

# 基于 Z 归一法的元音发音部位比较

——以彝语、藏语、汉语普通话元音为例

周学文

**[提要]** 基于文献对汉语普通话、少数民族语言、英语元音的三个共振峰 F1、F2、F3 的相对关系研究显示,元音 Z 归一法是元音归一化的理想方法,即:  $Z1 = \log F2 / \log F1$ 、 $Z2 = \log F3 / \log F2$  保持着稳定的关系,其标准差与平均值之比在 3% 左右,不同发音人 Z 值的波动幅度在 2% 左右。Z 归一法可用于比较不同语言元音声学空间的异同以及距离,作为语音演化和比较分析的量化依据。应用 Z 归一法对三个元音 a、i、u 的研究显示,藏语和彝语较接近,与汉语普通话的区别较大。

**[关键词]** 元音 元音归一 Z 归一法

Joos (1948) 提出,元音的舌位与元音共振峰相关,舌位高低与第一共振峰 F1 有关,舌位前后与第二共振峰 F2 有关。Delattre (1951) 对 Joos (1948) 的理论作了优化,建议用口腔“开度(上下门齿之间的距离)”代替舌位高度这一术语,并提出第三共振峰 F3 与唇的圆展和  $r$  化音色有关。此后,“开度”理论一直被语音学界认可并被大量的实验语音材料证实。但由于发音人的生理特征(声道、声带、口腔)差异以及语境、韵律等的影响,元音共振峰值的变化范围很大。例如元音/a/的共振峰典型值,汉语普通话男性发音人 F1 为 980 赫兹, F2 为 1150 赫兹;女性发音人 F1 为 1250 赫兹, F2 为 1350 赫兹;男、女发音人元音/a/的 F1 和 F2 差值为 200-300 赫兹,波动率(差值与共振峰之比)达到了 20% 以上(鲍怀翘、林茂灿 2014:113)。美国英语元音三个共振峰 F1、F2、F3 三组发音人(男、女、童)的平均值波动率大约在 10%~20% 之间(Peterson & Barney 1952)。即使对同一发音人所发的相同元音,其共振峰值在不同语境(如前后辅音、音节位置、语调等)的影响下,也有很大的波动<sup>①</sup>。

另外,对于不同语言(方言)的元音,即使使用相同的国际音标(IPA)字母标注,共振峰值的波动性也导致无法通过直接比较共振峰值来判别其发音的舌位差异。为此,语音学界提出了元音规整或归一(Vowel Normalization)的思路,目的是对元音特征进行提取,在保留元音的音位系统和社会语言学区别意义的同时,消除发音人的个性生理特征、发音特征、语境影响等差异,从而得到元音固有的、稳定的特征(锚)(Intrinsic Features)。

Gunnar (2004:201-202) 提出:“在语音研究中,我们经常被发音人的语言、方言、年龄、性别和生理制约甚至语境上下文引起的语音特征的变化无常所困扰。”其实,对语音中相对不

<sup>①</sup> 本文彝语、藏语、汉语普通话元音材料中,同一发音人相同元音的共振峰 F1、F2、F3 波动率达到了 15%~30%。

变关系的解码机制本质原理的探索，才是区别特征原理的根本原则。

Potter & Steinberg (1950) 指出，在元音感知中，人耳基底膜上某个空间的刺激模式 (pattern) 可以被识别为一个特定的声音，而不必考虑其与基底膜的相对位置。这就是共振峰比率理论 (Formant Ratio Theory) 的基本思想，即，元音的音色是基于相对共振峰，而不是基于绝对共振峰。虽然 F1 和 F2 主要决定了元音的音色，但基频 F0 和 F3 在元音感知中也起到一定的作用。共振峰比率理论正是基于在变化多端的共振峰中寻求不变特征的一类元音规整方法。该理论提出后，一些语音学家提出了一些基于该理论的归一方法，以下为几个著名的利用相对共振峰理论进行元音规整的公式 (Johnson 2005)。

Peterson (1961)  $\log(F_n) - \log(F_1)$ ,  $n=2,3,4$  公式 (1)

Sussman (1986)  $\log(F_n/F^*)$ ,  $n=1,2,3$ ,  $F^*=(F_1+F_2+F_3)/3$  公式 (2)

Syrdal & Gopal (1986)  $\text{Bark}(F_1) - \text{Bark}(F_0)$ ,  $\text{Bark}(F_2) - \text{Bark}(F_1)$ ,  $\text{Bark}(F_3) - \text{Bark}(F_2)$  公式 (3)

Miller (1989)  $\log(F_1/SR)$ ,  $\log(F_2/F_1)$ ,  $\log(F_3/F_2)$  (SR 为 F0 的均值) 公式 (4)

Gerstman (2011)  $F_i N = (F_i - F_{\text{imin}}) / (F_{\text{imax}} - F_{\text{imin}})$  公式 (5)

国内对元音归一的研究基本上采用了上述公式，如凌锋 (2006) 应用 Miller (1989) 公式，即公式 (4)，对苏州话元音作了研究；石锋等 (2010) 使用了与公式 (5) 类似的归一法。

笔者基于对多种少数民族语言元音材料的研究，经过对多种方法的比较和探索，提出了一种新的共振峰相对关系公式 (以下称“Z 归一法”或“对数商关系”， $Z_1 = \log F_2 / \log F_1$ ， $Z_2 = \log F_3 / \log F_2$ ) (Zhou & Long 2017)，并对多种语言 (彝语、藏语、维吾尔语、英语、汉语普通话等) 的元音共振峰进行了 Z 归一化测试，然后与国际上公开的归一化公式进行了比较；结果显示，Z 归一法模型性能最佳，远远优于其他公式。Nicholas Flynn 等在第 17 届国际语音学大会提出判断元音归一性能的指标  $SCV = [\sigma/\mu]^2$  (鲍怀翘、林茂灿 2014:413)，SCV 越小说明归一法越好，并首次对此标准进行了两处修正：一是评估归一值的收敛性，采用标准差与平均值比值 ( $\sigma/\mu$ ) 的百分比，该值越小越好，其实  $\sigma/\mu$  与  $[\sigma/\mu]^2$  作为收敛性的比较指标并无差别，只是前者更直观；二是增加了对不同发音人归一值比值的比较，不同发音人归一值的比值越接近于 1，说明归一化结果越好。

### 一 多种元音归一法性能比较

为了验证 Z 归一法的优越性能，本文选定了彝语北部方言男、女两位发音人以朗读语速发音的单音节词中元音/a/的共振峰数据，调用三个归一法 (Bark, Labov, Nearey2) (Johnson 2005)，得到归一结果。然后对这一组元音/a/的共振峰分别利用 Miller 法和 Z 归一法得到各自的归一结果。

结果显示，无论是收敛性  $\sigma/\mu$  还是不同发音人归一值的比值，Z 归一法都远远优于其他归一法。对收敛性  $\sigma/\mu$ ，Z 归一法介于 0.25 和 1.05% 之间，而其他方法的收敛性介于 2% 和 9% 之间；对发音人归一值的比值，Z 归一法得到的比值介于 0.98 和 0.99 之间，接近于 1，而其他方法的指标介于 0.7 和 1.2 之间，波动较大。

以下对美国英语 3 组 (男、女、童) 发音人 10 个元音共振峰的平均值数据，应用 Z 归一法和 Miller 法进行归一，并比较结果。见表 1:

表 1 美国英语元音（男、女、童）的 Z 和 M 归一值

元音	组	Z1	Z2	M1	M2	元音	组	Z1	Z2	M1	M2
/ɑ/	男	1.061	1.115	0.174	0.350	/ʊ/	男	1.273	1.033	0.708	0.108
	女	1.054	1.117	0.157	0.362		女	1.289	1.027	0.761	0.093
	童	1.041	1.116	0.124	0.364		童	1.261	1.035	0.712	0.120
	比例	1:0.99:0.98	1:1:1	1:0.90:0.71	1:1.03:1.04		比例	1:1.01:0.99	1:0.99:1	1:1.07:1.01	1:0.86:1.11
	σ/μ	0.96	0.09	16.76	2.11		σ/μ	1.10	0.40	4.06	12.64
/i/	男	1.382	1.035	0.928	0.119	/ε/	男	1.198	1.040	0.541	0.130
	女	1.383	1.022	0.954	0.074		女	1.209	1.032	0.582	0.108
	童	1.365	1.019	0.937	0.067		童	1.204	1.040	0.578	0.136
	比例	1:1:0.99	1:0.99:0.98	1:1.03:1.01	1:0.62:0.56		比例	1:1:1	1:0.99:1	1:1.08:1.07	1:0.83:1.05
	σ/μ	0.73	0.83	1.41	32.56		σ/μ	0.46	0.45	3.99	11.83
/u/	男	1.187	1.140	0.462	0.411	/ʊ/	男	1.138	1.114	0.365	0.342
	女	1.159	1.151	0.410	0.449		女	1.147	1.119	0.392	0.364
	童	1.165	1.145	0.435	0.445		童	1.146	1.118	0.401	0.371
	比例	1:0.98:0.98	1:1.01:1	1:0.89:0.94	1:1.09:1.08		比例	1:1:1	1:1:1	1:1.07:1.10	1:1.06:1.08
	σ/μ	1.26	0.48	5.97	4.80		σ/μ	0.43	0.24	4.85	4.22
/ɔ/	男	1.138	1.114	0.168	0.458	/ʌ/	男	1.096	1.098	0.269	0.303
	女	1.147	1.119	0.193	0.469		女	1.092	1.095	0.265	0.298
	童	1.146	1.118	0.193	0.477		童	1.093	1.102	0.272	0.325
	比例	1:1.01:1.01	1:1:1	1:1.15:1.15	1:1.02:1.04		比例	1:1:1	1:1:1	1:0.99:1.01	1:0.98:1.07
	σ/μ	0.43	0.24	7.82	2.04		σ/μ	0.19	0.32	1.31	4.65
/æ/	男	1.148	1.045	0.416	0.146	/ə/	男	1.164	1.031	0.440	0.098
	女	1.129	1.043	0.377	0.143		女	1.191	1.024	0.516	0.077
	童	1.120	1.046	0.361	0.156		童	1.186	1.023	0.512	0.074
	比例	1:0.98:0.98	1:1:1	1:0.91:0.87	1:0.98:1.07		比例	1:1.02:1.02	1:0.99:0.99	1:1.17:1.16	1:0.79:0.76
	σ/μ	1.26	0.15	7.35	4.59		σ/μ	1.22	0.42	8.74	15.76

表 1 是 Z 归一法和 Miller 归一法对美国英语 10 个元音的归一结果。美国英语元音三个共振峰 F1、F2、F3 的三组（男、女、童）平均值数据引自 Peterson & Barney (1952)。其中： $Z1=\log F2/\log F1$ ， $Z2=\log F3/\log F2$ ； $M1=\log (F2/F1)$ ， $M2=\log (F3/F2)$ 。表 1 显示，对所有 10 个元音的归一结果，不同组发音人的归一值比例（男:女:童），Miller 法的最大偏离（元音 /ə/）达到了 1:1.17:1.16 和 1:0.79:0.76；而 Z 归一法对所有 10 个元音的比例均在 1:0.98:0.98 和 1:1.02:1.02 之间，远远优于前者。

再观察 σ/μ 行，Z 归一法最大为 1.26%，其余都低于 1%；而 Miller 法最大为 32.56%，其余都大于 1%。因此，Z 归一法同样远远优于 Miller 法。

使用普通话元音 a 的共振峰数据（样本量约 150）进行归一，前述公式（5）的前两个归

一值的  $\sigma/\mu$  分别为 28% 和 35%，而 Z 归一法的  $\sigma/\mu$  均小于 2%。

## 二 Z 元音图及发音生理意义

Joos (1948) 提出分别以 F1 和 F2 为两个维度表示元音的声学元音图，其中 F1 和 F2 均为对数维度（鲍怀翘、林茂灿 2014:105），元音图的纵轴 F1 表示元音舌位高低，横轴 F2 表示元音舌位前后。后来出现一个维度或两个维度为线性（赫兹）等多个元音图的变体，由于各种元音图标度不同，元音图上元音之间的距离未必与实际发音的舌位距离对应，但元音之间的舌位相对关系（高低与前后）在各种元音图上是一致的。同理，根据元音共振峰计算出元音的 Z 值（归一）后，我们可以用 Z1 和 Z2 值画出元音的 Z 元音图。利用搜集到的汉语普通话和多种民族语言元音的材料，笔者生成了它们的 Z 元音图。例如图 1 为壮语的传统声学元音散点图和 Z 元音图，图 2 为汉语普通话的传统声学元音散点图和 Z 元音图；图 1 和图 2 分别基于一组相同的元音共振峰值，均提取于朗读语速的单音节词。壮语为北部方言，发音人为女性；汉语普通话发音人为男性，样本量为 200~300 之间。

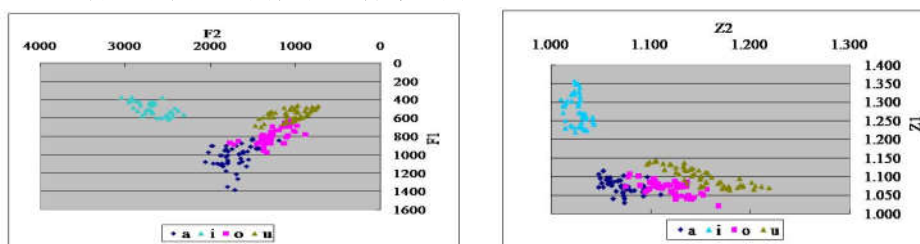


图 1 壮语元音声学元音图（左）与 Z 元音图（右）

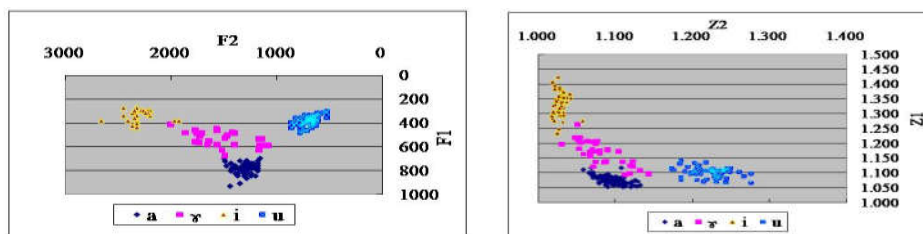


图 2 汉语普通话元音声学元音图（左）与 Z 元音图（右）

观察图 1（左）和图 2（左），F1 代表舌位高低，F2 代表舌位前后，各元音的舌位相对关系符合国际音标的规范；观察图 1（右）和图 2（右），Z1 代表舌位高低（开口度减小），Z2 代表舌位前后，与图 1（左）和图 2（左）中元音的相对位置关系一致。右图与左图最大的区别是后元音 u、o 的舌位高度比前元音的低很多，左图 u、i 的舌位高低几乎相同。实际上，就舌位的实际高度而言，这是传统声学元音图的一个缺陷。根据鲍怀翘、林茂灿（2014:93-94），汉语普通话元音的下颌开度与舌位高低一致，而汉语普通话元音的下颌开度实测数据显示，u 低于 i，o 低于 u；而图 1（右）和图 2（右）关于 u、i、o 的舌位高低关系与下颌开度实测完全一致，且 a 的舌位最低，也与下颌开度实测图一致。

另外，图 2（右）前高元音 i、后元音 u 的分布区域形状不同，i 的舌位高低变化大，前

后变化小，而 u 刚好相反，舌位高低变化小，前后变化大，揭示了由于元音舌位位置所限导致的舌位变化方向的规律。

Z 元音图远远优于一般声学元音图及其元音图变体之处，并非是元音的布局（舌位相对关系）与其一致，而是因为具有上文所述优异的收敛性（ $\sigma/\mu$  约为 1% 左右）和不同发音人的归一值接近（对多种语言测试，元音的归一值比值为  $1\pm 2\%$  左右）。正因如此，Z 元音图上每个元音的分布比较集中（因为  $\sigma/\mu$  较小），且与发音人的关系不大（因为发音人归一值比值为  $1\pm 2\%$  左右）。这样，语言的 Z 元音图就把每个元音“铆”在相对固定的一个椭圆形区域，图上 Z1 轴表示舌位高低、Z2 轴表示舌位前后，且元音之间的舌位关系与国际音标图上元音的相对舌位关系一致。如此，各个语言的元音就可以在 Z 元音图上就发音舌位进行比较（比较 Z1 和 Z2 值）。其他声学元音图之所以不能直接比较元音是因其横轴和纵轴没有进行归一或归一效果差（ $\sigma/\mu$  和发音人归一值比），因此元音在图上游移范围太大，没有一个稳定的区域。

### 三 彝语、藏语、汉语普通话元音发音部位比较

我们从彝语、藏语和汉语普通话的三个“标杆”元音 a、i、u 的 Z 图分布出发，探讨其舌位分布及差异。三种语言的语料均来自于男性发音人，朗读语速，彝语、藏语元音取自单音节词。汉语普通话大部分取自双音节词，少量取自多音节词；彝语为北部方言（乌尼乌且、王秀英 2003:10-45），藏语为卫藏方言。

表 2 彝语、藏语和汉语普通话元音的 Z 值

元音	Z 值	彝语			藏语			汉语普通话		
		$\mu$	$\sigma$	样本	$\mu$	$\sigma$	样本	$\mu$	$\sigma$	样本
/a/	Z1	1.078	0.007	93	1.066	0.015	65	1.076	0.013	139
	Z2	1.105	0.01		1.098	0.018		1.096	0.014	
/i/	Z1	1.311	0.022	89	1.321	0.022	32	1.334	0.042	46
	Z2	1.048	0.008		1.033	0.006		1.028	0.007	
/u/	Z1	1.155	0.056	87	1.162	0.029	24	1.101	0.017	54
	Z2	1.151	0.044		1.123	0.029		1.222	0.024	

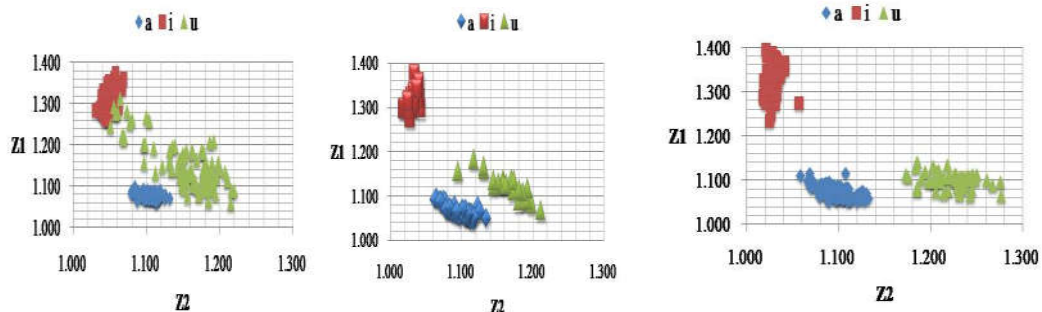


图 3 彝语（左）、藏语（中）、汉语普通话（右）a、i、u 元音的 Z 元音图

表2和图3显示,对于元音*i*,三种语言Z1、Z2的最大值、最小值都相差在2%以内,因此可视为在误差波动范围内,三种语言舌位高度和前后都接近,舌位变化都是高低变化大,前后变化小,但汉语普通话的舌位高低变化最大;对于元音*a*,三种语言Z1、Z2的最大值、最小值都相差在2%以内,三种语言舌位高低和前后都接近,舌位变化都是高低变化小,前后变化大;对于元音*u*,三种语言Z1、Z2的最大值、最小值相差分别为5.5%、8.8%,远远超过本文前述同语言、同元音的Z值,不论同一发音人或不同发音人,最大最小之比一般不可超过3%的规律,因此,可视为舌位有显著差别,对于舌位高低和前后,藏语、彝语接近,汉语普通话明显最低最后,藏语、彝语在两个维度的变化都大,汉语普通话仅在舌位前后变化大,舌位前后相当稳定。从三个元音的舌位总体比较来看,对于元音*i*、*a*,三种语言接近;而对元音*u*,藏语、彝语接近,与汉语普通话区别较大,汉语普通话的*u*的舌位明显靠下靠后。基于两个舌位维度的变化来看,也是藏语、彝语比较接近(除了彝语*u*的“唇颤”发音较特殊造成Z图上的分布范围较广外),与汉语普通话差别大。

总之,就三个元音的舌位和变化而言,藏语、彝语更接近,而与汉语普通话区别较大。

#### 四 结论和讨论

通过对多种语言元音共振峰数据的分析证实,元音存在稳定的共振峰对数商关系,可以应用对元音的“归一化”,即 $Z1 = \log F2 / \log F1$ 、 $Z2 = \log F3 / \log F2$ 。Z值介于1.0和1.5之间,一种语言内Z值标准差与平均值的比值一般在3%以内,即Z值收敛性很好;不同发音人归一值比值为 $1 \pm 2\%$ 左右,即,极大消除了发音人的差别。

元音对数商模型所基于的发音生理/感知的原理是什么呢?因为F1与舌位高低有关,F2与舌位前后及唇的圆展有关;那么F2/F1可解释为综合了舌位高低与前后信息,即口腔形状。取对数的意义是对线性值进行对数压缩,以符合听觉特性,类似人耳对音高值的按对数压缩以分出不同的级别(度或半音)。至于F3/F2,可视为用F3对F2的加权。总之,Z归一法模型综合了三个共振峰的信息,优于传统的用两个共振峰表示元音的方法。

Z归一法模型可以用于比较不同语言元音声学空间的异同以及距离,作为语音演化和比较分析的量化依据。从三个元音*a*、*i*、*u*的总体Z元音图的空间分布来看,藏语和彝语接近,与汉语普通话差别较大,特别是汉语普通话的*u*靠下靠后,与前两者明显不同;基于在两个舌位维度的变化来看,也是藏语、彝语接近,与汉语普通话不同。总之,基于三个元音发音舌位分析,藏语和彝语较接近,与汉语普通话的区别较大。

Z归一法虽然在收敛性和消除不同发音人差别方面表现优异,但就象由F1和F2构成的声学元音图一样,Z图也存在不同元音的重叠现象;经测试,重叠程度与声学元音图接近。要解决此问题,下一步可考虑增加Z3( $\log F4 / \log F3$ )或基频F0作为参数,与Z1和Z2一起形成三维元音图。增加参数将有助于增加不同元音间的区别度,解决元音重叠问题。

#### 参考文献

- [1] 鲍怀翘、林茂灿主编. 2014.《实验语音学概要》(增订版),北京:北京大学出版社。
- [2] 凌 锋. 2006.《元音相似度的量化计算及其应用》,第七届中国语音学学术会议暨语音学前沿问题国际论坛,北京,10月20-22日。

- [3] 石 锋、冉启斌、王 萍. 2010. 《论语音格局》, 《南开语言学刊》第 1 期.
- [4] 乌尼乌且、王秀英. 2003. 《现代彝语》, 成都: 四川民族出版社.
- [5] Delattre, Pierre. 1951. The physiological interpretation of sound spectrogram, *Publication of the Modern Language Association of America*, 66(5): 112-115. Berlin: Walter de Gruyter GmbH.
- [6] Gunnar, M. Fant. 2004. *Speech Acoustics and Phonetics*. Dordrecht: Kluwer Academic Publishers.
- [7] Johnson, Keith. 2005. Speaker normalization in speech perception. In D. B. Pisoni & R. Remez (eds.) *The Handbook of Speech Perception*, pp. 363-389. Oxford: Blackwell Publishers.
- [8] Joos, M. A. 1948. Acoustic phonetics. *Language*, 24(Suppl. 2): 1-136.
- [9] Miller, J. D. 1989. Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 85(5): 2114-2134.
- [10] Peterson, Gordon & Harold Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2): 175-184.
- [11] Potter, R. K. & J. C. Steinberg. 1950. Towards the specification of speech. *Journal of the Acoustical Society of America*, 22(6): 807-820.
- [12] Zhou, Xuewen & Long Congjun. 2017. An Efficient Vowel Normalization Algorithm-Logarithmic Quotient Model. Paper presented at O-COCOSDA 11.1-3, Seoul, South Korea.

## **A Comparison of Places of Articulation of Vowels Based on Z-Normalization: A Research on Yi, Tibetan and Mandarin Chinese Vowels**

**ZHOU Xuewen**

**[Abstract]** The paper examines previous studies on the relations of F1, F2 and F3 of vowels in Mandarin Chinese, some minority languages and English, and proposes an efficient vowel normalization model. Results prove that vowel formants keep stable relations of Logarithmic Quotient (Z-value,  $Z1=\log F2/\log F1$ ,  $Z2=\log F3/\log F2$ ). In terms of Z-values of vowels in one language, the ratio of standard deviation to average deviation keeps below 3% for most vowels, and the varying scope keeps below 2% for different speakers, both taking a lead in the vowel normalization field. Z-normalization can be used to compare the differences and distances in the acoustic space of vowels in different languages, for the purpose of serving as quantification basis for sound evolution and comparative analysis. Employing Z-normalization, this paper demonstrates that, in terms of vowel acoustic patterns and variations of the three vowels /a/, /i/ and /u/, Tibetan and Yi are more closely related while they are both quite distant from Mandarin Chinese.

**[Keywords]** vowel vowel normalization Z-normalization

(通信地址: 100081 北京 中国社会科学院民族学与人类学研究所)

【本文责编 普忠良】