

中图分类号:H211 文献标识码:A 文章编号:1004-8634(2000)04-0049-(07)

关于确定声调调形和调值方法的探讨

钱 瑶 高云峰

(上海师范大学 人文学院, 上海 200234)

摘 要: 本文包括三部分内容。第一,通过自相关函数的算法得出 F_0 曲线。第二,利用曲线拟合确定调形。第三,提出声调曲线的新的规整方法,确定调值。
关键词: 声调;基音周期;声调整

0 引言

汉藏语是世界上声调最丰富的语言,汉藏语言的田野调查中最困难的工作就是调形和调值的确定。近年来实验语音学和计算机语音处理的发展,使我们有可能通过计算机来自动处理这方面的工作,至少可以作为田野调查的一种重要的辅助手段。

On the Value of *The Quotations From Zhu Zi* in the Chinese Study

XU Shi-yi

(Humanities College, Shanghai Teachers University, Shanghai, 200234, China)

Abstract: *The Quotations From Zhu Zi* (Zhu Zi Yu Lei) is the valuable data that reflects the scholar's spoken language in the Song dynasty. The study of its language is less than that of its philosophy and ideology. This paper discusses its state that reflects the evolution from ancient Chinese to modern Chinese, examines its language feature, expounds its value on the study of Chinese vocabulary history and of the history of grammar.

Key words: *The Quotations From Zhu Zi*; Language; Value

收稿日期:2000-01-05

作者简介:钱 瑶(1972-),女,上海市人,上海师范大学语言研究所助教,语音学硕士研究生。

高云峰(1975-),男,北京市人,上海师范大学语言研究所汉语音系学硕士研究生。

1 F_0 曲线的机器确定

声调调形的物理基础就是语音的基频变化轨迹,即 F_0 曲线,因此首先需要在计算机上提取语音的基频并对基频轨迹进行平滑处理。

1.1 短时基音周期的估计

常用的基频估计方法有短时自相关函数法、短时平均幅度差函数法(AMDF)以及 FFT 倒频谱法。

以上三种方法中,各有优缺点。本文采用自相关函数法,这种方法的缺点是在实际应用中经常出现共振峰与声门波峰值的混杂现象,为了克服窗长影响和声道共振峰的干扰,我们采用这种方法时加入了“中心削波技术”,有效地消除了声道转移函数的影响,抑止了信号中其他因素带来的扰乱。

设 $S_w(n)$ 是一段加窗的语音信号,它的非零区间为 $n=0 \sim (N-1)$, n 的单位为帧,用 $R_w(l)$ 表示它的短时自相关函数,其计算公式如下:

$$R_w(l) = \sum_{n=0}^{N-l-1} S_w(n) S_w(n+l)$$

如果 $S(n)$ 是一个浊音性的周期信号,那么它的短时自相关函数也呈现出明显的周期性,而且短时自相关函数 $R_w(l)$ 的周期等于语音信号 $S(n)$ 的周期。相反,清音语音接近于随机噪声,它的短时自相关函数不具周期性且随着 l 的增大而迅速减小。因此我们可以利用这个特点来判断一个语音是浊音还是清音,还可以由此决定浊音基频 $R_w(l)$ 的第一个最大峰值的位置,也就是基音周期的位置。

中心削波可表示为:

$$y(n) = c[s(n)]$$

$$c[s] = \begin{cases} s - c_l, & (s > c_l) \\ s + c_l, & (s < -c_l) \\ 0, & (|s| < c_l) \end{cases}$$

$$c_l = 0.68 \min[IPK1, IPK2]$$

IPK1, IPK2 为每帧中前三分之一和最后三分之一范围内的峰值。

经过中心削波后峰起尖锐突出,用它来进行基频估计效果更好。

1.2 基频轨迹的平滑

具体的计算总会使个别帧的基频偏离正常轨迹(通常是偏离正常值的 2 倍或 1/2),所以必须对计算结果进行平滑处理。将当前时刻的基频与前一帧结果相减,若差的绝对值大于前一帧结果的 2/5,便需要将当前值减半、取 2/3 或加倍。用这种方法由前到后,由后到前处理两遍。最后用中值滤波和线性滤波的组合进行平滑,就可去掉那些“野点”。

除此以外,对基频的估计也会存在一些偏差。为了解决这个问题,本文采用相邻基音比较算法。它的根据是:语音信号及其基频变化是连续的,相邻两个语音段的基频变化不大。如果已知第 n 帧语音信号的基频为 $T(n)$,那么它的相邻频差的绝对值同 $T(n)$ 的比值应小于基频连续变化门限。门限与语音信号采样率有关,可以根据实际情况选择,根据笔者的经验取值为 0.1~0.2。

图 1 虚线是通过削波自相关函数求得的 F_0 轨迹,实线是经过平滑和相邻基音比较算

法后的轨迹。

2. 由 F_0 曲线确定声调的调形

确定调形有以下几个步骤。

2.1 去除弯头和降尾。

当声调的起点或结束频率很高的时候,声带由于惯性作用往往有一个由松到紧或从紧到松的过程,基频曲线中就会出现一个弯头段或降尾段的现象,弯头段现象比较严重,它会影响声调调形的判别。我们在 ASL 上用语音合成做过实验,去掉基音曲线中的弯头段与降尾段并不影响声调。

解决此现象可采取如下的方法:

(1)可根据弯头段与降尾段出现的位置来处理,弯头段出现在曲线的开始处,降尾段出现在曲线的尾部。我们把所得的 F_0 曲线在头尾处各截去 5%,对于弯头段与降尾段不明显的曲线这样处理也就可以了。

(2)对于弯头段与降尾段明显的曲线,例如高降调弯头段和高平调降尾段都很明显。我们在基频最大值和最小值的地方,通过声调曲线的走向判别弯头段和降尾段的确切位置予以去除。(普通话弯头段约 60ms,降尾段约 40~50ms)。

2.2 舒声调和促声调的判别。

舒声调和促声调的区别性特征是对长,当时长短到一定的程度调形不起作用。郭锦桴、鲍怀愨(1989)认为厦门话促声调的时长约是舒声调的 50%~70%。高云峰认为上海话的促声调时长约占舒声调时长的 17%~20%。

2.3 舒声调的调形判别。

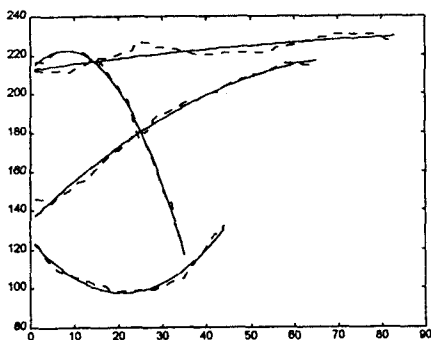


图 2

王士元等人(王士元 1983)曾于 1975 年做过北京话声调的阴平和阳平的范畴感知实验,通过改变基频曲线最低点同最高点连线的斜率进行听辨,得出当 ΔF 同时长之比小于 0.03 时是阴平,反之是阳平。

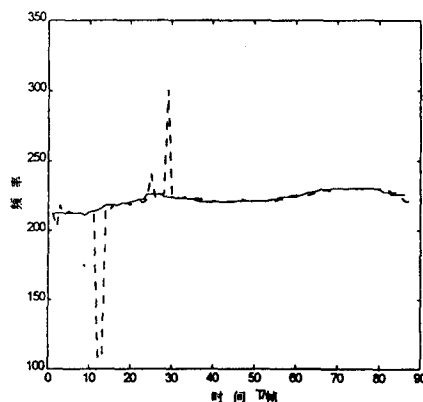


图 1

四条拟合曲线的二项式分别是:

$$y(x) = -0.0012x^2 + 0.3045x + 212.5058$$

$$y(x) = -0.0136x^2 + 2.1487x + 134.9612$$

$$y(x) = 0.062x^2 - 2.6215x + 125.3718$$

$$y(x) = -0.1451x^2 + 2.3429x + 212.9429$$

(x 代表时间/帧, y 代表频率)

Zue 曾于 1972 年对阳平和上声做过范畴感知实验,通过改变基频曲线最低点同最高点所持续的时长进行听辨,得出当 ΔF 同时长之比小于 0.15 时是阳平,反之上声。

本文采用曲线拟合的方法作调形判别。我们用二次多项式 $y(x) = ax^2 + bx + c$ 来近似拟合基频曲线,通过最小二乘法确定系数 a 、 b 、 c 。图 2 的四条曲线分别代表北京话的阴平、阳平、上声、去声,虚线表示 F_0 的实际曲线,实线是用二项式拟合的曲线。

我们用曲线拟合的判别根据是:

(1) 当 $|a|$ 和 $|b|$ 很小时, y 值受 x 值的影响很小,说明它是条平坦的直线,由此可以判别为平调。

(2) 在 x 连续的范围求 y 的导数 y' ,它代表了曲线的单调性。当 $y' > 0$ 时是单调上升,反之是单调下降。当曲线的绝大部分是上升或下降时,就可以判别为升调或降调。

(3) 在进行(2)时, $y' > 0$ 的 x 值的个数同 $y' < 0$ 的 x 值的个数差不多时,那么它既有单调上升又有单调下降部分。由此可以判别为曲折调。

(4) y'' 能决定曲线是凹的还是凸的。 y'' 也就是 $2a$, $a > 0$ 是凹的, $a < 0$ 是凸的。由此能判别是降升的曲折调还是升降的曲折调。

根据笔者所做的七组普通话声调的测试结果:

平调(阴平)的 $|a| < 0.005$, $|b| < 0.5$ 。

升调(阳平) $y' > 0$ 的 x 值数是 $y' < 0$ 的 3 倍以上。

降调(去声) $y' < 0$ 的 x 值数是 $y' > 0$ 的 3 倍以上。

上声调(上声) $y' < 0$ 的 x 值数同 $y' > 0$ 的比值在 2~0.5 之间。

笔者用以上数据进行三组普通话单字调声调调形的判别,判别正确率为 91%。

3 声调调值的确定

语音从一个人的口中发出,到另一个人听到,经过了编码、传输、解码三个阶段。编码阶段是发音器官的一系列动作,属于生理学的内容。语音在空气中的传输是声波,属于物理学的内容。解码阶段是人的听觉器官将声波转成脑电波,由人的语言中枢解译成语音信息,属于心理学的内容。我们上面讨论过的 F_0 曲线属于物理学的内容,它还不完全等同于我们心理上感受到的音高。所以,我们还必须把基频转换成心理上的音高。

那么,基频与心理平面上的音高有什么关系呢?音乐上的音高感知与语音上的音高是相当一致的。当两个音的基频相差一倍,人们感知到的音高相差一个八度,声学上叫作一个倍频程。我们举一个例子来说明这种关系,假设 $f_1 = 20\text{Hz}$,当 f_2 与 f_1 相差 1 个倍频程时, $f_2 = f_1 \cdot 2^1 = 40\text{Hz}$,当 f_2 与 f_1 相差 2 个倍频程时, $f_2 = f_1 \cdot 2^2 = 80\text{Hz}$,当 f_2 与 f_1 相差 n 个倍频程时, $f_2 = f_1 \cdot 2^n \text{Hz}$ 。由此我们可以得到倍频程的计算方法,当两个音的基频为 f_1 和 f_2 时, f_2 和 f_1 相差的倍频程 n 为:

$$n = \log_2 \frac{f_2}{f_1} = \log_2 f_2 - \log_2 f_1$$

一个倍频程作为一个音高单位还嫌太大,在音乐上把一个倍频程分作 12 等分,即十二平均律,每个等分为半音程。当 f_2 与 f_1 相差半音程的时候, $f_2 = f_1 \cdot (2)^{1/12}$,当 f_2 与 f_1 相差 n 个半音程的时候, $f_2 = f_1 \cdot (2)^{n/12}$ 。

决定调形的主要特征是音高的相对变化,并不是基频的绝对值。如下图中的两个声调,右边比左边的高,但是它们的调形是完全相同的。一个人说话的时候,声调曲线在

不断的变化,但是如果不发生语言环境的变化,音高的平均值通常没有太大的变化。这个平均值我们叫作调层。研究汉语的声调主要是研究调形,调层可能由许多非语言学的因素产生,例如女性的调层比男性的高,小孩比老人高,同一个人说的话因为各种说话环境不一样调层也会不一样。上图中的两个声调可能就是同一个语言中的同一个声调,一个是男人发的音,另一个是女人发的音。当一个人用同一个调层说话的时候,所有声调中的最高音与最低音的差值叫作调域。据吴宗济(1997)的实验,声调的调层因为不同的发音人会有不

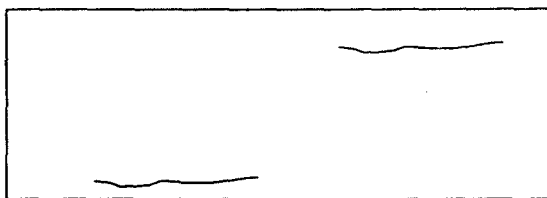


图 3

同,但调域的变化并不很大。我们对上海话的声调也做过类似的实验。我们用不同声调的字组成一组,把阴平调定作基调,让一个人用不同的基调发这组字。不同基调的音高可以相差十几个半音,但是每组读音的调域只在一二个半音内波动。正因为一个语言中的调域变化不大,所以调形的变化也不会很大。

所以,调形的规整主要是去除调层的影响。那么,我们通过什么方法能够去除调层的影响呢?让我们看下面两个图。

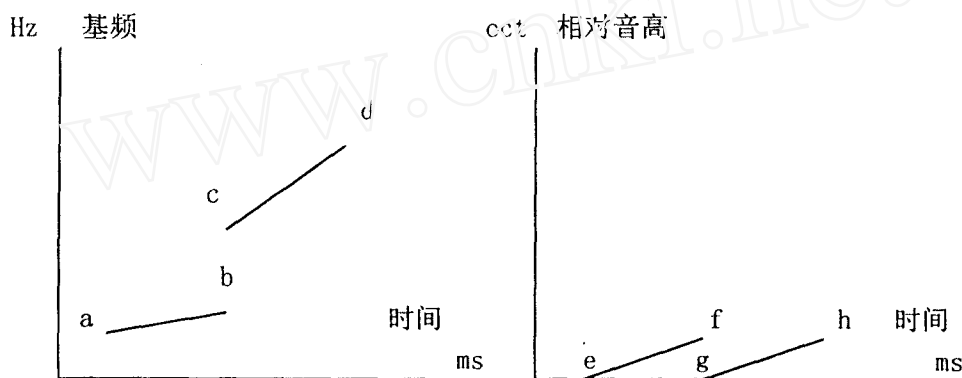


图 4

上图有两个升调,一个调的基频从 a 赫兹上升到 b 赫兹,另一个调从 c 赫兹上升到 d 赫兹。第一个调如果以 a 作为比较的基点,把 a 和 b 的频率转成倍频程分别是 $\log_2(a/a) = 0$ 和 $\log_2(b/a)$,画到右图上就是 e-f。另一个声调我们也以它的最低点 c 作为比较的基点,把它们分别转成相对于 c 的倍频程。假设 $c = ka$, $d = kb$,那么 c 和 d 的倍频程分别为 $\log_2(c/c) = 0$ 和 $\log_2(d/c) = \log_2(kb/ka) = \log_2(b/a)$,画到右图上就是 g-h,两者的相对音高一样。如果不考虑它们发音时间上的先后区别,g-h 就与 e-f 完全等同。由此可见,如果把两个不同调层的声调转成倍频程,就能去除调层的因素。

上面的例子说明,声调有两种不同的描写,潘悟云教授把它们叫作绝对音高域(Absolute Pitch Domain)和相对音高域(Relative Pitch Domain)。绝对音高域的横座标是时

间,单位是毫秒(ms),纵坐标是基频,单位是赫兹(Hz),简称为声调的 AP 值。相对音高域的横坐标也是时间,纵坐标是倍频程(Octive),或半音程,简称为声调的 RP 值。根据上面的讨论,RP 值与 AP 值有以下的关系:

$$RP_i = \log_2 \frac{x_i}{x_l} = \log_2 x_i - \log_2 x_l$$

其中的 x_l 为绝对调域中的最低点,也是倍频程计算的起点。 X_i 为声调 F_0 曲线上各点的频率值。调层只能是绝对声调域的内容,不同调层的声调到了相对音高域,调层的区别消除。调域可能是绝对音高域的值,单位是赫兹,也可能是相对音高域的值,单位是倍频程或半音程。在需要区别两者的地方,把前者叫作绝对调域,后者叫作相对调域。相对调域相同的两个音,绝对调域可能不一样。如上图中的 e-f 和 g-h 的相对调域相同,它们在对应的绝对音高域内是 a-b 和 c-d,绝对调域不一样。相对调域相同的声调,可以通过上面的方法进行规整。但是,不同的发音环境也可能导致相对调域的稍许变化,这就要对上面的数学处理方法作进一步的调整。我们可以把调形与调域的关系设想为画在一块橡皮上的图画,当橡皮拉长或压扁的时候,上面的图像也跟着改变。所以,如果通过调形与调域的比值来规整调形,就能去除调域差异给调形所带来的影响。为此,我们把 RP 法的公式作以下的修改:

$$RP_i = \frac{\log_2 \frac{x_i}{x_l}}{\log_2 \frac{x_h}{x_l}} = \frac{\log_2 x_i - \log_2 x_l}{\log_2 x_h - \log_2 x_l}$$

其中的 x_h 为绝对调域中的最高点, x_l 为绝对调域中的最低点,也是倍频程计算的起点。 X_i 为声调 F_0 曲线上各点的频率值。

下面我们来讨论如何确定 x_h 和 x_l 。一个语言中的 F_0 曲线画出来以后,调域最高值和最低值的确定并不是难事。例如北京话的基频曲线中可以把上声的最低调定为调域的最低点,把阴平或去声曲线的最高点定为调域的最高点。石锋(1994)也有类似的公式,但是本文同他不是一样的地方是对 x_h 和 x_l 的处理方法。如果用机器自动确定最高值和最低值都会碰上一些逸出值,影响到调域的确定。例如,去声调在下降过程中可能会降得很低,比上声调的最低点还要低。所以,本文采用统计学的方法来确定 x_h 和 x_l 。我们先求得同一调层中所有声调的均值,即调层值:

$$m = \frac{1}{n} \sum_{i=1}^n x_i$$

我们把基频 F_0 看作随机变量,标准方差 s 刻划了该随机变量的取值围绕平均数 m 的离散程度:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m)^2}$$

标准方差越小,随机变量的取值就越集中在它的平均数的周围。统计学告诉我们,用 $m+cs$ 和 $m-cs$ 可以近似包括偏离平均值的所有值。如果随机变量是正态分布的(影响随机变量的因素很多,但每个因素所起的作用都不太大),当 $c=2$ 时,97.7%的随机变量的值落在了 $m+cs$ 至 $m-cs$ 之间。也就是说,各个声调曲线上 97.7%的点落在 $m+cs$ 和 $m-cs$ 之间,所以可以把这两个值看作是调域的最高值 x_h 和最低值 x_l 。于是 RP_i 值的公

式可以改写如下:

$$rp_i = \frac{\lg x_i - \lg(m - cs)}{\lg(m + cs) - \lg(m - cs)}$$

有了规整后的 RP 值后,再根据赵元任先生“五度调符”的定调原则确定调值。

下图是同一个发音人用三个不同字发的普通话四声,左边是它们的基频域,右边是经过 RP 法规整以后的相对音高域。

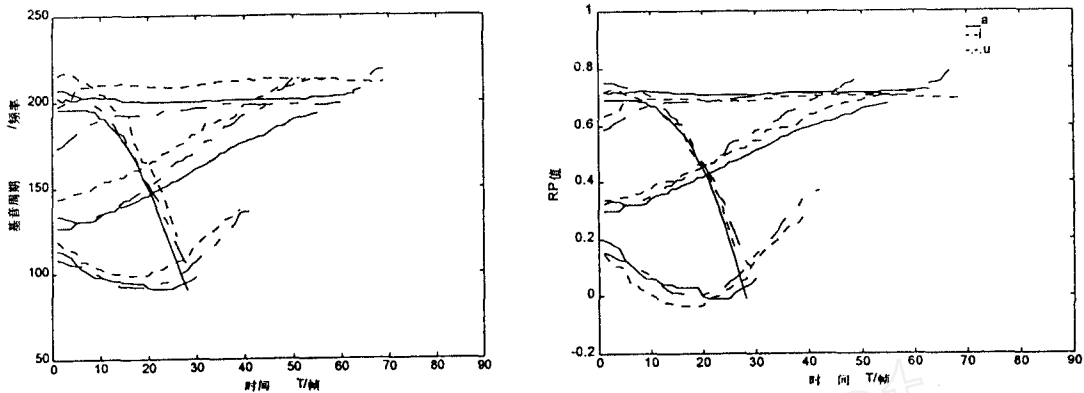


图 5

本文承蒙潘悟云教授的指导,在此谨致谢忱。

(责任编辑:敏 捷)

参考文献

- [1] 吴宗济.1989.实验语音学概要[M].高教出版社.
- [2] 石峰,廖荣蓉.1994.语音丛稿[M].北京语言学院出版社.
- [3] 郭锦桴,鲍怀翘.1989.厦门话声调的声学特性分析[J].语言论集.第三辑,文化艺术出版社.
- [4] 吴宗济.1997.从声调与乐律的关系提出普通话语调处理的新方法[C].庆祝中国社会科学院语言所建所45周年-学术论文集.中国语文编辑部.商务印书馆.
- [5] L·R·拉宾纳,R·W·谢弗.语音信号数字处理[M].科学出版社.
- [6] 胡光锐.语音处理与识别[M].上海科学技术文献出版社.
- [7] 数理统计与标准化[M].技术标准出版社.

① ASL, Analysis Synthesis Laboratory, Kay Elemetrics Corp.

On the Method of Determining Pitch Contour and Value

QIAN Yao GAO Yun-feng

(Research Institute of Languages, Humanities College, Shanghai Teachers University, Shanghai, 200234, China)

Abstract: This paper includes: 1. Acquire F_0 curve through the arithmetic of auto-correlation function. 2. Determine pitch contours by curve fitting. 3. put forward new normalized method of pitch contour to determine the pitch value.

Key words: Pitch; Normalization