

发音参数合成元音的初步研究

语言研究所 祖漪清

语音合成器，是人工研制成的，由一些参数控制的、能够发出与自然语音相类似的声音的装置。

语音合成的方法很多，从合成器的控制参数方面来看，大致可分为终端合成与发音参数合成两类。终端合成，是将声道看作一个系统，不考虑其内容，只研究其输出。可用输出信号的频谱形状、基频、共振峰频率等反映系统特性的物理参数作为合成器的控制参数。发音参数合成则是研究声道中声波传输的细节，以反映声道结构的生理参数作为合成器的控制参数。

研究语音合成的目的之一是探索语音产生的机理。由发音参数合成语音，可从另一角度验证语音分析的结果，它是语音研究不可缺少的手段。同时，对于语言教学、语言声学等应用领域，发音参数合成也具有重要的意义。

本研究参考现有的发音器官资料〔1, 2〕及国外关于这方面的研究，建立插值模型，对八个元音做了语音合成的初步尝试。

一、发音参数的插值模型

用发音模型合成语音的目的，是为了取得声道生理结构和与之对应的语音之间的联系，因此所选取的模型应当具有能够基本

反映声道生理结构的参数。同时，为了得到一些合成语音，还要求模型的这些参数便于定量计算。

1. 基本假设与模型结构

发音参数的插值模型参考了语言所语音室的X光照片及其有关研究〔3〕。真实声道在发音过程中变化十分复杂，而作为模型，则要求在本能反映真实情况的条件下尽可能得到简化。通过对资料的分析，假设：(1)声道中硬腭、软腭后咽壁不随发音动作而变化；(2)唇与齿的开口动作是一致的，因此，有了唇的位置，就能决定齿的位置；(3)声门位置固定，不随不同的发音条件而上、下移动；(4)上唇与下唇边缘在同一垂直线上。

这样，声道中可以变化的部分只有上、下唇，以及舌体。图1给出了示意图。直角坐标系的原点位于经过上、下唇边缘的垂线上。声道中固定不动的是上腭及后咽壁。硬腭由一段圆弧表示，与这段圆弧相连接的一段曲线表示软腭，后咽壁由一直线代表。斜率为-5。图中的参数R既代表上腭的那段弧线的半径，同时也是反映声道尺寸的一个参量。（图1见下页）

通过观察发音器官资料，我们知道，不同的发音人或同一个发音人在不同场合发出的同一个音时，舌位基本形状不尽相同。但舌位的高、低、前、后的基本趋势是相同的。这就是说，对应于一个元音，舌位有一个基本形状，这个基本形状又有一个范围，

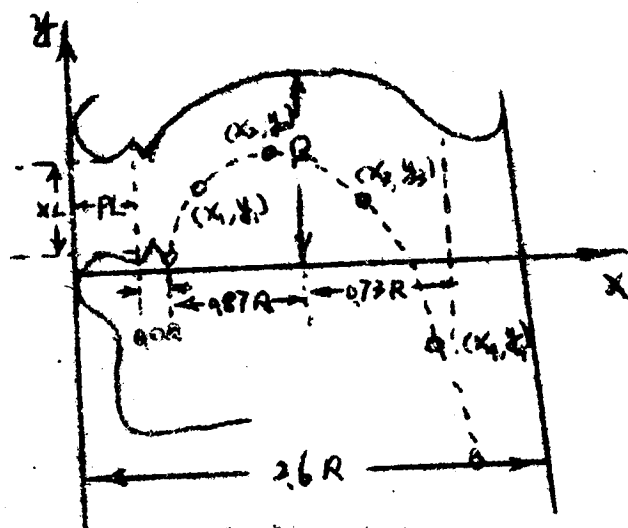


图1 发音参数模型示意图

只要舌位的变化不出这个范围，发出的音听起来就不会有区别意义。

由于上述原因，我们对舌位曲线的精度不作苛刻要求。参考X光照片，选取声道中五个位置的坐标作为模型参数，它们是：

- (1)、(2)：唇的伸展长度 PL 与唇的开口大小 XL ；
- (3)、(4)：舌体前部在坐标系中的位置 x_1 、 y_1 ；
- (5)、(6)：舌体中部在坐标系中的位置 x_2 、 y_2 ；
- (7)、(8)：舌体后部在坐标系中的位置 x_3 、 y_3 ；
- (9)、(10)：舌咽部在坐标系中的位置 x_4 、 y_4 。

2. 发音参数插值

如图2所示, 第3, 5, 7, 10点(即1, 2, 3, 4点为已知点。其它点通过插值确定, 插值的方法是在线性基础上加以修正。第1点由下齿位置决定:

$$x'_1 = 1.4PL \quad (1a)$$

$$y'_1 = 0.14R - xL \quad (1b)$$

第11点设为声门处:

$$x'_{11} = \begin{cases} 1.65R + 1.4PL & (\text{圆唇}) \\ 1.6R + 1.4PL & (\text{不圆唇}) \end{cases} \quad (2a)$$

$$y'_{11} = -0.11R \quad (2b)$$

其余各点见下列各式:

$$x'_q = \begin{cases} \frac{1}{2}(x_2 + x_3) & y_3 < y_2 \\ x_2 + |x_2 - x_3| & y_3 \geq y_2 \end{cases} \quad (3a)$$

$$y'_q = \frac{1}{2}(y_3 + y_2) \quad (3b)$$

$$x'_i = \frac{1}{2}(x'_{i-1} + x'_{i+1}) \quad (4a)$$

$$y'_i = \begin{cases} \frac{1}{2}(y'_{i-1} + y'_{i+1}) + D_{ii} & D_{ii} \leq w_i \\ \frac{1}{2}(y'_{i-1} + y'_{i+1}) & D_{ii} > w_i \end{cases} \quad (4b)$$

其中

$$D_{1i} = 0.055 [(y'_{2i+1} - y'_{2i-1})^2 + (x'_{2i+1} + x'_{2i-1})^2]^{1/2} \quad (5)$$

$$W_i = \frac{1}{2} |y'_{2i+1} - y'_{2i-1}| \quad (6)$$

$i = 2, 4, 6, 8$

经插值后的舌位节点共有 11 个，其中只有 3', 5', 7', 10' 点为原始输入参数。



图2 插值位置示意图

无上标的数字表示输入参数顺序。带有上标的数字表示插值后编排的顺序。

3. 舌位曲线的形成

将插值后形成的舌位节点两两直线连接，这些相互连接为折线段就代表舌位曲线。

4. 声道截面分割线

声道截面分割线同一些文献 [4] 中的矢状线意义相同。

如图 3 所示，上唇、上颚、后咽壁组成的固定曲线，以及下唇和舌位曲线，都等分为 10 个节点，将两组对应的节点连接起来，形成一组声道的分割线，以此获得声道面积函数。

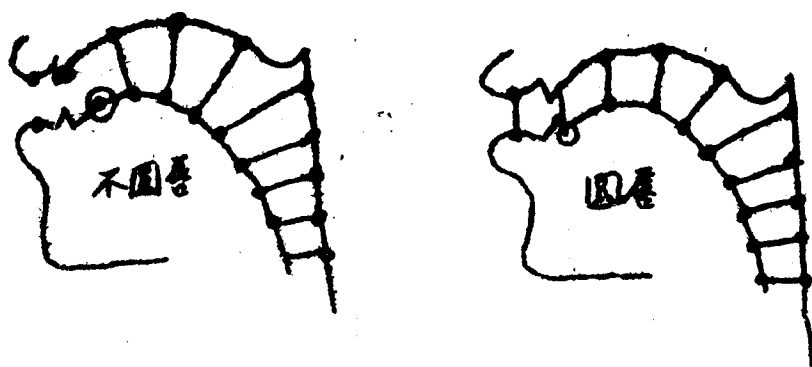


图3 声道分割线节点位置示意图

唇部至声门为曲线的总长度。声道上壁的10个节点位置基本保持不变。下唇与舌位曲线对于每一个音，都要重新确定位置。

5. 声道分割线与声道面积函数的关系

声道是一个三维形体，现有的实验手段还无力获得反映声道三维空间的材料，现仅能以X光照片作为参考。因此从声道二维图形求出分割线段到声道面积函数，只能经过一些实验，由经验得到。本实验得到的分割线段长度D与面积函数A的关系为：

$$A = D^{2.5} \quad \text{不圆唇} \quad (7a)$$

$$A = \begin{cases} D^{2.5} & (1-9\text{点}) \\ D^2 & (10\text{点}) \end{cases} \quad \text{圆唇} \quad (7b)$$

二、发音参数合成元音

发音参数合成结构如图4所示。

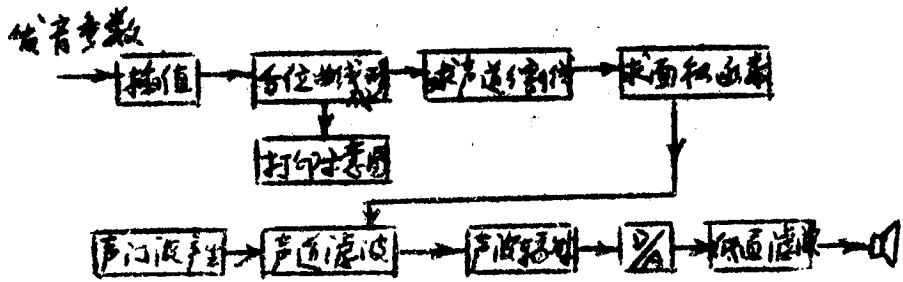


图4 发音参数合成过程图

由发音参数的输入到获得声道面积函数的方法前面已作了较详细的叙述，关于由面积函数合成元音，^[5]作以下几点说明：

1. 声门波^[6]

在一个周期内，声门波形为：

$$g(t) = \begin{cases} \frac{1}{2} \left(1 - \cos \frac{\pi t}{W_1} \right) & 0 < t \leq W_1 \\ \cos \left(\frac{\pi(t - W_1)}{2W_2} \right) & W_1 < t \leq W_2 \\ 0 & \text{其他} \end{cases} \quad (8)$$

如果将声门波周期归一化，则取 $W_1 = 0.3$ ， $W_2 = 0.2$ 。

2. 声道滤波^[6]

将声道看作一级联的无损声管，则由声道面积函数求声道滤波器系数由下列递推关系得到：

$$D_0(z) = 1$$

$$D_k(z) = D_{k-1}(z) + r_k^{-k} D_{k-1}(z^{-1}) \quad k=1, 2, \dots, N$$

$$D(z) = D_N(z) \quad (9)$$

其中 $r_k = (A_{k+1} - A_k) / (A_{k+1} + A_k) \quad k=1, 2, \dots, N \quad (10)$

面积函数 A_k 共有 $N+1$ 个。本实验中 $N=10$ 。

$D(z)$ 是声道转移函数的分母多项式，它具有如下形式：

$$D(z) = 1 + \sum_{k=1}^N \alpha_k z^{-k} \quad (11)$$

其中 α_k 为特定系数。

声道转移函数由下式决定（不考虑零点，并将分子取为 1）。

$$H(z) = \frac{1}{D(z)} = \frac{1}{1 + \sum_{i=1}^N \alpha_i z^{-i}} \quad (12)$$

3. 声波辐射 [7]

在语言声的频率范围内，嘴唇处的辐射可等效为无限障板上的活塞辐射。辐射阻的拉氏变换为：

$$Z_L(s) = \frac{5R_r L_r}{R_r + sL_r} \quad (13)$$

其中 $R_r = \frac{12\rho}{9\pi^2}$, $L_r = \frac{8a}{3\pi c}$, ρ 为声速。

4. 合成元音的基频与幅度变化

根据音调的自然特性，[8]基频和幅度都有弯头、降尾。它们

的变化设为正弦、余弦规律。

三、元音合成的实验结果

本实验对八个元音选取了发音参数，并合成了语音。

表1列出了八个元音的发音参数。

| 元音 | 嘴唇 | | 舌体前部 | | 舌体中部 | | 舌体后部 | | 咽部 | |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | x_1 | y_1 | x_2 | y_2 | x_3 | y_3 | x_4 | y_4 | x_5 | y_5 |
| i | 1.5 | 1.2 | 2.6 | 2.3 | 3.6 | 3.0 | 5.8 | 1.5 | 6.8 | 0.0 |
| e | 1.5 | 1.7 | 2.6 | 1.6 | 4.2 | 2.8 | 6.0 | 2.7 | 7.5 | -1.0 |
| ɛ | 1.5 | 2.2 | 3.1 | 0.8 | 5.2 | 2.2 | 6.3 | 1.5 | 8.0 | -0.9 |
| ɜ | 1.5 | 2.3 | 3.2 | 0.0 | 5.8 | 1.6 | 7.3 | 1.4 | 8.4 | -0.6 |
| ɔ | 2.0 | 1.5 | 2.6 | 1.2 | 5.8 | 2.3 | 7.4 | 1.2 | 9.2 | 0.0 |
| o | 2.0 | 0.8 | 3.2 | 2.0 | 5.8 | 3.0 | 7.0 | 2.5 | 9.2 | 0.0 |
| u | 2.0 | 0.5 | 3.0 | 2.1 | 5.8 | 3.1 | 7.2 | 2.4 | 9.1 | -1.2 |
| ɥ | 2.0 | 1.2 | 3.1 | 2.3 | 4.1 | 3.0 | 6.3 | 1.5 | 7.3 | 0.0 |

表1 发音参数表 (单位: cm)

表示上颚的圆弧半径 R 取 4 (cm)，这个值与真实声道尺寸接近，也同表1中的参数相匹配。

计算出的八组声道面积函数列于表3。其中第一点代表声门，第十点代表嘴唇。

合成元音共 70 帧，每帧 64 个点，采样频率为 10KHZ。

基频 F_0 为 200HZ，相对幅度 Δm 取 100000。

合成的语音信号经过 D/A 转换后，通过一个低通滤波器，截

止频率为4.7KHZ.

声门 → v

| A | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|------|------|------|------|------|------|------|------|------|------|
| i | 1.08 | 2.04 | 3.66 | 6.02 | 8.11 | 5.87 | 1.16 | 0.41 | 0.50 | 1.57 |
| e | 1.08 | 1.71 | 2.62 | 4.20 | 3.80 | 1.54 | 1.25 | 1.75 | 1.78 | 3.80 |
| ɛ | 1.08 | 1.05 | 1.24 | 2.85 | 5.10 | 5.45 | 6.54 | 7.40 | 5.45 | 7.10 |
| ɜ | 1.08 | 0.67 | 0.54 | 1.03 | 2.10 | 8.10 | 14.0 | 14.9 | 9.20 | 8.20 |
| ɔ | 1.08 | 0.46 | 0.21 | 0.34 | 3.19 | 6.59 | 5.31 | 5.17 | 5.52 | 2.25 |
| o | 1.08 | 0.42 | 0.11 | 0.17 | 1.13 | 1.85 | 1.82 | 1.51 | 1.93 | 0.64 |
| u | 1.08 | 0.50 | 0.18 | 0.99 | 1.33 | 1.27 | 1.37 | 1.04 | 1.05 | 0.25 |
| ʏ | 1.08 | 2.04 | 3.66 | 6.02 | 8.11 | 5.87 | 1.16 | 0.41 | 3.70 | 1.44 |

表2 面积函数表 (单位: cm²)

图5至图9为合成元音/i, a, u/的发音器官示意图和波形图。

四、实验结果的讨论

八个合成元音/i, e, ɛ, a, ɔ, o, u, ʏ/基本上与所设定的语音近似, 并与现有材料所示的发音器官结构有着基本正确的对应关系。但是合成元音音色还不是十分理想, 分析起来, 影响合成元音音色的主要原因有如下几个方面:

1. 模型本身不完善

由于实验手段的限制, 发音模型在一定程度上脱离了真实声道, 将其简单化了。选取的发音参数还不能很好地适用于所有元

音。另一方面，声管模型所作的线性近似也对合成造成了影响。

2. 辐射阻对合成元音的影响

对于八个元音，辐射开口等效半径均取为 1 cm，这对有些元音（如元音 / u /）不能说是合适的，这些元音的频谱在高频区产生了不合理的分配。

3. 实验者本身听力对合成元音音色的影响

实验者对有些元音的正确发音方法不够熟悉，加上实验过程中的听力疲劳，从主客上影响了合成语音音色。

如果进一步改进发音模型，抓住发音器官的关键部位，用发音参数合成元音一定会取得更佳效果。

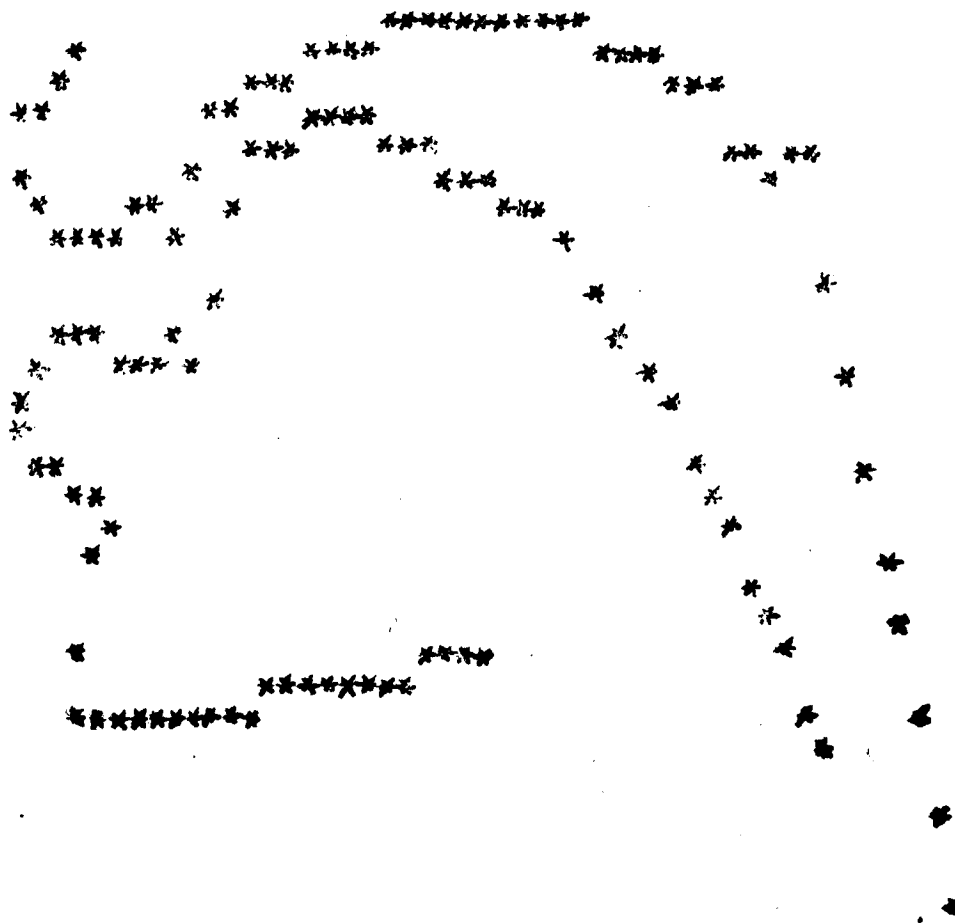


图5 元音 [ɪ] 的发音器官示意图

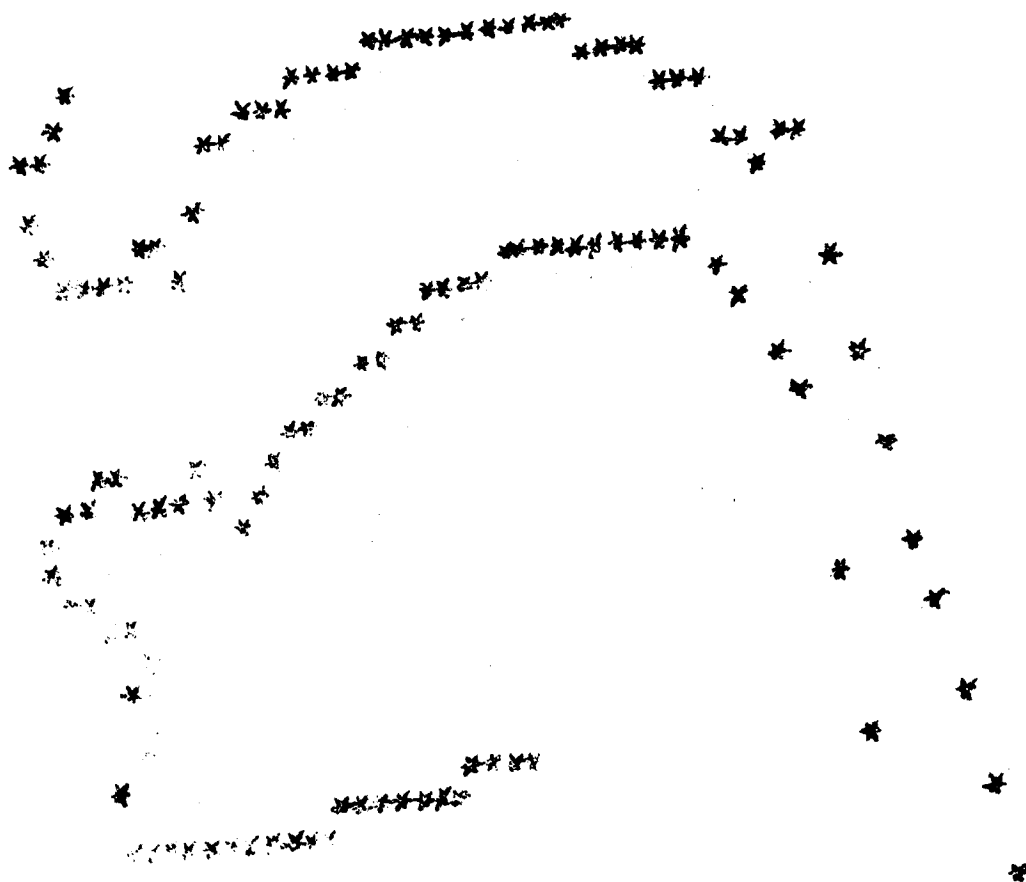


图6 元音 /a/ 的发音器官示意图



图7 合成元音 /a/ 的波形扩展图

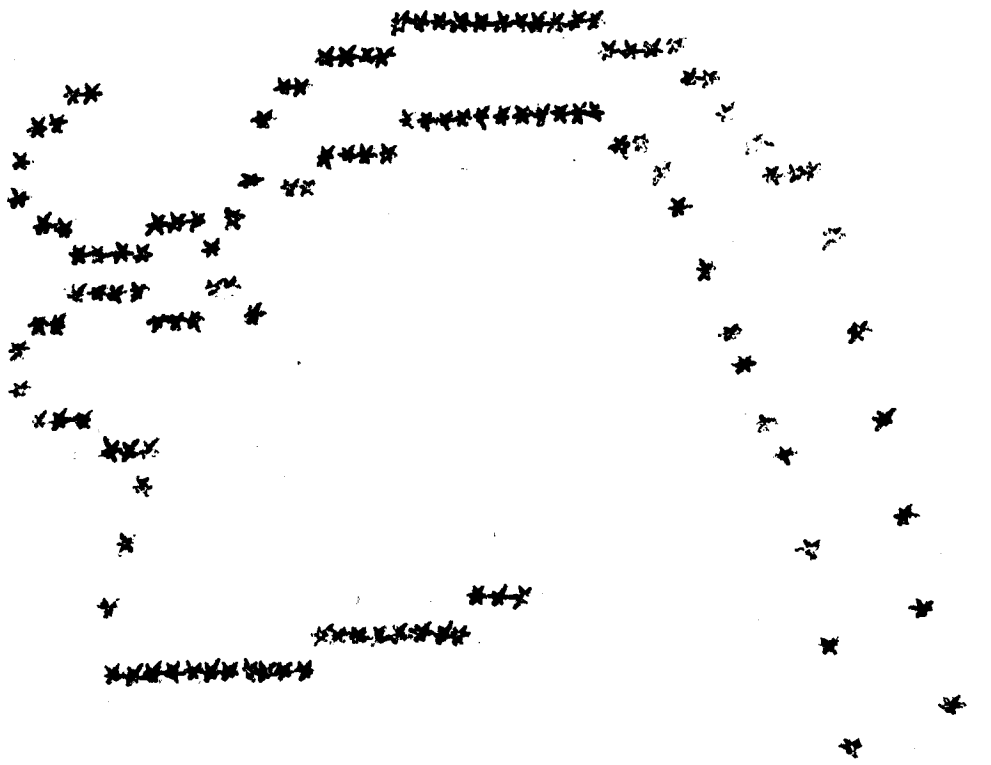


图8 元音[u]的发音器官位置图



图9 合成元音[u]的波形图

参 考 文 献

- [1] 周殿福、吴宗济, 《普通话发音图谱》(1963)
- [2] 鲍怀翘, “普通话单元音分类的生理解释”《中国语文》, 2, 117-127 (1984)
- [3] 吴宗济, “普通话元音和辅音的频谱分析及共振峰的测量”, 《声学学报》, 1, 1, 33-40 (1964)
- [4] Wood, S., (1982) “X-ray and model studies of vowel articulation”, Working Papers 23, 1982, Lund University.
- [5] G. Fant, “Acoustics Theory of Speech Production.”
- [6] L. R. Rabiner & R. W. Schafer, “Digital Processing of Speech Signal.” (1978)
- [7] J. L. Flanagan, “Speech Analysis, Synthesis, and Perception” (1972)
- [8] 林茂灿, “音高显示器与普通话声调音高特性”, 《声学学报》, 2, 1, 8-15 (1965)