



论 文

谱综合与汉语后元音/ γ /的知觉于水源^{①*}, 陈玉东^②, 吴洁茹^③

① 中国传媒大学计算机学院, 北京 100024

② 中国传媒大学对外汉语学院, 北京 100024

③ 中国传媒大学播音主持艺术学院, 北京 100024

* 通信作者. E-mail: yusy@cuc.edu.cn

收稿日期: 2009-09-28; 接受日期: 2009-12-16

中国传媒大学 211 工程三期重点学科建设资助项目

摘要 英语等很多语言中后元音的前两个共振峰相距小于临界距离 3.5 bark, 它们在语音知觉意义上等价于一个共振峰, 这个现象被称为谱综合. 谱综合后的谱峰与其语音性质之间有更直接的对应关系, 是判断语音性质的重要依据. 汉语单元音/ γ /是后元音, 但是其前两个共振峰相距通常大于临界距离, 那么, 汉语元音/ γ /的知觉线索是什么? 其前两个共振峰知觉上是否发生了谱综合? 文中开展了一系列辨识实验和评价实验, 研究两谱峰的距离、谱形状和相对幅度对/ γ /的语音性质的影响. 结果显示, 当前两个共振峰相距大于临界距离时, 只要呈现双峰形谱形状, 就被知觉为/ γ /; 当两峰相距在临界距离附近时, 显著的两峰形才能被知觉为/ γ /, 梯形和单峰形都不能被知觉为/ γ /; 当第二峰高于第一峰 10 dB 以后, 不被知觉为/ γ /, 且发生谱综合, 相反, 若第一峰与第二峰相对幅度在 -5 ~ 20 dB, 则被知觉为/ γ /, 且不会发生谱综合, 就是说, 谱综合现象关于两峰相对幅度是不对称的. 总之, 汉语单元音/ γ /虽然是后元音, 但是其知觉中并没有在前两个共振峰上出现谱综合现象. / γ /音色需要两个分开的共振峰. 这与西方语言的后元音和汉语其他后元音是不同的.

关键词谱综合
汉语单元音/ γ /
后元音
临界距离

1 引言

由于 Peterson、Barney^[1] 和 Klatt 等^[2] 杰出的工作, 元音共振峰位置 (尤其是前 2~3 个) 被认为是语音性质的主要决定因素. 进一步的研究指出前 2~3 个共振峰在语音性质的知觉上并不是独立的, 当两个相邻共振峰相距小于某距离时, 在语音性质的知觉上等价于一个处于两峰中间的单共振峰, 这个现象被称为谱综合、谱平均或者重心效应^[3~6]. 谱综合后的峰被称为知觉共振峰 (perceptual formant)^[4]. 但是当两个谱峰相距大于某临界距离时, 谱综合现象消失, Chistovich 等^[5] 指出, 这个临界距离是 3~3.5 bark.

文献 [7~10] 等研究使用两个共振峰合成元音, 其中第一共振峰与自然元音的相同, 第二共振峰是由原第二共振峰及其高次共振峰谱综合后的结果, 此时的第二共振峰被称为等价第二共振峰 (effective

second formant, 记为 F'_2). Escudier 等^[11]、Mantakas 等^[12]、Qiang 等^[13]给出了 F'_2 的计算公式. Schwartz 等^[14]利用 F'_2 公式研究元音系统的整体性质. 基于 F'_2 的概念, 如下假说被提出: 听觉系统在 3.5 bark 频率范围内存在一个大尺度谱综合, 而语音听觉是依据这个谱综合的结果. Sydal 等^[15]依据共振峰之间的距离, 以 3 bark 为分界对美国英语元音进行自动分类, 准确率是 85.7%. 这些研究结果表明, 谱综合后的知觉共振峰和声学共振峰相比与语音性质有更直接的关系.

Delattre^[3]、Sydal 等^[15]都指出, 美国英语后元音具有相距较近的前两个共振峰, 因而知觉上产生谱综合效应.

汉语单元音/ɤ/是后元音, 但是其前两个共振峰相距较大, 一般大于 3~3.5 bark, 这与西方语言不一致. 那么, 它的前两个共振峰的知觉是否发生了谱综合现象? 其知觉线索是什么?

本文的目的是研究哪些声学线索影响了汉语单元音/ɤ/的知觉, 以及在其前两个共振峰的知觉中是否发生了谱综合.

2 汉语元音/ɤ/的共振峰的分布

2.1 数据库

为了研究/ɤ/的共振峰的分布以及/ɤ/的语音性质, 建立了一个语料库, 由 48 个男声和 59 个女声的/a, o, u, ɤ/组成. 录音者全部是播音系 20~24 岁本科学生, 所录音经过专业语音研究者听音筛选. 语音信号录制的采样率是 16 kHz, 经 16 bit 模数转换, 录音在消声室内进行.

2.2 测量方法

汉语单元音/ɤ/在实际孤立发音时一般带有明显的动程, 其稳定段才是真正的元音/ɤ/. (详见文献[16]), 本文语料库的发音人是经过专业语言发音训练的, 并且在发音时被要求尽量较小动程. 发音结果经过 3 位语音学研究人员的听音筛选. 为了进一步减少动程的影响, 计算时只取较稳定的中间 1/3 段. 根据 LPC(linear predictive coding) 谱方程的根计算其共振峰频率, 计算采用 14 阶 LPC 谱分析, 0.97 预加重, 帧长 50 ms, 帧移 25 ms, 加汉明窗. 计算过程采用半自动方式, 经人工监督和选择, 避免自动计算引起的解根的错误^[17], 且舍弃那些明显错误的计算结果. 每个音的每帧计算结果的平均值作为该音测量结果.

2.3 测量结果

测量结果见图 1~4. 图 1 是男声元音/a, o, ɤ/的前两个共振峰的分布. 图 2 是男声/ɤ/的前两共振峰差的分布, 平均值为 4.63 bark, 标准差为 0.69 bark. 其中两峰相距小于 3.5 bark 的占 4.2%.

图 3 是女声元音/a, o, ɤ/的前两个共振峰的分布. 图 4 是女声/ɤ/的前两个共振峰差的分布, 平均值为 4.31 bark, 标准差为 0.75. 小于 3.5 bark 的占 15.79%, 小于 3 bark 的占 7.02%.

第三共振峰的测量结果见表 1. 由表 1 可见, 这 3 个元音的第三共振峰的分布很相近, /ɤ/与/a/和/o/的第三共振峰位置相差都在 10%左右范围内.

以上语料库测量结果显示, 3 个元音/a, o, ɤ/的前两个共振峰位置分布之间的关系, 即, /ɤ/的第一共振峰与/o/的第一共振峰分布在相同频率范围, 第二共振峰与/a/的第二共振峰分布在相同范围. /ɤ/的前两个共振峰的距离大部分大于 3~3.5 bark 临界距离, 少部分(男女声合计总平均在 10%左右)在临界距离附近. /ɤ/与/a/和/o/有相近的第三共振峰分布.

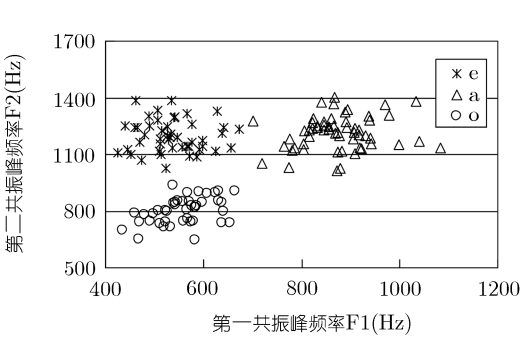


图 1 男声元音/a,o,γ/的前两个共振峰的分布

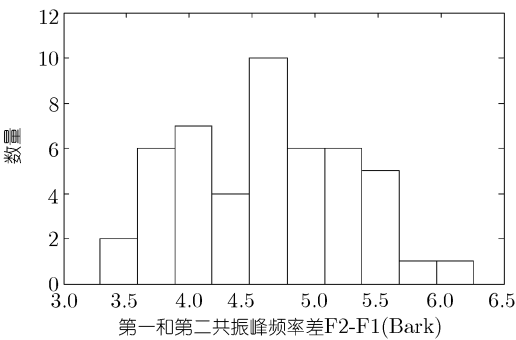


图 2 男声/γ/前两个共振峰差的分布

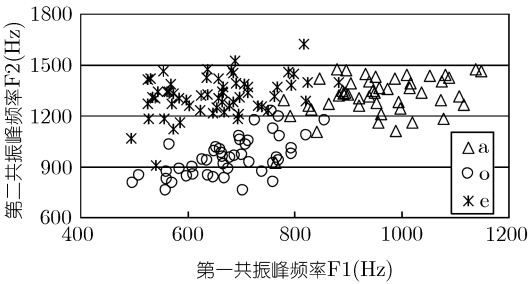


图 3 女声元音/a,o,γ/的前两个共振峰的分布

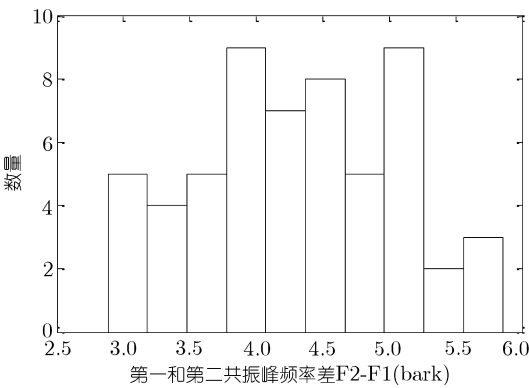


图 4 女声/γ/的前两个共振峰差的分布

表 1 第 3 共振峰测量结果

		/γ/	/a/	/o/
女声	平均 F3(HZ)	2105.96	1898.11	2267.73
	标准差 (HZ)	552.29	400.03	665
男声	平均 F3(HZ)	2773.93	3008	2882.08
	标准差 (HZ)	139.91	176.84	194.18

3 实验

由第 2 节数据库测量结果可见, /γ/的第三共振峰位置与/o, a/的相近, 故推论: 引起/γ/与/o, a/语音性质差异的, 是其前两个共振峰. 以下听觉实验 1 验证该推论.

3.1 实验 1: /γ/的前两个共振峰的语音性质

实验刺激 实验刺激由 3 部分组成. 第 1 部分, 随机选择自语料库的 10 个/γ/样本; 第 2 部分, 将以上 10 个样本信号通过 50 阶低通滤波器, 只保留其前两个共振峰, 滤波结果作为实验刺激; 第 3 部分, 随机选择 10 个/o/, 10 个/a/和 10 个/u/共 30 自然语音作为实验刺激, 这是为了避免连续听到过多的不自然语音使得被试产生听觉适应. 实验刺激共计 50 个. 全部刺激被调整具有相同的响度和时域包络形状.

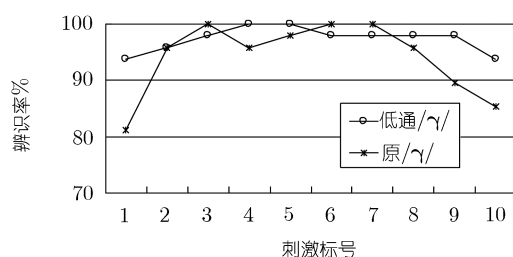


图 5 被试对 10 个随机选择的男声/ɤ/以及只保留其前两个共振峰的/ɤ/的辨识结果

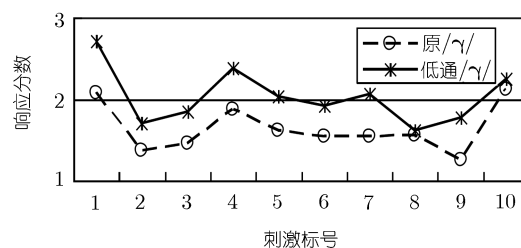


图 6 被试对/ɤ/刺激的典型性评价

实验方法 实验任务分为两部分. 第 1 部分, 被试依据听辨结果确定每个刺激的语音身份; 第 2 部分, 被试评价 10 个/ɤ/及其低通信号作为/ɤ/的典型性程度. 评价在 3 个等级上进行. 1 代表刺激是典型的/ɤ/, 2 代表刺激是/ɤ/但不典型, 3 代表勉强是/ɤ/. 每个刺激随机呈现 4 次, 每个刺激之间相隔 3 s.

被试为 12 名播音系大学生, 年龄 20~22 岁, 日常使用标准普通话, 听力正常, 有较好的语音鉴别能力, 无作为被试的经验. 在安静房间中使用监听耳机, 音量调至被试认为清晰 (此时大约 68 dB).

以下各实验的被试均与本实验相同, 不再赘述.

实验结果 实验结果如图 5, 6.

辨识实验结果显示, 原/ɤ/信号被知觉为/ɤ/的辨识率平均 0.974, 其标准差是 0.02, 低通滤波只保留前两个共振峰后的/ɤ/样本, 被辨识为/ɤ/的平均辨识率为 0.942, 标准差为 0.07. 配对 T 检验结果是: 双尾 T 检验显著性概率是 0.086, 大于 0.05, 差值 95%置信区间是 $-0.54419 \sim 6.80419$, 包括 0, 故可以认为只保留前两个共振峰对于/ɤ/的语音性质没有显著影响. 评价实验结果显示, 原/ɤ/平均得分 1.64, 标准差 0.29, 仅保留前两个共振峰的/ɤ/, 平均得分 2.03, 其标准差 0.33. 配对 T 检验结果是: 双尾 T 检验显著性概率是 0.000 小于 0.05, 差值 95%置信区间是 $-0.51431 \sim -0.26049$, 不包括 0, 故可以认为只保留前两个共振峰对于/ɤ/的语音性质的典型性有显著影响.

以上实验结果说明, 只保留/ɤ/的前两个共振峰对其语音性质没有决定性影响, 但是典型性程度降低了. 这是由于没有了高频成分, 听感不自然. 低标准差说明被试响应有较好的一致性, 以及响应结果的可靠性.

概括实验 1 的结果与上节语料库共振峰测量结果可以看到, /ɤ/的语音性质由其前两个共振峰决定.

第 2 节共振峰测量结果指出, /ɤ/的前两个共振峰之差分布范围比较大, 而且, 大部分大于临界值, 少部分 (约 10%) 在临界值附近的, 这意味着/ɤ/的知觉线索可能很复杂, 不只和两峰距离有关. 因此, 设计了实验 2 以研究前两个峰频率之差和谱形状与其语音性质之间的关系.

3.2 实验 2: /ɤ/的前两共振峰的距离和谱形状对其语音性质的影响

实验刺激 实验刺激采用经滤波处理的真实语音而不是合成元音. 这是因为 Chistovich^[5] 和 Nearey^[18] 都指出合成语音因缺少语音内部变化, 被试在判断实验刺激的语音性质时会受到音质的影响, 而不是仅依据语音性质, 所以使用经谱处理后的真实语音作为实验刺激, 有利于被试准确判断刺激的语音性质. 为了产生实验刺激, 一个自然孤立发音的男声/ɤ/被选择作为种子信号, 实验所有刺激均由该信号经谱处理得到. 该信号基频平均值 182.5 Hz, 最大值 184.2 Hz, 最小值 180.1 Hz, 标准差

为 0.67 Hz. 稳定段时长 500 ms, 总时长 650 ms. 该信号前两个共振峰由 8 个谐波组成, 它们的位置分别是 1.79, 3.53, 5.15, 6.62, 7.94, 9.1, 10.12 和 11.02 bark. 实验刺激是对该 8 个谐波进行幅度上的处理后得到的, 且保持高频谱成分不变. 刺激信号的前两谱峰的每一个谱峰由至少 3 个谐波成分组成, 且其中一个处在峰位置.

通过改变种子信号前 8 个谐波的相对幅度, 形成不同变化的前两峰距离和谱包络形状. 首先, 使得第一峰位置分别位于第 2, 3, 4, 5 次谐波上, 第二峰位置分别位于第 4, 5, 6, 7 次谐波上, 且第二峰频率高于第一峰的. 共计 10 种组合, 即如表 2 所列. 表 2 中每个单元中的数是两峰距离. 各峰斜率是 10 dB/谐波. 本文中用两峰的距离或者位置来表示这些刺激, 如刺激 3.09 或者刺激 3.53/6.62 表示表 2 中左上角单元的激励, 即第一峰在二次谐波 3.53 bark 处, 第二峰在 4 次谐波 6.62 bark 处, 两峰相距 3.09 bark; 然后, 改变经由上步骤得到的信号的两个谱峰间各谐波的高度以及各峰斜率, 从而形成不同的谱峰形状. 它们是: 双峰形, 其斜率分别为 3, 5, 10 dB/谐波, 记为, -3, -5, -10. 如图 7(a); 单峰形, 其基础部分斜率为 10 dB/谐波, 上部斜率分别为 +3 dB/谐波, +5 dB/谐波, 记为 +3, +5; 梯形, 其斜边斜率为 10 dB/谐波, 上底为水平, 记为 0. 如图 7(b) 所示. 共计 6 种形态. 值得注意的是共振峰斜率越大谱包络越锐.

本实验研究两个谱峰形成的整体谱包络形状, 而不是谱峰带宽, 与其语音性质的关系, 这是因为, 从文献 [5, 13, 14] 等研究结果看, 谱综合应该被看作是谱包络上的光滑处理, 而谱包络形状比谱峰带宽与光滑处理的关系更直接.

经过以上两步处理, 共计得到 $10 \times 6 = 60$ 个刺激.

实验过程 全部刺激信号被处理成同样的响度和同样的能量包络形状. 每个刺激信号随机呈现给被试 4 次, 每两个刺激之间间隔 3 s. 每呈现 100 个刺激 (约用时 10 min 左右) 休息 3 min, 整个实验历时约 30 min. 被试的任务是依据听觉结果确定每个刺激在汉语单元音 /a, o, γ , i, u, y/ 中的身份.

实验结果 虽然被试的选择范围包括 /i, y/, 但是没有被试的响应落入该范围. 实验结果见表 3. 其中行表示两峰位置及两峰距离, 列表示谱峰形状. 每个刺激的 4 个单元格分别为 / γ /, /a/, /o/ 和 /u/ 的识别结果. 表中加粗字体的为识别结果中大部分是 / γ /, 灰色底纹的为识别结果没有确定的语音身份, 即 4 个数字中没有有一个超过 50%.

由表 3 中的第 6~11 列可见, 除了刺激 4.41 以外, 当两峰相距大于等于 3.5 bark 才会被知觉为 / γ /, 且相距越远, 知觉到的 / γ / 比率越高; 由表中标记为 “+5” “+3” “0” 行可见, 当双峰谱形态形成梯形或者单峰形时, 无论两峰相距多大, 都不会被知觉为 / γ /. 对照标记为 “-3” “-5” “-10” 行可见两峰越突出 (即斜率越大, 或者说, 带宽越窄), 就越多被知觉为 / γ /. 由第 2~5 列可见, 在 3.09 bark 以下不能被知觉为 / γ /.

另外, 值得注意的是, 第 8 列中的刺激 3.53/7.94, 虽其两峰相距 4.41 bark, 但没有被知觉为 / γ /, 而表中其前后两列均被知觉为 / γ /, 这意味着两峰相距大于临界距离不是被知觉为 / γ / 的充分条件.

实验 3 研究两峰相对幅度对 / γ / 知觉的影响.

3.3 实验 3: 双峰相对幅度对元音 / γ / 知觉的影响

实验刺激 本实验中, 等步长改变实验 2 中双峰形态最显著的 (-10 dB/谐波)、能够被知觉为 / γ / 的激励的两峰相对幅度. 由于预实验证实, 当两峰相对幅度小于 ± 5 dB 时, 对其语音性质没有明显可知觉到的影响, 所以本实验两峰相对幅度分别设置为 ± 10 , ± 15 , ± 20 , ± 25 dB. 实验中所有刺激经过等响和等时域能量包络处理.

表 2 刺激信号两峰位置分布

		第二峰位置 (Bark)				
	谐波次数		4	5	6	7
		谱峰位置 (Bark)	6.62	7.94	9.1	10.12
第一峰位置 (bark)	2	3.53	3.09	4.41	5.57	6.59
	3	5.15		2.79	3.95	4.97
	4	6.62			2.48	3.50
	5	7.94				2.18

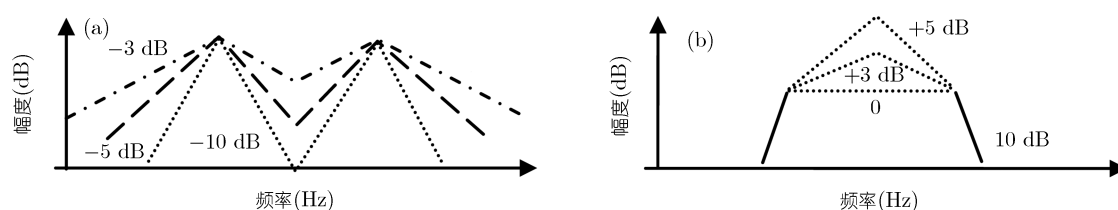


图 7 实验刺激的双峰谱包络形态

(a) 对应比峰形. 其中刺激的斜率分别是 3, 5, 10 dB/谐波. (b) 对应单峰形和梯形. 其中刺激的斜边斜率是 10 dB/谐波, 而顶部斜率分别是 3, 5 和 0 dB/谐波

实验过程 本实验为辨识实验. 实验中被试的任务是依据听辨结果为每个刺激指派语音身份. 每个刺激随机呈现 4 次. 每两个相邻刺激之间相隔 3 s.

实验结果 图 8 是两峰相对幅度与被试响应为/ɤ/的比率之间的关系. 从中可以看到, 刺激 4.41 除外, 当第一峰高于第二峰 -5 到 20 dB 时 (即横坐标介于 -20 到 5 之间), 辨识结果为/ɤ/. 当第二峰高于第一峰 10 dB 以上时 (即横坐标大于 10), 知觉结果为/ɤ/的比率很低, 一般为单元音/a/或者/o/.

为了测试/ɤ/的双峰知觉是否发生了谱综合, 设置了单峰听辨实验. 此时, 种子信号的前 8 个谐波只形成一个峰. 斜率是 -10 dB/谐波和 -5 dB/谐波. 单峰也可以被看作是双峰相对幅度变化的极限情况. 实验过程与本实验上部分相同.

图 9 是实验结果, 被试的听辨结果与谱峰位置的关系. 其中矩形表示被试响应结果为/u/, 圆环为/o/, 上三角为/a/, 星为/ɤ/. 划线表示峰斜率是 5 dB/谐波, 实线为 10 dB/谐波. 响应结果中只有 5.95%为/ɤ/, 这说明, 峰值位置从 3.53~10.12 bark 的频率范围内, 单峰不能被知觉为/ɤ/. 由于/ɤ/的前两个谱峰的分布是在该范围之内, 故/ɤ/的前两个共振峰没有发生谱综合. 在图 9 中还可以看到处在 7.94~10.12 bark 之间的单峰被知觉为/a/, 在 5.15~6.62 bark 为/o/, 在 3.53 bark 附近为/u/, 对照表 3 中实验 1 的结果可见, 对于/a, o, u/的知觉, 本实验的单峰刚好落在实验 1 的双峰之间, 故可以认为在这 3 个元音的知觉中发生了谱综合.

比较图 9 中的实线和划线, 它们表示谱峰斜率不同的刺激, 它们有几乎相同的变化规律, 这说明谱峰的突显 (谱包络的锐化) 程度对于语音身份没有显著影响, 这与文献 [2, 19] 的实验结果是一致的.

4 结论和一般性讨论

结论 概括本文的数据库测量结果和听觉实验结果, 结论如下:

表 3 被试对不同两峰距离和谱形状的反应结果

谱峰位置及 双峰距离	7.95/ 10.12	6.62/ 9.1	5.15/ 7.94	3.53/ 6.62	6.62/ 10.12	5.15/ 9.1	3.53/ 7.94	5.15/ 10.12	3.53/ 9.1	3.53/ 10.12
谱形状	2.18	2.48	2.79	3.09	3.5	3.95	4.41	4.97	5.57	6.59
+5	0	0	0.02	0.02	0	0.02	0	0.02	0.04	0
	1	0.91	0.15	0.06	1	0.67	0.04	0.94	0.38	0.88
	0	0.09	0.83	0.79	0	0.31	0.96	0.04	0.58	0.13
	0	0	0	0.13	0	0	0	0	0	0
+3	0	0	0	0.17	0	0	0.02	0	0.06	0.02
	0.94	0.98	0.25	0	1	0.71	0.06	1	0.75	0.96
	0.04	0.02	0.75	0.65	0	0.29	0.88	0	0.19	0.02
	0	0	0	0.17	0	0	0.04	0	0	0
0	0.04	0	0	0.1	0	0.08	0.08	0.13	0.25	0.17
	0.96	1	0.15	0.04	0.98	0.83	0.13	0.85	0.67	0.81
	0	0	0.85	0.46	0.02	0.08	0.75	0	0.08	0.02
	0	0	0	0.4	0	0	0.04	0	0	0
-3	0.02	0.06	0	0.02	0.17	0.4	0.19	0.75	0.63	0.63
	0.98	0.85	0.1	0.04	0.83	0.5	0.13	0.25	0.33	0.38
	0	0.04	0.9	0.35	0	0.08	0.63	0	0.04	0
	0	0.02	0	0.58	0	0	0.06	0	0	0
5	0	0.11	0	0	0.27	0.69	<u>0.31</u>	0.94	0.67	0.75
	1	0.89	0.1	0.04	0.71	0.31	<u>0.19</u>	0.02	0.31	0.25
	0	0	0.9	0.29	0.02	0	<u>0.33</u>	0.04	0.02	0
	0	0	0	0.67	0	0	<u>0.17</u>	0	0	0
10	0	0.21	0.02	0.04	0.63	0.96	<u>0.33</u>	0.98	0.58	0.88
	0.1	0.79	0.04	0.15	0.38	0.02	<u>0.27</u>	0.02	0.33	0.08
	0	0	0.9	0.21	0	0.02	<u>0.15</u>	0	0.02	0
	0	0	0.02	0.58	0	0	<u>0.25</u>	0	0.06	0.04

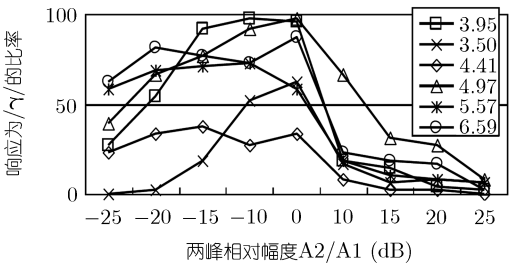


图 8 两峰相对幅度与 γ / 知觉的关系

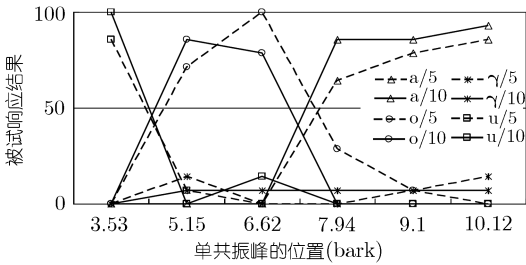


图 9 单峰位置与被试识别响应的关系. 矩形表示被试响应为 u /, 圆环为 o /, 上三角为 a /, 星为 γ /. 划线表示斜率是 5 dB/谐波, 实线表示斜率为 10 dB/谐波

数据库测量结果显示, /ɤ/的两峰相距较远. 对于本文的数据库和计算方法, 男声/ɤ/的两峰相距平均 4.68 bark, 女声的 4.31 bark, 略有差异.

概括数据库测量结果和听觉实验结果表明, 产生/ɤ/音色的两个谱峰, 第 1 个位置在/o/和/u/附近, 第 2 个在/a/附近. 离开这个位置不产生/ɤ/音色.

双峰相对幅度相等时, 若两峰相距大于 5 bark 左右, 只要谱形成双峰形态, 就可以产生/ɤ/音色. 但是当双峰相距较小, 在临界值附近时, 双峰间要形成明显的双峰形态, 就是说要求双峰之间的谐波成分很低, 才可以产生/ɤ/音色. 当双峰相距小于临界值时 (大约 3.1 bark), 不会产生/ɤ/音色.

当双峰相对幅度不等时, 第二峰大于第一峰超过 5~10 bark 以后, 不会产生/ɤ/音色. 而第一峰大于第二峰时, 在很大的范围 (20 dB 之内) 都会产生/ɤ/音色, 在 10bark 左右时, 达到最高. 在更高时, 会逐渐减弱, 超过 20 dB 以后, /ɤ/音色逐渐消失.

第二峰的幅度若高于第一峰超过 5~10 dB, 会发生谱综合.

汉语/ɤ/发音舌位是后元音, 但是知觉上并不发生谱综合, 且需要两个峰. 这与美国英语中后元音的结论不同.

一般性讨论 文献 [20] 在激励模式 (excitation pattern) 上以二次函数拟合峰位置的测量结果以及在此基础上的听觉实验结果表明, 汉语单元音/u, o, a, y, i/的知觉是由特定的单峰位置决定的. 该计算方法是一种大尺度 (大于 1 bark) 光滑算法, 因此得到的峰位置是几个相邻的声学谱包络峰综合后的结果. 本文实验结果显示, 汉语/ɤ/是由相距较远的两个共振峰峰决定的, 且这两个峰不能被综合为一个, 所以这两个共振峰不会被变换为一个单一的听觉谱峰. 概括以上这两个结论得到: 汉语单元音的知觉是由听觉谱上的 1 个或者 2 个峰 (不能看作是共振峰, 应看作是听觉谱峰) 引起的. 从信息编码的角度看, 这可以看作是听觉系统的编码方式: 一位编码和二位编码. /u, o, a, i, y/使用一位编码, 即由一个峰的位置决定其语音性质, 而/ɤ/使用二位编码, 语音性质由两个峰的位置共同决定.

当两个峰的位置相距不是很大时 (此时可以认为是/ɤ/音色不很明显), 决定/ɤ/音色知觉的不只是前两个共振峰的位置, 还需要峰突出作为知觉线索, 越是凸出的峰, 就越有利于/ɤ/音色的知觉. 所以从声学谱上看, 影响/ɤ/语音性质知觉的线索除了共振峰位置外, 还有在一定条件下的谱锐化程度作为第 2 线索. 这可以被解释为, 由于/ɤ/的知觉需要两个独立的共振峰, 且听觉系统存在一个大尺度谱综合——平滑化处理, 所以越是尖锐的谱峰越不容易被平滑掉的.

刺激 3.53/7.94 的双峰相距虽然很大 (已经达到 4.41 bark), 但是没有被知觉为/ɤ/, 而双峰相距小于该值或大于该值的刺激均可以被知觉为/ɤ/. 该刺激的第二峰位置很低, 在 7.94 bark, 而所有被知觉为/ɤ/的刺激的第二峰均高于该位置, 这意味着/ɤ/的两个峰除了要求相距较大外, 位置也是特定的. 该结论支持 Chiba 和 Kajiyama^[21] 提出的空间模式理论 (space pattern theory) 的观点: “如果共振峰的中心处于对每个元音来说特定的频率范围, 则元音由其相对共振峰所决定.” (见文献 [22]).

刺激 3.53/7.94 虽然未被知觉为/ɤ/, 仍不能认为在其知觉中发生了谱综合. 此时, 4 个响应结果中最多的不过占 1/3, 仅比随机水平 1/4 稍高, 所以该刺激不能在汉语元音中产生确定的知觉对象. 4 种响应结果显示被试在知觉时依据的线索分为 4 种, 第 1 种, 依据双峰且未发生谱综合, 知觉结果为/ɤ/; 第 2 种, 依据双峰且谱发生综合, 知觉结果为/o/. 因为双峰的平均位置为 5.74, 在/o/的位置; 第 3 种, 依据第一峰知觉为/u/. 第 4 种依据第二峰, 知觉为/a/.

美国英语后元音以及汉语后元音/u, o, a/(央偏后)的前两共振峰相距较小, 知觉上发生谱综合现象^[3,20]. 但是, 汉语/ɤ/发音上是后元音, 其前两峰相距较大, 没有谱综合现象发生. 所以, 在汉语中不能依据前两峰的距离区分元音的前后.

对于产生谱综合两峰相对幅度是不对称的, 在第一峰高于第二峰时, 一般不产生谱综合, 而第二峰高于第一峰很容易产生谱综合, 这与谱重心假说不相符, 但与文献 [23, 24] 的结论是一致的。

参考文献

- 1 Peterson G E, Barney H L. Control methods used in a study of the vowels. *J Acoust Soc Am*, 1952, 24: 175–184
- 2 Klatt D H. Speech processing strategies based on auditory models. In: Carlson R, Granstrom B, eds. *The Representation of Speech in the Oerioheral Auditory System*. Amsterdam: Elsevier, 1982. 181–196
- 3 Delattre P C, Liberman A M, Cooper F S, et al. An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 1952, 8: 195–210
- 4 Chistovich L A, Sheikin R L, Lublinskaya V V. ‘Centers of gravity’ and the spectral peaks as the determinants of vowel quality. In: Lindblom B, Ohman S, eds. *Frontiers of Speech Communication Research*. London: Academic, 1979. 143–157
- 5 Chistovich L A. Center auditory processing of peripheral vowel spectral. *J Acoust Soc Am*, 1985, 77: 789–805
- 6 Chistovich L A, Lublinskaja V V. The center of gravity effect in vowel spectra and critical distance between the formants. *Hearing Res*, 1979, 1: 185–195
- 7 Fant G. Acoustic analysis and synthesis of speech with application to Swedish. *Ericsson Tech*, 1959, V1: 1–106
- 8 Carlson R, Granstrom B, Fant G. Some studies concerning perception of isolated vowels. *Quarterly Progress Status Reports STL-QPSR 2/3*, Speech Transmission Laboratory. 1970
- 9 Carlson R, Fant G, Granström B. Two-formant models, pitch and vowel perception. In: Fant C G M, Tatham M A, eds. *Auditory Analysis and Perception of Speech*. London: Academic Press, 1975. 55–82
- 10 Bladon A, Fant G. A two-formant model and the cardinal vowels. *Speech*. In: *Speech Transmission Laboratory Quarterly Progress Report*. Stockholm, 1978. 1–8
- 11 Escudier P, Schwartz J, Boulogne M. Perception of stationary vowels: internal representation of the formants in the auditory system and two-formant models. In: *Franco-Swedish Seminar, Société Française d’Acoustique (SFA)*. Grenoble, 1985. 143–174
- 12 Mantakas M, Schwartz J L, Escudier P. Vowel spectrum processing and the large scale integration concept. In: *7th FASE Congress*. Edinburgh, 1988.
- 13 Xu Q, Jacewicz E, Feth L L, et al. Bandwidth of spectral resolution for two-formant synthetic vowels and two-tone complex signals. *J Acoust Soc Am*, 2004, 115: 1653–1664
- 14 Schwartz J L, Escudier P. A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception. *Speech Commun*, 1989, 8: 235–259
- 15 Syrdal A K, Gopal H S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *J Acoust Soc Am*, 1986, 79: 1086–1100
- 16 Zhang J. The consonantal properties of the unrounded high-mid back vowel /ɤ/ in Mandarin Chinese. *J Acoust Soc Am*, 1996, 99: 2548–2574
- 17 Vallabha G K, Tuller B. Systematic errors in the formant analysis of steady-state vowels. *Speech Commun*, 2002, 38: 141–160
- 18 Nearey T M. Static, dynamic, and relational properties in vowel perception. *J Acoust Soc Am*, 1989, 85: 2088–2113
- 19 Van Veen T M, Houtgast T. Spectral sharpness and vowel dissimilarity. *J Acoust Soc Am*, 1985, 77: 628–634
- 20 Yu S Y. Excitation patterns and phonetic identities of Chinese vowels. *Sci China Ser F-Inf Sci*, 2009, 52: 1936–1946
- 21 Chiba T, Kajiyama M. *The Vowel: its Nature and Structure*. Tokyo: Tokyo-Kaiseikan, 1941. 193–194
- 22 Potter R K, Steinberg J C. Toward the specification of speech. *J Acoust Soc Am*, 1950, 22: 807–820
- 23 Assmann P F. The perception of back vowels: centre of gravity hypothesis. *Q J Exp Psychol*, 1991, 43A: 423–448
- 24 Beddor P S, Hawkins S. The influence of spectral prominence on perceived vowel quality. *J Acoust Soc Am*, 1990, 87: 2684–2704