

# 基于熵的二语语音习得评价研究 —以日本学习者习得汉语声母为例

冯晓莉<sup>1</sup>, 高迎明<sup>1</sup>, 林炳怀<sup>2</sup>, 张劲松<sup>1\*</sup>

<sup>1</sup>北京语言大学 信息科学学院, 北京市100083

<sup>2</sup>腾讯科技有限公司智能平台产品部

fengxiaoli314@163.com, gaoyingming1@sina.com,  
binghuailin@tencent.com, jinsong.zhang@blcu.edu.cn

## 摘要

本文引入“熵”对学习者的二语音素发音错误的分布情况进行了量化研究。通过对不同音素及不同二语水平学习者音素错误率和错误分散度的分析发现: 1. 错误率与错误分散度有较高的相关性, 二者的差异反映出错误分布的差异性; 2. 错误率类似的音素中, 与母语音素相似度越高的音素错误分散度越小; 3. 较初级水平, 中级水平学习者音素错误率下降而错误分散度上升。由此可见, 熵可以在错误率基础上可以进一步揭示学习者母语音系及二语水平对音素发音错误分散度的影响。

**关键词:** 二语语音习得; 发音错误分散度; 熵

## An Entropy-based Evaluation of L2 Speech Acquisition: The Preliminary Report on Chinese Initials Produced by Japanese Learners

Xiaoli Feng<sup>1</sup>, Yingming Gao<sup>1</sup>, Binghuai Lin<sup>2</sup>, Jinson Zhang<sup>1\*</sup>

<sup>1</sup>School of Information Sciences, Beijing Language and Culture University, Beijing, 100083

<sup>2</sup>Smart Platform Product Department, Tencent Technology Co., Ltd, China

fengxiaoli314@163.com, gaoyingming1@sina.com,  
binghuailin@tencent.com, jinsong.zhang@blcu.edu.cn

## Abstract

This study introduced “entropy” to quantify the dispersion of second language (L2) learners’ phonetic errors. By comparing the error rates and error dispersion of different phones and different proficiency levels of learners (elementary level (EL), intermediate level (IL) and advanced level (AL)), we found that: (1) There is a high correlation between error rate and error dispersion, and the difference between them reflects the difference of error distribution; (2) The greater the difference between target and native phones, the higher the pronunciation error rate and the greater the error dispersion; (3) Compared with the EL speakers, the IL speakers’ phone error rate decreased while their error dispersion increased. In summary, entropy can reflect the differences of L2 pronunciation distribution resulting from the influence of learners’ native language and L2 proficiency.

**Keywords:** L2 speech acquisition, pronunciation error dispersion, entropy

## 1 引言

在第二语言语音教学研究中,评价指标是评估二语学习者语音习得效果不可或缺的工具。现有的评价指标主要从发音正确与否、发音质量好坏两个方面对学习者的发音进行评估。针对发音正确与否,研究者主要通过判断学习者产出的语音是否正确(Chen et al., 2016; Jia et al., 2006)及目标语音错成了什么(Jia et al., 2006; Jouvét et al., 2015),以确定目标语音整体习得水平、语音错误类型及不同语音范畴间可能存在的混淆情况等。对发音质量好坏的评估则主要以知觉打分(如使用Likert量表)、声学测量等方式对学习者的自然度(Tsurutani and Luo, 2013)、口音度(Jesney, 2004)、类母语度(Sun and van Heuven, 2007)、可懂度(Crowther et al., 2015)、声学参数分散度(Xie and Jaeger, 2020; Smith et al., 2019)及不同语音范畴在声学上的交叠程度(Wang et al., 2006; Bohn and Flege, 2011)等进行评测,以判断学习者的发音偏离标准音的程度或距离。如果说发音质量好坏的评估是对学习者的发音偏离标准音“量”的测量,发音正确与否即是对学习者发音是否出现“质”的错误的评判。

针对学习者发音的正误,目前的研究主要采用正确率(Jia et al., 2006)、错误率(Chen et al., 2016)、混淆矩阵(Jia et al., 2006; Jouvét et al., 2015)等指标进行量化评估。如Chen et al. (2016)使用大规模中介语语音语料库考察了欧洲多母语背景学习者汉语声、韵、调的习得情况,结果显示学习者音段错误率为5%,声调错误率为32%。Flege et al. (1997)对学习者的英语元音[i]、[ɪ]、[e]、[æ]进行了评估,研究发现不同母语背景学习者产出的元音的正确率存在较大差异,如德国学习者产出的元音[i]的正确率可以达到100%,而西班牙学习者只有57%。此外,混淆矩阵作为一种多维向量,同时呈现了目标语音的正确率及各种错误类型的比率分布,如Jia et al. (2006)以混淆矩阵的形式呈现了三组不同二语水平的中国学习者对英语单元音[i]、[ɪ]、[e]、[æ]、[ɑ]、[ʌ]、[u]的习得情况,结果发现与学习者母语音素发音相同的[i]、[u]的正确率最高,而学习者母语中不存在的[e]、[æ]和[ɑ]、[ʌ]两组对立音位存在较大程度的互相混淆,正确率较低。在该类研究中,正误率(正确率和错误率)指标可以对不同语音单元的习得程度进行对比,混淆矩阵在正误率指标的基础上可以进一步揭示目标语音单元的发音错误分布及不同语音单元间互相混淆的程度和方向性。

然而,正误率指标并不能精准反映二语学习者语音习得状况的所有面貌。例如,Jouvét (2015)在考察法国学习者产出德语元音时发现,虽然[e:]和[a:]两个音素的正确率相同,都是83%,但是[e:]主要有[i:]、[ɛ:]、[ɪ]三种错误类型,对应百分占比分别为4%、2%和4%,而[a:]只有与母语音素发音相似的[a]一种主要错误形式,对应百分占比为16%。又有,Jia (2006)在研究中国学习者产出英语单元音[ʌ]时发现,随着学习者在目的语环境生活时间的延长,学习者在音素[ʌ]上的正确率始终在50%左右,其错误形式却由多种逐渐集中到[ɑ]一种主要错误上。因此,具有相似正误率的不同语音单元或者不同学习阶段的同一语音单元,其发音错误分布可能存在较大的差异,而这种差异往往可以反映出学习者产出的目标语音是否存在确定的错误类型。本文综合考虑学习者产出的目标语音的错误类型的数量及百分占比,提出发音错误分散度:发音错误分散度越小,目标语音实际产生的错误类型越少、百分占比越集中。发音错误分散度可以揭示出相似正误率的语音在发音错误对象上的确定性,从而为二语研究者及二语教学者在发音偏误研究及二语语音教学策略提供一定的参照。显然,正误率指标难以对学习者的发音错误分散度进行量化分析,混淆矩阵尽管可以详细呈现学习者在目标语音上发音错误的分布情况,却无法对发音错误分散的分散情况进行总体量化。

针对上述问题,本文引入信息论中的“熵”来对二语语音习得效果中的发音错误分散度进行量化研究。首先,我们将每个目标语音单元产出作为一个随机变量,每一种实际产出形式作为变量的取值,其概率倒数的对数对应着该取值的不确定性,所有可能取值不确定性的加权和反映该目标语音单元产出的平均不确定性,即“熵”。熵越小说明语音单元产出对象越集中(熵为0时即表示变量只存在一种产出形式),熵越大说明语音单元产出形式越分散。

为了验证熵在反映二语语音发音偏误分散度中的作用,本文使用日本学习者产出的汉语普通话音节首辅音(声母)为实验语料,从不同音素间的横向对比及不同二语水平学习者(EL、IL、AL)音素发音错误的纵向发展两方面开展了实验研究。下文研究方法部分对本研究使用的实验语料及评价指标进行了介绍;实验结果部分分别对不同音素及不同二语水平学习

根据《Creative Commons Attribution 4.0 International License》许可出版

\*通讯作者:张劲松(jinsong.zhang@buc.edu.cn)

致谢:本工作得到中央高校基本科研业务专项资金(20YJ040002)、北京语言大学梧桐创新平台(19PT04)、语言资源高精尖中心项目“面向智能语音教学的汉语中介语语音多模态语料库研究”(KYR17005)、教育部规划基金项目(18XJJA740001)及研究生创新基金项目(21YCX178)的资助。

者的音素错误率和发音错误分散度进行了对比分析；讨论和结论的部分分别对实验结果进行了讨论并得出本研究的结论。

## 2 研究方法

### 2.1 实验语料

#### 2.1.1 发音人

本研究采用了BLCU-SAIT汉语中介语语音语料库中日本学习者的发音数据作为实验语料。发音人为55位日本学习者（男性17，女性38），年龄19-35周岁（平均年龄22.8周岁，标准差：4.2）。所有发音人均在日本出生长大，开始学习汉语的年龄均在18岁之后。根据发音人的HSK（汉语水平考试）等级，55位发音人被分为初、中、高三个等级，其中初级水平（HSK3级及以下）13人；中级水平（HSK4-5级）22人；高级水平（HSK6级）20人。

#### 2.1.2 音素分布

每位发音人需完成BLCU-SAIT汉语中介语语音语料库中283个双音节词和103个句子的录音，双音节词和句子的设计均考虑了汉语拼音方案声韵组合的合法性及多样性，具体录音文本设计可参照王玮(2019; 2020)。本研究使用语音数据20921条，其中包含汉语声母93280个，录音文本中汉语声母频次分布如图1所示。

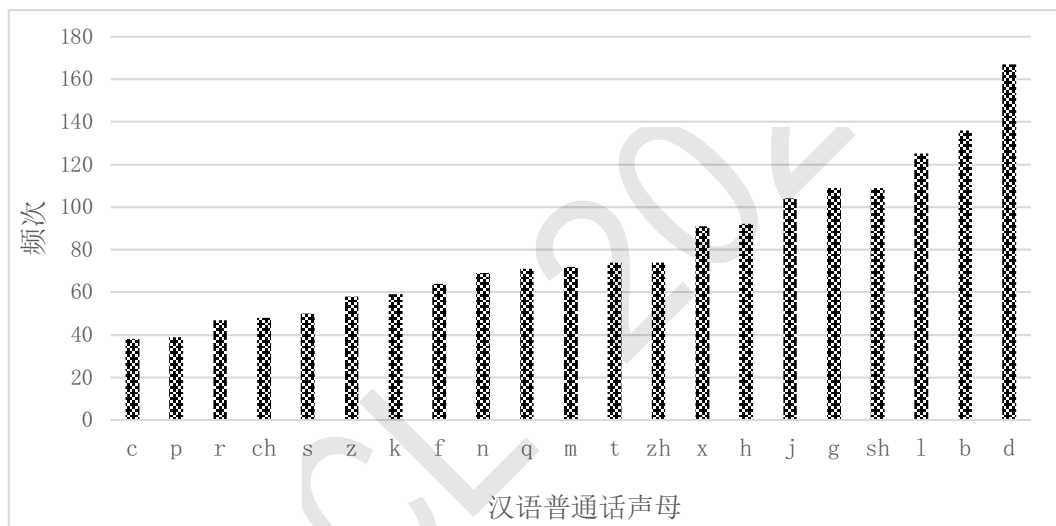


Figure 1: 录音文本声母频次分布

#### 2.1.3 数据标注

本研究所使用的全部数据都进行了声韵母发音错误的人工标注。标注规范的设定参照了曹文、张劲松(2009)语料库标注方案中对音段发音偏误趋势标注的理念，标注内容不仅涵盖了传统语料库标注中替换式、删除式、插入式等错误种类，还对二语学习者中介语发音中“似A似B”的发音错误趋势进行了标注。标注符号为汉语拼音符号和附加符号（使用汉语拼音不足以描述清楚的发音错误，使用附加符号在中进行标记）。具体标注方案见王玮(2020)。

标注员为经过系统培训的语言学专业本科生和研究生，均来自中国北方方言区，普通话达到二级甲等及以上水平。所有标注员都可以对汉语普通话声母、韵母进行清晰的识读、辨别和区分，具备本标注任务所需语音学基本知识，如使用严式音标记音、正确描述语音发音属性等，并且可以熟练使用语音标注软件Praat(Boersma and Weenink, 2021)。标注过程由项目负责人将待标注数据进行等量分包然后随机分配给标注员，待标注完成后由三位有经验的质检员按照30%的比例对标注数据进行抽检，抽检正确率达到90%及以上方为合格（标注合格的标准为能够准确找出学习者的发音错误并使用正确的符号进行标注，漏标、错标均为标注不合格），否则需要将全部数据重新进行分配，进行再次标注，直至标注合格为止。

## 2.2 评价指标

### 2.2.1 错误率

在以往的研究中，错误率是衡量二语学习者的语音习得水平必不可少的指标，发音错误率的计算方式如公式(1)所示。

$$p = \frac{N_{err}}{N_{sum}} \times 100\% \quad (1)$$

其中， $N_{sum}$ 表示该目标语音样本的数量， $N_{err}$ 表示目标语音出现发音偏误的样本的数量， $p$ 表示目标语音的发音错误率。

### 2.2.2 发音偏误分散度

二语学习者在产出目标语音时实际发音可能包括正确发音和多种错误发音类型，并且每种发音类型所占百分比具有不确定性。为了对学习者的这种发音分散情况进行量化，本研究提出使用信息论中熵的计算方法对二语学习者发音分散度进行量化，量化方式如公式(2)所示。

$$H(x) = - \sum_{i=1}^n p_i \log_2 p_i \quad (2)$$

将每个目标语音单元的产出 $x$ 作为一个随机变量，其实际产出形式 $P_i$ 存在 $n$ 种取值，每种取值对应的概率为 $p_i$ ，其概率倒数的对数 $-\log_2 p_i$ 对应着该取值的不确定性。所有可能取值不确定性的概率加权和反映该目标语音单元产出的平均不确定性，即“熵”。熵越小说明语音单元产出对象越集中（熵为0时即表示变量只存在一种产出形式），熵越大说明语音单元产出形式越分散。

基于上述“熵”的计算方法，当对目标语音单元 $x$ 的实际发音进行正误二元判断时， $i$ 仅存在两种取值，即正确和错误，对应的概率分别为 $p_{err}$ 和 $p_{corr}$ ，此时目标语音的产出形式的分布情况对应熵的基准值 $H(x)_{base}$ ，量化方式如公式(3)所示。

$$H(x)_{base} = -(p_{err} \log_2 p_{err} + p_{corr} \log_2 p_{corr}) \quad (3)$$

熵 $H(x)$ 与熵的基准值 $H(x)_{base}$ 的差即为由发音错误的分散性造成的熵的增加，即由错误类型的增加或不同错误类型的百分占比均衡化导致的发音不确定性的增加。因此，通过计算 $H(x)$ 与 $H(x)_{base}$ 的差值我们可以得到发音错误分散度 $\Delta H(x)$ ，计算公式如(4)所示。

$$\Delta H(x) = H(x) - H(x)_{base} \quad (4)$$

## 3 实验结果

本研究结合音素错误率和发音错误分散度两个指标，使用日本学习者产出的汉语中介语语音语料，对学习者的汉语声母的发音错误情况进行了研究。首先，本研究所使用的55名日本学习者汉语中介语语音语料中声母的整体错误率约为10%。通过上述熵的量化公式，我们首先将21个汉语声母作为21个随机变量，每个声母在全部声母中的百分比为其对应的概率值为 $p_i$ ，使用公式(2)可以得到正则语音的整体熵值为4.28。以同样的方式，我们将55名日本学习者在21个声母的每一个实际产出对象作为一个随机变量，每一个随机变量占全部语音产出形式的百分比为其对应的概率值 $p_i$ ，通过(2)可以计算出学习者整体发音熵 $H(x)$ 为4.88。以全部声母的错误率和正确率为随机变量，使用公式(3)可以计算出学习者整体发音基准熵 $H(x)_{base}$ 为4.69，整体发音熵 $H(x)$ 与整体发音基准熵 $H(x)_{base}$ 的差值即为整体发音错误分散度 $\Delta H(x)$ ，取值为0.19。

以下将使用上述量化方法分别对日本学习者21个汉语声母的错误率和发音错误分散度以及初、中、高三个水平发音人的平均音素发音错误率和发音错误分散度进行解析。

### 3.1 日本学习者汉语声母错误率和发音错误分散度对比研究

图2展示了日本学习者汉语21个声母的错误率和发音错误分散度的分布情况，柱状图对应声母发音错误率，曲线图对应各个声母的发音错误分散度。

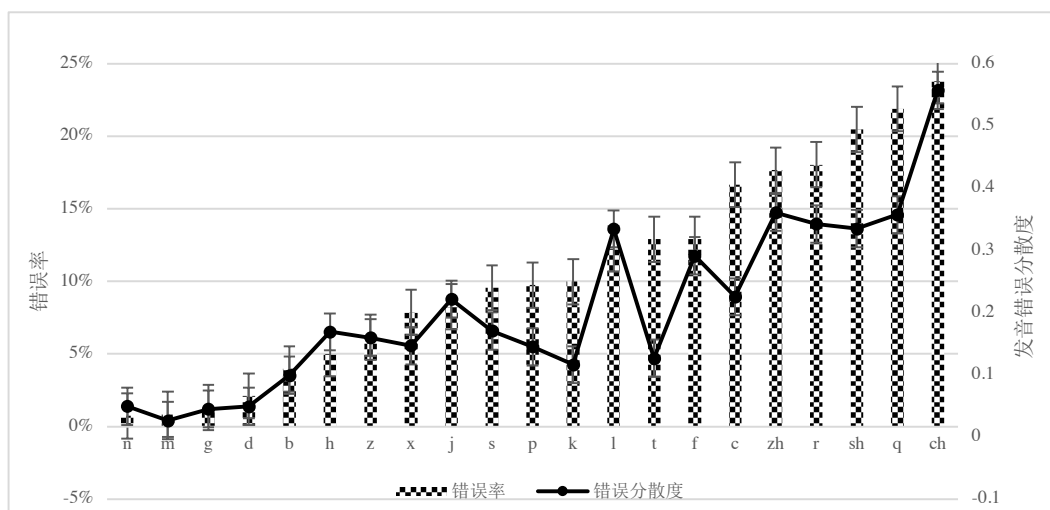


Figure 2: 日本学习者不同汉语声母发音错误率及发音错误分散度

根据图2首先可以看出，音素错误率和发音错误分散度整体分布趋势比较一致，即发音错误率较低的音素对应的音素发音错误分散度也较低，反之亦然。相关性检验结果显示，声母发音错误率与发音错误分散度的相关系数为0.91。同时，根据上述结果可以发现，与学习者母语发音相似度较高的m[m]、n[n]、b[p]、d[t]、g[k]的错误率和发音错误分散度最低，与学习者母语音素差异较大的zh[ts]、ch[tʂʰ]、sh[s]、r[z]等声母的发音错误率和发音分散度最高。

其次，对比不送气声母b[p]、d[t]、g[k]、j[tɕ]、z[ts]、zh[tʂ]与其对立的送气声母p[pʰ]、t[tʰ]、k[kʰ]、q[tʰ]、c[tʂʰ]可以发现，日本学习者产出不送气声母的错误率和发音错误分散度都明显小于与之对立的送气声母，这可能是由于日语的辅音仅存在清浊的对立而不存在送气与不送气对立(朱春跃, 2001)，日本学习者对送气这一声学线索不够敏感(Holt and Lotto, 2006)。

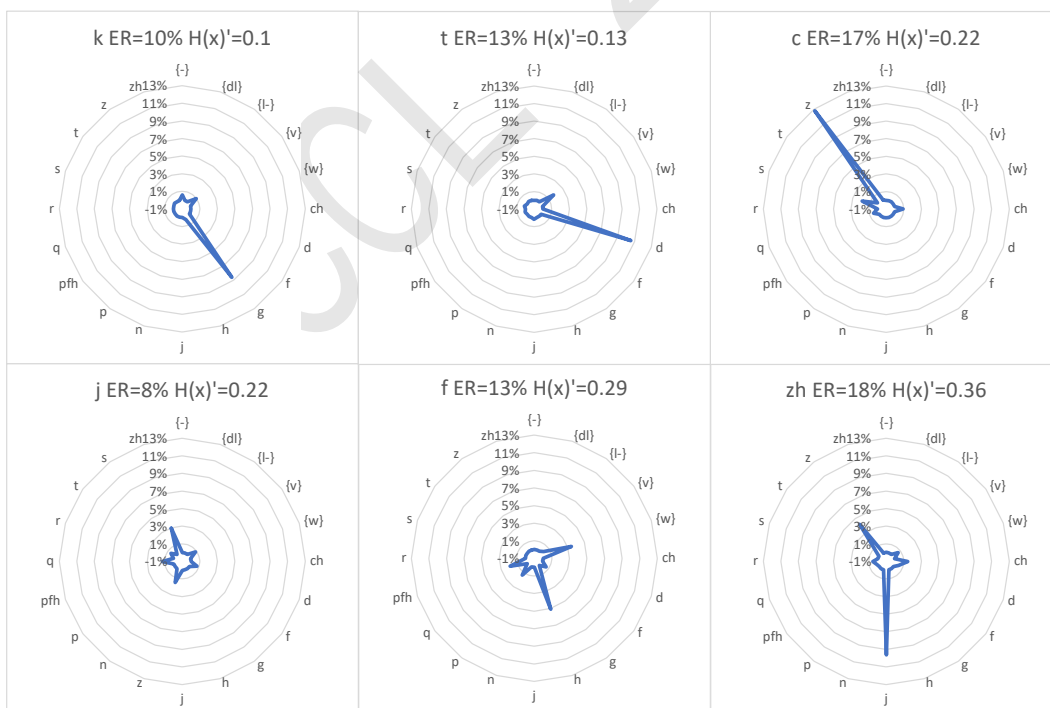


Figure 3: 错误率类似而错误分散度差异较大的音素错误分布情况示例

第三，学习者发音错误率和错误分散度的差异性主要通过发音错误率相似而错误分

散度差异较大的声母得以体现，如k[k<sup>h</sup>]和j[tɕ]、t[t<sup>h</sup>]和f[f]、c[ts<sup>h</sup>]和zh[tʂ]等声母的错误率类似而错误分散度上却存在较大的差异。雷达图3详细成仙了上述音素的发音错误对象，可以看出k[k<sup>h</sup>]、t[t<sup>h</sup>]、c[ts<sup>h</sup>]的错误对象主要是与之对立的不送气辅音g[k]、d[t]、z[ts]，发音错误分布非常集中；而与g[k]、d[t]、z[ts]在错误率上接近的j[tɕ]、f[f]、zh[tʂ]音素的发音错误类型明显更多也更分散，结合图2可以看出j[tɕ]、f[f]、zh[tʂ]的发音错误率明显高于g[k]、d[t]、z[ts]，由此可以看出，错误率类似的音素对应的发音错误的分布情况可以使用熵进行有效的量化。同时，由本研究中日本学习者产出汉语普通话声母的发音错误情况可以看出，当目标语音为学习者母语发音类似音素的送气对立音时，目标音素的发音错误分散度低于与其错误率接近的其它音素，并且错误类型主要为与目标语音对立的不送气音。

综上所述，本章节主要得到如下结果，第一，二语音素与学习者母语音素越接近时，学习者的发音错误率越低，发音错误分散度整体也越小。第二，日本学习者产出不送气声母的错误率和发音错误分散度都明显小于与之对立的送气声母。第三，当目标音为与学习者母语类似音素的送气对立音时，目标音素的发音错误分散度明显低于与其错误率接近的其它音素。由此可以得出，使用熵对发音错误分散度进行量化，可以在音素错误率的基础上进一步揭示出目标音素发音错误对象的分布情况：发音错误分散度越小，目标音素的错误类型越少、主要错误类型的百分占比越集中。

### 3.2 初、中、高水平日本学习者汉语声母错误率和发音错误分散度对比研究

本章节从发展的角度出发，分别对初、中、高三组不同二语水平的日本学习者产出汉语声母的错误率和发音错误分散度进行了考察。以下图表和公式中分别将初级、中级、高级三个等级简写为EL、IL、AL。首先根据分组计算每个音素的错误率和发音错误分散度，通过取平均得到每组音素的平均错误率和平均错误分散度，如图4所示。

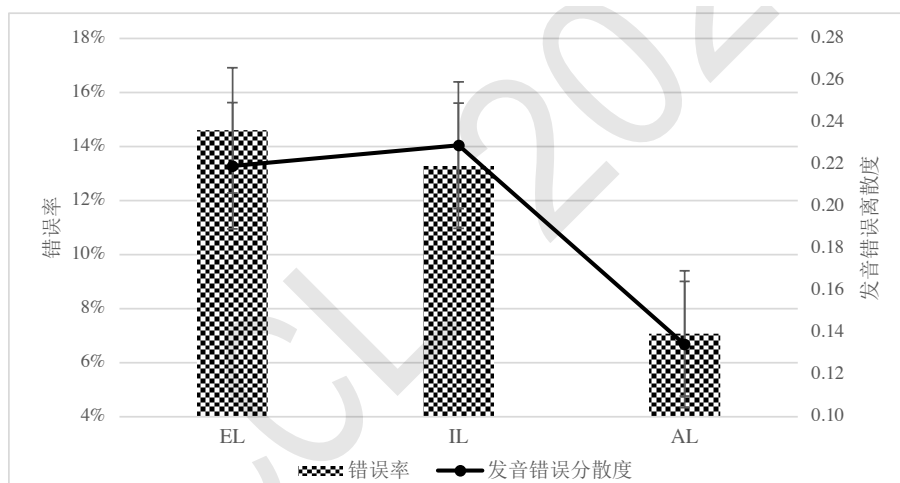


Figure 4: 初、中、高水平日本学习者平均声母错误率及评价发音错误分散度

根据图4可以看出，从初级水平到高级水平，学习者音素平均错误率整体逐步下降，而平均发音错误分散度却呈现先微弱上升后下降的发展趋势，相关性检验结果显示平均错误率与平均发音错误分散度的相关系数为0.966。运用Shapiro-wilk分别对每组数据进行正态性检验，结果显示，初、中、高三个水平的音素错误率均满足正态分布的条件 ( $P_{EL}=0.115>0.05$ ,  $P_{IL}=0.276>0.05$ ,  $P_{AL}=0.419>0.05$ )；三组发音错误分散度中，除初级水平不满足正态分布的条件 ( $P_{EL}=0.033<0.05$ )，IL、AL均为正态分布 ( $P_{IL}=0.287>0.05$ ,  $P_{AL}=0.136>0.05$ )。针对满足正态分布和不满足正态分布的组别，我们分别使用配对样本T检验和Wilcoxon检验考察不同组别错误率和不同组别发音错误分散度差异的显著性。检验结果显示，从发音错误率来看，与初级水平相比，中级水平学习者声母发音错误率下降不显著 ( $t=-1.706$ ,  $P=0.103>0.05$ )，而中级水平到高级水平学习者发音错误率显著下降 ( $t=4.38$ ,  $P<0.001$ )；从发音错误分散度来看，与初级水平相比，中级水平学习者声母发音错误分散度上升不显著 ( $z=-0.574$ ,  $P=0.569>0.05$ )，而中级水平到高级水平学习者发音错误率显著下降 ( $t=4.38$ ,  $P<0.01$ )。

根据上述组间数据对比可以看出,从初级水平到中级水平,虽然学习者在发音错误率和发音错误分散度两个指标的差异均不显著,而两指标在变化趋势上却呈现出相反的方向,即从初级水平到中级水平,学习者音素错误率下降而错误分散度呈微弱上升的趋势;从中级水平到高级水平,发音错误率和发音错误分散度均显著下降,即从中级到高级水平,学习者的音素发音的准确性有显著进步且目标音素的发音不确定性显著下降。

## 4 讨论

本文针对第二语言语音习得效果的评价问题,在错误率这一指标的基础上提出了发音错误分散度,并且尝试使用“熵”的计算方法对二语学习者发音错误的分布情况进行量化。本章节我们将分别从不同音素错误率与发音错误分散度以及不同二语水平学习者音素错误率与发音错误分散度两方面对结果进行讨论。

### 4.1 不同音素错误率与发音错误分散度研究

第二语言语音习得相关理论和研究指出,学习者二语语音习得的效果往往受到其母语经验的影响(Kuhl, 1993; Flege, 1995; Best and others, 1994; Best and Tyler, 2007)。Flege (1995)在SLM (speech learning model) 中提出了一系列二语语音习得的假设,其中包括: 1.学习者母语与目的语发音的相似度对二语语音习得有重要的影响; 2.二语语音范畴的构建会受到等价归类 (equivalent classification) 机制的影响。以上假设分别从学习者一语对二语的影响和学习者二语语音的内在加工机制两方面对不同音素的习得结果进行了预测,根据预测学习者对不同的二语语音往往采用不同的学习策略: 对与母语中发音一致的“相同音素” (identical phones) 及发音类似的“相似音素” (similar phones), 一般直接使用母语中的近似音代替二语中音素的发音; 对于母语中不存在的“陌生音素” (new phones), 只能预测该类音素可能存在学习困难, 却无法预测这类音素可能会被发成什么音(鲁健骥, 1984)。由于“陌生音素”与学习者母语中的音素差异较大, 学习者往往可以比较容易感知到其与母语音素的差异性, 因此在初期学习时可能存在一定的困难, 通过不断练习学习者有可能最终可以构建起新的语音范畴; 对与母语发音接近的“相似音素”, 由于难以感知其与母语音素的差异性, 最终可能会被“等价分类” (category classification) 到与之相似的母语的语音范畴中, 形成复合型语音范畴 (composite L1-L2 phonetic category) (Flege and Ocke-Schwen, 1997)。显然, 上述预测不仅对不同二语语音的习得结果进行了预测, 对不同的语音学习者可以出现的发音错误状态也进行了预测, 如果学习者对不同音素的习得情况符合上述预测, 那么“相同音素”的错误率最低, 语音产出对象最集中, 即直接由母语中对应的“相同音素”进行替换, 此时发音错误分散度最小; “相似音素”的错误率较“相同音素”会有所上升, 由于产出语音主要由母语中对应的“相似音素”进行替换, 发音错误的分散度也继续保持较低的状态; “虽然在SLM中陌生音素可能不是最难习得的, 其错误对象却存在最大的不确定性, 因为在学习者母语中不存在与之对应的替换对象, 在新的语音范畴建立之前, “陌生音素”可能存在多种不确定发音对象, 发音错误率也保持在较高水平。

本研究使用发音错误率和发音错误分散度两个评价指标呈现出日本学习者产出汉语21个声母的情况。根据上述实验结果可以发现, 与日语音素具有相同发音部位和发音方法的鼻音m[m]、n[n], 清塞音b[p]、d[t]、g[k]等音素可被归入“相同音素”, 其对应错误率和发音错误分散度最低; 日语中不存在的zh[tʂ]、ch[tʂʰ]、sh[ʃ]、r[ʒ]等音素可归入“陌生音素”(廖序东, 2017)(?), 其对应的发音错误率和发音错误分散度最高, 其余“相似音素”的发音错误率和发音错误分散度整体居中。此外, “相似音素”中p[pʰ]、t[tʰ]、k[kʰ]与“相同音素”b[p]、d[t]、g[k]为送气对立音, 它们虽然在错误率上与s[s]、l[l]、f[f]等音素接近, 其发音错误分散度却明显低于s[s]、l[l]、f[f]等音素, 根据发音错误分布雷达图3显示, p[pʰ]、t[tʰ]、k[kʰ]的错误对象主要为其不送气对立音b[p]、d[t]、g[k], 即日本学习者主要使用b[p]、d[t]、g[k]来替换p[pʰ]、t[tʰ]、k[kʰ]。“陌生音素”的错误分布有更大的不确定性, 发音错误分散度表现为最高。上述实验结果不仅印证了SLM对二语语音习得结果的预测, 也对学习者对不同音素的错误模式进行了探索, 对音系对比、二语语音习得研究及二语语音教学策略的探索都能提供一定的参照。

## 4.2 不同二语水平音素错误率与发音错误分散度研究

Slinker (2013)指出, 中介语被认为是一种动态的、不稳定的语言系统, 这往往是因为学习者在不同的学习阶段使用不同的学习策略来获得系统的规则和信息。Flege et al. (2021)在SLM-r (speech learning model-revised) 也指出, 当学习者初次接触目的语时往往使用母语中存在的语音进行替换发音, 随着学习的不断深入, 学习者可能会逐渐感知到二语语音与母语语音的差异, 从而尝试构建新的语音范畴。

本文通过对比不同水平日本学习者汉语声母产出的情况可以看出, 从初级水平到中级水平, 学习者的音素错误率呈下降趋势而发音错误分散度呈现微弱上升趋势, 虽然错误率和发音错误分散度的变化均没有呈现出显著性差异, 我们可以尝试对错误率下降和发音错误分散度上升的现象进行解读: 初级水平学习者二语发音错误率较高, 在学习上主要使用母语中存在的语音替换二语中的语音, 由于错误的对象主要为母语中的音素集合, 因此错误的分散度较小; 中级水平学习者, 一方面音素错误率有一定的下降, 同时学习者感知到二语语音与母语语音的差异性, 尝试为二语语音构建新的语音范畴, 在这一过程中可能会做更多更大胆的尝试, 在尝试构建新的语音范畴的过程中也可能伴随更多类型的发音错误, 从而造成发音错误分散度的增加。从中级水平到高级水平, 学习者音素发音的错误率和发音错误分散度均呈现显著性下降, 说明学习者的音素发音取得整体进步, 不仅整体错误发音比例降低, 而且发音的不确定性减小, 错误的种类更集中, 这在一定程度上可以反映出中介语语音系统逐渐趋于稳定的发展趋势(Selinker and Rutherford, 2013)。

上述解析只针对于本研究的实验结果, 这种发音错误率于错误分散度发展的不同步性是否普遍存在于二语习得的发展过程中, 以及不同音素的发展变化有怎样的具体表现有待于进一步探索。

## 5 总结与展望

本研究通过对日本学习者汉语不同声母发音错误率和发音错误分散度的分析与对比, 主要有如下发现: 第一, 音素错误率与错误分散度两个参数存在较大的相关性, 其差异性较好的反映出二语学习者发音错误的分布情况, 即错误率相似的两音素, 发音分散度越小错误类型越集中; 第二, 二语音素与学习者母语音素的差异性对学习者的音素发音错误率和发音错误分散度都有较大的影响: 二语音素与学习者母语音素差异越大, 发音错误率越高, 错误分散度也越大; 第三, 送气音素p[p<sup>h</sup>]、t[t<sup>h</sup>]、k[k<sup>h</sup>]的发音错误分散度明显低于与其具有类似错误率的其它音素。对比初、中、高三个水平学习者对不同音素的习得情况可以发现, 与初级水平学习者相比, 中级水平学习者的音素发音错误率下降而发音错误分散度微弱上升, 中级水平到高级水平学习者的发音错误率和发音错误分散度都显著下降。上述发现证明了使用“熵”可以对二语学习者的音素发音错误分散度进行有效量化, 同时发音错误分散度这一指标在错误率的基础上可以进一步揭示不同音素的发音错误分布情况及不同二语水平学习者音素的习得规律。

本文引入“熵”对二语语音发音错误的分散度进行量化, 对学习者的母语音系和二语学习经验对音素习得效果的影响进行了初步探索, 为二语语音习得效果的研究提供了新的视角。然而当前的研究结果是否符合不同母语背景的学习者, 学习者在感知上会有什么样的错误模式等问题则需要更深入且广泛的研究来探索。

## 参考文献

- Catherine T Best et al. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech perception: The Transition from Speech Sounds to Spoken Words*, 167(224):233-277.
- Catherine T Best and Michael D Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, pages 13-34.
- Paul Boersma and David Weenink. 2021. Praat: Doing phonetics by computer (version 6.2.04).
- Ocke Schwen Bohn and James Emil Flege. 2011. Perception and production of a new vowel category by adult second language learners. *Second-Language Speech: Structure and Process*, pages 53-74.



- Nancy F Chen, Darren Wee, Rong Tong, Bin Ma, and Haizhou Li. 2016. Large-scale characterization of non-native mandarin chinese spoken by speakers of european origin: Analysis on icall. *Speech Communication*, 84:46–56.
- Dustin Crowther, Pavel Trofimovich, Kazuya Saito, and Talia Isaacs. 2015. Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL quarterly*, 49(4):814–837.
- James Emil Flege and Ocke-Schwen Bohn. 2021. The revised speech learning model (slm-r). *Second Language Speech Learning: Theoretical and Empirical Progress*, pages 3–83.
- James Emil Flege and Bohn Ocke-Schwen. 1997. Perception and production of a new vowel category. *Second-Language Speech: Structure and Process*, 13:53.
- James Emil Flege, Ocke-Schwen Bohn, and Sunyoung Jang. 1997. Effects of experience on non-native speakers' production and perception of english vowels. *Journal of Phonetics*, 25(4):437–470.
- James E Flege. 1995. Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 92:233–277.
- Lori L Holt and Andrew J Lotto. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, 119(5):3059–3071.
- Karen Jesney. 2004. The use of global foreign accent rating in studies of l2 acquisition. *Calgary, AB: University of Calgary Language Research Centre Reports*, pages 1–44.
- Gisela Jia, Winifred Strange, Yanhong Wu, Julissa Collado, and Qi Guan. 2006. Perception and production of english vowels by mandarin speakers: Age-related differences vary with amount of l2 exposure. *The Journal of the Acoustical Society of America*, 119(2):1118–1130.
- Denis Jouviet, Anne Bonneau, Jürgen Trouvain, Frank Zimmerer, Yves Laprie, and Bernd Möbius. 2015. Analysis of phone confusion matrices in a manually annotated french-german learner corpus. In *Workshop on Speech and Language Technology in Education*.
- Patricia K Kuhl. 1993. Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, pages 259–274. Springer.
- Larry Selinker and William E Rutherford. 2013. *Rediscovering interlanguage*. Routledge.
- Bruce L Smith, Eric Johnson, and Rachel Hayes-Harb. 2019. Esl learners' intra-speaker variability in producing american english tense and lax vowels. *Journal of Second Language Pronunciation*, 5(1):139–164.
- Lei Sun and Vincent J van Heuven. 2007. Perceptual assimilation of english vowels by chinese listeners: Can native-language interference be predicted? *Linguistics in the Netherlands*, 24(1):150–161.
- Chiharu Tsurutani and Dean Luo. 2013. Naturalness judgement of l2 mandarin chinese-does timing matter? In *INTERSPEECH*, pages 239–242.
- Hongyan Wang, Vincent J Van Heuven, et al. 2006. Acoustical analysis of english vowels produced by chinese, dutch and american speakers. *Linguistics in the Netherlands 2006*, pages 237–248.
- Xin Xie and T Florian Jaeger. 2020. Comparing non-native and native speech: Are L2 productions more variable? *The Journal of the Acoustical Society of America*, 147(5):3322–3347.
- 黄伯荣, 廖序东. 2017. 现代汉语 (增订六版). 高等教育出版社.
- 曹文, 张劲松. 2009. 面向计算机辅助正音的汉语中介语语音语料库的创制与标注. *语言文字应用*, (4):10.
- 王玮, 张劲松. 2019. 汉语中介语语音库的文本设计. *世界汉语教学*, 33(1):13.
- 朱春跃. 2001. 语音详解. 外语教学与研究出版社.
- 王玮. 2020. 大规模汉语中介语语音库设计和标注研究. 博士学位论文, 北京语言大学.
- 鲁健骥. 1984. 中介语理论与外国人学习汉语的语音偏误分析. *语言教学与研究*, (3):13.