语言科学
ISSN 1671－9484　CN 32－1687/G

2004 年 3 月
第 3 卷 第 2 期(总第 9 期)3－19

基频归一化

——如何处理声调的随机差异?

朱晓农

香港科技大学人文学部　香港

提要　一个语言信号的物理性质表现形式是无限多的。归一化的主要目的是消减人际随机差异,提取恒定参数,在语际变异中找到共性,从而使得人际比较和语际比较的研究成为可能。本文比较六种归一化方法,以及各种选点、定频域的方法,引入了判断方法好坏的标准指数和离散系数。结果表明对数 z-score (LZ) 方法最好,与此相应的频域定义以均值和标准差为好,用以归一化的采样点以选用语音目标处的时点为佳。

关键词　声调　基频　归一化　人际差异

中图分类号　H017　　**文献标识码**　A　　**文章编号**　1671－9484(2004)02－0003－17

1　引言

本文旨在探讨对声调的主要载体——基频——的归一化处理。负荷声调信息的除了基频以外,有时还涉及时长 (duration) 和发声 (phonation)。(Zhu 1999) 不过,基频是任何声调系统都不可或缺的最普遍、最重要的区别因素;时长和发声只是在某些较为特殊的声调系统中起作用。

尽管基频负载着声调的主要区别特征,目前在描写声调时却只能通过对基频的感知——音高——来进行印象式 (impressionistic) 描绘,因而语言调查人员在记录声调时都有一种难以一下确定的感觉。例如某个听感上的高升调,不同的调查人员可能会分别描写成 [24]、[25]或[35]。他们可能都对,因为声调描写中的这种不确定性实际上反映了两个基本事实:其一,不同的发音人可能有发音差别;其二,描写工具 (五度制) 本身并未对这些差别有精确定义。(朱晓农 2002)

上述第一点是显而易见的,一个语言信号 (linguistic signal) 的物理性质表现形式多种多样,简直可以说是无限多;不但人各不同,甚至同一个人说两次都不一样。最近二三十年来语音学的最大成就也许就是从实验上证明了这一经验常识。

但是,另一方面,这个语言信号,不管男女老少谁说,也不管是尖叫乱喊还是一字一顿,在听者耳中的语言内容是一样的。这说明在感知层面有不变的范畴存在,这种感知范畴有可能用作音韵学层面的对立特征。归一化的主要目的就是消除人际随机差异,提取恒定参数,即滤掉个人特性,获得具有语言学意义的信息。(Rose 1993)正如 Disner (1980) 对元音归一化和费国华 (Rose 1987) 对声调归一化

[收稿日期]2003 年 5 月 27 日　　[定稿日期]2003 年 9 月 10 日

所作的的论证，归一化的数学程序可以看作是模拟听者把一个听到的音归于某个音类的感知过程。归一化的目标是归聚相类的音，同时分开不相类的音。个人各自的归一化（self-normalization）的物理含义就是以本人的频域作为坐标，以显示本人的各个声调在此空间中的分布。经过归一化以后，原先各人看似杂乱无章的基频曲线便会显示出惊人的一致性。因此，基频归一化的首要作用在于，把对声调的感觉描绘建立在标准化的定量描写的基础上。

归一化的另一作用是消减录音时的发音风格（正式、随意、紧张等）差异。不同发音人的不同风格一直是个很大的麻烦，它可能破坏材料之间的可比性。使用归一化后这个问题可以得到极大改善，只要发音人在一次录音中保持一贯的风格就行，因为一种特定风格可以看作某个有个人特点的发音人。

归一化不但能提取同一个语言中的某个音无数多表面变异底下的不变性，而且还能比较不同语言中的音，如果相似，看它们相似到多大程度，又有什么差异；如果不同，看它们在可比的语音空间中各自分布在什么区域。

总之，如果不对测量到的语音进行归一化处理，人际随机差异就无法减少，人际比较和跨语言比较就无法进行。归一化能够在人际差异中找到常量，在语际变异中找到共性，从而使得人际比较和语际比较的研究成为可能。

自 20 世纪 60 年代以来提出过很多种归一化的转换方法。归一化先是用于元音的分类和比较（Fant 1966，Gerstman 1968，Harshman 1970，Labanov 1970，Wakita 1977），然后引入到声调研究中（Jassem 1971，Menn & Boyce 1982，Earle 1975，Takefuta 1975，Rose 1982，Rose 1987，Rose 1989，Ladd et al 1985，沈炯 1985，石锋 1990）。归一化一般有两个步骤，一是在坐标上作平移，一是压缩或扩大频域。基本公式如下：

$$[A] \qquad x' = \frac{x_i - x_1}{x_2}$$

其中 x' 表示归一化了的标准值，x_i 代表要被归一化的原始测量值，x_1 和 x_2 代表两个参照值，x_1 用来平移，x_2 用来压缩或扩大频域。

下面讨论六种归一化方法：z-score，频域分数，频域比例，对数半音差比，对数 z-score，对数频域比例。前三种是线性的，后三种是对数的。头两种 z-score 和频域分数很常用，还有一类国内一些学者亦有使用，其余是我发展出来的。比较的结果表明由我设计的对数 z-score（LZ）法最好。讨论所用的例子是上海话 11 位发音人的五个声调的基频，数据见文末附录一表 11。Dart（1987）对朝鲜语中硬（fortis）：软（lenis）对立的研究表明，不同发音人在实现同一音韵类或某个音韵对立时发音的策略可能很不同，因此必须使用“足够数量的受试人”。从统计学角度来看，如果有三十个人的样本当然很理想，但处理这样一个样本对本研究来说太大，难以完成。我们决定取的样本大小为 11 个人，这样既能做统计，时间又许可。这五个声调的调值可描写为阴平 T1 高降 [52]，阴去 T2 中微升 [34]，阳去 T3 低升 [14]，阴入 T4 高短调 [44]，阳入 T5 低升短调 [13]。（朱晓农 1996，Zhu 1999）附录二是有关实验的背景情况。

2 六种归一化方法介绍

第一种 z-score 转换法（如 Jassem 1971，Menn & Boyce 1982，Rose 1987，Rose 1989）“把一个基频观察值表达为距离基频均值的某个量度的倍数”（Rose 1987：347）。归一化的基频值（z 值）用标准公式 [B] 计算：

$$[B] \qquad z_i=\frac{x_i-m}{s}$$

其中 x_i 是观察值，m 是给定样本的均值，s 是标准差（下文简作 SD）。例如后文表 7 中，发音人 M1 的阴平 T1 在 5%时点处的基频值是 179Hz，总平均值和 SD 分别为 141Hz 和 16Hz（表 9）。可以算出 179Hz 的归一化值为 (179−141)/16=2.38.

第二种是"频域分数（FOR）转换法（如 Earle 1975，Takefuta 1975，Rose 1982，Ladd et al. 1985）把某个观察值表为两个域限定义值之差的分数"(Rose 1987：347)。计算公式[C]如下：

$$[C] \qquad v_i=\frac{x_i-x_L}{x_H-x_L}$$

其中 v 是归一化值，x_H 和 x_L 分别是确定频域上限和下限的值。要计算 M1 的 T2 在 20% 时点的基频值（139Hz）的 v 值，先找出在 T1 的 5%时点处的基频最大值 179Hz（归一化时不考虑在 0% 时点处的基频）和出现在 T3 谷点处的最低值 120Hz，然后算出 FOR 值是 0.32（=(139−120)/(179−120)）。

第三种是频域比例（POR）转换法，把某个观察值表为用均值和标准差定义的频域的比值，计算公式 [D]如下：

$$[D] \qquad g_i=\frac{x_i-(m-cs)}{(m+cs)-(m-cs)}$$

其中 m 是均值，s 是标准差，c 是常数，可根据需要取 $0<c\leqslant 3$ 之间的值，一般以取 2 和 2.5 为常见。仍以 M1 为例，他的 T1 在 5%时点处的 g 值是 1.09（=(179−(141−2×16))/((141+2×16)−(141−2×16))）。

公式 [D] 看似像 [C]，但两者频域定义的策略不同。FOR 法中频域是由两个基频值确定的，而 POR 法中的频域是由均值和标准差定义的，也就是由全体参与归一化的基频值决定的。[D] 中的 g 值可以是一个频域的分数，比如在 $(m\pm 2s)$ 之间，也可以是个倍数，比如如果频域小于 $(m\pm 0.5s)$。

第四种是对数半音差比（ratio of logrithmic semitone distances，LD）转换法，把观察值表为某个以半音定义的特定频域的分数。半音差距定义为：$D=12\cdot\log_2\frac{x_2}{x_1}=\frac{12}{\log_{10}2}\cdot\log_{10}\frac{x_2}{x_1}$(Johan't Hart et al. 1990:24)，或 $x_2=(\sqrt[12]{2})^n\cdot x_1$(Baken 1987：127)。归一化公式 [E]如下：

$$[E] \qquad d_i=\left(\frac{12}{\log_{10}2}\cdot\log_{10}\frac{x_i}{x_L}\right)\div\left(\frac{12}{\log_{10}2}\cdot\log_{10}\frac{x_H}{x_L}\right)$$

$$=\left(\log_{10}\frac{x_i}{x_L}\right)\div\left(\log_{10}\frac{x_H}{x_L}\right)=\frac{\log_{10}x_i-\log_{10}x_L}{\log_{10}x_H-\log_{10}x_L}$$

归一化值被表示为两个半音差距的比值。其中 x_H 和 x_L 分别为频域的上限和下限。以上面 M1 为例，T_1 在 5%时点处的基频的 LD 归一化值是 1.07（=[$\log_{10}(179)-\log_{10}(141-2\times16)$]÷[$\log_{10}(141+2\times16)-\log_{10}(141-2\times16)$]），如果 x_H 和 x_L 分别定义为 $(m+2s)$ 和 $(m-2s)$。

根据不同的频域定义法，LD 可分为两种：LD_{FOR} 用两个极端值，LD_{POR} 用全体值。石锋（1990）的求 T 值的公式与 LD_{FOR} 相仿。沈炯（1985）使用的 D 值与五度值转换的公式与 LD_{POR} 相仿。

第五种是对数 z-score（LZ）转换法，把观察值表为均值的某个离散量度的倍数，用的是对数形式，计算公式 [F]如：

$$[F]\qquad z'_i=\frac{y_i-m_y}{s_y}=\frac{\log_{10}x_i-\frac{1}{n}\sum_{i=1}^{n}\log_{10}x}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\log_{10}x_i-\frac{1}{n}\sum_{i=1}^{n}\log_{10}x_i)^2}}$$

其中 $y=\log_{10}x_i$，m_y 和 s_y 分别是 $y_i(i=1,2,\cdots,n)$ 的算术平均值和标准差，m_y 因而是原始基频值的对数几何均值。再以 M1 为例，对数基频值总平均是 2.147，标准差是 0.048（见表 9）。T1 在 5% 处的基频 179Hz，相应的 LZ 值就是 $(\log_{10}(179)-2.147)\div 0.048=2.21$（见表 10）。

第六种是对数频域比例法（LP()R），把观察值表为以对数均值和标准差定义的频域的比例，计算公式 [G] 如下：

$$[G]\qquad g'=\frac{y_i-(m_y-cs_y)}{(m_y+cs_y)-(m_y-cs_y)}$$

其中 c 是常数，$y_i=\log_{10}x_i$，m_y 和 s_y 分别为 $y_i(i=1,2,\cdots,n)$ 的均值和标准差，m_y 是原始基频值的对数几何均值。例如，M1 的 T1 在 5% 处的基频值为 179Hz，相应的归一化了的 g' 值为：$[(\log_{10}(179)-(2.147-2\times0.048)]\div[(2.147+2\times0.048)-(2.147-2\times0.048)]=2.21$。

上述有些方法并不是完全独立的，例如 z-score 和 P()R 有如 [H] 的关系：

$$[H]\qquad g=\frac{x-(m-cs)}{(m+cs)-(m-cs)}=\frac{x-m+cs}{2cs}=\frac{1}{2c}\cdot\frac{x-m}{s}+\frac{1}{2}=\frac{1}{2c}z+\frac{1}{2}$$

式中 z 值和 g 值的区别在于斜率（$1:0.5c$）和截距（$0:0.5$）。常数 0.5 对公式无碍，只是在坐标上垂直移动了半个单位。也就是说，平均 z 值在纵轴上是 0，而平均 g 值是 0.5。斜率（0.5c）表示在纵轴上压缩或扩大曲线，不过它不改变任何一对归一化前后的 g 值和相应 z 值的相对比例。因此 z-score 法和 P()R 法可以用一个系数（$0.5c$）和一个常数（0.5）加以换算，它们实质上是等价的。这一点如果在 P()R 法中用（$m\pm0.5s$）做频域就更明显了，此时，$g_{0.5s}=z+0.5$，连斜率都相等了。所以 P()R 法只是 z-score 的一个变体。由上可见，所有用且仅用均值 m 和标准差 s 作为归一化参数的线性归一化方法实质上都是等价的。对数的 LZ 和 LP()R 转换法有同样的关系，见 [I] 中的推导：

$$[I]\qquad g'=\frac{y_i-(m_y-cs_y)}{(m_y+cs_y)-(m_y-cs_y)}=\frac{1}{2c}\cdot\frac{y_i-m_y}{s_y}+\frac{1}{2}=\frac{1}{2c}z'+\frac{1}{2}$$

所以 LP()R 可以看作 LZ 法的变体。因此，对 z-score 和 LZ 法的讨论也分别适用于 P()R 和 LP()R 法。LD_{POR} 法和 LZ 法之间的关系更复杂。但可以找到一个表达两者换算关系的等式 [J] 如下所示：

$$[J]\qquad d_i=\frac{z'_i-\frac{\log_{10}(m_x-2s_x)-m_y}{s_y}}{\frac{\log_{10}(m_x+2s_x)-\log_{10}(m_x-2s_x)}{s_y}}$$

综上所述，实际上有四种归一化方法：z-score（= P()R），频域分数 F()R，对数 z-score（= 对数 P()R）LZ，以及对数半音差比 LD。在作比较之前，先介绍几个用于评价归一化方法的基本概念。

3 标准指数和离散系数

为了比较由各种归一法得到的结果，我们使用标准指数（normalisation index，NI）作为指针。NI 越大，结果越好，其他条件一样的话。归一化的目标是归聚转写时相类的音，同时分开不相类的音。否则 NI 值再大也没意义（Disner 1980）。标准指数定义为归一化前后的离散系数之比（Rose 1987），如

下［K］式所示：

$$[K] \qquad NI=\frac{\text{归一化前的基频值的离散系数}}{\text{归一化后的标准值的离散系数}}$$

离散系数（dispersion coefficient，DC）反映了对不同发音人的测量值的归聚程度。Earle（1975：132）和 Rose（1981：143）所用的定义实际上是等价的，离散系数可以定义为如［L］：

$$[L] \qquad DC=\frac{\text{标准点上的人际均方差}}{\text{样本方差}}$$

其中标准点（normalisation points）指用于归一化处理的采样点，采样点（sampling points）指的是在该处进行基频、音强测量的百分数时刻点。样本方差（sample variance，SV）是采样点上的个人平均基频值的总方差。人际均方差（the mean between-speaker variance，MBSV，$\overline{s^2}$）用公式［M］算：

$$[M] \qquad \overline{s^2_{BS}}=\frac{1}{n}\sum_{i=1}^{n}s_i^{\,2}$$

其中 n = 标准点的个数，i = 第 i 个标准点。一个样本的离散度由其方差（s^2）表示，用公式［N］算：

$$[N] \qquad s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-m)^2$$

其中 n＝样本中的个例数，所有人的基频值或标准值，x_i＝样本中第 i 个实例，m＝样本的算术平均。

4　定义频域的策略

使用 FOR，POR，LD 等转换法会引出确定频域的问题。而不同的频域定义策略会引出不同的归一化结果。Rose（1981）选择高降调在 30％和 90％处的基频均值作为频域定义的上下两个值。Earle（1975：100）选择所有声调中的最高和最低取样点。为了找到哪种策略能得到最好结果，我试用表 1 中多种频域定义值对文末表 11 中 11 个人所有声调的基频均值进行归一化。

FOR 和 LD_{FOR}	1）	$f_{max}-f_{min}$
	2）	$(f_{max}+s_{f_{max}})-(f_{min}-s_{f_{min}})$
	3）	$(f_{max}+2s_{f_{max}})-(f_{min}-2s_{f_{min}})$
	4）	$f_{20\%\ of\ T1}-f_{80\%\ of\ T}$
	5）	$f_{10\%\ of\ T1}-f_{60\%\ of\ T}$
POR 和 LD_{POR}	6）	$m\pm 2s$
	7）	$m\pm s$

表　1

此处 f_{max} 是 T1 高降调的最高基频值，这一般也是所有声调中的最高点（不算 0％处的值），$s_{f_{max}}$ 是距离 f_{max} 一个标准差，f_{min} 是 T3 低凹调的最低点的基频均值，这一般也是各人的最低基频值（不算 100＋％ 处的值），$s_{f_{min}}$ 是距离这均值一个标准差，$f_{20\%\ of\ T1}$ 是 T1 的 20％ 处的基频均值，等等。

表 2（见下页）中的标准指数（NI）反映了归一化的结果，标准点是从 5％到 100％处，但不包括受到喉塞尾干扰的 T4 的 100％ 时点。表中的指数与表 1 中的频域定义法是对应的。例如，使用 $m\pm s$ 这个频域（表 1 第 7 行）得到的标准指数 13.8（线性 POR 法）和 14.0（半音 LDPOR 法）在表 2 中也是第 7 行。表 2 表明 POR 转换法（6，7）远远好于 FOR 法（1－5），前者的 NI 值高好多，这是预计中的。费国华（Rose 1987：348）发现 z-score 转换法产生的归一化结果好于 FOR 法，“z-score 法避免了强制归拢的循环性，因为决定归一化参数（NP）的是许多基频值的分布，而不是仅仅两个值。此外，这公式的‘均方

根'root-mean-square 基础能保证减少的是全局分布上的人际差异（相比之下，若用 FOR 归一法，而且频域定义值正好出现在同一个人际采样点，会人为地引起人际差异的局部波动）。"

		线性	对数	
1)	FOR	10.4	10.5	LD_{FOR}
2)		9.4	9.8	
3)		7.5	8.2	
4)		6.7	6.3	
5)		4.1	4.0	
6)	POR	13.8	13.4	LD_{POR}
7)		13.8	14.0	

表 2 使用不同的频域确定策略的不同归一化方法所得到的标准指数

与 z-score 一样，POR 转换法避免了这两个不利，即缺乏均方根基础和所用基频值太少。由于线性和对数的 FOR 频域定义的方法（表 2 中 1—5）所产生的 NI 值都不及使用均值和标准差的方法，以下不再考虑的 FOR 和 LD_{FOR} 转换法。

5 选择标准点的策略

有关测量声调基频如何选择采样点及其理由的详细讨论见 Zhu(1999:第 3 章)。简单地说：在入声短调的时长上每三分之一处取一个点；舒声长调每五分之一取一个点，再加 5% 时点；对 T2、T3 和 T5 再加一个 100+% 时点。这 100+% 时点上的基频测量值不用于归一化，因为它们反映的是音节尾喉塞音的声带关闭阶段，(Zee & Maddieson 1980; Rose 1987)"反映了由辅音引起的基频波动在不同的人身上所产生的很大的不同效应，所以最好略去不顾"。(Rose 1993:197)排除阴入 T4 在 100% 时点的基频值的理由与前升调类相同，T4 也有个喉塞尾。排除所有声调在 0% 时点的测量值是因为起点基频受到声母辅音的影响是人各不同的 (Rose 1993:197)。排除阴平 T1 在 100% 时点的测量值的理由下面再说。八种选点法在表 3 中，三种归一法与八种选点法的结果列在表 4 中。POR 和 LPOR 归一化结果也在表中括号内，由于分别与 z-score 和 LZ 等价，它们的 NI 值与 z-score 和 LZ 的相同。

1) [0—100%]	所有五个声调从 0% 到 100%点的所有的值
2) [0—100%\T1]	所有五个声调从 0% 到 100%点的所有的值，除了 T1 在 100% 处的值
3) [0—100%\T4]	所有五个声调从 0% 到 100%点的所有的值，除了 T4 在 100% 处的值
4) [0—100%\T1+4]	所有五个声调从 0% 到 100%点的所有的值，除了 T1 和 T4 在 100% 处的值
5) [5—100%]	所有五个声调从 5% 到 100%点的所有的值
6) [5—100%\T1]	所有五个声调从 5% 到 100%点的所有的值，除了 T1 在 100% 处的值
7) [5—100%\T4]	所有五个声调从 5% 到 100%点的所有的值，除了 T4 在 100% 处的值
8) [5—100%\T1+4]	所有五个声调从 5% 到 100%点的所有的值，除了 T1 和 T4 在 100% 处的值

表 3

分别比较表 4(见下页)中的 3)与 1)，7)与 5)，结果表明，排除了 T4 在 100% 处的值改善了归一化结果，用 z-score 法 NI 值增加了 1.3 (=12.9—11.6,{3),1)})和 2.1 (= 13.8—11.7,{7),5)})，用 LD 法增加了 1.6 (=12.8—11.2,{3),1)})和 2.1 (=13.4—11.3,{7),5)})，用 LZ 法增加了 1.6 (= 13.3—11.7,{3),1)})和 2.3 (=14.0—11.7,{7),5)})。排除所有声调在 0%处的值（分别比较 5)—8) 与 1)—4)得出更好的 NI 值。选点法 7)的结果 (13.8/13.4/14.0) 高于第 3)行的结果 (12.9/12.8/13.3) 0.9—0.6。第 8)行的结果 (12.5/12.3/12.7) 高于第 4)行的结果 (12.0/11.5/12.2) 0.5—0.8。另两

对选点策略 5) 和 1),6)和 2) 结果差不多。

上述两个改善是预料之中的,因为 T4 在 100% 处和所有声调的起点处的基频是依人、依音段不同而不同,因而有更大的随机差异。因此可以摒弃表 4 中第 1) 到第 6)种选点标准。在最后两行中,尽管第 7)行的选点策略产生最好的归一化结果 ,但有理由让我们选择第 8)行的选点策略。阴平 T1 的基

		Z-score(=POR)		LD	LZ(=LPOR)	
1)	[0−100%]	11.6	(11.6)	11.2	11.7	(11.7)
2)	[0−100%\T1]	10.6	(10.6)	10.0	10.5	(10.5)
3)	[0−100%\T4]	12.9	(12.9)	12.8	13.3	(13.3)
4)	[0−100%\T1+4]	12.0	(12.0)	11.5	12.2	(12.2)
5)	[5−100%]	11.7	(11.7)	11.3	11.7	(11.7)
6)	[5−100%\T1]	10.5	(10.5)	10.0	10.0	(10.0)
7)	[5−100%\T4]	13.8	(13.8)	13.4	**14.0**	(14.0)
8)	[5−100%\T1+4]	12.5	(12.5)	12.3	**12.7**	(12.7)

表 4　各种转换法的标准指数,归一化点不同

频降尾降得如此之低,大大超过阳去 T3 的凹点,在 11 位发音人中除了 M6 都是如此。可以证明 T3 的凹点是喉门主动控制的最低点,T1 的终点不是主动控制的结果。也就是说,T3 的凹点高度是基频的目标,说者想要把他们的基频降到这个特定高度再往上升。不同的是,T1 的基频终点仅仅是降调的自然结果,而不是有意要把喉门控制降到某个高度,它可以在说者的低频区中相当宽的一个范围内任何一点上停下来,说话者并未故意调整声带紧张度和喉下气压以操纵 T1 的基频终点。费国华 (Rose 1993) 说,T3 的凹点基频是个人际常数,而 T1 的终点值是依人而定的变量。这一观点在下文将得到证明:T1 的最低点终点处的标准差大约是 T3 的最低点凹点处的四倍之大,可见 T1 的终点基频是个随机变量,而不是说者刻意追求的语音目标,所以在提取声调的不变性质的程序中应把它排除。

现在还剩下表 4 第 8)行的三种方法需要讨论。LZ 法的结果 (NI=12.7) 最好,其次 z-score 法 (12.5),最差的是 LD 法 (12.3)。不过在淘汰后两种方法前还要讨论一下 LD 法和 z-score 法。

6　再论半音 LD 转换法

半音 LD 转换法有些很特别的性质是其他方法没有的。首先是频域定义和由此得到的标准指数的关系。在线性 POR 和对数 LPOR 公式中改变系数 c,不会改变最终的标准指数,见 [H] 和 [I] 中的公式,又见表 2 中的两个 NI 值 13.8。

$$[\mathrm{H}]\qquad g=\frac{x-(m-cs)}{(m+cs)-(m-cs)}=\frac{x-m+cs}{2cs}=\frac{1}{2c}\cdot\frac{x-m}{s}+\frac{1}{2}=\frac{1}{2c}z+\frac{1}{2}$$

$$[\mathrm{I}]\qquad g'=\frac{y_i-(m_y-cs_y)}{(m_y+cs_y)-(m_y-cs_y)}=\frac{1}{2c}\cdot\frac{y_i-m_y}{s_y}+\frac{1}{2}=\frac{1}{2c}z'+\frac{1}{2}$$

但 LD 法中改变系数就会改变结果。图 1(见下页)中画出标准指数作为频域的函数,频域是由标准差定义的。可以看到,在 LD (5−100%\T4) 公式中频域从 $m\pm4.97931887s$ 逐渐减小为 $m\pm6.35\mathrm{E}{-15}s$,产生了一个标准指数 NI 的拱形曲线,顶点 (NI=13.98) 在 $m\pm1.03382s$ 处。那就是说,如果 [E] 中频域上下限确定值 x_H 和 x_L 各自用 $m+1.03382s$ 和 $m-1.03382s$ 来定义的话,就能得到一个最高的标准指数 NI=13.98。由此可见,用 LD 转换法,随着所选频域的不同,归一化结果也不同。第二个特点是各发音人相同的 LD 值的可比性问题。用其他转换法如 z-values 和 LZ 法,发音人的归一化值的均值是 0,用 POR 和 LPOR 法是 0.5,但是 LD 归一化值的均值人各不同。例如,经 LD 归一化 ($c=2$,标准

点：5－100%\T4）后发音人的平均 d 值见表 5：

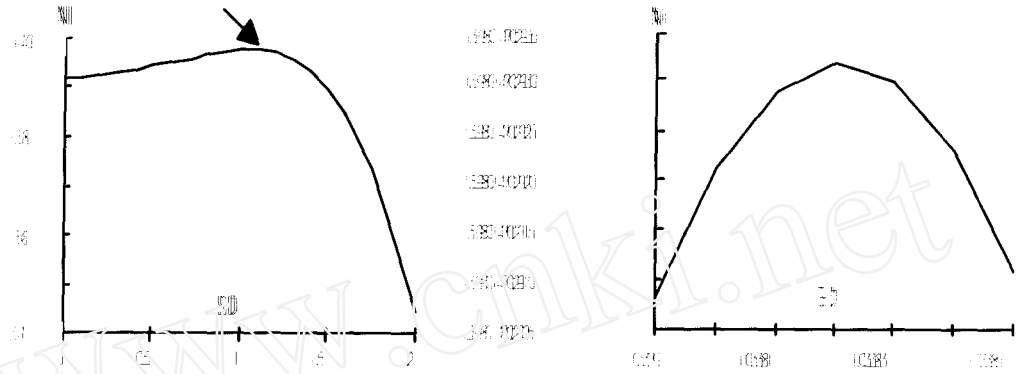

图 1 LD 转换法中，标准指数作为频域的函数

Sp	M1	M2	M3	M4	M5	M6	W1	W2	W3	W4	W5
$\bar{d}$	0.55	0.55	0.54	0.53	0.54	0.56	0.56	0.54	0.55	0.54	0.54

表 5

当所有个人的 d 值画在一张图上，只有他们两个频域定义点上的 LD 值能排在同一个高度上。那就是说，在这种情况下，所有人在 $m+2s$ 和 $m-2s$ 点上的原始基频值经归一化后各自定为 1 和 0。但如果用 $m\pm1s$ 来定义频域，那么在 $m+1s$ 和 $m-1s$ 处的基频值会标准化为 1 和 0，如此等等。这对于人际比较来说显然是不利的。当然如果作［R］中的坐标转换，也可以把各人的 LD 均值（$\bar{d}$）排在同一高度，但代价是频域两端就无法对齐了（$\bar{D}$＝总均值，i＝第 i 个发音人，j＝第 j 个标准点）。

［R］ $d'_{ij}=d_{ij}-(\bar{d}_i-\bar{D})$

经过坐标转换以后，不同频域定义法所得的所有标准指数都改善了，如果原来的指数较小，那么改进的就很大。转换以后指数在 $s=2$ 时最高，如选点是［0－100%\T4］，NI＝14.0；如选点是［5－100%\T1＋4］，NI＝12.7。也就是说，现在的结果与 LZ 的结果（见表 4）一样好。但现在各个发音人取自同处的基频值，除了均值，变得无法比较。例如，M1 在高于均值 1SD 的经坐标转换后的基频值不同于 W1 在该处的值。尽管误差不大，但依然是个理论上的缺陷。因此我决定不用 LD 法。现在只剩下下一节要讨论的最后两种方法了。

7 对数 z-score 法与线性 z-score 法比较

对数 z-score（LZ）比 z-score 转换法更好，理由有如下三条。

首先，LZ 归一化结果稍好于 z-score（12.7 ＞ 12.5），见前文表 4。

第二，统计上更有利，这实际上反映了发音机制。各发音人五个单字调的基频分布都不是正态的，而是有点正偏差，见图 2a（见下页）中的典型样图。经过对数转换，基频值的分布显得正态了（图 2b），因而更适合于统计处理。下页表 6 给出各人基频分布的偏差（定义为标准三阶中心矩）因子，这些因子画在图 3（见下页）中。从表中看到经过对数转换，偏差平均减少了 0.251（＝0.383－0.132），0.132 是对数偏差因子绝对值的均值。所有人的偏差因子都大大减小了，只有 W5 仅减小 0.013（＝0.127－|－0.114|）。用 Hz 为单位的原始基频值的偏差因子很大，表明五个声调的基频分布不均衡，调型曲线更多的是挤在低频区。同样的正偏差也存在于其他声调语如越南语（Earle 1975：153），Pakphanang 泰语（Rose 1994），以及非声调语如波兰语（Jassem 1971），英语（Menn & Boyce 1982：379）。这也许反映

了基频变化的对数性质。Fujisaki (1983:52) 提出一个假说:主要由声带前端移位引起的声带肌肉的线性拉长导致基频对数增长。这种基频的对数变化表达在 LZ 量度上:基频在低频区的一个小增长,在对数尺度上与高频区的一个大增长有同等的差距。就是说,如果基频值小,转换后的对数值上升就快,但随着基频值的增加,对数值的上升就越来越慢。

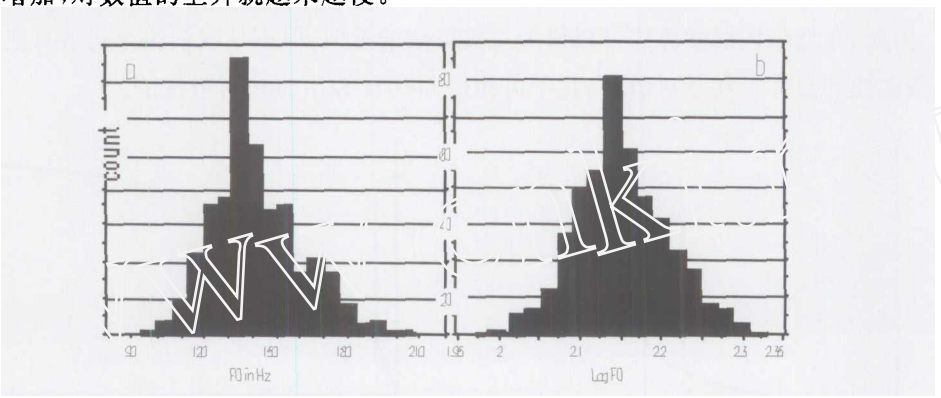

图 2　M5 五个声调的基频分布,单位 (a) Hz ,(b) log 基频

	M1	M2	M3	M4	M5	M6	W1	W2	W3	W4	W5	平均
Hz	0.550	0.299	0.101	0.466	0.812	0.296	0.456	0.503	0.349	0.252	0.127	0.383
Log (Hz)	0.172	−0.041	−0.040	0.138	0.521	−0.013	0.122	0.241	0.014	0.031	−0.114	0.132

表 6　原始基频值和对数基频值的偏差因子

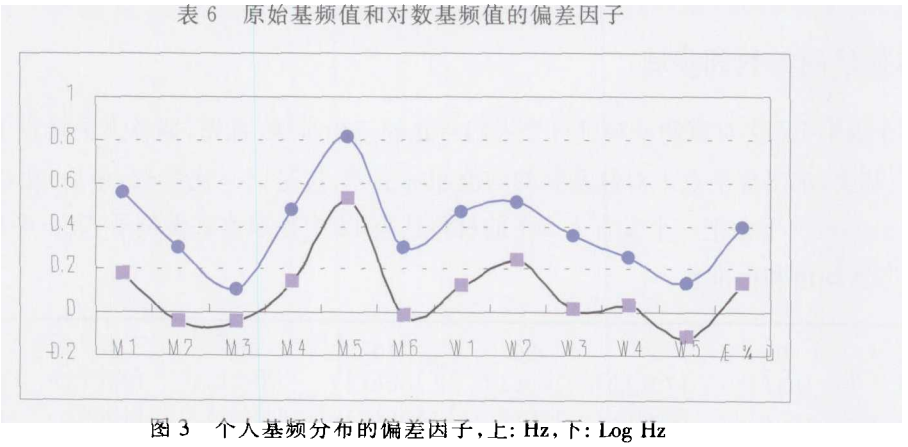

图 3　个人基频分布的偏差因子,上: Hz,下: Log Hz

第三是如何从听感上解释 mel 标度。在音乐中,半音阶与频率是线性变化的,例如,如果第一个八度音程是从 100 到 200Hz,那么第二个就是从 200 到 400Hz,参看上文中半音的定义。但是,语言中听感音高和基频的关系如何并不很清楚。mel 标度在 1000Hz 以上无疑是对数性质的。但是,在 1000Hz 以下的低频区内它是对数性的还是线性的,存在不同看法,如 Menn & Boyce (1982),费国华 (Rose 1987) 认为是线性的,而 Johan't Hart et al. (1990) 认为是对数性的。确定这一点对于声调处理很重要,因为语言一般远低于 500Hz。Koenig 阶设计成 1000Hz 以下是线性的,以上是对数性的。这只是为使用方便,并不能很好地对应 mel 标度。Fant (1973:48) 为共振峰提出一个对数公式 ($mel = k\log(1 + f/1000)$),作为"一个在技术上对 mel 标度有用的近似",并认为"比 Koenig 阶更逼近 mel 标度"。1000Hz 以下听感音高与频率的线性关系,只能在 Koenig 阶上读出,而不能在 mel 标度上读出。

我不同意类似 Menn & Boyce (1982:380) 对 mel 标度在低频区的解释。他们说:"根据 Beranek (1954) 对 Stevens & Volkmann (1940) 结论的图标解释,在 100 到 400Hz 段,等高听感音高的 mel 标度基本上是线性的。"重新审核 Beranek (1954:398) 图 13.8 的结果让我确信,所谓"线性"是个错觉,可

能是由频率横轴的对数标度以及图中相对稀少的数对（mel：Hz）引起的。图 4a 是 Beranek 的原图，横轴是对数基频，纵轴 mel 标度。在 100－400Hz 之间只有两对 Hz：mel 的数值，即 160Hz ：250mel；394Hz ：500mel，因此只能画一条直线。如果如图 4b 那样把频率横轴改为线性，那么这曲线就变成对数增长了。当然，在低于 500Hz 的低频区，等高音高的 mel 标度大体上是线性，那是由于对数算法的性质决定的。因此，在低频区线性算法可以作为对数算法的近似，但不是我们拒绝对数算法的理由。由于图 4b 的整条曲线的总体形状是对数的，我宁可相信 500Hz 以下也是对数性的。

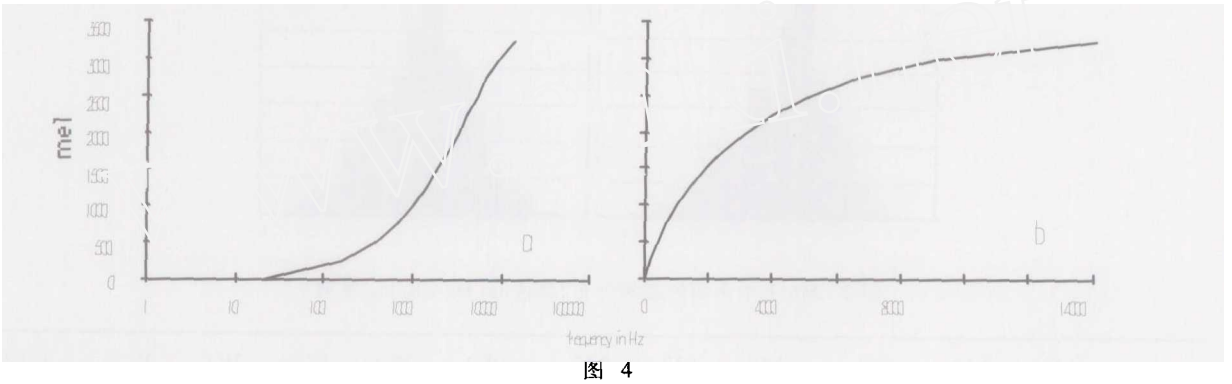

图 4

a. Beranek 图：横轴是对数频率，纵轴是 mel 标度　　b. 纵轴不变，横轴改为线性频率

综上所述，LZ 转换法是所有六种归一化方法中的最佳选择。

8 LZ 转换法的参数和步骤

下面介绍用 LZ 法对声调基频进行个人归一化的三个步骤：首先，把各人采样点上的基频均值换算成对数值；其次，求出标准点上对数基频的均值和标准差，这是归一化参数；最后，用这些参数求出对数基频值的 z-score。下面把一个发音人 M1 的材料从表 11 中提取出来做例子，表 7 中是他的五个声调采样点上的基频均值和标准差。

%	0	5	20	40	60	80	100
T1	183(15)	179(12)	169(13)	156(12)	138(13)	120(12)	104(14)
T2	156(10)	146(8)	139(8)	136(7)	138(7)	143(8)	152(12)
T3	136(15)	130(13)	121(9)	120(9)	124(10)	132(11)	143(17)
%	0	33	67	100			
T4	177(23)	166(15)	156(13)	130(15)			
T5	142(10)	130(6)	131(9)	136(9)			

表 7　M1 五个声调 N 上的基频均值（和标准差），单位 Hz，n＝18

步骤一：把采样点上的基频均值换算成常用对数值，见下页表 8；

步骤二：求出个人归一化参数，即在 22 个标准点（除去 0%点，以及 T1 和 T4 的 100% 点，即表 8 中带阴影的值）上的对数基频的均值（log$m$）和标准差（log$s$）。这个均值是原始基频值的对数几何平均值。M1 的对数基频均值是 2.147，标准差为 0.048，见下页表 9 带框的两个值。该表给出全部 11 位发音人的归一化参数，表上部是单位为 Hz 的均值和标准差，用于线性转换；下半部是对数值的均值和标准差，用于对数转换；中间一行是几何均值；

步骤三：用参数 log$m$ 和 log$s$ 求出对数基频值的 z-score。例如，要把 M1 的 T2 在 5% 处的基频值 146Hz 转换成 LZ 值，可以用如下公式：$lz = (\log_{10}146 - \log_{10}m) \div (\log_{10}s) = (2.164-2.147) \div$

0.048 = 0.35，即表 10 中带框的值。该处基频（146Hz）的对数值 2.164 比他对数基频均值 2.147 高出 0.35 个标准差（SD=0.048）。即在该处的原始基频 146Hz 等于 10 的 0.35SD 次方乘以其几何均值 140.3Hz，即：$146 = 10^{0.35\times0.048}\times140.3$。M1 五个声调的全部 LZ 转换值列在表 10 中。

%	0	5	20	40	60	80	100
T1	2.262	2.253	2.228	2.193	2.140	2.079	2.017
T2	2.193	2.164	2.143	2.134	2.140	2.155	2.182
T3	2.134	2.114	2.083	2.079	2.093	2.121	2.155
%	0	33	67	100			
T4	2.248	2.220	2.193	2.114			
T5	2.152	2.114	2.117	2.134			

表 8　M1 五个声调基频以 10 为底的对数值

	M1	M2	M3	M4	M5	M6	W1	W2	W3	W4	W5
m（Hz）	141	148	105	105	117	134	216	226	230	208	245
s（Hz）	16	28	11	9	12	22	33	26	29	21	26
几何 *m*	140.3	145.0	104.8	105.0	115.9	132.6	213.6	224.2	228.1	207.3	243.5
log *m*	2.147	2.161	2.020	2.021	2.064	2.123	2.330	2.351	2.358	2.317	2.386
log *s*	0.048	0.083	0.046	0.038	0.045	0.071	0.065	0.049	0.054	0.044	0.047

表 9　各人基频和对数值的均值和标准差，n=22

%	0	5	20	40	60	80	100
T1	2.37	2.21	1.66	0.96	−0.12	−1.37	−2.72
T2	0.97	0.35	−0.12	−0.25	−0.14	0.16	0.74
T3	−0.31	−0.71	−1.31	−1.39	−1.10	−0.57	0.15
%	0	33	67	100			
T4	2.11	1.50	0.94	−0.72			
T5	0.13	−0.67	−0.63	−0.28			

表 10　M1 五个声调基频的 LZ 转换值

还要交代一下，表 9 中给出的是四舍五入到整数值（上半）或 3 位小数（下半），而不是真用来计算的精确值。我使用 MicroExcel 计算，可以储存最多达 15 位小数。因此，实际计算结果与表 10 中的 LZ 值之间有些微小误差。

9　归一化的效果

最后让我们来看一下归一化的效果。文末表 11 列出 11 个发音人所有五个声调的基频均值和标准差。以阴平 T1 为例，11 个人的基频均值图标于图 5（见下页）。

从图 5 可以看到这 11 个人的阴平降调高低平斜，急降缓降，差异颇大。如果自变量改为实际时长，差异更大。记得几年前有一次我作一个声调报告时，有位刚接触到实验语音学的听众急切地发问：他测到的每条曲线都不一样，实验语音学有用吗？不用还好，用了更添乱。我用本文开头的话安慰他，原始基频材料的确如此，乱蓬蓬地像一堆茅草，不但难以说明问题，似乎真如他所说更添乱了。这就是为什么实验要有多方面的控制，（Zhu 1999：第 3 章）数据要归一化处理。一句话，要做得系统而彻底。

文末表 11 中 11 位发音人的 T1 原始基频数据经归一化处理后的 LZ 值在表 12 中给出，如下页图

6、图 7 所示。图 6－7 中不但各人的曲线密切吻合在一起（T1 还不是吻合得最紧密的，最紧密的是阴去 T2），而且还分出了两种降调形状：直降和缓降。详细讨论请看朱晓农（Zhu 1999）。〔1〕

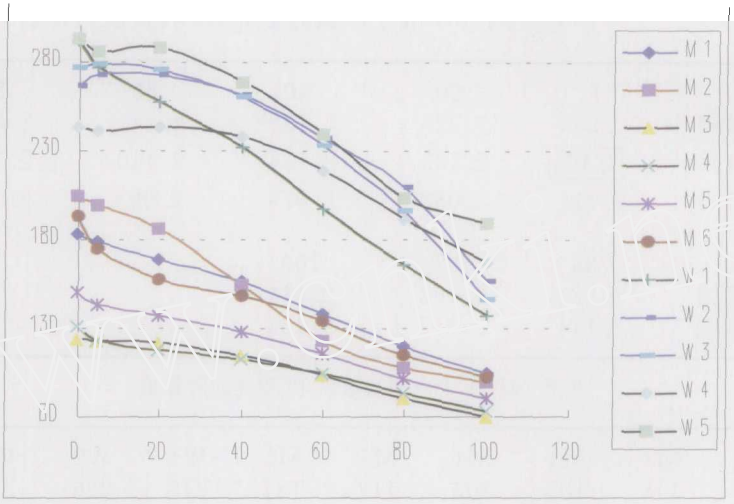

图 5 上海话阴平 T1:11 位发音人的基频曲线，横轴是等长的相对时长

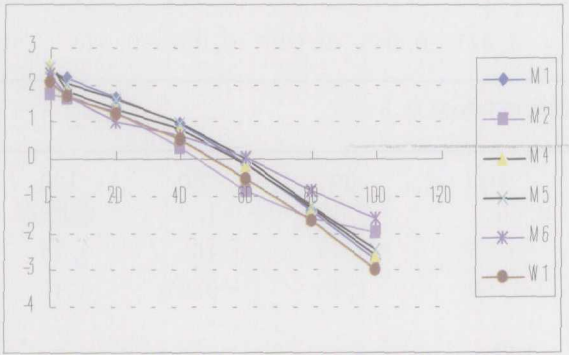

图 6 上海话阴平 T1 直降型：
6 位发音人的 LZ 归一化曲线

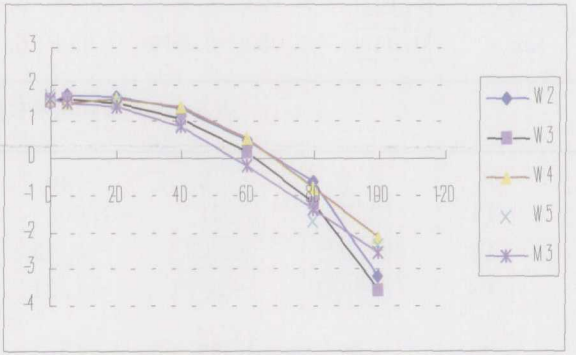

图 7 上海话阴平 T1 缓降型：
5 位发音人的 LZ 归一化曲线

至于语际比较，可以以吴语台州片 14 个点的阴去声为例，这 14 个点的采样大致均匀分布在台州各处。图 8 是他们的 LZ 归一化了的基频曲线，基本上是个半高到高的平调，个别有些小变体。用此图展示阴去声高平调是台州片的区域特征可以令人接受，以此方式来进行跨方言比较可以令人满意。

10 结语

鉴于目前语音分析软件逐渐普及，在方言调查和语音研究中越来越多的人使用实验手段，所以有必要探讨有关实验、测量、数据处理等工作中的方法、原则问题。有一点必须强调：语言实验工作必须做得系统而彻底。如果随意用一些语图，那么你想说明什么都是可能的，因为如本文开头所说，一个语音信号有各种随机变异的物理形式。

本文比较了六类归一化方法，以及各种选点、定频域的方法，引入了判断方法好坏的标准指数和离

〔1〕 有一点需作说明：在《语言科学》编辑部审稿过程中，本文的一位匿名审稿人敏锐地注意到两种降调类型与性别相关：图 6 中 6 条直降型曲线中有 5 个是男性，而图 7 中 5 条缓降型曲线中有 4 条是女性。仅从本文的数据来看，男性与直降型，女性与缓降型相关程度的确很高。此问题的确证有待于更多的统计资料。

散系数。结果表明我的对数 z-score（LZ）方法最好，与此相应的定义频域的参数以均值和标准差为好，而用以归一化的采样点（标准点）以选用语音目标处的时点为佳。

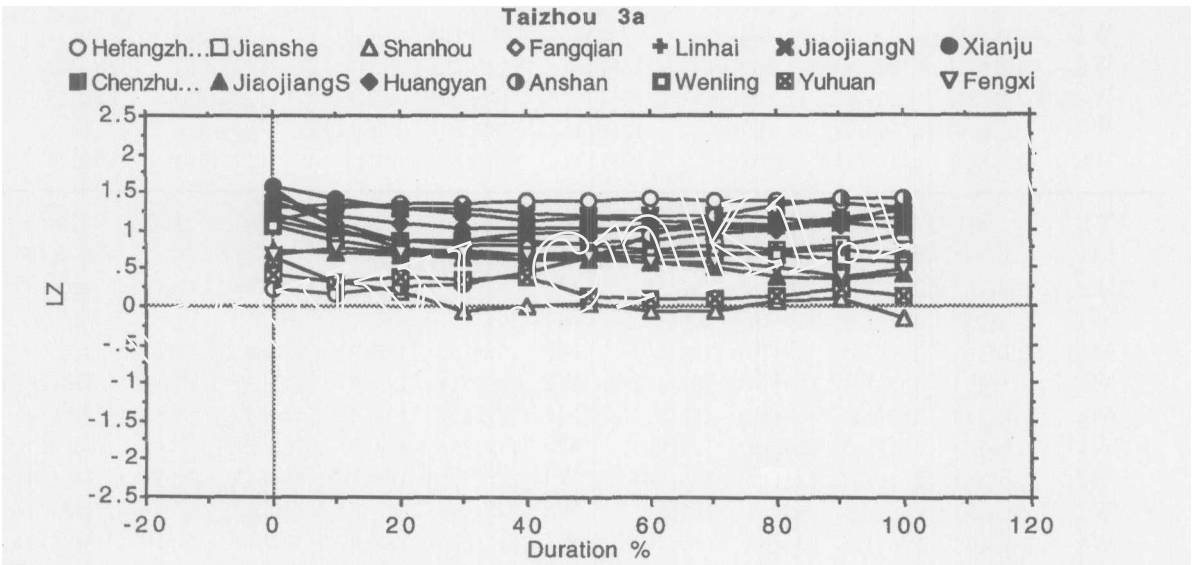

图 8　吴语台州片 14 个方言点的阴去声经 LZ 归一化处理后的基频曲线

附录一：上海话五个声调的基频值和阴平归一化后的标准值

T1	0%	5	20	40	60	80	100
M1	183(15)	179(12)	169(13)	156(12)	138(13)	120(12)	104(14)
M2	204(15)	199(17)	186(21)	154(16)	123(11)	107(9)	99(9)
M3	124(7)	123(6)	122(7)	115(6)	103(7)	90(8)	80(6)
M4	131(14)	123(12)	118(11)	113(9)	104(6)	93(5)	83(7)
M5	150(8)	143(8)	137(9)	128(7)	116(5)	101(4)	90(6)
M6	193(23)	175(17)	157(13)	148(10)	134(9)	115(8)	102(11)
W1	292(18)	277(15)	257(20)	232(16)	197(17)	166(16)	137(22)
W2	266(22)	273(23)	272(24)	261(26)	237(23)	209(14)	156(20)
W3	277(12)	279(12)	276(16)	260(15)	233(11)	196(10)	146(16)
W4	243(15)	241(13)	243(16)	238(17)	219(14)	191(14)	167(15)
W5	293(27)	286(26)	288(17)	268(18)	239(16)	203(12)	189(18)

T2	0%	5	20	40	60	80	100	100+
M1	156(10)	146(8)	139(8)	136(7)	138(7)	143(8)	152(12)	127(8)
M2	166(18)	158(15)	146(13)	141(12)	144(14)	152(16)	166(18)	129(18)
M3	119(11)	113(10)	107(8)	102(8)	103(8)	107(7)	112(8)	96(9)
M4	119(7)	111(6)	104(5)	103(6)	104(6)	106(7)	110(7)	89(9)
M5	131(9)	125(8)	118(9)	115(10)	115(10)	120(11)	126(14)	100(16)
M6	182(19)	159(15)	140(13)	131(14)	128(12)	130(13)	155(19)	#
W1	262(14)	243(16)	223(17)	218(15)	218(17)	222(20)	229(19)	168(33)
W2	243(20)	232(10)	224(11)	224(11)	227(12)	233(12)	239(11)	187(30)
W3	255(19)	243(10)	234(10)	228(9)	228(10)	234(14)	244(22)	208(29)
W4	228(16)	214(9)	206(7)	202(8)	204(8)	212(10)	218(10)	182(18)
W5	268(14)	258(12)	247(14)	245(16)	248(16)	254(16)	261(19)	211(32)

T3	0%	5	20	40	60	80	100	100+
M1	136(15)	130(13)	121(9)	120(9)	124(10)	132(11)	143(17)	116(13)
M2	121(9)	113(8)	108(8)	112(10)	120(11)	135(15)	160(21)	135(17)

M3	99(5)	94(4)	88(3)	87(4)	93(4)	102(5)	114(8)	99(8)
M4	101(11)	95(6)	90(7)	92(6)	98(6)	107(8)	117(9)	100(6)
M5	118(7)	109(5)	101(5)	100(6)	104(7)	109(8)	115(10)	90(9)
M6	122(11)	110(7)	102(6)	102(7)	106(8)	114(11)	154(17)	#
W1	217(13)	200(11)	180(9)	173(8)	175(10)	192(14)	216(15)	156(23)
W2	197(11)	192(8)	188(9)	189(9)	197(12)	213(15)	233(15)	178(26)
W3	222(24)	204(18)	190(16)	191(16)	196(19)	209(19)	223(20)	182(16)
W4	220(16)	198(11)	180(9)	175(11)	182(11)	198(13)	214(12)	172(21)
W5	228(19)	215(15)	203(11)	208(11)	222(14)	242(17)	270(14)	219(20)

T4	0%	33	67	100	T5	0%	33	67	100	100+
M1	177(23)	166(15)	156(13)	130(15)	M2	140(14)	128(14)	143(11)	177(14)	150(16)
M2	196(16)	186(15)	185(15)	162(13)	M3	102(8)	96(6)	101(6)	108(6)	96(8)
M3	132(9)	122(7)	117(7)	100(10)	M4	104(7)	95(7)	99(7)	108(8)	95(8)
M4	128(10)	116(10)	113(10)	102(10)	M6	138(17)	115(9)	121(8)	151(10)	#
M5	141(12)	132(15)	129(16)	101(14)	W2	202(11)	195(6)	205(9)	225(9)	179(18)
M6	178(17)	156(14)	154(12)	138(9)	M1	142(10)	130(6)	131(9)	136(9)	114(10)
W1	280(18)	281(15)	259(19)	192(26)	M5	117(7)	103(7)	105(7)	112(8)	95(8)
W2	250(20)	255(15)	244(17)	180(26)	W1	212(19)	188(10)	194(11)	213(12)	165(24)
W3	289(19)	286(22)	259(22)	193(18)	W3	222(16)	204(9)	210(10)	228(11)	188(12)
W4	245(9)	238(10)	233(8)	196(19)	W4	219(22)	183(9)	187(7)	205(8)	171(10)
W5	294(10)	278(13)	270(14)	233(16)	W5	231(11)	222(10)	226(7)	234(11)	200(19)

表 11 上海话 T1 至 T5 的基频均值（和标准差），单位 Hz，n=18

	0%	5	20	40	60	80	100	*n*
T1s M1	2.37	2.21	1.66	0.96	−0.12	−1.37	−2.72	
M2	1.78	1.66	1.30	0.32	−0.85	−1.60	−1.98	
M4	2.50	1.81	1.34	0.83	−0.14	−1.35	−2.60	
M5	2.46	2.02	1.61	0.97	0.00	−1.28	−2.46	
M6	2.29	1.71	1.02	0.66	0.08	−0.87	−1.61	
W1	2.09	1.73	1.23	0.56	−0.54	−1.67	−2.96	
m (*s*)	2.25(0.27)	1.86(0.21)	1.36(0.24)	0.72(0.25)	−0.26(0.36)	−1.36(0.28)	−2.39(0.50)	6
BSV		0.045	0.058	0.065	0.129	0.080		
MBSV=0.076		SV=1.438		DC=5.3%	NI=77.2%/5.3%=**14.6**			
T1d W2	1.51	1.72	1.69	1.33	0.50	−0.64	−3.20	
W3	1.56	1.63	1.53	1.06	0.15	−1.22	−3.58	
W4	1.59	1.50	1.59	1.38	0.55	−0.80	−2.13	
W5	1.70	1.49	1.55	0.89	−0.18	−1.69	−2.36	
M3	1.60	1.50	1.41	0.84	−0.18	−1.40	−2.57	
m (*s*)	1.59(0.07)	1.57(0.10)	1.55(0.10)	1.10(0.25)	0.17(0.35)	−1.15(0.43)	−2.77(0.60)	5
BSV		0.010	0.010	0.061	0.124	0.185		
MBSV=0.078		SV=1.175		DC=6.6%	NI=101.3%/6.6%=**15.3**			

表 12 经 LZ 转换的 T1 基频值

附录二：实验程序

附录一中的上海话单字调录音是 1992 年 7 月至 9 月间在澳洲国立大学（ANU）文学院语言学系的语音实验室中进行。下面简单介绍测试字表、发音人、录音情况、实验所用仪器和测量步骤。

测试字表

本项研究中使用的单音字都是上海话中实有的语素，见表 13。大部分字是自由语素，个别是黏着语素，如：便 bi 。测试字的声母有两个，或是不送气清辅音 p，或是哗声清辅音 b（关于哗声，请看 Zhu 1999）。测试字的韵母包括以下六个：i，u，a，iq，oq，aq。上海话中派发音人区分有摩擦的紧 ii [i$_z$] 和普通 i [i] ，如：比 pii ：扁 pi。受试字用的是普通 i

[i]。其所以选择这六个韵母是为了保持高低前后元音的平衡，因为元音高低会影响基频高低（Lehiste 1970）。

T1	阴平/长	边 pì [pì]	播 pù [pā]	爸 pa [pà]
T2	阴去/长	扁 pi [pi]	布 pu [pu]	摆 pa [pa]
T3	阳去/长	便 bi [p̤i]	步 bu [p̤u]	败 ba [pa]
T4	阴入/短	笔 piq [pI]	北 poq [puʔ]	八 paq [paʔ]
T5	阳入/短	别 biq[p̤Iʔ]	薄 boq [p̤uʔ]	白 baq [p̤aʔ]

表 13　字表

发音人

本文使用的 11 个发音人的年龄在录音时（1992）最小 25 岁，最大 43 岁。他们都是上海人。表 14 给出他们的个人资料。第 4 栏是方言变体情况，发音人有 5 个说的是中派 I，6 个中派 II，前者比后者多 3 个韵母。三个人的教育程度是中学，其余都是大学。最后一栏是录音时读字表的风格，由我作为本地人作出判断。所有 5 个女发音人读字表时都很正式，有两个（W2，W5）有些紧张。男发音人有 3 个读得很正式，尤其 M6 和 M2，其余 3 个较随意，尤其 M3 和 M5。

发音人编号	姓名缩写	年龄变体	中派程度	教育	在澳年数	发音风格
M1	HSG	30	I	大学	2	刻意
M2	ZXN	40	I	大学	2.5	刻意
M3	ZT	34	II	中学	4	随意
M4	XZL	36	I	中学	4	随意
M5	CR	39	I	大学	3	随意
M6	XWY	42	II	大学	5.5	刻意
W1	WYH	28	II	大学	1	刻意
W2	WXW	25	II	大学	3	紧张
W3	LSB	43	I	大学	4.5	刻意
W4	DCM	30	II	大学	3	刻意
W5	HYQ	35	II	中学	3.5	紧张

表 14　发音人情况

录音

所有录音使用的都是 Nagra 4.2 盘对盘的磁性录音机，录音速度为每秒 7.5 英寸，Nakamichi CM300 cardioid 麦克风，音强水平手动调整。录音前，发音人先进行练习，直到觉得读起来没问题。这被证明很有用，因为发音人从小到大都是用普通话读书的。很多人一开始很不习惯用上海话念书，再加有些字在自然语言中从不或很少单独使用。有两个人（M3 和 W5）练习了 15 分钟，但录音时还有不少读错。我让他们多念了两遍，以保证每人每个字至少有六个发音正确的读例。实验没用负载句，所有测试字都是以单字形式读的。每人读 6 遍，所以一共有 990 个读例（= 5（声调）×3（元音）×6（遍）×11（人））。测试字写在 3 张 A4 大小的纸上。为了消除开始高、结尾低的"页调"（page intonation）效应，每张纸上开头和结尾处加两个充数字（dummy words），发音人被告知每个字读完稍停两三秒，以避免可能的连读效应。为保证发音人不知道实验目的，字表中间还插入了一些声韵配合不同的充数字。尽管告诉发音人读字表时尽可能保持自然，但各人的发音风格无法保持一致。有人正式点，有人有些紧张。不过，由于进行了归一化处理，不影响结果。

声学仪器和测量步骤

单字调的声学信息是从仿真式仪器上获得。我试过 ILS（Signal 技术公司的 Interactive Laboratory System）的基频提取程序，发现它在好几种情况下不工作。为保持材料之间的可比性，我决定对所有声调测量都用仿真式仪器。所有语图都是用 Voice-Print Laboratories 系列 700 Spectrum Analyser 做的。基频曲线从窄带（45Hz 滤波）图上测量，辅以同时性的宽带（300Hz 滤波）图定时。

用以提取时长和基频的测量程序基本上是按照费国华（Rose 1982，1990）所描写的。声调起点"等同于发声起点。在这种情况下，经常要忽略第一个明显的声门脉冲，因为它的振幅不够大，人耳听不到。这是 Lisker & Abramson（1963:416）采用的方法"（Rose，1982:7－10）。这个方法实际上等同于 Baken（1987:376）的方法，他把起点定义为 F2

的第一个声门直条清晰可见处。降调基频终点是在宽带图上的基频直条有规律成比例的间隔结束处。升调基频常常末尾附上一条听不见的基频降尾，它的终点定在窄带图的基频峰点处。降尾是音节末喉塞尾的反映（Rose 1989）。在宽带图上获得的起终点再转到窄带图上。

语图用笔和尺在短调每三分之一处，长调每五分之一处（再加 5%时点）定出采样点。这样的取样率大约是 25Hz，足以提取整段声调上的基频信息。此外，升调末尾的降尾点（标为 100＋%时点）也测量。基频测量用的是 Rose（1987:237）所用的那种连接计算机的数字化测量器，计算机软件由"Macquarie 大学言语听觉和语言研究中心设计，经修改每图可测量多达 100 测量点。在 90% 置信度精确度估计为±5Hz。"

参考文献

沈 炯 1985 北京话声调的音域和语调，载林焘等《北京语音实验录》，73－130 页，北京：北京大学出版社。

石 锋 1990 论五度值记调法，载《语音学探微》，27－52 页，北京：北京大学出版社。

朱晓农 1996 上海音系，《国外语言学》第 2 期，29－37 页。

朱晓农 2002 论分域四度标调制，初稿内容曾提交"中国东南部方言比较研究第九届国际研讨会"（杭州，2002）。

Baken, R. J. 1987 *Clinical Measurement of Speech and Voice*. Boston: College-Hill Press.

Beranek, Leo Leroy 1954 *Acoustics*. New York: McGraw-Hill.

Dart, Sarah 1987 An aerodynamic study of Korean stop consonants: measurements and modelling. *Journal of the Acoustical Society of America* 81:1, 138－147.

Disner, S. 1980 Evaluation of vowel normalization procedures. *Journal of Acoustic Association of America* 67:1, 253－261.

Earle M. A. 1975 An Acoustic Phonetic Study of Northern Vietnamese Tones. *Speech Communications Research Laboratory monograph* No. 11. Santa Barbara, California.

Fant, Gunnar 1966 A note on vocal tract size factors and non-uniform F-pattern scaling. *Speech Transmission Laboratory Quarterly Progress and Status Report* 4/66 (Stockholm: Royal Institute of Technology). Reprinted in Fant, 1973, *Speech Sounds and Features*, Cambridge, Mass.: MIT Press.

Fant, Gunnar 1973 *Speech sounds and features*. Cambridge, Mass.: MIT Press.

Fujisaki, Hiroya 1983 Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (ed) The Production of Speech, 39－55. New York: Springer-Verlag.

Gerstman, L. H. 1968 Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics AU*－16, 78－80.

Harshman, Richard 1970 PARAFAC: Models and conditions for an 'explanatory' multi-model factor analysis. Working Papers in Phonetics 16. UCLA Phonetics Laboratory.

Hart, Johan 't; Rene Collier and Antonie Cohen 1990 *A Perceptual Study of Intonation: An Experimental Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Jassem, Wiktor 1971 Pitch and compass of speaking voice. *Journal of International Phonetic Association* 1:2, 59－68.

Labanov, B. M. 1971 Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49:2, 606－608.

Ladd, D. R.; K. Silverman; F. Tolkmitt; G. Bergmann and K. Scherer 1985 Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America* 78, 435－444.

Lehiste, Ilse 1970 *Suprasegmentals*. Cambridge, Mass.: MIT Press.

Lisker, Leigh and Arthur S. Abramson　1964　A cross-linguistic study of voicing in initial stops: acoustical measurements. *Word* 20:3, 384－422.

Menn, L. and S. Boyce　1982　Fundamental frequency and discourse structure, *Language and Speech* 25, 341－383.

Rose, Phil　1981　An Acoustically based Phonetic Description of the Syllable in the Zhenhai Dialect. PhD dissertation, The University of Cambridge.

Rose, Phil　1982　Acoustic characteristics of the Shanghai-Zhenhai syllable types. In David Bradley (ed), *Papers in South-East Asian Linguistics*, No 8: *Tonation*, *Pacific Linguistics* A－26, 1－53.

Rose, Phil　1987　Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication* 6, 343－351.

Rose, Phil　1989　Phonetics and phonology of Yang tone phonation types in Zhenhai. CLAO 18:2, 229－245.

Rose, Phil　1993　A linguistic-phonetic acoustic analysis of Shanghai tones. *Australian Journal of Linguistics* 13, 185－220.

Stevens, S. S. and J. Volkmann　1940　The relation of pitch to frequency: a revised scale. *American Journal of Psychology* 53, 329－353.

Takefuta, Y.　1975　Method of acoustic analysis of intonation. In S. Singh (ed) *Measurement Procedures in Speech Hearing and Language*, 363－378. Baltimore: University Park Press.

Wakita, Hisashi　1977　Normalization of vowels by vocal tract length and its application to vowel identification. *IEEE Transaction on Acoustics, Speech, and Signal Processing ASSP*－25, 183－192.

Zee, Eric and Ian Maddieson　1980　Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis. *Glossa* 14:1, 45－88.

Zhu, Xiaonong　1999　*Shanghai Tonetics*. Muenchen, Germany: Lincom Europa.

作者简介

朱晓农，男，原籍浙江乌镇，1952 年生于上海。澳洲国立大学（Australian National University）语言学系博士。研究兴趣：实验语音学，音韵/系学，方言学，方法论，语言和逻辑。主要论著有《北宋中原韵辙考》，*Shanghai Tonetics*。

F0 Normalization: How to Deal with Between-speaker Tonal Variations?

Zhu Xiaonong

Hong Kong University of Science and Technology, Hong Kong

Abstract　This paper compares six mathematical transforms for F0 normalisation using eleven speakers of Shanghai, a dialect of Wu Chinese. Various strategies for defining range and selecting normalisation points are also exploited. The logarithmic z-score transform with normalisation points at F0 targets is found to be superior due to its higher normalisation indices, its advantages in statistics and, perhaps, its accordance with perception and production.

Keywords　tone　fundamental frequency　normalization　between-speaker variation