

复合元音的指数式动态模型

社会科学院语言研究所

杨顺安

一、前 言

复合元音是一种音色上有动态变化的元音，揭示复合元音的动态变化模式，对语音学的研究与教学，对语音合成和自动语音识别都是很有意义的。

根据语图仪的测量数据，我们已对普通话九个二合元音的头两个共振峰 F_1 和 F_2 的数据进行过处理，归纳出两种用多项式表示的归一化的动态模式〔1〕。按此模式合成的二合元音，音色较自然，但就合成应用而言，此种模型还不够方便，欠缺变通性，也没有反映出协同发音 (co-articulation) 现象，而且不能推广到三合元音的合成中。

本文依据上述测量数据，提出一种复合元音的指数式动态模型，属于合成研究，效果较好。

二、指数式动态模型

一般来说，二合元音的共振峰的动态变化，大多是起始段和

收尾段比较平稳，中间过渡段变化急剧。这种情况很象两段。指数的负幂 $e^{-\alpha t}$ 的变化形式，于是，我们设定二合元音共振峰的动态变化有如下模型（图 1）：

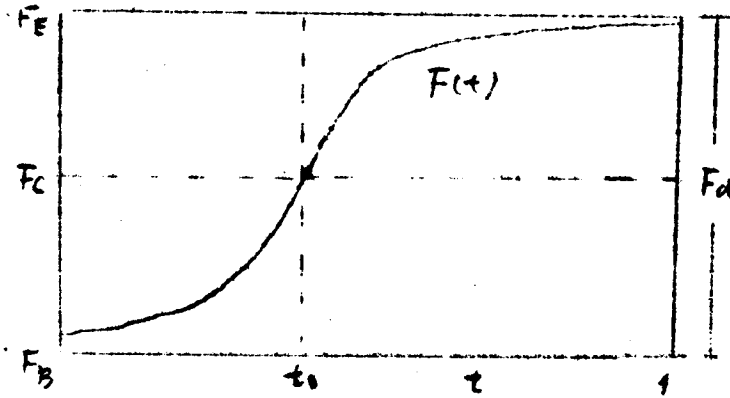


图 1 指数式动态变化模式

$$F(t) = F_C \mp 0.5 F_d [1 - e^{\pm \alpha (t_0 - t)}] \quad (1)$$

式中，中值频率 $F_C = 0.5 (F_B + F_E)$

$$\text{动程 } F_d = F_E - F_B$$

F_B 是起始目标值

F_E 是收尾目标值

t_0 是分界时长

α 是过渡速率因子

t 是归一化时长

这样，只要有了 F_B 、 F_E 、 t_0 和 α 这四个参数，就可以按式

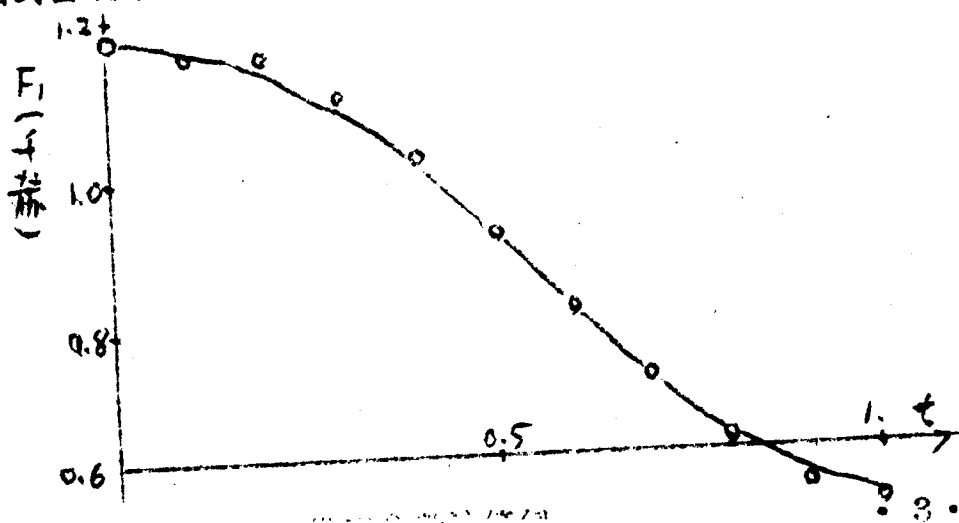
(1) 计算出某二合元音某共振峰的动态特性。可以看出，按此

模型, F_B 和 F_E 是作为只能接近而不能达到的两个目标值。

现在的问题是, 要根据共振峰的某一组实测数据, 设法确定出这四个参数, 如果按式 (1) 计算出的共振峰频率与实测值比较接近, 则可以说, 这种指数式模型是可行的。

四个模型参数是采用正交试验法 [2] 确定下来的。首先, 根据经验为某一组实测共振峰频率选定四个模型参数的初值及其上下两水平值, 这样形成了四参量三水平的试验表。接着按正交表的搭配, 分别按式 (1) 进行九组计算, 从中选出效果最佳一组, 如果按此四参数计算, 与实测值的平均误差已足够小 (例如小于 5%), 则可以认为这套参数就是所要求的, 如果误差较大, 则在此基础上往下计算。

图 2 的小圆圈是 $\angle a_1 /$ 和 $\angle i a_2 /$ 的某一共振峰的实测值, 细实线是分别确定了模型参数后, 按式 (1) 计算的曲线, 可以看到曲线与实测值拟合得较好。



表一列举出九个二合元音的 F_1 和 F_2 的介界时长 t_0 和过渡速率因子 α 的数值以及误差 E 。可以看到，对前响二合元音（ e i 例外）， t_0 居中， α 较小，说明它们的两音段较均匀，过渡较平缓；对 ia 和 ua 等后响二合元音， t_0 较小， α 较大，说明它们的前音段较短，过渡急速。

表一 九个二合元音的 t_0 和 α 值

	ai		ei		ao		ou		ia		ie		ua		uo		üe	
	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2
t_0	.55	.43	.27	.19	.52	.49	.50	.44	.2	.2	.45	.4	.2	.25	.35	.42	.35	.23
α	3	4	3.1	4.2	1.9	2.1	1.9	2.3	5.0	5.6	3.4	3.4	6.1	5.8	3.8	2.8	3.5	3.5
E	1.8	.8	1.4	1.5	2.8	1.1	2.1	1.6	2.9	2.6	2.6	0.6	3.1	2.8	3.3	1.4	3.0	1.8

现在，我们可以很方便地把二合元音的指数式动态模型加以推广。对三合元音来说，考虑到三音素间的协同发音效应，其动态变化可用下式描述（图3）：

$$\begin{aligned}
 F(t) &= F_{Bm}(t) + F_{mE}(t) - F_m \\
 \left. \begin{aligned}
 F_{Bm} &= F_{c1} \mp 0.9 F_{d1} [1 - e^{\pm \alpha_1 (t_{c1} - t)}] \\
 F_{mE} &= F_{c2} \mp 0.5 F_{d2} [1 - e^{\pm \alpha_2 (t_{c2} - t)}]
 \end{aligned} \right\} (2)
 \end{aligned}$$

式中各符号的意义参见图3。这样一来，对一个三合元音的某一共振峰的动态变化而言，需要三个目标值 F_B 、 F_M 、 F_E ，两个分界时长 t_{01} 和 t_{02} 以及两个过渡速率因子 α_1 和 α_2 共七个参数来表征。

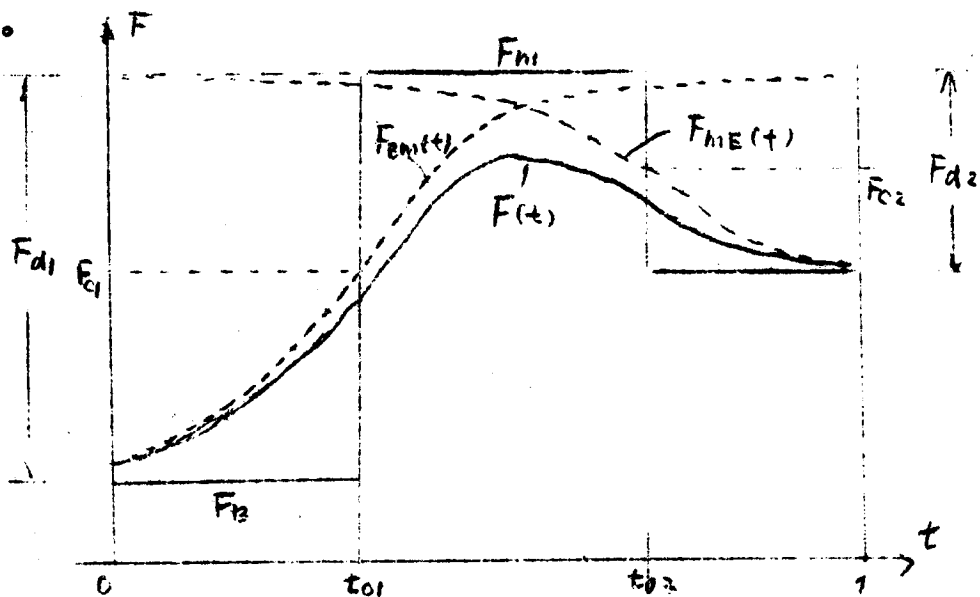


图3 三合元音的指数式过渡

三、合成验证及应用

为了验证该模型的可行性，我们在一台小型计算机上，编制了级联式共振峰合成程序，进行了合成听辨试验。合成流程如图4所示，用于合成的元音音素的头三个共振峰的频率如表二（ F_4 和 F_5 分别固定为 3500 和 4500 赫）。

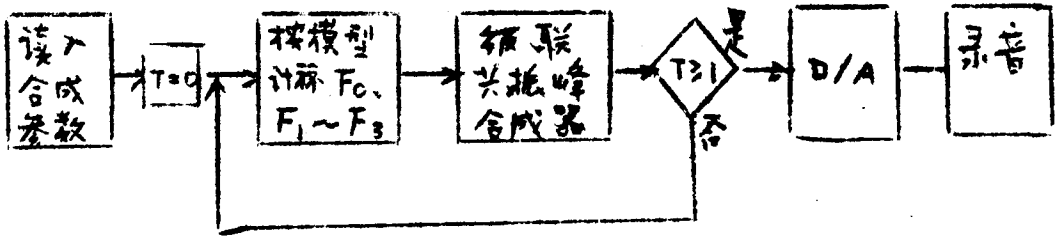


图4 复合元音的合成流程图

表二 目标元音的共振峰频率(赫)

	/i/	/e/	/A/	/o/	/u/	/ɜ//	/ə//
F ₁	270	520	1070	500	360	300	500
F ₂	2350	2030	1200	750	600	1890	1500
F ₃	3050	2720	2600	2700	2700	2250	2500

为了合成出比较自然的带声调的复合元音音节，基频 F_0 由一种归一化的字调模型动态地产生出来 [3]，声门波形采用如下函数形式 [4]：

$$g(t) = \begin{cases} 0.5[1 - \cos(\pi t / t_p)] & 0 \leq t \leq t_p \\ \cos[\pi(t - t_p) / 2 t_N] & t_p \leq t \leq t_p + t_N \\ 0 & t_p + t_N \leq t \leq T_c \end{cases}$$

式中各符号的意义如图 5。

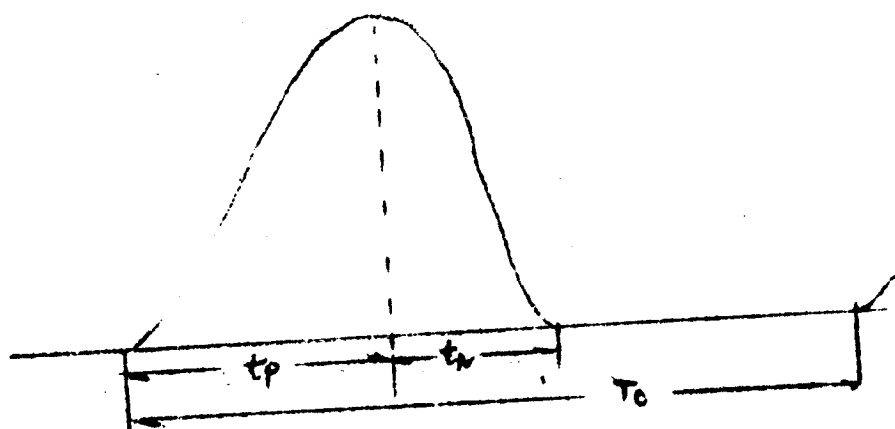


图 5 声门波形图

我们分别按该模型的式(1)和式(2)合成出普通话的九个二合元音和四个三合元音,初步试听,其音色过渡变化比较自然。

对于合成应用来说,该模型与多项式模型相较,有两个特点:第一,所需输入的目标值较少;第二,能自动反映协同发音效应。例如要合成/aɪ/和/ia/,用多项式模型则需要[a-I]和[i-A]四套真正目标值,而利用指数式模型所具有的“接近而不能达到”目标值的特点,就只需要[A]和[i]两套目标值,合成/aɪ/时,分界时长居中,过渡速率因子不大,始端达不到[A],靠前达到[a],终端达不到[i],只达到较低较后的[I];合成/ia/时,分界时长靠前,过渡快,

始端较接近〔i〕，而终端逼近〔A〕。在某种意义上，用指数模型合成复合元音，是一种音位参数的合成。要合成十三个普通话的复合元音，只需/i、e、A、o、u、 \bar{u} 、 \bar{y} /七个音位目标值。例如，以/iAu/为目标值，可以合成出〔la \bar{o} 〕（腰）。如果用多项式模型，至少需要十几个音素目标值。

参考文献：

- 1、杨顺安、曹剑芬，“普通话二合元音的动态特性”，语言研究，1984年，第一期。
- 2、马希文，《正交设计的数学理论》，人教版，1981年。
- 3、杨顺安，“普通话字调模型及其合成中的应用”。1984年
- 4、A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowel", J. Acoust Soc. Am., 49, 583-590(1971).

“第二届应用声学学术会议（84.11，重庆）。