ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 38,No. 1,Jan. 2023,pp. 93－100
DOI:10. 16337/j. 1004-9037. 2023. 01. 007

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于声学模型共享的零资源韩语语音识别

王皓宇[1]，JEON Eunah[1]，张卫强[1]，李 科[2]，黄宇凯[2]

(1. 清华大学电子工程系,北京国家信息科学技术研究中心,北京 100084;2. 北京海天瑞声科技股份有限公司,北京 100083)

摘 要：精准的语音识别系统通常使用大量的有标注语音数据训练得到,但现有的开源大规模数据集只包含一些广泛使用的语言,诸多小语种则面临着训练数据不足的问题。声学模型共享方法给出了这个问题的一种解决方法,它利用不同语种间的相似性,可以实现不需要小语种语音数据的语音识别。本文将声学模型共享方法扩展到韩语语音识别上,利用汉语声学模型构建韩语和汉语之间的音素映射关系。在不使用任何韩语语音数据的情况下构建的语音识别系统在Zeroth测试集上的字错误率达到了27.33%。同时本文还测试了不同映射方式之间的差异,结果表明这种共享模型的音素映射应当采用将目标语言词汇映射为源语言音素的方式。

关键词：语音识别;零资源语音识别;韩语语音识别

中图分类号：TN912 **文献标志码**:A

Zero Resource Korean ASR Based on Acoustic Model Sharing

WANG Haoyu[1], JEON Eunah[1], ZHANG Weiqiang[1], LI Ke[2], HUANG Yukai[2]

(1. Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;2. Beijing Haitian Ruisheng Science Technology Ltd., Beijing 100083, China)

Abstract: A precise speech recognition system usually is based on a large amount of training data with handcrafted transcription, which sets a barrier to the recognition of many low-resource languages. Acoustic model sharing, which is based on the similarity of certain rich and low resource language pair, provides a new method to solve the problem and helps to build an automatic speech recognition (ASR) system without any training data of the given low resource language. This paper expands the method to Korean speech recognition. Specifically, we train an acoustic model on Mandarin data, and lay down a set of mapping rules between Mandarin and Korean phonemes. A character error rate (CER) of 27.33% is achieved on Zeroth Korean test set without using any Korean speech data. Moreover, we also test the difference between source-to-target and target-to-source phoneme mapping rules, and prove that the latter is more appropriate for acoustic model sharing.

Key words: speech recognition; zero resource speech recognition; Korean speech recognition

---

基金项目:NSFC-通用技术基础研究联合基金重点项目(U1836219)。
收稿日期:2021-10-19;**修订日期**:2021-11-04

引　言

语音识别技术一直是人工智能技术最为重要的研究领域之一，也是实现人与计算机之间自然、便利交互的基石。近几年来，蓬勃发展的深度学习技术为语音识别领域带来了生机，包括深度神经网络-隐马尔可夫混合模型(Deep neural network-hidden Markov model, DNN-HMM)[1]和端到端语音识别模型[2]在内的诸多新技术不断发展。同时随着计算机运算能力、训练集数据规模的不断扩大，让语音识别系统的准确率进一步提升，目前最精确的语音识别系统已经可以在较为清晰的英语数据集上达到1%左右的词错误率[3]。然而，准确的语音识别系统的背后是大量数据的支撑。对同一个模型来说，识别正确率和所需训练集的大小呈近似的对数关系[4]。在工业界，一个错误率极低的语音识别系统常常需要上万小时的训练数据[5]。不过，对世界上的大多数语言来说，要收集这种量级的数据是相当困难的。根据最近几年的一项统计[6]，目前世界上仍在使用的语言有7 100余种，但其中约40%都属于濒危语言，其使用者不足1 000人，因此可以说世界上的大多数语言都是小语种。这些小语种普遍面临着现有训练数据不足、收集训练数据困难等问题，因此如何为小语种快速地、低成本地构建精度足够的语音识别系统，成为语音识别领域一直关注的问题。

一般来说，解决小语种的语音识别需要利用不同语种间的相似性。不同语言之间的差异主要体现在顶层的单词构成和句子结构上的不同，但在底层的发音层面上则通常差异不大。因此，可以利用资源丰富的常用语言训练一个基础的语音识别模型，并将这个模型的底层部分与目标小语种的语音识别模型进行共享，以降低小语种语音识别模型的训练难度。这种共享可以在音素、模型的参数或是提取的特征等多个层面进行。Das等[7]构建了英语和土耳其语的共享音素集，并设计了英语迁移到土耳其语的迁移学习模型，证明了音素级别的共享对小语种音素识别的作用。模型的底层参数同样可以在常用语言和小语种之间进行共享。Huang等[8]构建了一个共享隐层参数的多语言识别模型，不同语言之间使用相同的底层参数，只有最终的softmax分类层保持独立。结果表明，共享模型仅使用3 h左右的英语训练数据就可以达到28%左右的词错误率(Word error rate, WER)，优于只使用这些数据从头训练的模型。近年来，以Wav2vec为代表的一系列预训练模型[9]给基于特征共享的小语种语音识别提供了新的思路。Yi等[10]利用在英语数据上预训练的wav2vec2模型作为语音特征提取器，在其基础上进行了日语、德语等多个语种的低资源语音识别实验。实验表明，预训练模型提取的特征能为小语种的语音识别带来20%～50%的相对增益。

尽管小语种语音识别已经有很多研究，但鲜有研究关注零资源情况下的语音识别。最近，Prasad等[11]在这一方向展开了一些研究。他们直接使用了常见语言的声学模型，结合低资源语言的语言模型、发音字典以及两种语言间的音素映射关系，实现了仅需来自专家的语言学知识，而无需任何标注数据的零资源的语音识别。在他们的实验中，语言相似度的判断由专家进行，发音词典和映射规则同样由专家编写。实验表明，这种直接的声学模型替换在选择的常用语言和目标小语种相近时会取得很好的效果。他们在宿雾语和吉尔吉斯斯坦语中都达到了18%左右的WER，这是相当出色的表现。

本文将这种基于声学模型共享的零资源语音识别系统进一步扩展到韩语语音识别方面，使用与韩语较为相似的汉语训练了语音识别模型，在不使用任何有标注韩语语音数据的情况下达到了27.33%的字错误率(Character error rate，CER)。同时，本文分别构建了汉语音素到韩语以及韩语音素到汉语这两种不同的映射关系，比较了两种不同映射方式的优劣。

1　韩语发音特点与韩语-汉语音素映射关系

1.1　韩语发音特点

在语言学分类上，韩语属于表音文字中的全音素文字，其书写符号体系以元音音素和辅音音素作为

基本单元,同时这些基本单元与语音间又存在着直接的对应关系。需要说明的是,这种对应关系并不意味着可以混淆语音和音素这两个概念,相反语音和音素的概念有明显的区别。音素是一个语言中区分语义的最小声音单位,音素是说话者和听者所识别的基本概念单元,而语音则是音素物理性的实现。韩语音素可分为辅音和元音两种,共包括19个辅音和21个元音。其中,韩语元音又可以分为10个单元音和11个由半元音和单元音组合而成的双元音。韩语辅音和元音如图1所示。

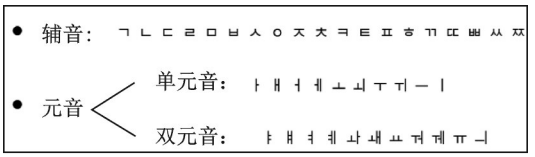

图1 韩语的辅音和元音

Fig.1 Consonants and vowels in Korean

1.2 韩语的音节与发音

韩语中另一个重要的概念是音节。韩语音节基本上由初声、中声和终声组成,不过一个完整音节中并不一定会包含全部三者。韩语的音节可以在没有终声的情况下用初声和中声来实现,也可以在没有初声的情况下用中声和终声来实现,中声单独也可以成音节。韩语的音节结构如图2所示。

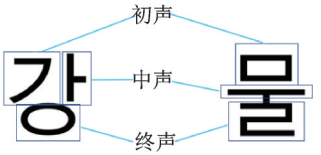

图2 韩语的音节结构

Fig.2 Syllable structure of Korean

韩语的19个辅音均可以作为初声,21个元音均可以作为中声,而终声则由14个基本辅音和13个复收音组成。图3展示了构成韩语音节的全部音素。韩语的音素和音节结构是构建韩语语音识别系统时最基本的要素。

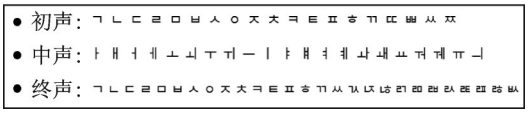

图3 构成韩语音节的音素

Fig.3 Phonemes of Korean

根据音素所处的位置或环境,音素的发音发生改变的现象称为音韵变化。韩语是一种音韵变化较多的语言,这种音韵变化的发生是为了发音的方便,体现了韩语发音的经济性。由于音韵变化现象的存在,韩语文字的表示和实际的发音经常有差异。图4展示了这种差异的一个例子。

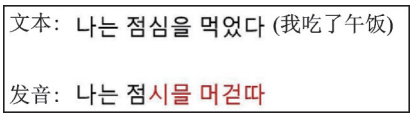

图4 韩语句子和其句子发音的差异

Fig.4 Example of pronunciation difference in Korean

韩语中的音韵变化主要分为以下5类:(1)一个音韵变成另一个音韵的交替现象;(2)一个音韵向另一个音韵靠拢的同化现象;(3)原来的音韵消失的脱落现象;(4)添加没有的音韵的添加现象;(5)两个音韵合为一的缩略现象。特别是如图5所示,音韵变化的交替现象中的“音节的终声规则”会发生在收音的位置上,“ㄱ、ㄴ、ㄷ、ㄹ、ㅁ、ㅂ、ㅇ”在收音上发音时按照原来的音价发音,但不属于此的其他收音的辅音则会被更换为“ㄱ、ㄷ、ㅂ”之一。脱落现象中的“辅音简化”也会在收音位置上发生,此时复收音中会脱落掉一个辅音,只发剩下的一个辅音。由于这些现象的存在,在收音的位置上只有“ㄱ、ㄴ、ㄷ、ㄹ、ㅁ、ㅂ、ㅇ”的7个辅音才会发音。因此,在韩语语音识别系统的构建过程中,需要应用这些复杂和多样的音韵变化规则,把韩语字母中的音素映射到表示其真实发音的符号上。

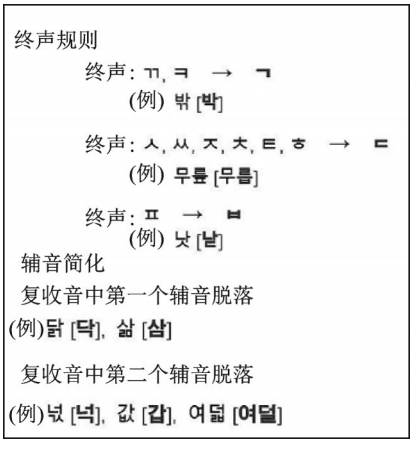

图5 音节的终声规则和辅音简化

Fig.5 Trailing consonant rules and consonant simplificatio

1.3 韩语-汉语音素映射关系

基于声学模型共享的零资源语音识别需要建立源语言和

目标语言之间的音素对应关系。在本文实验中,源语言为汉语,目标语言为韩语。通过比较汉语音素及韩语音素的国际音标(International phonetic alphabet, IPA)[12-13],本文建立了这两种语言的音素映射。这种映射关系可以是双向的,既可以用汉语的音素表示韩语中的单词,也可以反过来用韩语的音素表示汉语中的单词。为了探究两种映射方法对最终的零资源识别效果的影响,本文分别用这两种不同方法构建了汉语和韩语的发音词典。

1.3.1 汉语词用韩语音素表示

为了构建使用韩语音素集的汉语发音词典,需要将汉语词用韩语发音表示。在具体的实现过程中,本文采用了两步映射的策略,首先根据已有的汉语发音词典将汉语词表示为汉语自身的音素,之后再建立汉语音素到韩语音素的对应关系。大致的对应关系根据汉语音素与韩语音素的IPA构建,同时对于因两种语言的差异而发生的例外情况,本文也采用了特殊的处理方式。汉语音素到韩语音素的映射关系如图6所示。

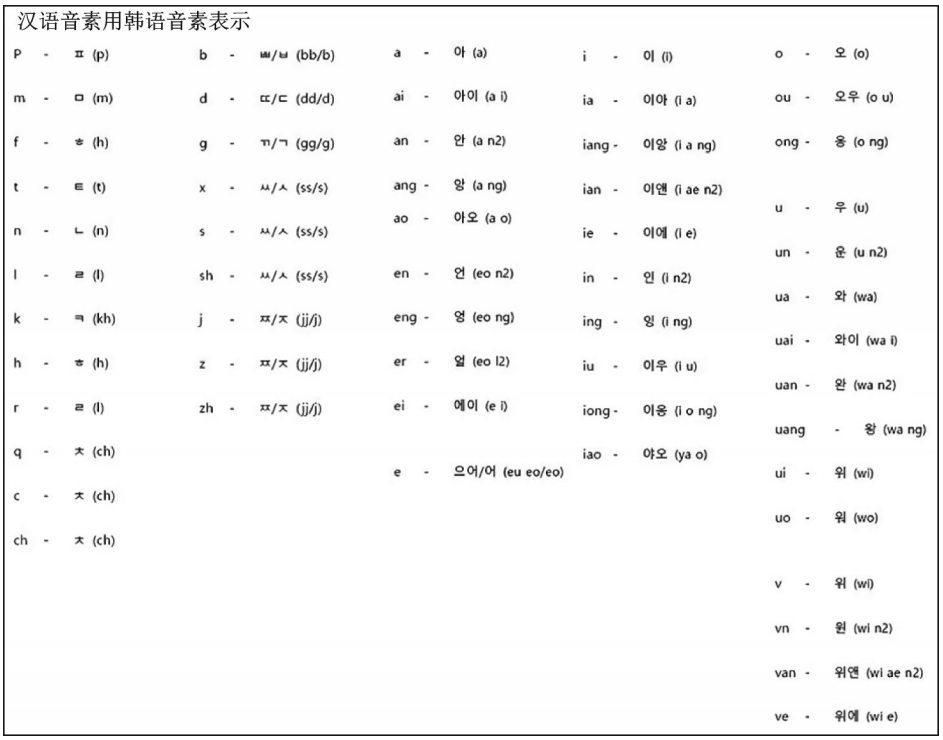

图6 汉语音素到韩语音素的映射关系

Fig.6 Phoneme mapping rules from Chinese to Korean

这种映射关系包含以下4种情况:

(1)汉语音素和韩语的音素一一对应,发音基本相同的情况。例如,汉语的声母中“p”“t”与韩文的辅音“p”“t”对应,汉语的单韵母中的“a”“u”与韩语的元音“a”“u”相对应。

(2)一个汉语音素发音相似的韩语音素可能超过2个。例如,在韩语的音素中,虽然没有与汉语的声母“f”相匹配的音素,但发音相似的韩语音素有辅音“h”与“bb”。另一个例子是汉语的音素中的“ia”,这是由两个单韵母结合而成的复合韵母。该韵母可以表示为韩语元音中包含两个以上的单元音的音节,也可以表示为包含双重元音的音节。

(3)汉语的同一个音素,根据声调所对应的韩语音素可能会不一样。例如,汉语声母中的“b”,从一声到四声时发音为韩语的辅音“bb”,而轻声时发音变为“b”。这是清音即不送气音“b”的音素为轻声时变成浊音的现象。

(4)无法用汉语发音韩语音素的情况。例如韩语音素中在收音位置的辅音“b2”和元音“oe”等。为了能正常地对韩语进行识别,实验所用的音素集除了那些在汉语中存在的韩语音素之外,也必须包含那些和汉语音素集不重合的韩语音素。因此,那些无法用汉语发音的剩下的韩语音素同样需要在汉语发音字典中有一个近似的对应。

考虑到两种语言之间音素映射关系的复杂情况,本文对映射关系进行了反复修改后构建了发音字典,并进行了相应的语音识别实验,通过比较识别率,最终选择了结果更优秀的对应关系。对于那些无法用汉语发音的剩下的韩语音素,本文在不包含无法用汉语发音的韩语音素的情况下,先构建了发音词典,之后从识别结果中找出与这种音素最相似的音素或发音后单独提取,并将其部分转换为韩语音素符号,之后重新添加到发音字典中。无法用汉语发音的韩语音素的标记方法如图7所示。

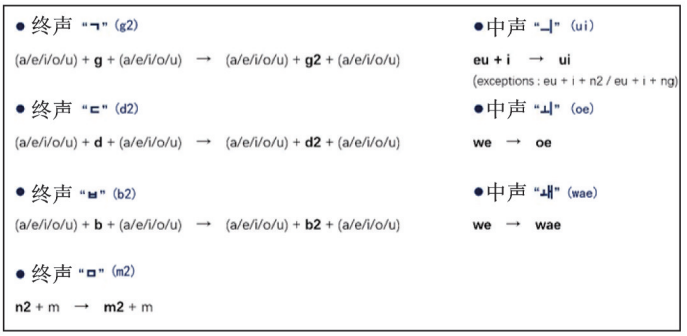

图7 无法用汉语发音的韩语音素的标记方法

Fig.7 Mapping the out-of-vocabulary Korean phonemes

1.3.2 韩语词映射到汉语音素

本文同样使用1.3.1节中提到的两步策略将韩语词映射到汉语音素,即首先通过已有的韩语发音词典将韩语词映射为韩语音素,再构建韩语音素到汉语音素的映射关系。另外,对于因两种语言的差异而发生的例外情况,本文同样进行了单独的处理。韩语音素到汉语音素的映射关系如图8所示。

1.3.3 韩语音素到汉语音素的映射关系

首先,确定两种语言间辅音的对应关系。在本文实验中,韩语辅音中“ㅅ、ㅆ”和“ㅈ、ㅉ、ㅊ”这2组辅音映射被映射为2个音素,分别是“s、x”和“j、z”。除了这5个辅音之外,其余音素都各自与1个汉语音素相对应。

其次,为了用汉语音素表现韩语元音和终声,本文将复合韵母拆分,用分离的音素表现韩语音素。例如,将“ian”这一个复合韵母分离为“i”“a2”“n2”,将每个音素作为一个音素使用。通过这样的过程,韩语的所有元音都可以用中文音素中相似的发音来标记,韩语终声中的“ㄴ、ㄹ、ㅇ”可以映射到“n2、r2、ng”,而韩语终声中的“ㄱ、ㄷ、ㅁ、ㅂ”则直接映射到汉语声母“g、d、m、b”。

最后,汉语声母中的“zh、ch、sh、r、f”,以及与

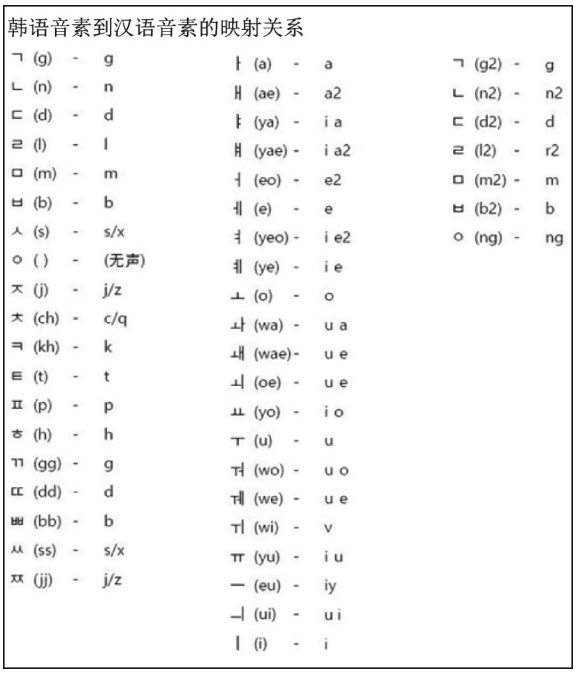

图8 韩语音素到汉语音素的映射关系

Fig.8 Phoneme mapping rules from Korean to Chinese

“zh、ch、sh、r”一起使用的“ix、iz”是很难用韩语发音的音素。因此,在用汉语音素表示的发音词典中,这些音素很少出现,仅在一两个词中被使用。

2 声学模型共享实验与不同映射关系的对比分析

2.1 数据集

本研究中使用的汉语训练数据是Aishell[14]提供的训练数据集。该数据集包含178 h的汉语语音语料库(400名说话人),以及转录的文本文件。韩语训练和测试数据则来自Zeroth[15]。Zeroth的训练集部分包含95.7 h、共181个说话人的标注语音,测试集则包含10名说话人的1.2 h语音语料以及对应的标注。另外,Zeroth数据集还包括韩语语言模型,发音字典和Morpheme-based Segmenter(Morfessor)[16]。Morfessor是一个对文本数据进行语素分割的工具。通过Morfessor可以将韩语语料库进行分割,这是Zeroth提供的发音词典和语言模型的基础。

2.2 Zeroth基线系统

本文使用Kaldi工具[1],利用Zeroth数据集[15]训练基线系统。解码时使用的测试集来自于Zeroth项目中预训练的3-gram语言模型。模型结构方面,本文分别使用了GMM-HMM(Gaussian mixture model)和TDNN-HMM(Time delay neural network-Hidden Markov model)两种模型结构,其中GMM模型使用了说话人自适应方法,高斯个数为40 000,决策树叶子节点个数为4 200。TDNN声学模型参数与后文声学模型共享实验中的设置一致。Zeroth基线系统的字错误率表现如表1所示。

表1 基线系统的字错误率表现

Table 1 CER of the baseline systems

模型结构	CER/%
GMM-HMM	17.74
TDNN-HMM	7.57

2.3 基于声学模型共享的韩语识别实验

2.3.1 实验设置

本文同样使用Kaldi工具来构建声学模型共享的语音识别系统。本文使用的声学模型结构为11层的TDNN,每层隐层维度为1 280,瓶颈特征维度为256,时延为3。模型输入为40维的MFCC特征。在损失函数方面,本文使用了交叉熵损失函数和最大互信息损失函数的组合,使用Kaldi提供的LF-MMI(Lattice free-Maximum mutual information)流程进行训练。语言模型方面,为了模拟真实的小语种识别情况,本文使用Zeroth项目的训练集文本来训练语言模型。Zeroth训练集文本包括22 263个短句,共448 466个词。本文使用SRILM工具,以wb-discount平滑方法训练了3-gram语言模型。解码时,本文使用了较低的声学模型权重来更充分地利用语言模型信息。训练和解码过程中使用的部分关键超参数参见表2。

表2 实验的部分关键超参数

Table 2 Some important hyper-parameters

超参数	设定值
Epoch	4
Batch Size	128
起始学习率	0.001
终止学习率	0.000 1
交叉熵损失权重	0.1
TDNN维度	1 280
TDNN瓶颈特征维度	256
TDNN时延	3
L2正则化系数	0.000 5
解码声学模型权重	0.4

2.3.2 不同映射方法的结果比较与分析

本文分别按照2.2节所述构建了2种不同的映射关系。基于这2种不同的映射关系分别构建了不同的发音词典,训练了汉语声学模型,并对Zeroth测试集数据进行了解码。实验结果如表3所示。其中,kr2zh表示将韩语词汇映射为汉语音素,zh2kr则表示将汉语词汇映射为韩语音素。kr2zh方法取得了27.33%的识别字

错误率,与TDNN模型有较大差距,但接近使用95 h有标注数据训练的GMM-HMM模型的结果。同时,实验结果表明映射方式的不同对汉语声学模型识别韩语语音的效果有着较大的影响。在使用zh2kr映射时,模型的字错误率为40.27%,识别的精度较低;而使用kr2zh映射时,模型的字错误率降低到27.33%,识别的精度得到了大幅度的下降。

表3 不同映射方式的字错误率结果

Table 3 CER under different mapping rules

映射方式	CER/%
kr2zh	**27.33**
zh2kr	40.27

本文认为,zh2kr相比kr2zh方法有较大差距的原因主要是将汉语音素映射为韩语音素的过程中引入了过多的错误标注。为了确保训练时对齐结果的正确性,在zh2kr流程中,本文为汉语中存在而韩语中不存在的音素同样指定了近似的对应关系。不过,尽管本文在构建汉语音素到韩语音素的过程中已经通过迭代修改的方式尽量寻找与汉语音素匹配的汉语词,但由于两种语言之间的差异,这种映射关系仍旧是不够精准的。举例来说,本文最终使用韩语音素ㅈ(ㅉ)、ㅊ、ㅅ(ㅆ)、ㄹ、ㅎ来表达汉语声母中的zh、ch、sh、r、f,但实际它们的发音之间仍有较大差距,这也就使得模型最终学习到的这些韩语音素对应的汉语声学特征和测试集中真正的韩语声学特征的分布有较大差异。换句话说,在使用这种映射关系时,本文其实人工引入了很多的领域漂移。另一方面,在kr2zh方法中,这些在源语言中存在而目标语言中不存在的音素无需被强行映射,因此避免了这个问题。

3 结束语

本文通过声学模型共享的方法,在不使用任何韩语数据的情况下构建了韩语语音识别系统,并在Zeroth韩语测试集上实现了27.33%的字错误率,证明了此方法在韩语语音识别上的潜力。另外,本文还比较了将源语言词汇表示为目标语言音素以及将目标语言词汇表示为源语言音素这两种不同的映射方式。结果证明,后者在实际使用时不会面临源语言中部分音素无法被精确映射的问题,从而令这种方法更适用于声学模型共享情况。实验中仍然存在着一些有待进一步探索的问题。首先,现阶段本文仍旧依赖于专家知识来设计映射关系,而数据驱动的方式可能有助于构建更为精准的对应关系。另外,韩语中仍然存在着一些和汉语对应较差的音素,如果能够训练多语言的声学模型,就有可能实现更低的识别错误率。最后,近年来出现的预训练模型为各种小语种的语音识别带来了较大的增益,如果能将预训练模型提取到的特征与声学模型共享方法结合起来,也有可能会进一步降低识别的错误率。

参考文献:

[1] POVEY D, GHOSHAL A, BOULIANNE G, et al. The Kaldi speech recognition toolkit[C]//Proceedings of the IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. [S.l.]: IEEE Signal Processing Society, 2011: 4960-4964.

[2] WATANABE S, HORI T, KARITA S, et al. Espnet: End-to-end speech processing toolkit[J]. Interspeech, 2018. DOI: 10.21437/Interspeech.2018-1456.

[3] GULATI A, QIN J, CHIU C C, et al. Conformer: Convolution-augmented transformer for speech recognition[J]. Interspeech, 2020. DOI: 10.21437/Interspeech.2020-3015.

[4] MOORE R K. A comparison of the data requirements of automatic speech recognition systems and human listeners[C]// Proceedings of the Eighth European Conference on Speech Communication and Technology. Geneva, Switzerland: [s. n.], 2003: 67-72.

[5] ROHIT P T S. End-to-end models for automatic speech recognition[EB/OL]. (2018-01-01) [2021-05-30]. http://interspeech2018.org/program-tutorials.html.

[6] EBERHARD D M, SIMONS G F, CHARLES D. Ethnologue: Languages of the world(24th edition)[EB/OL]. (2021-01-01) [2021-05-30]. http://www.ethnologue.com/statistics.

[7] DAS A, HASEGAWA-JOHNSON M. Cross-lingual transfer learning during supervised training in low resource scenarios [C]//Proceedings of the Sixteenth Annual Conference of the International Speech Communication Association. [S.l.]: ISCA, 2015. DOI:10.21437/interspeech.2015-700.

[8] HUANG J T, LI J, YU D, et al. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers[C]//Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. [S.l.]: IEEE, 2013: 7304-7308.

[9] SCHNEIDER S, BAEVSKI A, COLLOBERT R, et al. Wav2vec: Unsupervised pre-training for speech recognition[J]. Interspeech, 2019. DOI:10.21437/Interspeech.2019-1873.

[10] YI C, WANG J, CHENG N, et al. Applying wav2vec 2.0 to speech recognition in various low-resource languages[EB/OL]. (2020-01-01)[2021-05-30]. https://arxiv.org/pdf/2012.12121.

[11] PRASAD M, VAN ESCH D, RITCHIE S, et al. Building large-vocabulary ASR systems for languages without any audio training data[C]//Proceedings of the Interspeech. Graz, Austria: [s.n.], 2019: 271-275.

[12] HONG H-J, KIM S-H, CHUNG M-H J M. A phonetics based design of PLU sets for Korean speech recognition[J]. Malsori, 2008 (65): 105-124.

[13] JO J E. Contrastive study between the Chinese and Korean phonemes for the teaching of the Chinese pronunciation[D]. Seoul: Ewha Womans University, 2007.

[14] BU H, DU J, NA X, et al. Aishell-1: An open-source mandarin speech corpus and a speech recognition baseline[C]//Proceedings of the 2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA). [S.l.]: IEEE, 2017: 1-5.

[15] JO L. Zeroth project[EB/OL].(2018-01-01)[2021-05-30]. https://github.com/goodatlas/zeroth.

[16] CREUTZ M, LAGUS K. Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0[M]. Helsinki, Finland: Helsinki University of Technology Helsinki, 2005.

作者简介:

王皓宇(1997-),男,硕士研究生,研究方向:低资源语音识别,E-mail: w-hy21@mails.tsinghua.edu.cn。

JEON Eunah(1996-),女,本科,研究方向:多语种语音识别、语音合成,E-mail: tee16@tsinghua.org.cn。

张卫强(1979-),**通信作者**,男,副研究员,研究方向:语音与音频信号处理、机器学习,E-mail: wqzhang@tsinghua.edu.cn。

李科(1981-),男,研究方向:语音信号处理,E-mail: like@speechocean.com。

黄宇凯(1983-),男,研究方向:语音信号处理,E-mail: huangyukai@speechocean.com。

(编辑:张黄群)