

元音响度顺序的生理和声学分析

殷治纲

(中国社会科学院语言研究所, 北京 100732)

文 摘: 本文结合语音学、音系学和生理、声学知识, 详细探讨了影响元音“响度顺序原则(SSP)”的各种因素。该研究通过分析不同发音人低、中、高三类舌位的元音响度听辨结果与绝对声能的关系, 及其与听觉灵敏度加权声能之间的关系, 证明元音响度值不仅与其声强(声能), 还与元音频谱能量分布特性, 以及人耳对不同频段的听觉灵敏度特性都有关。此外, 研究还发现, 主要频谱能量分布于人耳听觉敏感区的元音, 有可能实现以相对较小声强获得相对较大听感响度的“经济性”效果。

关键词: 元音; 响度顺序原则; 响度; 声强; 声能; 频谱; 听觉; 听觉灵敏度; 等响度曲线

中图分类号: H012; H017; H018; O428

1 引言

在日常说话时, 我们不难发现, 即使在其他发音条件都相同时(如同样的说话力量, 同样的声调和韵律位置等), 某些音听起来也会比另外一些音“响”一些, 如“嗒”/da/一般比“嘀”/di/响。这一规律, 在音系学和语音学上被称为音节的响度顺序原则(SSP, Sonority Sequencing Principle)^{[1][2][4]}。详细说来, 即在其他条件相同时, 元音比辅音响, 低元音比高元音响, 浊辅音比清辅音响, 送气音比不送气音响。^[5]更详细描述可见下表(从上向下按响度从大到小的顺序排列)。

表 1 音节的响度顺序

类别	子类
元音	低元音>中元音>高元音
半元音	
流音	
鼻音	
擦音	浊>清
塞音	浊>清

响度顺序原则虽然早已被大家认可并广泛使用, 但以往语音学文献中却较少分析其背后的声学原理。本文将元音为例, 结合不同元音类别(低元音, 中元音, 高元音)的声学特点和人耳的听觉特性, 对“元音固有强度”^{[3][7]}问题进行简略探讨。

2 元音响度顺序和有关基本概念

2.1 元音响度顺序和响度概念

按照元音的响度顺序原则, 一般认为在其它条件相同时, 低元音响度要大于中元音, 中元音响

度要大于高元音。

所谓低元音, 是指发音时开口度大, 舌位低的元音, 如图 1 中的/a/元音。高元音是指发音时开口度小, 舌位高的元音, 如图 1 中的/i/和/u/。中元音是指发音时开口度和舌位介于高、低元音之间的元音, 如央元音/ɜ:/。



图 1 元音舌位图

要说明低、中、高不同元音的响度差异, 首先要了解“响度”这一概念。响度是指听觉上感到的声音强弱的程度^[6]。当一个音的响度大于另一个音时, 我们就会认为该音更“响”。

响度概念有两个特点:

- 1) 响度是对声音强度的反映。
- 2) 响度是一个主观心理量。

所以也可以说, 响度是对客观声音强度的主观感觉量。一般来说, 声音强度越大, 人感知的响度也越大。不过, 由于人耳听觉特性和主观感受的原因, 响度大小又不完全等同于音强大小。其主要原因是人耳对不同频率声音的灵敏度不同。

下面就声音强度和人耳听觉特性这两方面的知识作一下回顾。

2.2 声音强度和响度的关系

和声音强度有关的物理量是声压、声能、声强。

声压是指介质中有声场时的压强与无声场时的压强之差，其单位是帕斯卡 (pa)。声能是声波动、势能总和，单位是瓦(W)。声强是单位时间内流过某单位面积的声能平均量。对于平面波来说，声强 I 正比于声压有效值的平方，其单位是 W/m²。

响度和声强之间存在正相关关系，即在听觉区域内，随着声音强度的逐渐增加，主观上对这个声音也会产生由弱到强的感觉。但是响度和声强并不是完全线性相关的，因为强度是声音的客观物理量，而响度是主观心理量。

响度受人耳听觉特性等因素影响。

2.3 人耳听觉特性和等响度曲线

人耳有其听觉特性。其中一个特性是人耳在听觉区域内对不同频率声音的响度感知灵敏性不同。

人的听觉区域是指人可听闻声音的声强范围和频率范围。

人耳听觉的声强范围下限即听阈（闻阈），一般声压约为 10⁻¹⁵—10⁻¹⁴Pa。人耳听觉范围的声强上限是痛阈，其最大声压一般在为 10—10²Pa。

人耳可以听到的频率范围一般为 20-20000Hz。

图 2 中斜线区域即为人的听觉区域。其中最下面曲线为听阈曲线，最上面的曲线为痛阈曲线；最左侧为频率下限 20Hz，右侧为频率上限 20000Hz。

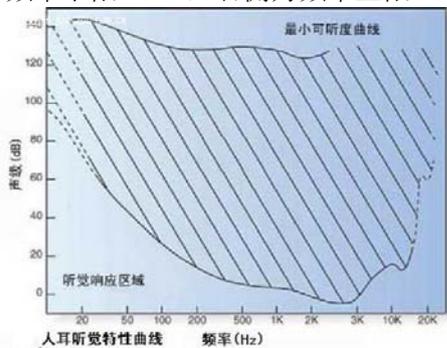


图 2 人耳的听觉区域

人耳在听觉区域内对不同频率声音的听觉灵敏度不同。为此，人们统计出了等响度曲线图，如图 3 所示。

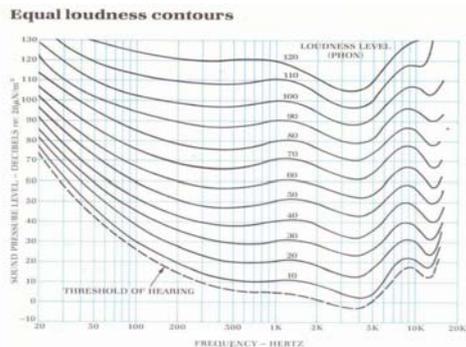


图 3 等响度曲线图

在同一条等响度曲线上，具有不同声强和频率

的声音的响度级是一样的。响度级的单位是“方” (Phon)。计算一个纯音的响度级的方法是用它和 1000Hz 的某个声级的声音响度作对比，如果感觉响度一样，就把 1000Hz 的这个声级定为该频率纯音此时的响度级。如果把一系列频率的纯音和 1000Hz 的该声级的纯音作等响度的声级比对，就可以画出一条等响度的声级曲线。^[7]改变 1000Hz 纯音的参考声级，就可以画出一组等响度曲线，形成图 3 的等响度曲线图。

在图中 13 条等响度曲线中，最下面的虚线曲线为听阈曲线，而数值为 120 方的最上方的那条曲线为痛阈曲线。

等响度曲线有以下特点：

1) 低于 500Hz 时，等响度曲线随着频率的减小而升高。这说明在此频率范围内，频率越低，人耳听觉越不灵敏，需要给以更高的声音强度才能让人耳感知到。尤其在 100Hz 以下频段，人耳的听觉灵敏度变得较差，这一现象叫做“低频钝化”现象。

2) 在 500-1000Hz 这段频率范围内，人耳听觉灵敏度趋于稳定。

3) 在 1000-4000Hz 这段频率范围内，等响度曲线随着频率的减升高而降低，即听觉变的更为敏感。在 3000-4000Hz 时，达到人耳的最大灵敏度。在这个频率范围内，声音强度为 1000Hz 的十分之一时就能听见。

4) 5000Hz 以后，听觉灵敏度再次下降。

3 不同元音的声学频谱特性

在此，我们以/a/、/C/、/i/和/u/为例来说明一下低、中、高三类元音的声学特性。

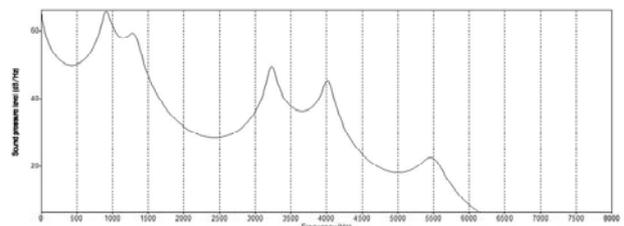


图 4a 低元音/a/的二维频谱图

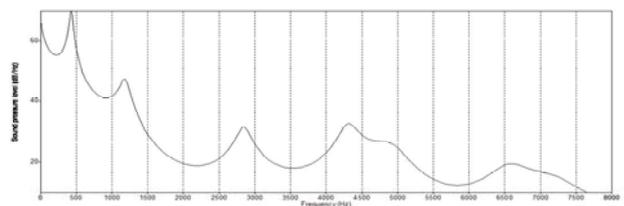


图 4b 中元音/C/的二维频谱图

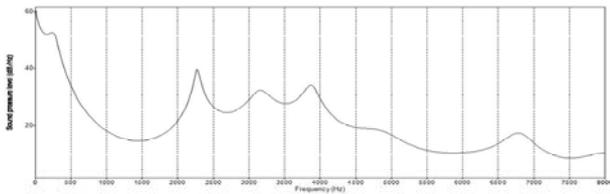


图 4c 高元音/i/的二维频谱图

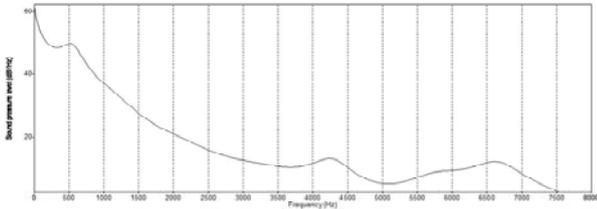


图 4d 高元音/u/的二维频谱图

由于声音的高频衰减特性，语音的声源频谱能量总体呈现从低频到高频逐渐减弱的趋势。同时，又由于不同元音的舌位和口腔形状不同，除了最低频区外（基频 F0 附近），在其他频率处还形成了各自不同的共振区域（共振峰）。一般认为，元音音色和其第 1、2 共振峰关系最大。其中，第一共振峰 F1 主要与舌位高低有关，第二共振峰 F2 与舌位前后有关（唇形对它们也有影响）。这样，不同元音的共振峰分布特点形成了各自不同的能量集中区。具体说来，特点如下（以成年发音人为例）。

低元音/a/的发音舌位低，略前，故其 F1 值最高，约在 700-1000Hz 左右（见图 4a）；F2 值略高，一般在 1100-1400Hz 左右。F1 和 F2 共同在 1500Hz 以下形成能量集中区。此外，/a/在 3000-4500Hz 还有另一个较弱的能量集中区。

中元音/C/的舌位高低和前后均居中，其 F1 和 F2 的值也都适中，F1（约在 400-600Hz）和 F2（约在 1200-1600Hz）形成能量依次递减的两个能量集中区。

高元音/i/的舌位高前，其 F1 值低（300hz 左右），F2 值高（1500-2000hz 左右），分别在低频和中频区形成了 2 个能量集中区。

高元音/u/的舌位高而后，并伴有圆唇动作，故其 F1 值和 F2 之均低，1000hz 以下的低频区是其 主要能量集中区。

4 实验与数据分析

为了更深入探讨元音响度顺序问题，我们请 2 位成年男性发音人和 2 位成年女性发音人分别录了 /a/、/C/、/i/和/u/等不同舌位高低的元音，以作进一步的分析。录音时，采用相同的录音设备和音量，并保持相同的录音距离。每人发音用程度也

尽量保持稳定。

4.1 响度听辨结果

我们请 5 位听辨人对以上声音数据进行了响度听辨实验。听辨结果符合音系学上元音响度顺序，即所有听辨人都将以上元音的响度归为明显的 3 类，分别是“响亮”（低元音/a/）、“较响”（中元音/C/）、“不太响”（高元音/i/和/u/）。

表 2 元音响度听辨结果

	/a/	/C/	/i/	/u/
响度等级	3	2	1	1
听辨一致性	100%	100%	100%	100%

4.2 元音声能与响度听辨结果的关系

然后，我们对发音人录音稳定段数据（取 50 毫秒）进行了频谱总能量统计。同时，我们还按等响度曲线灵敏度分段特性分了 6 个频段，并分段统计了各频率分段的能量（4.4 会用到），结果如表 3。

所有发音人的元音绝对频谱能量顺序均符合响度听辨顺序，即频谱能量从高到低依次为/a/ > /C/ > /i/和/u/。

表 3a. 男发音人 M1 频谱能量数据

M1 频段能量 单位: 10 ⁻⁵ Pa ² sec	/a/	/C/	/i/	/u/
0-300Hz	44.44	55.58	30.02	50.74
300-800 Hz	105.44	242.66	1.15	20
800-1500 Hz	248.01	19.51	0.07	0.3
1500-2500 Hz	2.12	5.91	0.2	0.01
2500-5000 Hz	9.76	2.41	1.78	0.02
5000-6000 Hz	0.03	0.08	0.03	0.02
总能量	409.78	320.83	33.26	71.08

表 3b. 男发音人 M2 频谱能量数据

M2 频段能量 单位: 10 ⁻⁵ Pa ² sec	/a/	/C/	/i/	/u/
0-300Hz	2.34	3.57	4.4	3.24
300-800 Hz	9.01	55.69	0.16	3.18
800-1500 Hz	59.09	1.29	0.01	0.12
1500-2500 Hz	0.32	0.02	0.01	0.01
2500-5000 Hz	1.89	0.11	0.13	0.01
5000-6000 Hz	0.01	0.01	0.00	0.00
总能量	72.65	60.68	4.79	6.55

表 3c. 女发音人 F1 频谱能量数据

F1 频段能量 单位: 10 ⁻⁵ Pa ² sec	/a/	/C/	/i/	/u/
0-300Hz	23.76	14.67	62.26	35.25
300-800 Hz	10.68	56.27	2.05	18.59

800-1500 Hz	59.55	3.66	0.11	0.17
1500-2500 Hz	9.66	0.16	0.06	0.06
2500-5000 Hz	1.6	0.61	1.35	0.05
5000-6000 Hz	0.13	0.04	0.05	0.01
总能量	105.39	75.42	65.88	54.12

表 3d. 女发音人 F2 频谱能量数据

F2 频段能量	/a/	/C/	/i/	/u/
单位: 10^{-5} Pa ² sec				
0-300Hz	12.66	14.91	44.26	36.75
300-800 Hz	17.2	35.75	1.68	11.25
800-1500 Hz	167.8	6.46	0.07	0.02
1500-2500 Hz	16.49	0.47	0.04	0.01
2500-5000 Hz	1.47	0.19	1.99	0.00
5000-6000 Hz	0.02	0.01	0.1	0.00
总能量	215.64	57.78	48.15	48.03

我们对上述发音人声能数据以各人最大声能值为标准做归一后，和响度感知结果作相关性分析，分析结果证明响度和声能大小呈正相关关系，即声能越大，响度越大。其结果如下。

表 4. 响度听辨结果和声强相关性统计

在 .01 水平（双侧）上显著相关		声能	响度
声能	Pearson 相关性	1	.868**
	显著性（双侧）		.000
	N	16	16
响度	Pearson 相关性	.868**	1
	显著性（双侧）	.000	
	N	16	16

4.3 对元音声能与开口度大小关系的分析

从上面的分析，我们可以发现不同元音的开口度大小（舌位高低）与其声能之间存在比较系统地对应关系，即开口度越大，声能越大。在此，我们初步分析一下这种现象背后的原理。

一般认为，人的发音过程可以分解为 3 个阶段：

1. 声源（声带）激励阶段：该阶段产生的声源能量按照每增大一个倍频程，能量衰减 12 分贝的规律分布；
2. 口腔调制/共振阶段：发音人控制发音器官的形状，对声源产生的声音信号进行调制，使其具有不同的频谱特性，产生不同的语音；
3. （口唇）辐射阶段：这是语音信号经过口唇辐射到外部的阶段。该阶段按照每增大一个倍频程能量增加 6 分贝的特性再次改变语音信号的能量分布。

在以上 3 个阶段中，元音的开口度大小和语音辐射阶段的关系最为密切，因为开口度越大，语音

能量可以尽可能多地辐射出去，其辐射特性越好。因此，这就不难理解元音声能与其开口度大小存在正相关特性的原因了。

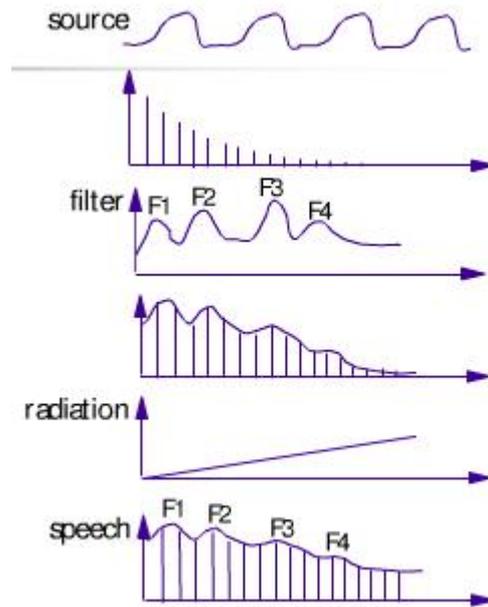


图 5 发音的 3 个过程

4.4 需要关注的问题

尽管从上面的分析中，我们可以基本了解不同元音开口度大小会影响其声强，但是进一步分析发现，元音的响度值关系并不完全等同于能量或者声强关系，即某些元音的响度值相对其能量值存在数据“偏移”的现象。

我们利用 Praat 语音分析软件计算了各语音数据的响度值，结果如下：

表 5a. 元音响度值

单位: sones	/a/	/C/	/i/	/u/
M1	4.47	4.15	3.42	3.16
M2	3.79	3.33	2.82	2.61
F1	4.19	3.72	3.39	2.78
F2	4.04	3.74	3.36	2.94

表 5b. 元音频谱强度值

单位: 10^{-5} Pa ² sec	/a/	/C/	/i/	/u/
M1	409.78	320.83	33.26	71.08
M2	72.65	60.68	4.79	6.55
F1	105.39	75.42	65.88	54.12
F2	215.64	57.78	48.15	48.03

以上四个元音的响度顺序从大到小是/a/、/C/、/i/、/u/。

值得注意的是：高元音/i/的强度一般近似于或者低于另一个高元音/u/，但是其响度值却一般大于/u/。也就是说，相对于声强关系，/i/元音的响度值发生了“正向偏移”。

另外，相对于最不响的/u/元音，/C/元音响度值也体现出这种正向偏移特性，即其声强值有时虽然只比高元音/i/、/u/略大，但是响度值却偏向于更响的元音/a/。

4.5 元音响度值与声强值差异的分析

在前面介绍中，我们知道响度是一个心理声学量，它并不等同于声强。具体说来，我们认为前面提到的某些元音的响度值关系和声强值关系间存在的差异，是由于声音的频谱能量分布特性会在一定程度上影响响度值。

在此，我们不妨进一步看看响度值的计算问题。从语音响度计算原理看，语音某时刻响度值应等于该音此时频谱能量函数 $E(f)$ 与人耳响度函数 $S(f)$ 的卷积，即响度值 = $E(f) * S(f)$ 。

实际响度计算方法大致经历了以下几个阶段：

1) ISO 532A 标准：传统响度算法，基于倍频程或者倍频程的某分数倍。直接将倍频程声压级与倍频程等响曲线对比，得到频带响度，再算出总响度。

2) ISO 532B 标准：是 Zwicker 的响度计算方法，是基于 Bark 频带理论的；

3) ANSI S3.4-2005 标准：是 Moore 的响度计算方法，基于 ERB 频带理论。

三类方法的关键区别在于频带划分方法的选择，也就是如何地模拟人耳基底膜的滤波特性。

从上面响度值计算公式可以发现，即使元音频谱频谱总能量相等或相近，但如果它们的频谱分布特性不同，也可能形成不同的听感（响度）差异。

对于/i/、/u/来说，由于/i/在人耳听觉敏感的3000—4000Hz区域比/u/有更多能量分布，所以在总能量差异不大时，/i/比/u/听起来要更响一些。

5 结论

本研究证明，在其他条件相同时，元音响度顺序原则依赖于不同元音的响度值大小序列。

详细说来，其响度顺序主要受以下两方面因素影响：1) 元音开口度大小（舌位高低）会影响其最终辐射出的声音能量；声音能量相差较大时，哪个元音的能量值大，哪个元音更“响”；2) 在声音能量相差不大时，哪个元音的声音能量更多处于人耳听觉敏感频率区域，哪个元音更“响”。

本实验结果也可以作为解释元音和其它音（如辅音）的响度顺序的参考。

参考文献

- [1] Selkirk, E., On the major class features and syllable theory. In Aronoff & Oehrle (eds.) *Language Sound Structure: Studies in Phonology*. Cambridge: MIT Press. 1984, 107-136.
- [2] Clements, G. N., The role of the sonority cycle in core syllabification. In J. Kingston and M. E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press. 1990, 283-333.
- [3] Lehiste, *Suprasegmentals*, M.I.T. Press. 1970, 106-153
- [4] 王洪君, 汉语非线性音系学, 北京大学出版社, 2008
Wang Hongjun, *Non-linear phonology of Chinese*, Peking University Press, 2008. (in Chinese)
- [5] 林焘, 王理嘉, 语音学教程, 北京大学出版社, 1992
Lin Tao, Wang Lijia, *The Course of Phonetics*, Peking University Press, 1992. (in Chinese)
- [6] 中国社会科学院语言研究所, 现代汉语词典, 商务印书馆, 2001
The Institute of Linguistics, CASS, *Modern Chinese Dictionary*, The Commercial Press, 2001. (in Chinese)
- [7] 吴宗济, 林茂灿, 实验语音学概要, 高等教育出版社, 1989
Wu Zongji, Lin Maocan, *The Summarize of Experimental Phonetics*, Higher Education Press, 1989. (in Chinese)
- [8] 殷治纲, “嗯、啊”类话语标记研究, 中文计算技术与语言问题研究, 电子工业出版社, 2007. (in Chinese)
Yin Zhigang, *The research of 'ng/a' type of discourse markers*, *The Research of Chinese Computational Linguistics*, Electronic Industry Press, 2007. (in Chinese)

(原载 NCMMSC2011 中国西安 2011 年 10 月)

Physiological and Acoustic Analysis of Vowel Sonority Sequencing Principle

Yin Zhigang

Linguistics institute, Chinese Academy of Social Sciences, Beijing100732

Abstract: This article researched in detail the influential factors of vowel sonority sequencing principle (SSP). According to the perceptual results of the sonority degree of low, middle and high vowels and the correlation between perception, and the absolute sound energy and the weighted sound energy based on the aural sensitivity, the research proved that the vowel sonority was related

Report of Phonetic Research 2011

not only to the sound energy, but also to the characteristics of frequency spectrum of vowels and the human's aural sensitivity to frequencies. In addition, the study showed that the vowels whose main energy concentrated on the sensitive range of hearing could be perceived louder than those whose main energy located outside.

Key words: vowel; sonority sequencing principle (SSP); sonority; intensity, sound energy; frequency spectrum; hearing; aural sensitivity; equal-loudness contour