

# 关于汉语音节知觉空间的实验研究\*

周迅溢 王 蓓 杨玉芳

(中国科学院心理研究所 100101 北京)

2001 年 6 月 13 日收到

2002 年 6 月 4 日定稿

**摘要** 通过心理物理实验方法建构汉语音节知觉的多维空间结构, 寻求有关汉语音节知觉的客观表现。结果表明, 在声学特征层面上, 音高和时长是音节知觉结构的主要维度; 在韵律层面上, 句中位置、韵律词长度等指标比较直观地反映了音节在知觉多维空间中的分布。

PACS 数: 43.70

## Experimental research on the perceptive space of Chinese syllables

ZHOU Xunyi WANG Bei YANG Yufang

(Institute of Psychology, The Chinese Academy of Science Beijing 100101)

Received Jun. 13, 2001

Revised Jun. 4, 2002

**Abstract** The perceptive multi-space structure of Chinese syllables is studied by psychological-physical experiment. The results indicate that  $F_0$  and duration are two main dimensions of the perceptive structure of syllables. Prosodic information such as position of syllable in sentence and number of syllables in a prosodic word is the main factors of syllable distribution in perceptive multi-space of syllables.

## 引言

处于连续语流中的音节, 由于同时负载了各种句法和韵律信息, 其发音模式与音节单念时相比发生了很大的变化, 但听者仍能把它们正确地知觉为某一个音节, 也就是说, 这些能引起不同知觉效应的音节变体在一定程度上是可以相互替代的。研究表明, 处于不同语流环境中的某一音节, 其声学特征有一个动态的变化范围, 在此范围内, 听者都能接受该音节的这种变化, 并获得正确的知觉结果。如果从物理角度对音节的变异性大小作评价, 可以用不同的声学参数作为评价的指标, 从而建立一个根据音节声学参数为维度建构的多维物理空间。如果从知觉角度研究人对音节判断的机制, 也可以设想, 人对音节的知觉判断也是依赖于若干个心理指标, 而这些心理指标分别对应于某个或某几个相关的声学参数, 并同样地建立起一个根据这些心理指标为维度的多维心理空间。本文通过探讨处于语流中的音节在物理空间分布和在心理空间分布的规律, 以期找到汉

语音节的声学特征、韵律线索与音节知觉机制的对应关系。

Singh 等<sup>[1]</sup>通过对英语辅音的考察, 提出了 22 个英语辅音在知觉空间的分布。关于汉语辅音知觉结构的研究, 吕士楠、张家禄等<sup>[2,3]</sup>通过建构汉语 21 个辅音的知觉空间, 研究了汉语辅音的主要知觉特征。这些研究着重于探讨音节内部元音和辅音的知觉结构, 本文则是在音节层面研究知觉结构。此外, 由于我们具体考察的是某个音节在各种不同语句环境中的差异性, 将这些音节从语句中切分出来, 它们在音段内容方面保留了绝大部分的相同特征; 相比之下, 它们在超音段特征方面的变化要显著的多。体现超音段特征变化的主要是音高、时长和音强 3 个特征, 本文将重点对它们进行考察。另外, 句中音节负载着各种语义、语用信息和韵律信息, 根据已有研究, 音节所在韵律词的声调模式、句子语调模式、音节在韵律词中的前后位置、所在韵律词长度等等, 都将影响音节的声学表现和知觉效应。可以归纳为以下变量: 轻重、词中和句中位置等。本文将通过实验

\* 973 国家重点基础研究发展规划资助项目 (G1998030505-3)

探讨这些变量与声学参数之间的关系。

## 1 实验设计

本研究主要采用音节相似度判定的知觉实验,包括 4 个实验:实验 1 为语流中阴平声音节相似度听辨实验;实验 2 为合成阴平声音节相似度听辨实验;实验 3 为语流中去声音节相似度听辨实验;实验 4 为合成去声音节相似度听辨实验。选择语流中的阴平声音节和去声音节作为实验材料进行考察,是因为:阴平声没有音域的展开,其音高模式较为简单;去声具有音域的展开,且与阳平声、上声相比,其音高变化较大。另外,通过高音点和低音点就能比较准确地确定去声音节的音高曲线走向,基频的动态变化范围也有较大的离散度。实验 3 与 4 选用相应的合成音节作为实验材料,实验目的是考察在除去自然语流中句法、语境等因素的影响,仅保留音高、时长、音强 3 个声学参数差异的情况下所得到的实验结果,通过比较,达到确证并精确化实验 1 和实验 3 所得音节知觉空间模型的目的。最后对实验 1 和实验 3 的音节知觉空间,还进一步探讨了韵律、句法信息对音节特征的影响。

## 2 实验材料

实验 1 和 3 所选音节全部来自标准汉语语料库的语句。分别选取音节“gong1”和“ju4”为实验音节。根据音节的声学参数及语境信息(包括词中、句中位置),选出“gong1”和“ju4”各 20 个作为实验 1 和 3 的实验材料,以编号 1~20 来表示。实验 2 和 4 分别选用合成音节“gong1”和“ju4”,通过改变音节的音高、时长和音强,得到 20 个合成音节作为实验材料。为保证实验结果的可比较性,实验 1 和 2 的音节在各声学参数上的离散度比较接近,两组材料总体上分布情况差异不大。实验 1 与实验 2 实验材料的声学参数比较见表 1。实验 3 与实验 4 实验材料的声学参数比较见表 2。

表 1 实验 1 和实验 2 实验材料的声学参数的比较

	来自实验	全距	最小值	最大值	平均值
时长	实验 1	128	91	219	151.45
	实验 2	108	97	205	157.15
平均音高	实验 1	161	95	256	168.15
	实验 2	120	102	222	163.95
音强	实验 1	8.48	68.45	76.93	72.71
	实验 2	7.80	67.30	75.10	71.40

表 2 实验 3 和实验 4 实验材料的声学参数的比较

	来自实验	全距	最小值	最大值	平均值
平均音高	实验 3	90	108	198	149
	实验 4	92	112	204	155
高音点	实验 3	116	118	234	173
	实验 4	124	120	244	177
低音点	实验 3	80	92	172	125
	实验 4	96	98	194	132
音域	实验 3	92	9	101	48
	实验 4	97	1	98	45
时长	实验 3	141	73	214	141
	实验 4	130	93	223	154
音强	实验 3	11.30	64.42	75.72	69.24
	实验 4	11.42	64.11	75.53	69.66

## 3 被试

每个实验都有 12 名被试参加,共 48 名被试。他们是来自北京高校的大学生,年龄在 20 岁左右,身体健康,听力正常,无耳疾,来自北方地区,说标准普通话。

## 4 实验任务

4 个实验的任务都是一样的。将 20 个音节两两配对,得到 190 个音节对。将这 190 个音节对以随机的顺序排列,并依次呈现给被试。被试根据两音节听觉相似度的大小做出判断。3 个判断选择为:“两个音节听起来稍有差别”,“两个音节听起来有较明显差别,能区分开”和“两个音节差别很大,听起来完全不一样”。为避免顺序效应,被试做完 190 个音节对的反应后,将这 190 个音节对的每对音节内的两个音节位置互换,组成另 190 个音节对,并以同样的方式呈现给被试,获得被试的听辨反应。

## 5 实验结果与分析

### 5.1 知觉距离矩阵的获得

采用多维标度(Multidimensional Scaling)方法<sup>[4]</sup>对实验结果做统计分析。以打分制记录被试每次的判断结果:以 0 表示“这两个音节听起来稍有差别”;以 1 表示“这两个音节听起来有较明显差别,能区分开”;以 2 表示“这两个音节差别很大,听起来完全

不一样”。所有 12 名被试对音节对  $ij$  (表示该音节对由音节  $i$  和音节  $j$  组成) 共作出 24 次判断, 取其平均值作为该音节对的判断结果, 以  $P_{ij,k}$  表示。

以  $D_{ij}$  表示音节  $i$  与音节  $j$  的知觉距离:

$$D_{ij} = P_{ij,k}/2.$$

显然,  $0 < D_{ij} < 1$ 。由所有音节的知觉距离得到音节的知觉距离矩阵:

$$\begin{bmatrix} D_{11} & D_{12} & D_{13} & \cdots & D_{1j} & \cdots & D_{1n} \\ D_{21} & D_{22} & D_{23} & \cdots & D_{2j} & \cdots & D_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ D_{i1} & D_{i2} & D_{i3} & \cdots & D_{ij} & \cdots & D_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ D_{n1} & D_{n2} & D_{n3} & \cdots & D_{nj} & \cdots & D_{nn} \end{bmatrix}$$

位于矩阵对角线上的元素  $D_{ii}$  代表同一音节的知觉距离, 显然  $D_{ii} = 0 (1 \leq i \leq 20)$ 。处于矩阵中对称位置的元素  $D_{ij}$  和  $D_{ji}$  代表同一音节对的知觉距离, 即  $D_{ij} = D_{ji}$ 。因此, 该知觉距离矩阵是一个对称矩阵。

## 5.2 声学分析

对所得知觉距离矩阵可以做任意维度的分析。依据实验假设, 被试对音节对相似度的知觉判断取决于相关声学参数的综合差异。如果知觉判断主要取决于  $k$  个声学参数, 那么对知觉距离矩阵做  $m$  维标度所得到的  $k$  维知觉空间模型, 将会与以这  $k$  个声学参数为维度所得到的  $k$  维物理空间结构有很好的 consistency。

### 5.2.1 实验 1 与实验 2

对实验 1 结果作二维标度分析, 计算各维度 (维度  $x$  和  $y$ ) 与各声学参数的相关系数, 见表 3。

表 3 二维标度各声学参数与维度  $x, y$  的相关系数

	$x$	$y$
平均音高	0.88***	0.03
时长	0.07	0.79***
音强	0.43	0.13

注: \* 表示  $P < 0.05$ , \*\* 表示  $P < 0.01$ , \*\*\* 表示  $P < 0.001$

从表 3 的相关性分析看, 平均音高与维度  $x$ 、时长与维度  $y$  都达到了显著的高相关。继续做三维标度, 同样计算维度  $x, y, z$  与各声学参数的相关系数, 见表 4。

表 4 三维标度各声学参数与维度  $x, y$  的相关系数

	平均音高	时长	音强
$x$	0.945***	0.069	0.438
$y$	0.058	0.786***	0.153
$z$	0.069	0.037	0.079

表 4 中, 三维分析结果与二维情况下有很好的 consistency。可见, 平均音高和时长是影响音节“gong1”知觉结构的最主要的两个物理参数。各声学参数与维度  $z$  的分析表明, 平均音高、时长和音强与维度  $z$  都不存在相关。由此认为, 在分离出平均音高和时长这两个知觉结构的声学参数以后, 维度  $z$  代表了其他因素对音节知觉距离的整合影响, 而这些因素对音节知觉距离的影响要远小于前两个因素。

如果将知觉结构压缩至一维表示, 并对其作相关性分析, 此时平均音高与维度显著相关 ( $r = 0.839^{**}$ ), 而时长呈现低相关。可见, 当将知觉结构压缩至一维时, 平均音高成为最重要的知觉判断的物理因素。

实验 2 在仅改变合成音节“gong1”的平均音高、时长、音强的情况下做多维标度分析, 减少了影响音节差异的自变量。对这些合成音节同样做二维和三维分析, 维度与声学参数的相关见表 5。其中, 只列出相关显著的结果。

表 5 合成音节维度与声学参数相关系数

	二维标度	三维标度
$x$ 维与平均音高	0.955**	0.825**
$y$ 维与时长	0.976**	0.751**

从表 5 中可见, 实验 2 中用合成音节得到的结果与实验 1 基本是一致的: 二维和三维标度中, 维度  $x$  与平均音高显著相关; 维度  $y$  与时长显著相关。可见, 音节的知觉空间与物理空间有很好的对应关系。影响音节“gong1”的知觉距离的两个声学参数是平均音高和时长。

### 5.2.2 实验 3 与实验 4

去声音节的变化较阴平声音节复杂, 影响其知觉差异的声学参数也更丰富。我们对实验 3 的结果作三维标度分析, 计算维度  $x, y$  和维度  $z$  与各声学参数的相关系数, 见表 6。

从表 6 的相关性分析看, 在维度  $x$  上, 音域与维度  $x$  相关最显著, 高音点次之; 时长与维度  $y$  显著相关; 在维度  $z$  上, 低音点、平均音高与之有很高的相关性。

实验 4 仅改变合成音节“ju4”的高音点、低音点、时长和音强(高音点和低音点改变,音域和平均音高也发生了相应的改变)。维度为 1 时,该维度与声学参数的相关系数见表 7。

从表 7 中可见,合成语音的结果与实验 3 有很好的-致性:与维度相关最高的两个声学参数是音域和高音点。对实验 4 进行二维分析,两个维度与声学参数的相关系数见表 8。

表 8 中,对合成的“ju4”进行二维分析,在维度  $x$  上,仍然是高音点和音域与该维度的相关最高;维度  $y$  上,时长与该维度的相关最高。

对实验 4 进行三维分析,三个维度与声学参数的相关系数见表 9。

表 9 中三维分析的结果与二维的一致:高音点和音域与维度  $x$  显著相关,时长与维度  $y$  显著相关。对于新增加的维度  $z$ ,平均音高和低音点与其显著相关。

表 6 三维标度各声学参数与维度的相关系数

	$x$	$y$	$z$
平均音高	0.375*	0.069	0.533**
高音点	0.544**	0.121	0.385*
低音点	0.053	0.063	0.568***
音域	0.751***	0.201	0.011
时长	0.121	0.723***	0.09
音强	0.000	0.2	0.116

表 7 合成音节“ju4”一维相关分析

	相关系数	$P$ 值
平均音高	0.28	0.08
高音点	0.56	**
低音点	0.14	0.38
音域	0.81	**
时长	0.21	0.19
音强	0.28	0.08

表 8 合成音节“ju4”二维相关分析

	$x$	$y$
平均音高	0.14	0.1
高音点	0.47**	0.07
低音点	0.24	0.19
音域	0.81**	0.09
时长	0.31	0.61**
音强	0.27	0.27

表 9 合成音节“ju4”三维相关分析

	$x$	$y$	$z$
平均音高	0.17	0.09	0.69**
高音点	0.51**	0.06	0.42**
低音点	0.21	0.18	0.63**
音域	0.83**	0.12	0.08
时长	0.31	0.62**	0.18
音强	0.26	0.24	0.09

从实验 3 和 4 的结果可以看出,对于去声音节,音高与音节知觉空间的关系比较复杂,不是平均音高一个变量所能描述的。这也正是汉语音节的声调特点,不仅要考虑音高的值,而且要考虑音高曲线的变化。“ju4”的知觉主要取决于 3 个维度:维度 1 对应音域和高音点,维度 2 对应时长,维度 3 对应低音点和平均音高。其中,音域与维度 1 的相关最高,高音点次之。平均音高是高音点与低音点的平均值,而音域是高音点和低音点的差值,可以认为它们是体现高音点与低音点变化的综合变量。但是从实验结果来看,高音点、低音点和平均音高、音域对不同维度相关性的贡献是不同的。高音点与音域的关系更为密切,低音点与平均音高的关系更为密切,前两者构成知觉空间的第一维度,而后者构成知觉空间的第二维度。这也说明,实际语流中高音线和低音线的变化并不完全是一致的。

通过对阴平声音节“gong1”和去声音节“ju4”的知觉空间与物理空间的比较可以看出,影响音节知觉空间的声学参数是音高和时长,音强起很弱的作用或者不起作用。

### 5.3 韵律分析

韵律分析主要考察音节所处句中、词中位置以及音节所在韵律词长度等韵律线索对音节知觉空间的影响。首先对实验 1 中 20 个阴平声音节“gong1”组成的知觉空间做关于韵律信息的分析。表 10 是按音节在句中所处位置的简单分类。

表 10 实验 1 音节按句中位置的划分

所处位置	句首部分	句中部分	句末部分
音节归类	10,12,18	1,3,5,6,8,11, 13,14,15,16,20	2,4,7,9,17,19

实验 1 中“gong1”的知觉空间平面图见图 1。

从图 1 中可以发现,代表音节 10,12,20 的点位于切分线  $x = x_1$  左侧,其中 10 和 12 是处于句首部分的音节;代表音节 2,4,7,9,17,19 的点位于切分线

$x = x_2$  右侧，正好是处于句末部分的所有音节；其余点在两切分线之间。可见，维度  $x$  与音节 gong1 所在句中的位置有很大相关性，而这一维度在声学参数上与平均音高有很高的相关性。对此的解释是，句中位置在很大程度上是与句子的语调模式以及节奏和韵律结构相联系的，渐降的语调模式很大程度上导致了句中位置的这种影响作用。音节 14 的  $x$  维度坐标也说明了这一点。它紧跟在一个句中大停顿之后，正好位于音高重置的位置，从声学表现来说，音高值变大；从知觉结构图来看，则其  $x$  维度坐标左移，靠近  $x = x_1$  直线。以上的变化关系说明，即使将音节从语境中分离出来，其所包含的句中位置信息仍然能间接地被听者知觉出来，即句中位置的差异确实导致了音节的知觉差异。

除音节所处句中位置外，重音也是影响音节在  $x$  维度分布的一个因素。如音节 20 处于语句重音位置上，引起其在知觉距离图  $x$  维度上左移。从声学表现来说，是因为重音的影响使音节“gong1”的音高值增加。

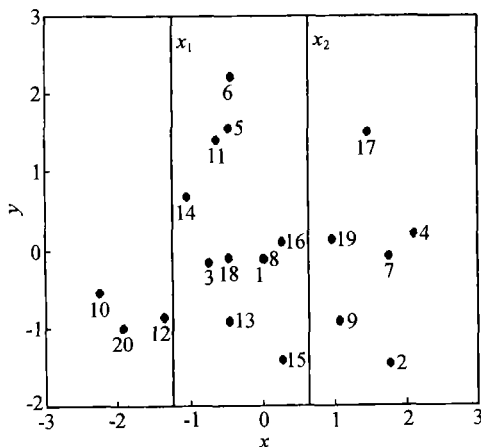


图 1 实验 1 二维标度分析平面坐标图

表 11 实验 1 音节按所在韵律词长度的划分

所在韵律词长度	1	2	3	≥ 4
音节归类	6	3,4,7,8 11,14,17,19	10,18	1,2,5,9,12 13,15,16,20

如果按音节所在韵律词长度做分类(见表 11)，则韵律词长度与  $y$  维度有较好的相关性(参见图 1)，随着韵律词长度的缩短， $y$  维度坐标值变大。由声学表现与知觉结构的分析结果，该维度与音节时长有一致性，可见，音节所在韵律词长度的信息主要通过时长在知觉距离上表现出来。此外，还有其他一些因素同样对  $y$  维度的知觉距离做出贡献，如音节在韵律词、韵律短语中的位置、重音，并影响了其中某些音节在

$y$  维度的分布。例如音节 5 代表的 gong1 处于四字韵律词(“干部职工”)的末尾位置，紧跟其后出现的是一个明显的停顿。由于处于词末、短语末的音节比相应起首位置上的时长延长，导致该音节时长偏大，在  $y$  维度坐标也偏大。

对实验 3 中音节“jiu4”的知觉空间分析发现，句中位置、韵律词长度仍是影响 ju4 知觉结构的两个主要变量，所不同的是，句中位置的影响主要体现在维度  $z$  而不是维度  $x$  上，即与由声学特征平均音高(低音点)所决定的维度相对应。对此的解释与实验 1 一致，即句中位置很大程度上反映了语句渐降语调模式的变化。对于去声音节来说，即使它本身的高音点、低音点分别受所在韵律词的声调模式、重音、节奏、停顿等各种韵律现象的影响，但总体上来说，其平均音高值还是能大致体现出与音节在句中位置的变化规律。也可以认为，句中位置或者语调是决定音节平均音高的重要因素。表 12 和表 13 分别列出了音节“ju4”按句中位置和所在韵律词长度的划分。图 2 则显示了实验 3 在三维标度下维度  $x$  和  $z$  的平面投影图。

表 12 实验 3 音节按句中位置的划分

所处位置	状语前 置部分	靠近句 首部分	句中部分	靠近句 末部分
音节归类	2,4,7,8, 11,12,17	9,16 9,16	3,6,10,13, 14,18,19,20	1,5,15

表 13 音节按韵律词长度分类结果

所在韵律词长度	1	2	3	≥ 4
音节归类	2,4,8	1,6,7,10,12,14, 15,16,19,20	3,5,11,17	9,13,18

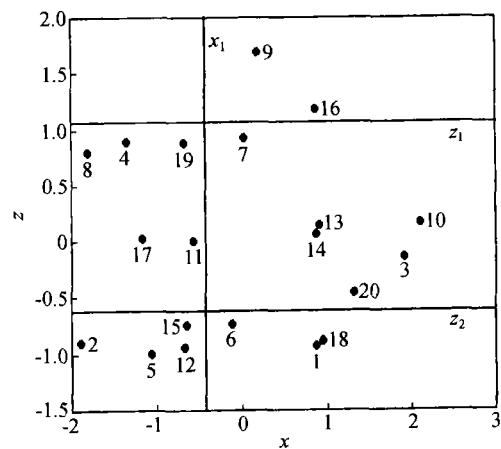


图 2 实验 3 三维标度  $x-z$  平面投影图

值得注意的是,处于前置状语部分的音节,在  $x$  维度均集中在左侧 ( $x < x_1$ )。这是因为,音节“ju4(据)”在这些插入语中都是作为一个介宾短语的介词出现,可能正是由于虚词词性的原因,在语义上不被突出,导致其音域(对应于维度 1)变窄,从而使它们在知觉  $x$  维度上表现为坐标值偏小。而且,作为虚词的音节 2,4,8,11,17(在句中均为介词“据”)在  $y$  维度的位置也相对于同等词长度的音节偏低,其相应的声学表现为时长变小。可见,词性对音节在知觉结构中的分布也起到了一定的作用。

与  $x$  维度的知觉差异对应的韵律信息比较复杂,如:音节所在韵律词的声调模式、句重音、短语重音位置等。韵律词内搭配和韵律词前后环境相似的音节,在该维度上也较为接近。

## 6 讨论与总结

本文采用多维标度方法,对汉语音节的知觉特征、知觉空间进行了实验探讨,结果表明:

(1) 平均音高和时长是影响阴平声音节“gong1”知觉结构的两个主要物理参数。而音域(高音点)、时长、平均音高(低音点)构成了影响去声音节“ju4”知觉结构的 3 个主要物理参数。影响知觉结构的主要声学相关物均为音高和时长,其中又以音高的变化最为重要。音强不是决定音节知觉距离的主要声学参数。

(2) 从韵律信息与音节知觉多维空间的分析来看,音节在句中的位置是影响音节知觉结构的重要因素之一。从韵律表现角度来看,句中位置与音节知觉结构的这种对应关系,其实正是节奏和语调信息影响音节知觉结构的反映。此外,语句重音对音节在

该维度的分布也起到了一定作用。

(3) 音节所在韵律词的长度是影响音节知觉结构的重要因素。音节在韵律词中的位置、所在韵律词的重音模式也对音节在该维度的分布起作用。此外,音节所代表字的词性也表现出了影响作用,如果是被弱化的虚词,该音节在知觉空间的位置就将产生较大变化。

综上所述,汉语音节的知觉多维空间与由若干典型声学参数为维度构成的物理空间有很好的对应关系。有关音节知觉的客观表现可以归纳为相关的一组声学参数,或者是相关的一组句法和韵律信息符号,以此建构的音节知觉模型基本上可以反映音节在真实知觉多维空间中的分布趋势。本研究带来的一个启示,就是根据这一原则拟合的音节知觉模型,对汉语语音合成系统的多样本音库设计,对样本选择算法提供了理论上的依据。此外,本研究是以大规模语料库为研究对象的,选用基于大规模容量的标准语料库的音节作为实验材料,所获得的研究结果更具普遍意义,也更能反映真实言语交流中的客观规律。

## 参 考 文 献

- 1 Singh S, Woods, Becker D R, G M. Perceptual structure of 22 prevocalic English consonants. *Journal Acoustic of Social America*, 1972; 52: 1968—1713
- 2 吕士楠,张家喆,齐士钤. 汉语辅音知觉混淆研究中的多维标度方法. *声学学报*, 1981; 7(6): 63—68
- 3 张家喆,齐士钤,吕士楠. 汉语辅音知觉结构初探. *心理学报*, 1981(1): 76—85
- 4 Kruscal J B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964; 29: 1—26