

普通话二合元音的动态特性

杨顺安 曹剑芬

摘 要

复合元音是一种有动态变化的元音，分析它们的动态特性，无论是对研究语音产生的理论，还是对语音识别和语音合成，都是很有必要的。

一般认为，普通话有九个二合元音，它们以其动态特性而有别于单元音。本文根据语图分析测得的数据，采取时间和频率归一化的方法，对音长和共振峰频率数据进行了处理，获得了表征各二合元音动态特性的曲线和有关参数。

初步分析表明，二合元音可以大体分为三个部分：起始段、过渡段和收尾段；平均说，起始段较短，过渡段较长，而收尾段与过渡段时长间有近似的反比关系；与几种外语相较，普通话二合元音的动态特性又有自己的特点。文中最后归纳出可用于二合元音合成的两种基本模式。

一 引言

对于复合元音的解释和定义，在历来中外语言学论著中，是不尽相同的。这些差异往往是由于着眼点不同而产生的：有的着眼于生理特性，有的着眼于声学特性和感知特性，有的则从语言学的角度来下定义。尽管这样，基本认识还是一致的，那就是：复合元音是在一个音节内的元音滑移；发音时，舌、唇等发音器官的部位在不断变化着，声道（即共鸣腔体）形状也随之变化；所发出的音，听起来音色也是不断变化着的。总之，复合元音是一种有动态（dynamic）变化的元音。

早先，Lehiste等^①，Holbrook等^②和Gay^③曾用语图仪对美国英语的几个二合元音（diphthong）进行过测量和分析。Lehiste的定义说：“二合元音是一个包含有两个目标（target）位置的元音音节核”，它与滑音（glide）之间的区别，在于后者只有一个目标位置。他们将二合元音分为三段：起始（onset）段、过渡（transition）段和收尾（offset）段。在起始段和收尾段，其音色接近或达到目标位置。从这些文章列举的语图和测量数据来看，英语二合元音中，起始段和收尾段大多是长而平稳，发得比较“到家”。Borzzone对14个西班牙语二合元音的语图测量也得到类似的结果。^④

关于普通话复合元音，从60年代初就进行过初步的声学分析^{⑤⑥}，得到了复合元音的动程图（图1）。在该图中，以元音的第一共振峰（ F_1 ）为纵座标，以第二共振峰（ F_2 ）为横座标，

组成了元音的声学空间。从图中可以明显看出复合元音的动程变化和起讫位置。此

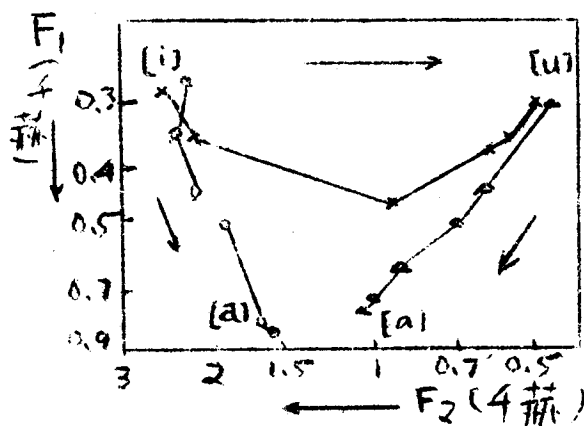


图1 普通话复合元音动程图

后，又陆续报

告了复合元音的一些研究结果。^⑦

作为上述研究的继续，我们对男女四个发音人所发的九个二合元音的语图数据进行了较仔细的处理和分析，目的在于进一步了解二合元音动态特性的细节，归纳出一些对语音研究和对语音合成和识别有益的规律。

二 测量和数据处理

四位成年发音人（男女各半），在消声室中以中等音量和速度发音。被分析的二合元音取自普通话自然音节中的零声母字，包括四个前响二合元音（ ai 、 ao 、 ei 、 ou ）和五个后响二

合元音 (ia、ie、ua、uo、ue)，一般用阴平字。录音后，用语图仪进行分析测量。由于本试验的重点是考察音色的动态变化，所以只对反映音色的第一共振峰 (F_1) 和第二共振峰 (F_2) 做测量和分析。从宽带谱图中，每50毫秒 (必要时每25毫秒) 测量一组 F_1 和 F_2 的频率值，共得 $9 \times 2 \times 4 = 72$ 组二合元音的原始数据。

由于不同发音人在发同一音时，无论音长和音色都不尽相同，为了得到比较一般性的结果，对原始数据采取时间归一化、频率归一化和平均的方法进行了处理。所谓归一化 (normalization) 就是将不同人次的某一组参数，都统一地定为一个单位。以下以 /ai/ 的数据为例，将处理过程作一简要说明。

(1) 根据 /ai/ 的 F_1 的数据可画出图2所出的四条 F_1-t 曲线。

(2) 时间归一化。将各曲线时长都定为一个单位，各测量点的归一化时间 (t) 等于该时刻比音长。这样作出时间归一后的 /ai/ 的 F_1-t 曲线 (图3)。

(3) 插值。在归一时间 $t=0, 0.1, 0.2, \dots, 0.9, 1.0$ 11 点处，就各曲线数据进行抛物线插值 (如图3中小实圆点)。

(4) 平均得时归一平均曲线。就同一归一时刻的四个频率求出算术平均值 \bar{F}_1 ，作出时归一平均曲线 \bar{F}_1-t (如图3中粗实线

所示)。

(5) 频率归一。作出时频归一平均曲线。

根据时归一平均值，按下式计算归一化频率值 f (参看图 4)

$$f = \frac{\bar{F} - \bar{F}_B}{\bar{F}_E - \bar{F}_B} = \frac{\bar{F} - \bar{F}_B}{\bar{F}_d}$$

式中 \bar{F} 是时归一平均值， \bar{F}_B 是起始点 ($t=0$) 的 \bar{F} 值， \bar{F}_E 是末尾点 ($t=1$) 的 \bar{F} 值， \bar{F}_d 是该共振峰频率的变化范围。

图 5 是如此作出的 /a i/ 的 F_1 的时频归一平均曲线。由图看到，这条曲线是上升的，而实际上，/a i/ 的 F_1 是先大后小的。这是因为此时的 $\bar{F}_B > \bar{F}_E$ ， $\bar{F}_E - \bar{F}_B$ 和 $\bar{F} - \bar{F}_B$ 总为负，而且 $t=0$ 时 $\bar{F} - \bar{F}_B = 0$ 、 $f = 0$ ，而 $t=1$ 时， $\bar{F} - \bar{F}_B = \bar{F}_d$ 、 $f = 1$ ，所以曲线上升。即不论 F 的实际值是小而大，或大而小，归一曲线总是上升的。

(6) 由于发音器官的运动同其产生的共振峰之间并非线性关系，所以一个二合元音的几个共振峰的变化模式也可以不同。为了表征一个二合元音的动态特性，我们可以将 F_1 和 F_2 的频时归一曲线加以平均，这样就得到一个二合元音的动态特性曲线。

将 /a i/ 的 F_2 的数据，经过上述方法处理后，再与 F_1 的数据逐点平均，可得 /a i/ 的动态特性曲线 (图 6)。

(7)下面。我们就一个二合元音的动态特性曲线。来定义一些表征动态特性的参量(参见图7)。

(a) 起始 (on-set) 段长 t_{on} —— 频率变化不超过 10% 的起始音段的归一化时长;

(b) 收尾 (off-set) 段长 t_{off} —— 频率变化不超过 10% 的收尾音段的归一化时长;

(c) 过渡 (transition) 段长 $t_t = 1 - (t_{on} + t_{off})$

(d) 过渡速率 $v_t = 0.8 / t_t$

(e) 前音 (front) 段长 $t_f = t_{on} + t_t / 2$

(f) 后音 (back) 段长 $t_b = t_{off} + t_t / 2$

(g) 交界频率 f_c —— 前、后音段交界点对应的归一化频率值;

(h) 中界频率 f_h —— 时长一半时对应的归一化频率值。

以图 6 中的 /ai/ 为例。可测算得: $t_{on} = 0.19$,
 $t_{off} = 0.21$, $t_t = 0.60$, $v_t = 1.33$, $t_f = 0.49$,
 $t_b = 0.51$, $f_c = 0.52$, $f_h = 0.54$ 。

以上参量, t_{on} 和 t_{off} 分别反映出该二合元音起始和收尾相对稳定部分的长短; t_t 代表该复合元音变化部分的长短; v_t 反映着动态变化的程度; t_f 和 t_b 可以分别表征该二合

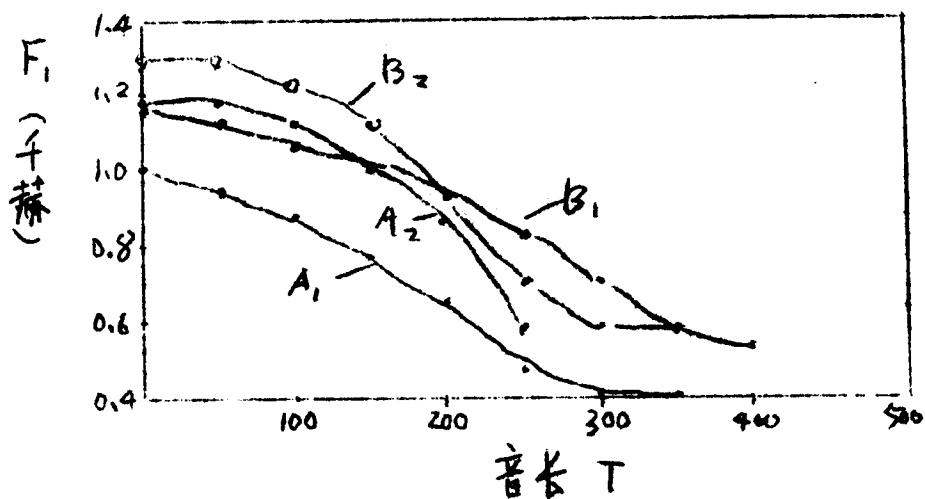


图2 /ai/ 的 F_1 的测量数据图

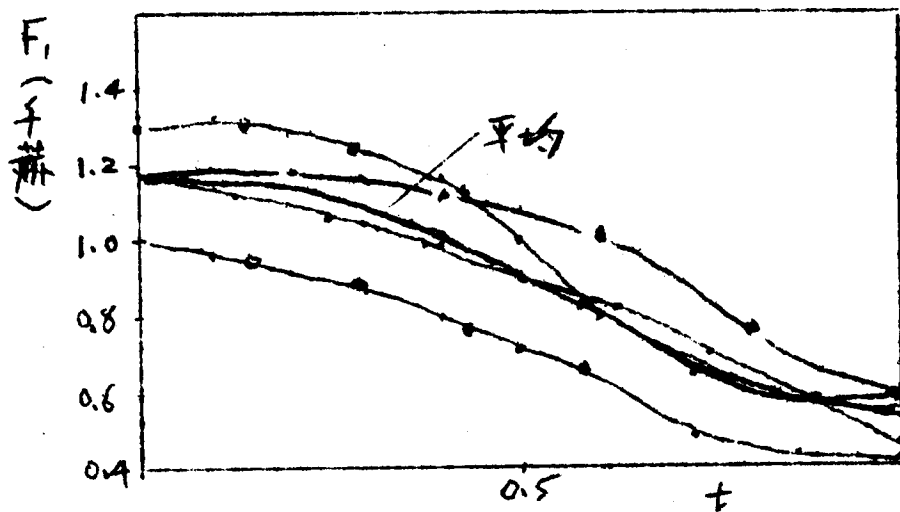


图3 /ai/ 的 F_1 时归曲线

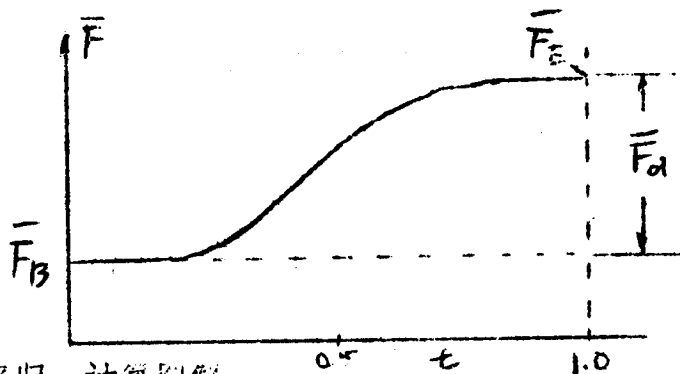


图4 频率归一计算图解

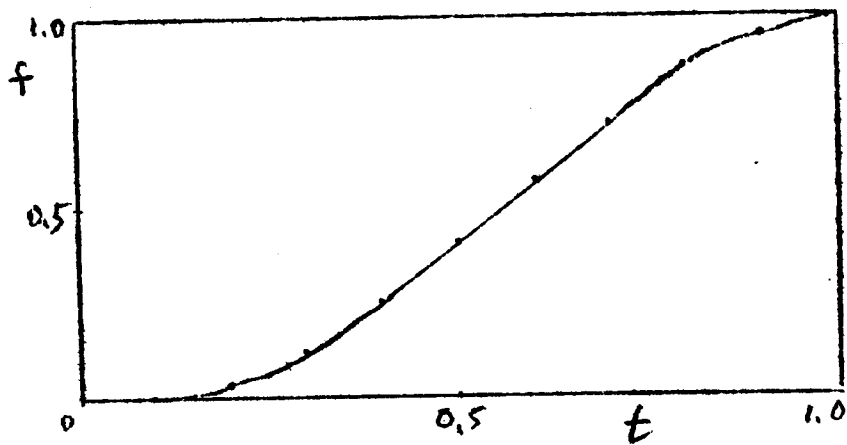


图5 $|a_i|/F_i$ 的频时归一曲线

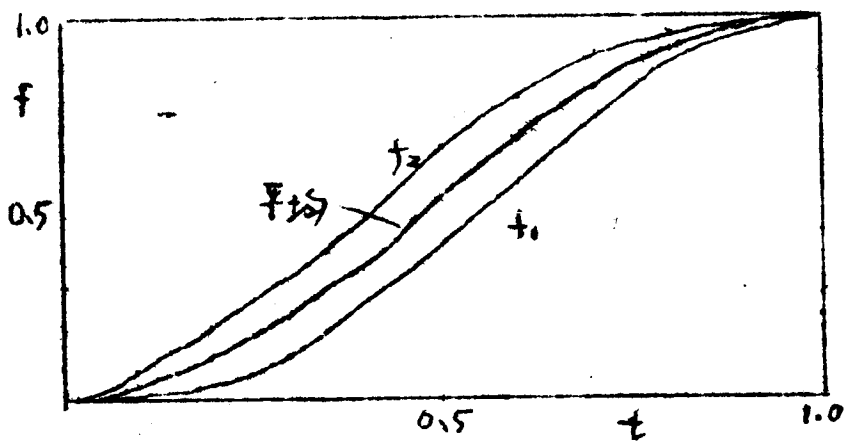
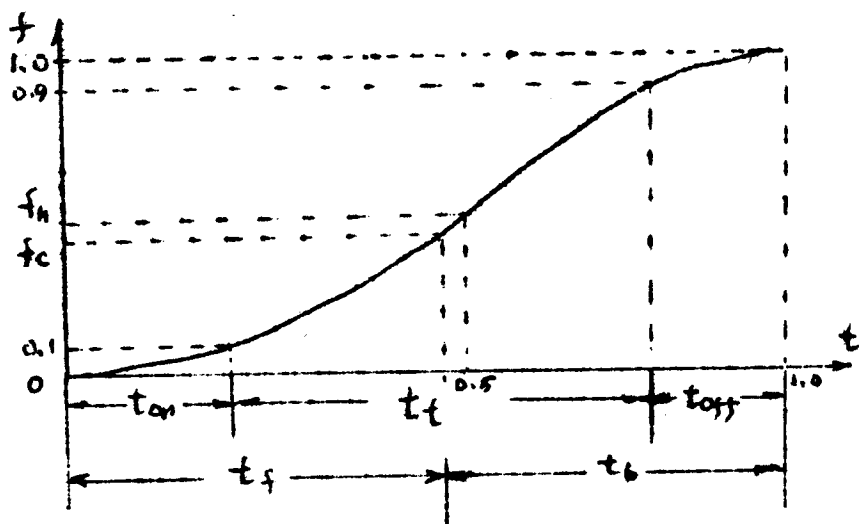


图6 $|a_i|$ 的动态特性曲线



• 8 • 图7 动态特性参量

元音在音色上的偏向程度，譬如 t_b 大于 t_f ，则意味着该合元音的音色偏向后一目标元音； r_o 和 r_h 也反映着音色上的偏向程度。

三 测试结果和分析

1. 时归一平均曲线

各二合元音的 F_1 和 F_2 的时归一平均曲线如图 8 所示。可以看到，共振峰频率基本上是依时变化的，变化的速率各处不一，一般说，中间急速，两端徐缓，个别音（如 /ia/ 的 F_1 和 F_2 ，/ei/ 的 F_2 ，/uo/ 的 F_1 ，/ua/ 的 F_1 和 F_2 ）收尾段相当平稳。

2. 频时归一平均曲线

九个二合元音，共作出 18 条频时归一平均曲线，这些曲线大体有三种模式（图 9）：一种近似于直线，如 /ou/ 的 F_1 和 F_2 ，/ao/ 的 F_1 和 F_2 ，/ie/ 的 F_1 和 F_2 ；另一种是后半段平稳，前半段激变，如 /ia/、ua 和 /ei/ 的 F_1 和 F_2 ，/ue/ 的 F_2 ；第三种介于这二者间。九个二合元音中，/ao/、/uo/、/ua/、/ie/、/ia/ 的 F_1 和 F_2 的变化趋势大体相似，两根曲线近似重合；而 /uo/、/ue/ 的两根曲线，差别甚大。

3. 动态特性曲线

九个二合元音的动态特性曲线如图10所示，从曲线形状来看，/ai/、/ie/、/ou/、/ao/的动态曲线近于直线；/ia/、/uo/、/ua/的曲线，后段平稳，前段激变；/ei/、/ue/介于二者间，稍偏向后者。

4. 动态特性参量的分析

按上面的定义，分别测算得九个二合元音的动态特性参量，整理归纳列入表1中。在该表中，还将前响二合元音和后响二合元音的参量分别平均，最后总平均。由表3的数据可以得出如下几点印象：

1) 一般说，二合元音的收尾段都比起始段长，特别是后响元音/ia/和/ua/，收尾段时长约为起始段的六倍。

2) 九个二合元音中，过渡段长从0.5到0.76，平均约0.56，即是，二合元音中，在一半的时间内，音质是迅速改变着。

3) 若以过渡段长为基准，画出各二合元音的各段时长的比例关系图(图11)，则可以看出过渡段、时长与收尾段时长之间，有明显的相关性，过渡段短者，其收尾段就长。如果以收尾段长为自变量，过渡段长为变量，画出函数图(图12)，可以看到，二者之间有近似的反比关系。利用统计学中的公式，

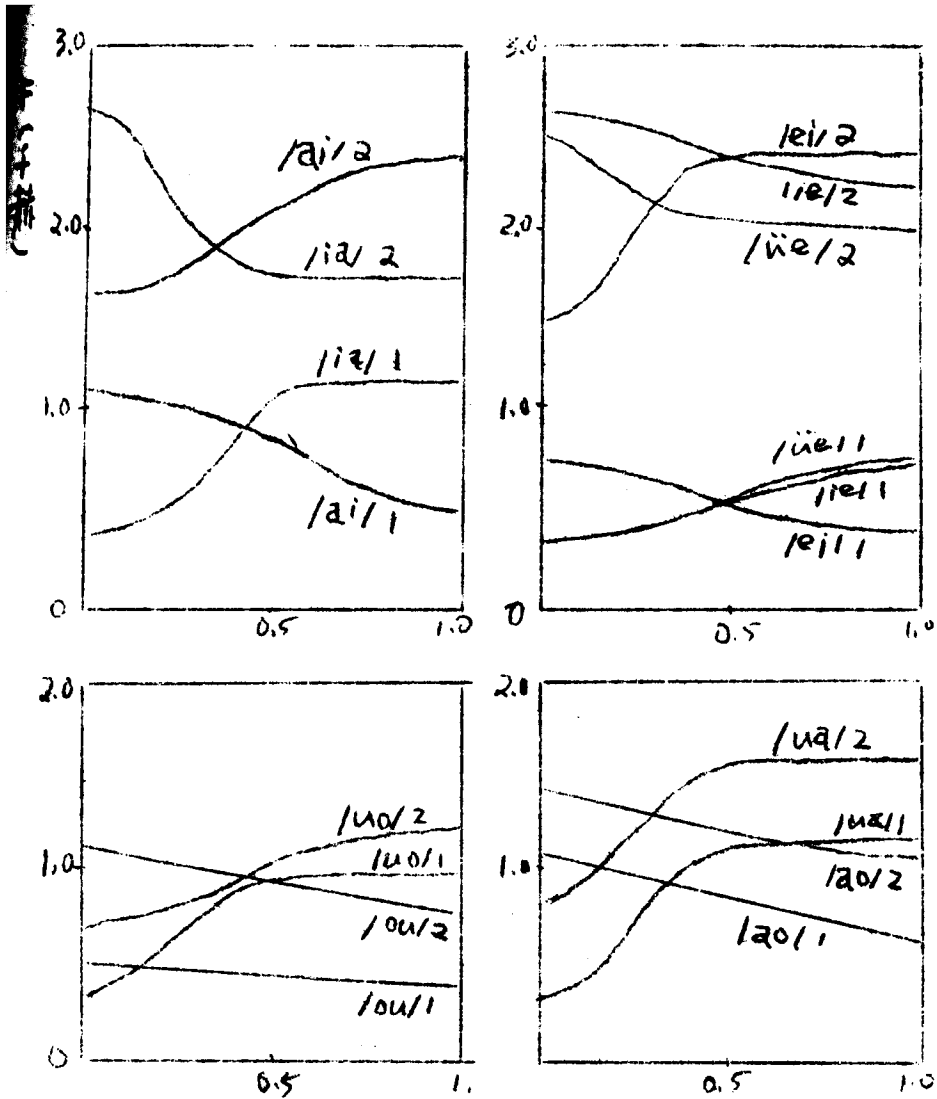


图8 时归一平均曲线 (/ai/1、/ai/2 分别表示 /ai/ 的第一、二共振峰)

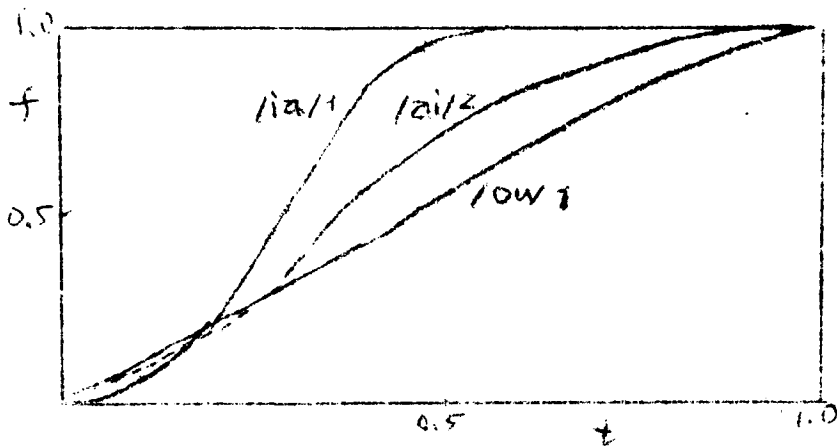


图9 频时归一曲线的三种模式

图 10 动态特性曲线

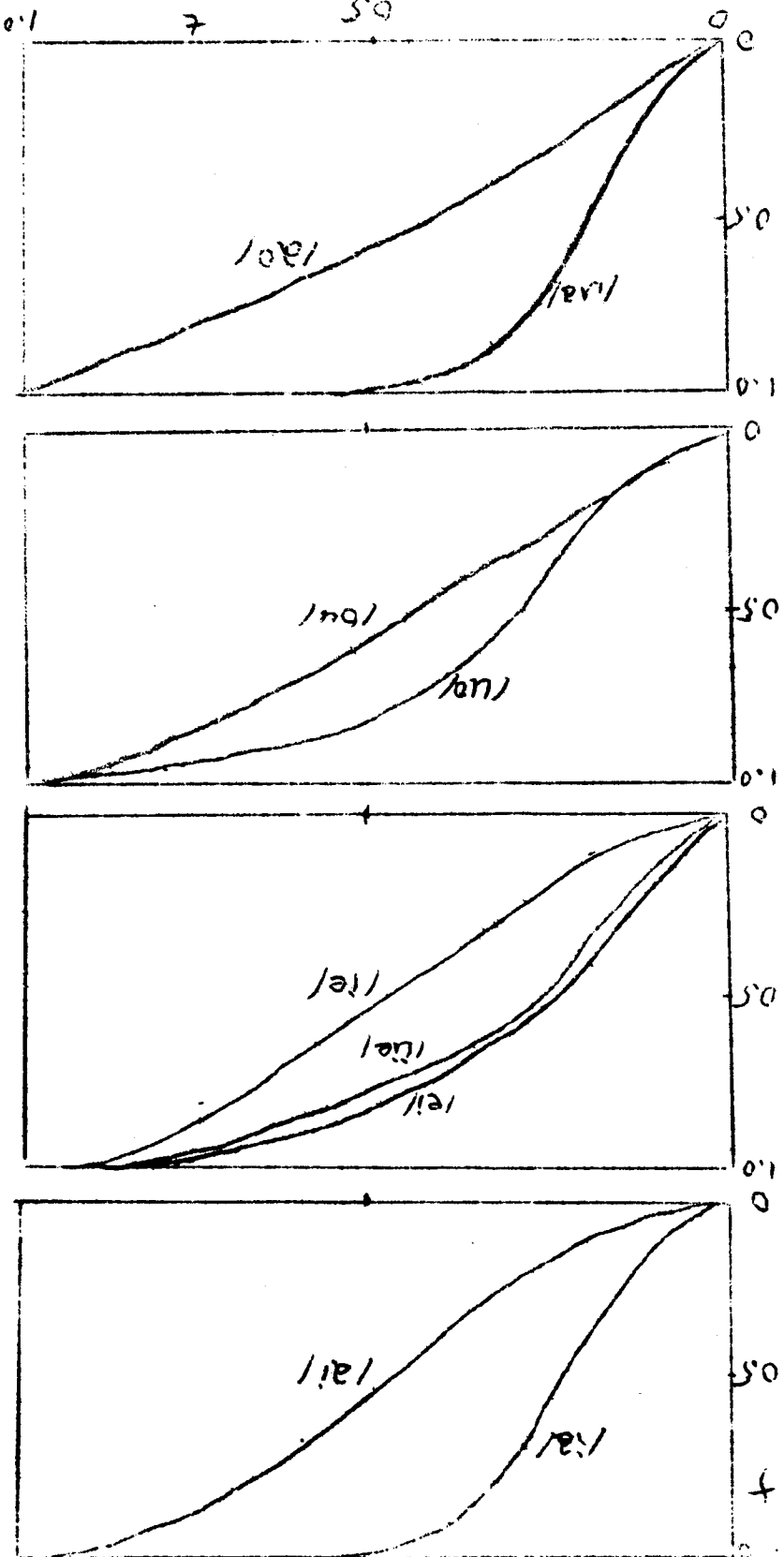


表1 二合元音动态特性参量表

参 重 二 合 元 音	起始 段长 t_{on}	收尾 段长 t_{off}	过渡 段长 t_c	过渡 速率 \dot{t}_c	前音 段长 t_f	后音 段长 t_b	交界 频率 f_c	中界 频率 f_h
/ai/	0.19	0.21	0.60	1.33	0.49	0.51	0.52	0.54
/ao/	0.08	0.16	0.76	1.05	0.41	0.59	0.52	0.61
/ei/	0.06	0.40	0.54	1.48	0.33	0.67	0.65	0.84
/ou/	0.10	0.17	0.73	1.10	0.465	0.535	0.52	0.56
/ia/	0.09	0.61	0.30	2.67	0.24	0.76	0.49	0.97
/ie/	0.15	0.19	0.66	1.21	0.48	0.52	0.50	0.53
/üe/	0.08	0.32	0.60	1.73	0.38	0.62	0.67	0.78
/uo/	0.13	0.35	0.52	1.54	0.39	0.61	0.66	0.81
/ua/	0.10	0.60	0.30	2.67	0.25	0.75	0.50	0.96
前响	0.108	0.235	0.658	1.24	0.424	0.576	0.553	0.638
后响	0.11	0.414	0.476	1.68	0.348	0.652	0.564	0.810
平均	0.109	0.334	0.557	1.49	0.382	0.618	0.559	0.733

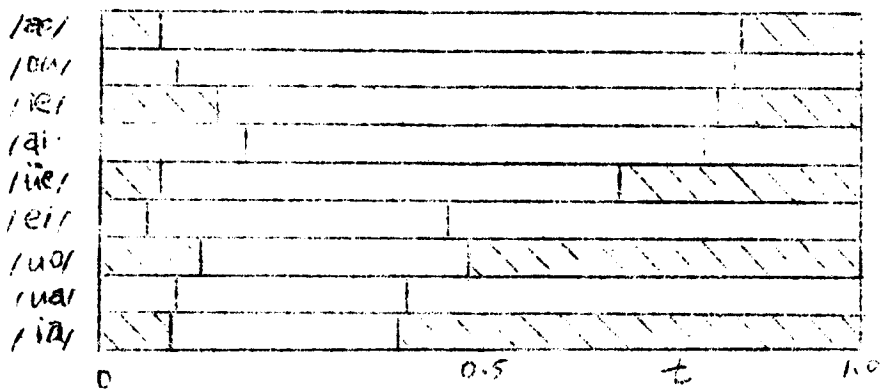


图11 二合元音各段时长比例

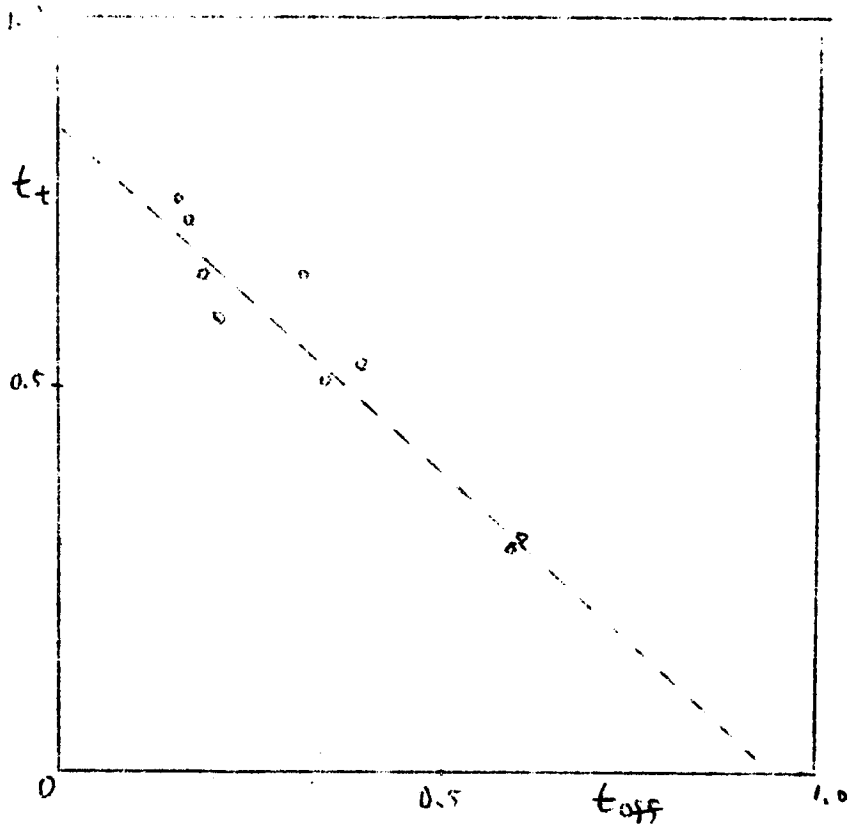
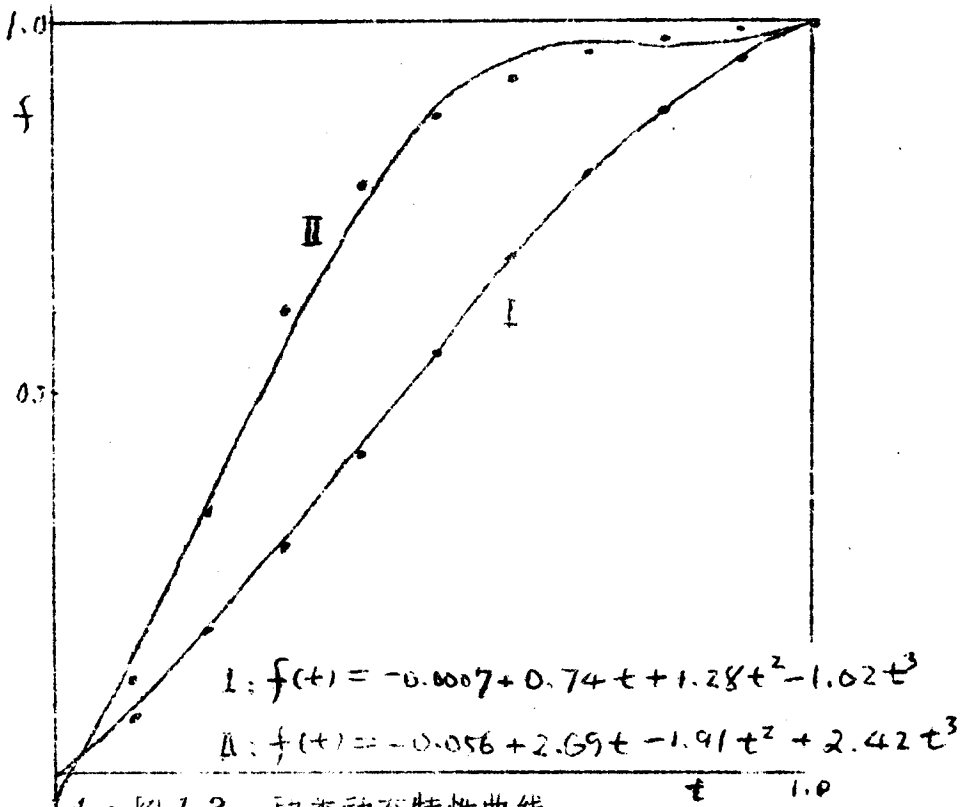


图 12 过渡段长随收尾段长的变化



14 • 图 13 两类动态特性曲线

可以计算出二者的相关系数 r 和线性回归系数 a 、 b ：

$$r = S_{xy} / \sqrt{S_{xx} \cdot S_{yy}}$$

$$a = S_{xy} / S_{xx}$$

$$b = \bar{t}_e - a \bar{t}_{off}$$

式中 \bar{t}_e 是 t_e 的平均值， \bar{t}_{off} 是 t_{off} 的平均值

$$S_{xx} = \sum (t_{off} - \bar{t}_{off})^2$$

$$S_{yy} = \sum (t_e - \bar{t}_e)^2$$

$$S_{xy} = \sum (t_{off} - \bar{t}_{off})(t_e - \bar{t}_e)$$

代入表 3 中的数值，可算得 $r = -0.971$ ， $a = -0.916$ ，

$b = 0.853$ 。此相关系数已非常接近于负一，即是二者间有十分近似的反比关系；

4) 九个二合元音的后音段都比前音段长，特别是后响元音的后音段更甚。

四 结 语

根据以上曲线、数据及其分析，我们对普通话二合元音的动态特性有如下几点认识：

1. 二合元音的动态特性主要体现于元音共振峰（主要是 F_1 、 F_2 ）的滑移之中。依据动态特性曲线，二合元音可以大体上分为起始段、过渡段和收尾段。不同二合元音，这三段的时长比例

关系是不相同的，平均而言，过渡段最长，收尾段其次，起始段最短，过渡段是二合元音复合信息的主要载信者。而且，过渡段时长与收尾段时长有近似的反比关系。

2. 九个二合元音的动态特性曲线可归为两大类：线性变化型和前急变后平稳型。一般说来，前响二合元音的动态曲线可归于第一类（/eɪ/例外），后响的归于第二类（/iə/、/ue/例外）。每一类曲线可以用一满足最小乘法的多项式来拟合（图13）。这种归类和拟合是否合适，还有待语音合成来检验。

3. 与英语和西班牙语相比，普通话二合元音的后音段要长一些。如果采取本文的处理方法，对文献〔3〕和〔4〕的数据进行计算，则可算得五个英语二合元音的后音段的归一化平均时长为0.48，十四个西班牙语的为0.56，而九个普通话二合元音的后音段平均时长达0.62。

参考文献:

- 1、Lehiste, J. and Peterson, G., "Transitions, glides and diphthong", J. Acoust. Soc. Am., 31, 268-277(1961).
- 2、Holbrook, A. and Faibanke, G., "Diphthong formants and their movements", J. Speech Hearing Res., 5, 1, 38-58(1962).
- 3、Gay, F., "Effect of speaking rate on diphthong formant movement", J. Acoust. Soc. Am., 44, 6, 1570-3(1968).
- 4、Manrique, A., "Acoustic analysis of the Spanish diphthongs", Phonetica, 36, 3, 194-206(1979).
- 5、关宗济等, 《普通话单音节语图参考册》(社会科学出版社 编印中)。
- 6、关宗济、曹剑芬, "实验语音学知识讲话(三)", 中国语文, 1979年, 第4期, 314-320。
- 7、曹剑芬, "普通话复合元音特性分析", 北京市语言学会第一届年会, (1982)。
- 8、Stevens, K. N., "The quantal nature of

speech; Evidence from articulatory-acoustic data", in «Human Communication, A unified view», David, E. E. and Penes, P. B. ed. McGraw-Hill, New York, 1972.