方法皆以此思路为前提，从语音模型和环境因子两方面，实现汉语到日语系统的知识迁移，从而可以实现如下两点目标：

**（1）用较少的日语语料即可构造较好的日语语音识别系统**

**（2）可继承汉语识别系统在环境和口音上的鲁棒性。**

## 发明要点

1. **利用汉语声学模型提高日语声学模型能力**

不论中文和日文，其底层的人类发声机理是一致的，人耳对不同语音信号的前端处理也是一致的。对于中文和日文，这一发音特性更为接近。利用这一特性，将汉语的特征提取（即前端处理）部分和日语特征提取部分共享，可实现从汉语模型到日语模型的知识迁移。具体如图（2）所示。图中所示一个声学信号处理系统包括两个模块：特征提取和声学建模。这两个模块都是由人工神经网络组成，其中特征提取模块由DNN或CNN构成， 声学模型模块由RNN构成。因为汉语有大量训练语料，包含了信道、说话人、口音等多种复杂性，因此汉语的特征提取模块对这些非语音因素极为鲁棒。我们用汉语的特征提取模块来增强日语特征提取模块的能力，增强方法包括（1）将日语特征提取模块直接替换成汉语特征提取模块，直接复制汉语的特征提取模型；（2）目标函数约束法。这一方法将日语训练数据同时经过汉语和日语特征提取模块，并训练过程中将两者在各个隐层节点上的偏差作为约束项加入训练的目标函数，即：

$L\left(x;w\right)=H\left(x;w\right)+\sum\_{x}^{}||h\_{C}\left(x\right)-h\_{J}\left(x\right)||^{2}$（1）

其中,$ H\left(x;w\right)$为传统神经网络训练的目标函数，如交叉熵；$h\_{C}(x)$和$h\_{J}(x)$分别为训练样本x在汉语和日语两个特征提取网络中所有隐藏结点的激发值向量。式（1）意味着当训练日语声学模型时不仅考虑到声学模型输出目标的最优化（即H(x;w)），而且考虑特征提取结果尽可能接近汉语的特征提取模块输出结果。这意味着汉语中学到的特征提取知识以约束项的方式传递给了日语特征提取模块。这一方法比直接替换法更能平衡模型分类误差最小化的要求和向汉语学习的要求。

声学模型

声学模型

特征提取

特征提取

日语模型

汉语模型

**图2. 基于汉语模型的日语声学模型增强方法**

1. **利用跨语言因子增强日语声学模型**

不论是中文还是日文，其中一些因子是跨语言的，包括环境因子，说话人因子等。利用汉语数据学习得到的这些因子可以直接应用到日语模型中。本发明首先利用汉语数据训练一个提取语言无关因子（环境、信道、说话人）的因子模型，再利用该模型在日语识别系统的训练和解码过程中生成语言无关因子。因为这些因子代表环境的噪音、信道等信息，因此可提高语音识别的性能。

可采用多种方法提取跨语言因子。（1）i-vector模型，该模型是全因子(total variance)模型，可将所有与语音长时特性相关的因子表征成一个低维特征向量。因为发音内容具有短时特性，因此该因子中不包含发音内容信息，具有语言无关性。（2）CNN/RNN自动编码器(Auto Encoder，AE)模型。该模型将语音信号表达成一串特征序列，用CNN/RNN压缩成一个低维特征向量，再通过CNN/RNN重新生成原始特征序列。同i-vector类似，CNN/RNN压缩成的低维特征向量不表达短时发音特征，因此具有语言无关性。

日语模型

声学模型

特征提取

**中文数据**

i-vector CNN/RNN

**图3. 利用跨语言因子对日语声学模型进行增强的方法**

## 发明优势

1. 可基于较少的日语语音数据实现日语系统的模型训练。
2. 继承了汉语模型对噪声和口音的鲁棒性，系统实用性强。