

言语感知

南京大学

张安琪

2022/09/23

语音学

- 发音语音学 (Articulatory Phonetics)
 - 研究大脑和声道如何共同产生语音
- 声学语音学 (Acoustic Phonetics)
 - 研究语音的物理机制
- 听觉语音学 (Auditory Phonetics)
 - 研究耳朵和大脑如何感知语音

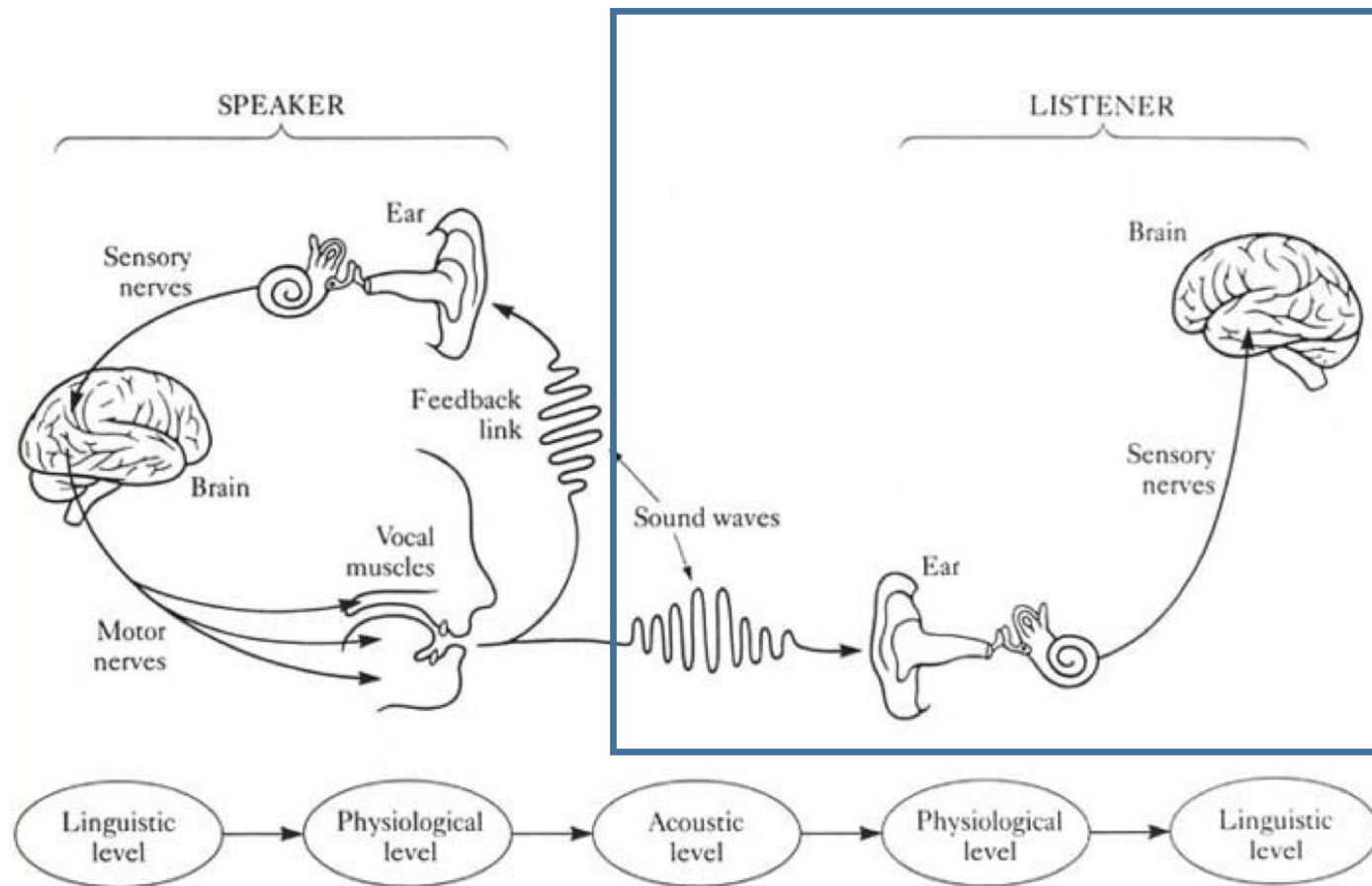
语音学

- ✓ • 发音语音学 (Articulatory Phonetics)
 - 研究大脑和声道如何产生语音
- 声学语音学 (Acoustic Phonetics)
 - 研究语音的物理机制
- 听觉语音学 (Auditory Phonetics)
 - 研究耳朵和大脑如何感知语音

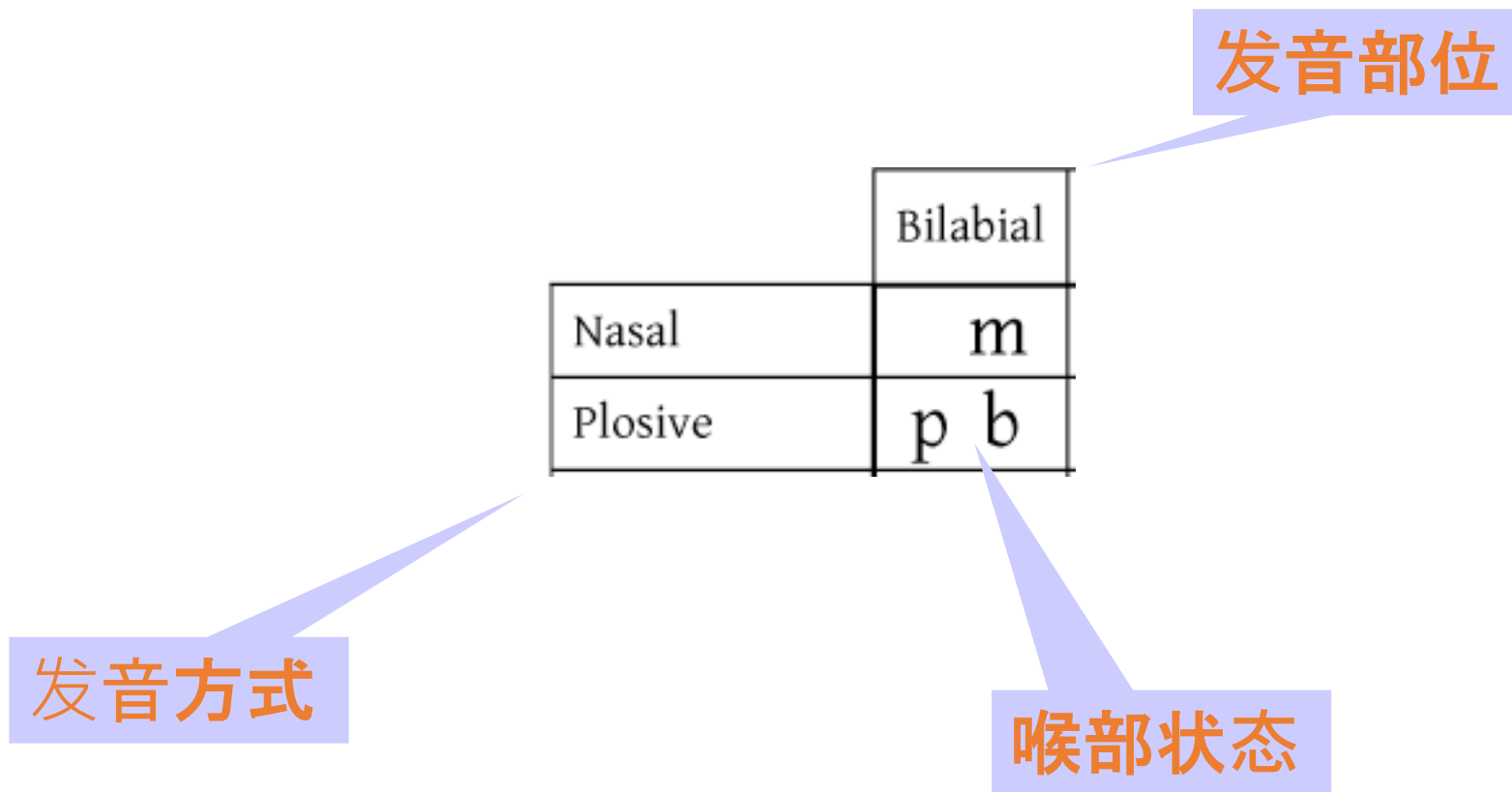
语音学

- ✓ • 发音语音学 (Articulatory Phonetics)
 - 研究大脑和声道如何共同产生语音
- 声学语音学 (Acoustic Phonetics)
 - 研究语音的物理机制
- 听觉语音学 (Auditory Phonetics)
 - 研究耳朵和大脑如何感知语音

本课关注



音位差异 (Phonological Contrasts)



音位差异

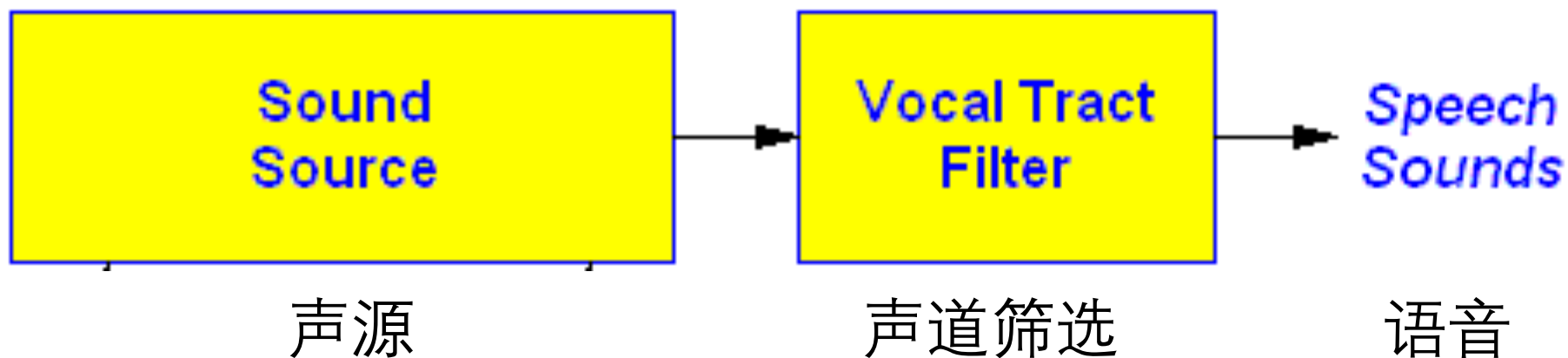
我们用**最小对立体**（**minimal pairs**）来鉴别音位差异：

- 喉部状态（有声化）：**cab** vs. **cat** / **sing** vs. **zing** / **which** vs. **witch**
- 部位：**bag** vs. **gag** / **sell** vs. **shell** / **sin** vs. **sing**
- 方式：**tell** vs. **sell** / **bail** vs. **mail** / **cat** vs. **catch**

➤ 差异何在？从何而来？

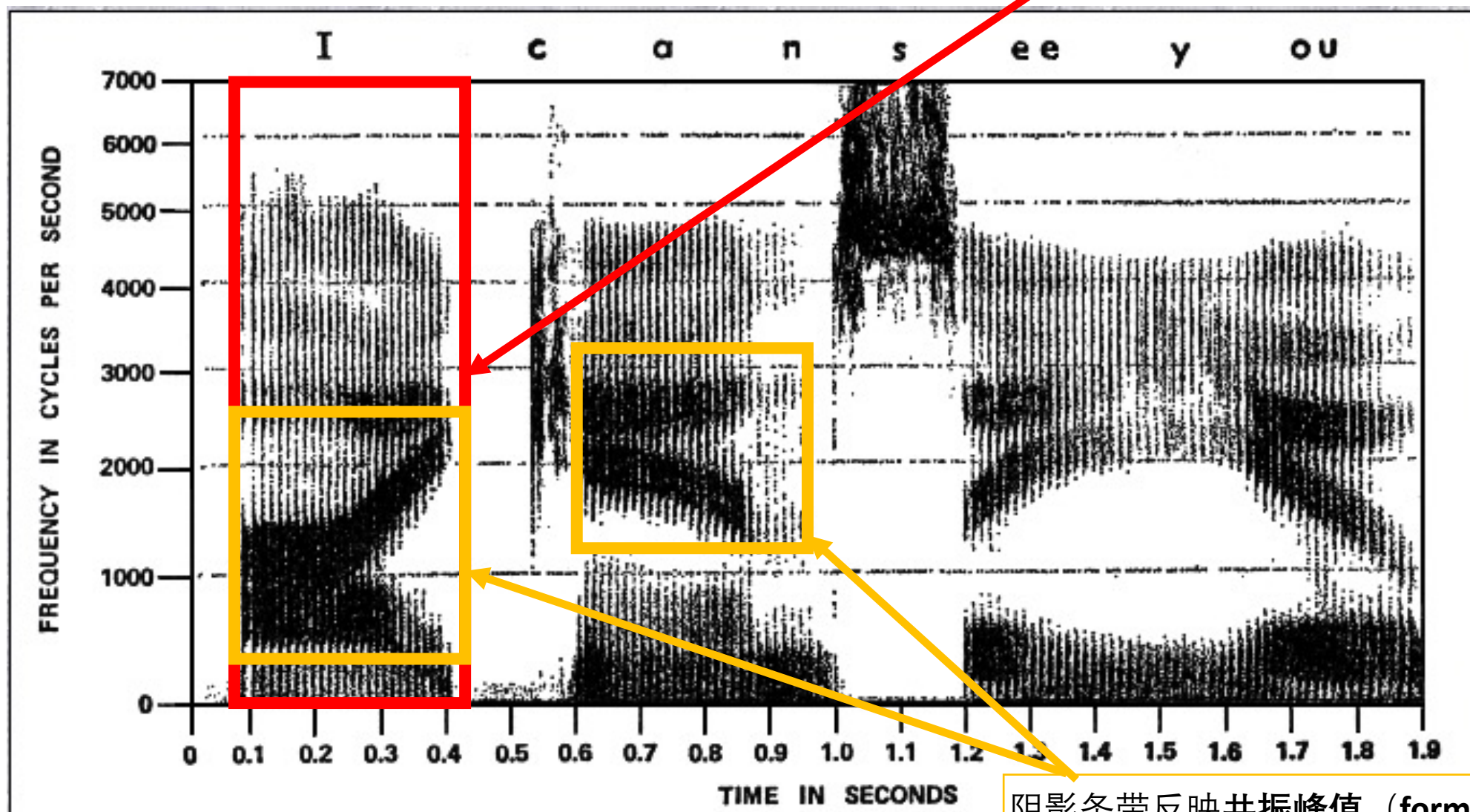
声学语音学 (Acoustic Phonetics)

来源筛选理论 (Source-Filter Theory)

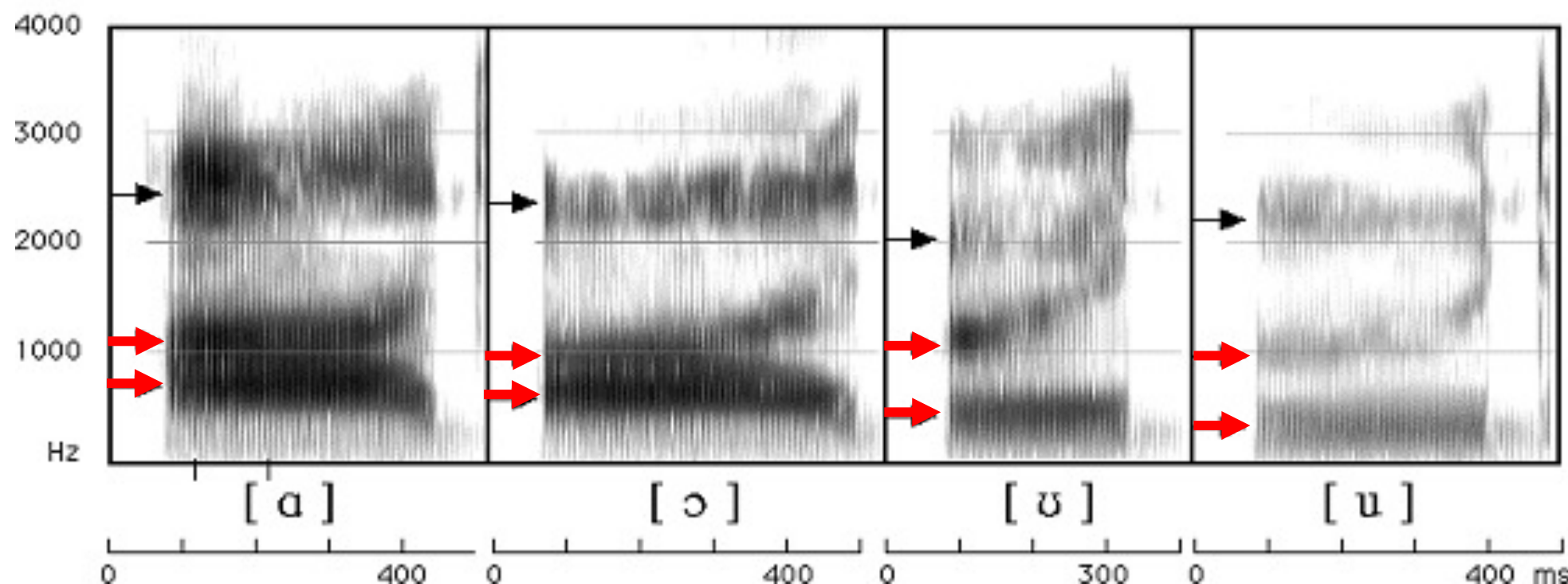
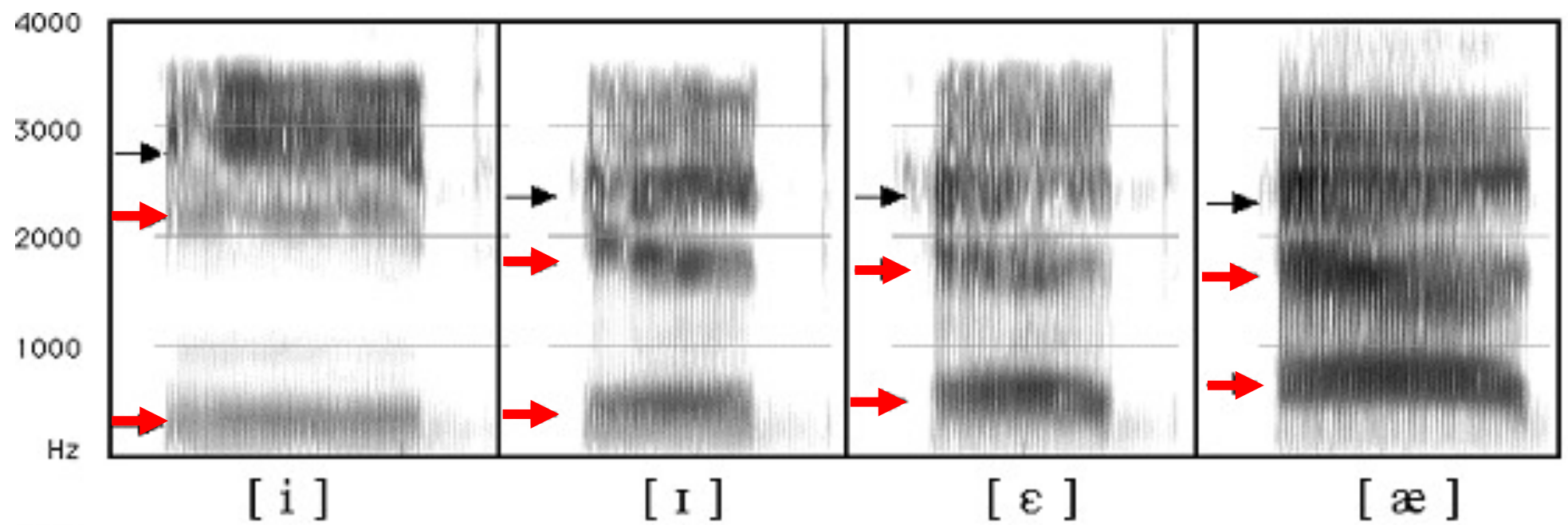


声学语音学

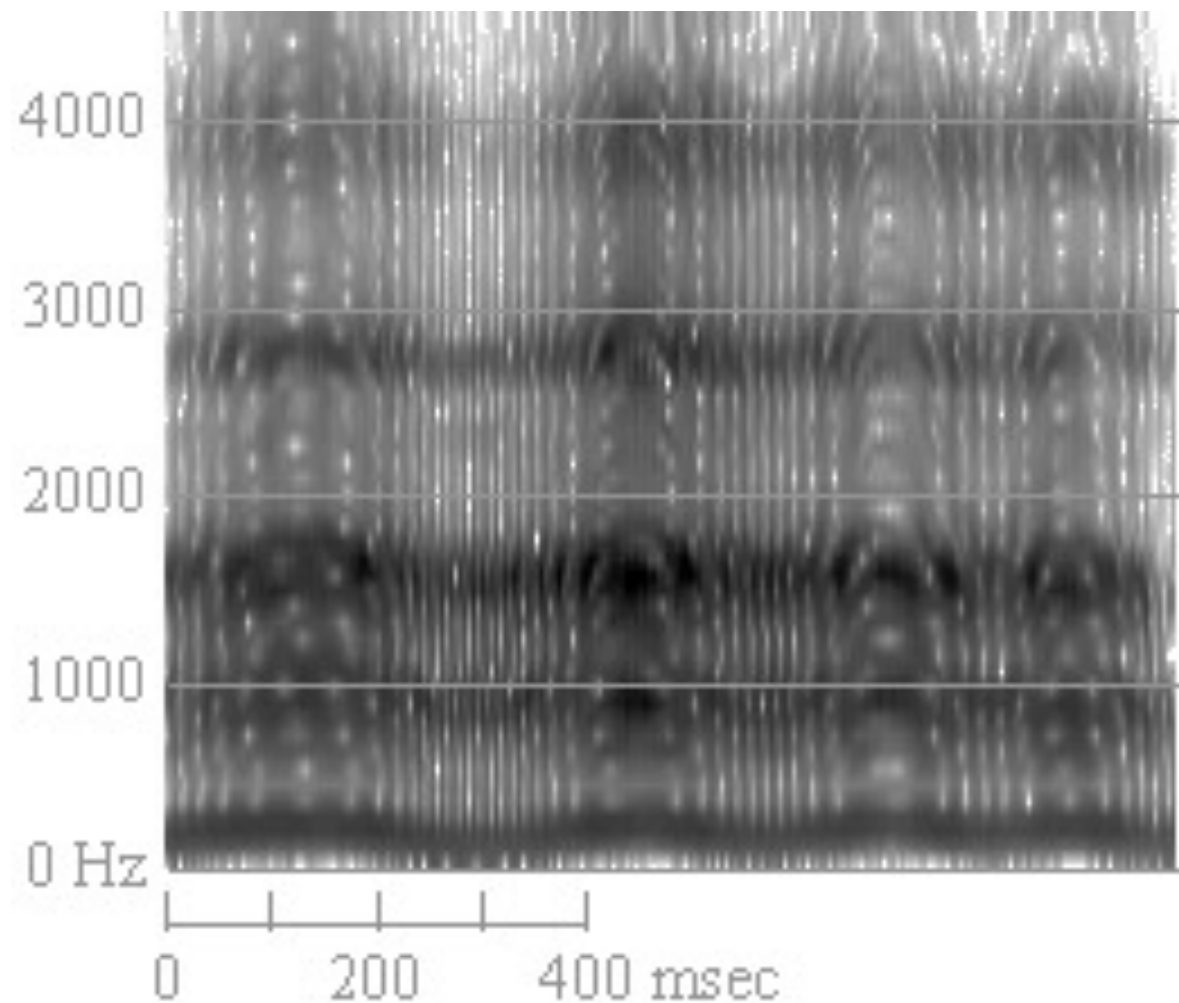
垂直条纹反映喉部搏动 (glottal pulses)



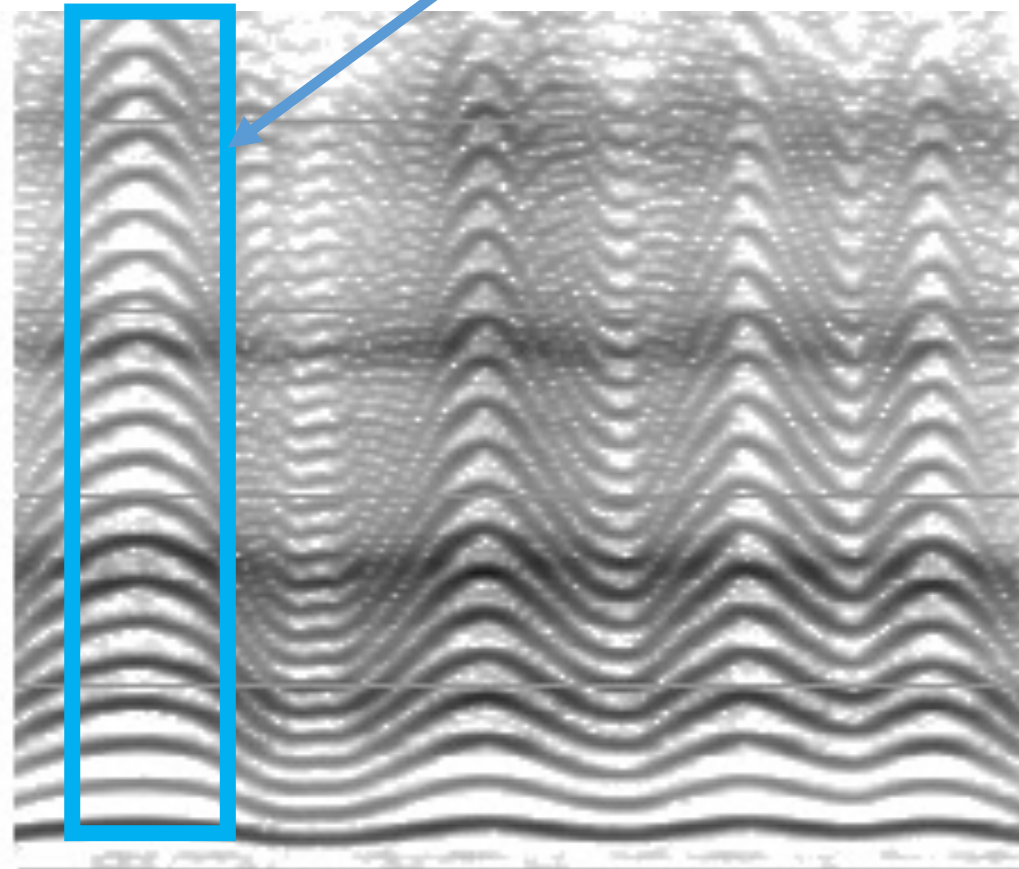
阴影条带反映共振峰值 (formants)



声学语音学



水平条纹反映谐波 (harmonics)



声学语音学

- **稳态共振峰 (steady state formants)**，或称谐振频率 (resonate frequency)，与元音形状 (vowel shape) 以及辅音发声情况 (consonant voicing) 相关联
 - 第一共振峰关联于元音的舌位高低 (vowel height)
 - 第二共振峰关联于元音的舌位前后 (vowel backness)
- **共振峰过渡 (formant transitions)** 与发音方式及部位相关联
- 短时的**声谱线索 (short-term spectral cues)** 包括谐波谱 (harmonic spectrum) 中的所有振幅线索 (amplitude cues) 和频率线索 (frequency cues)，与发音部位相关联
- **基频 (Fundamental frequency)** 和语调 (intonation) 相关联

你听到了什么？



Sample 1

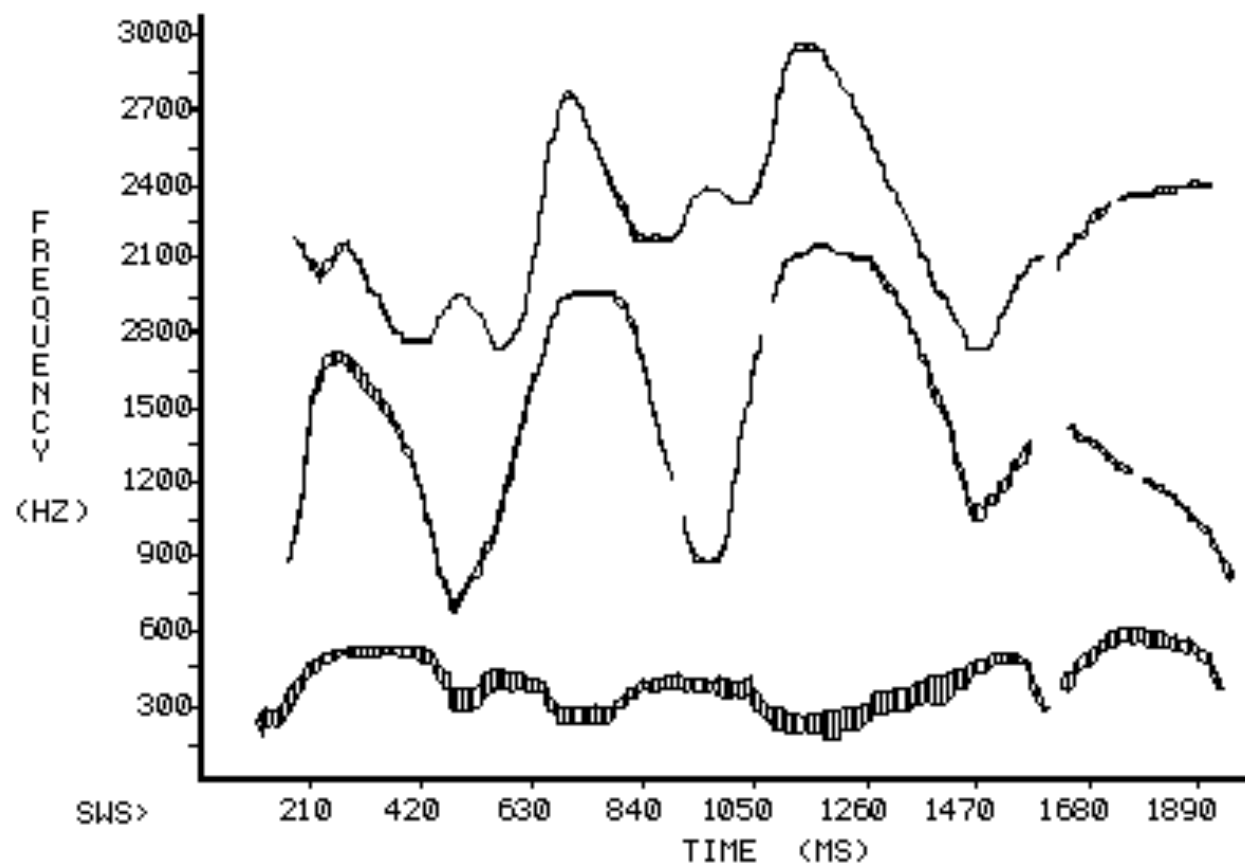


Sample 2

正弦波语音 (Sine-wave speech)

- 合成语音由两到三个时间变化 (time varying) 的正弦波模式 (sine-wave pattern) 构成
 - 保持共振峰的中心频率 (但没有共振峰的结构)
 - 能量的集中遵循共振峰的时间变化模式, 但能量局限在共振峰的中心频率

正弦波语音






◀ Tone 3  

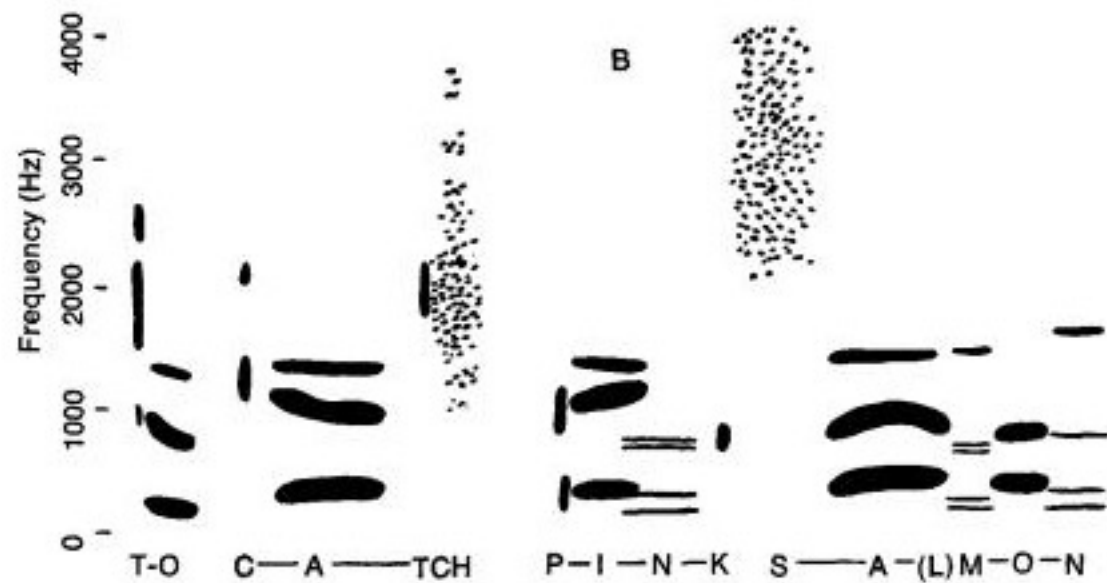
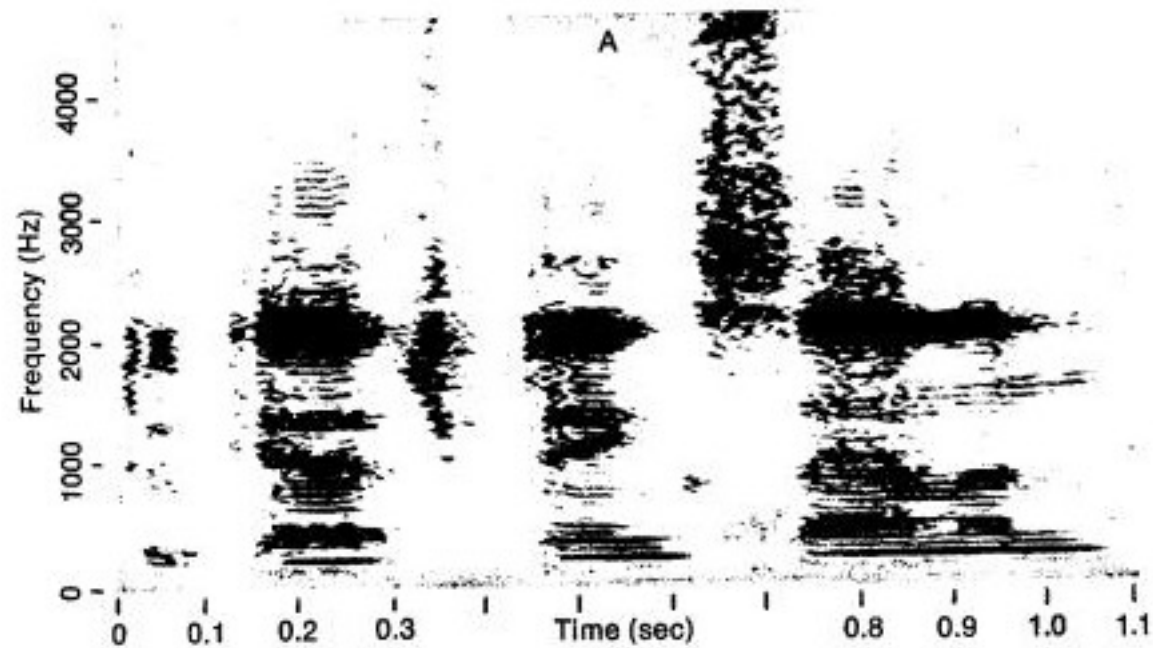
◀ Tone 2  

◀ Tone 1  




Remez, et al. 1981

Original Sentence  

Tone 1 + Tone 2 + Tone 3  



短语 “to catch the pink salmon” 的声谱图，上方是真人发声，下方是简化的人工语音。From Liberman et al., 1952, p498.

- Sentence 1: • I read a book today 
- Sentence 2: • Kick the ball straight,
and follow through 
- Sentence 3: • Rice is often served in
round bowls 

SWS揭示出的言语感知事实

- 感知并不（仅仅）依赖于原始的发声线索
- 我们对于语音的期望会影响我们听到这些语音的方式
 - “由上到下的” 处理方式 (Top-down processing)
- 我们有很强的感知学习能力。随着经验的积累，我们越来越容易辨别不同语音

可见，音位差异并不仅仅在于发声信号（acoustic signals）的不同。那么，当我们听到语音的时候，我们感知到了什么？

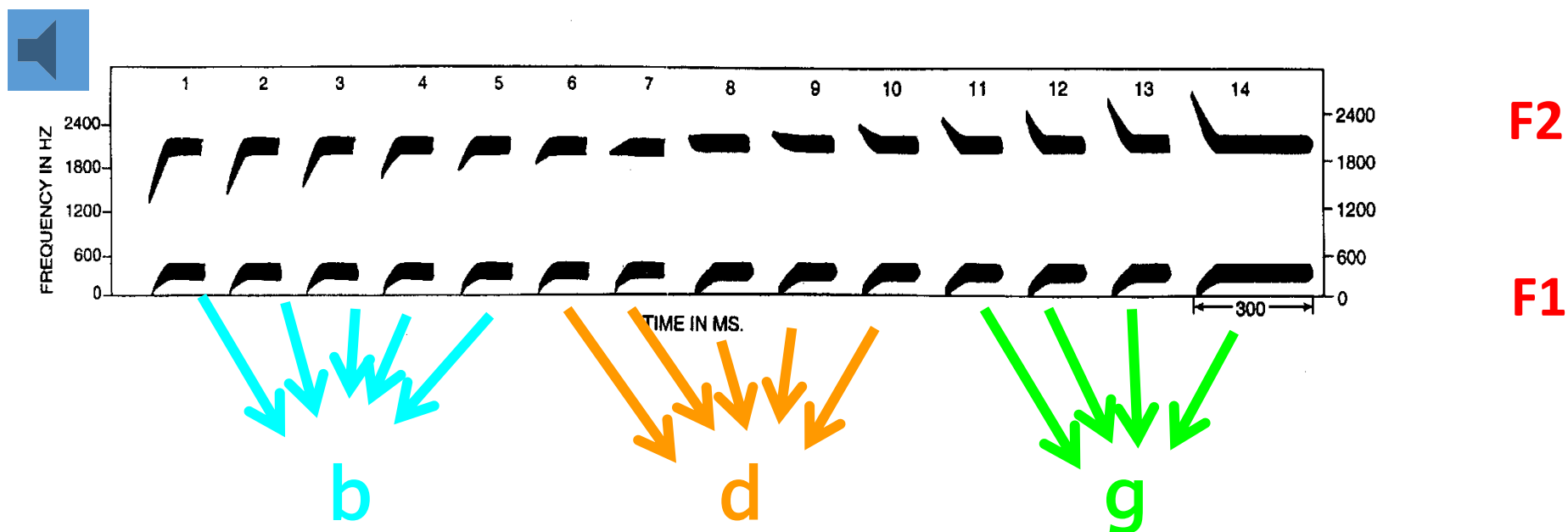
- 语音被组织成**离散的（discrete）意识单元**（音位phoneme），这些意识单元被表征为抽象特征的束（bundle of abstract features）
- 这些表征对应于**理想化的（过度发音的）发音“目标”（idealized (hyperarticulated) production ‘targets’）**，这些目标在常规的语音中可能会受到语音还原过程（phonetic reduction process）的影响

感知差异

- 同等变化 (equal changes) 如果是在与之没有关联的发声维度上发生的, 那么就是听觉印象 (auditory percepts)
- 可以通过两类实验研究不同音位之间的感知 “边界” :
 - 识别 (**Identification**) 实验: “你听到的是语音A, 还是语音B?”
 - 辨别 (**Discrimination**) 实验: “这些是相同的语音吗?”

测试三个闭塞音在发音部位上的差异

/b/ vs /d/ vs /g/:



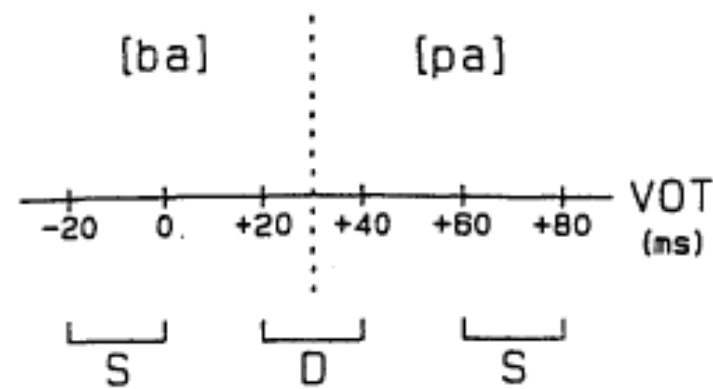
Lieberman, Harris, Hoffman, and Griffith (1957)
Journal of Experimental Psychology 54, 358-368
Picture courtesy of S. Hawkins

小结：范畴感知 (Categorical Perception)

物理连续体的组成部分被感知为离散的范畴 (discrete categories)。我们可以把这些范畴的抽象单元称为**音位** (phoneme)。

“语言特有” 还是 “人类特有” ？

- 日本鹌鹑与南美栗鼠的发声
- (Kuhl & Miller, 1984; Diehl et al., 1988)
- 语言学关注的某些发声区别可能是共通的，例如天生就有能力区分与一般感知边界（general perceptual boundaries）相符合的差异（contrasts），而不是语言学本身特有的边界（linguistic specific per se）



范畴感知的母语效应

- 成年人对语音的范畴感知依赖于听音者母语中的音位差异 (phonological contrasts)
 - 例如，对于你听见的/b-p/ VOT 连续体，讲英语的听音者的识别和辨别边界约为+25 ms VOT

嗓音起始时间 (Voice Onset Time) (w.r.t. cons. Release)

-100 0 +100 (msec)

Spanish:



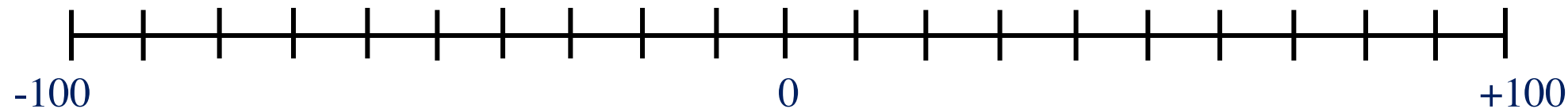
English



Thai



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21



/b/

/p/

范畴感知的母语效应

- 显然，成年人对于语音的感知是基于语言的
(**language-specific perceivers** of speech)
- 区分差异的能力是习得的吗？
- 我们是否能失去或学会忽略某些区别？

我们如何“认识”这些范畴？

- 区分差异的能力是习得的吗？
- 我们能否失去或学会忽略某些区别？

婴儿感知



- **高振幅吸吮法 (High-Amplitude Sucking Procedure)**
(Eimas et al. 1971)
- 受试对象是年龄很小的婴儿（两到三个月大）

高振幅吸吮法

- 目标：比较婴儿对两种合成语音（synthetic speech sound）的区分度，这两种合成语音的区别在于其嗓音起始时间（VOT）所具有的固定差异。实验有如下两个条件：
 - 条件1：两种刺激应该分别位于成年人音位边界的两端（条件20D）
 - 条件2：两种刺激必须来自同一个音位范畴（图表中的条件20S）

高振幅吸吮法

受试对象：

- 一个月大和四个月大的婴儿
- 把每个年龄段的八个婴儿随机分到两个条件下
- 把每个年龄段的十个婴儿分到控制条件下

高振幅吸吮法

流程：

- 在消声室中进行
- 婴儿斜躺在椅子上；可以调节位置的机械臂把奶嘴放到他/她嘴里
- 压力传感器测量奶嘴内的气压，并将信号传输到计算机
- 实验程序检测吸吮状况，计算其振幅

高振幅吸吮法

- 当婴儿感知到两类刺激的不同之处时，他们通常会在刺激改变之后的几分钟内吸吮的次数，以此表示对新情况的反应
- 当刺激来自于不同的成年人音位范畴时，婴儿能注意到的差异更大

高振幅吸吮法

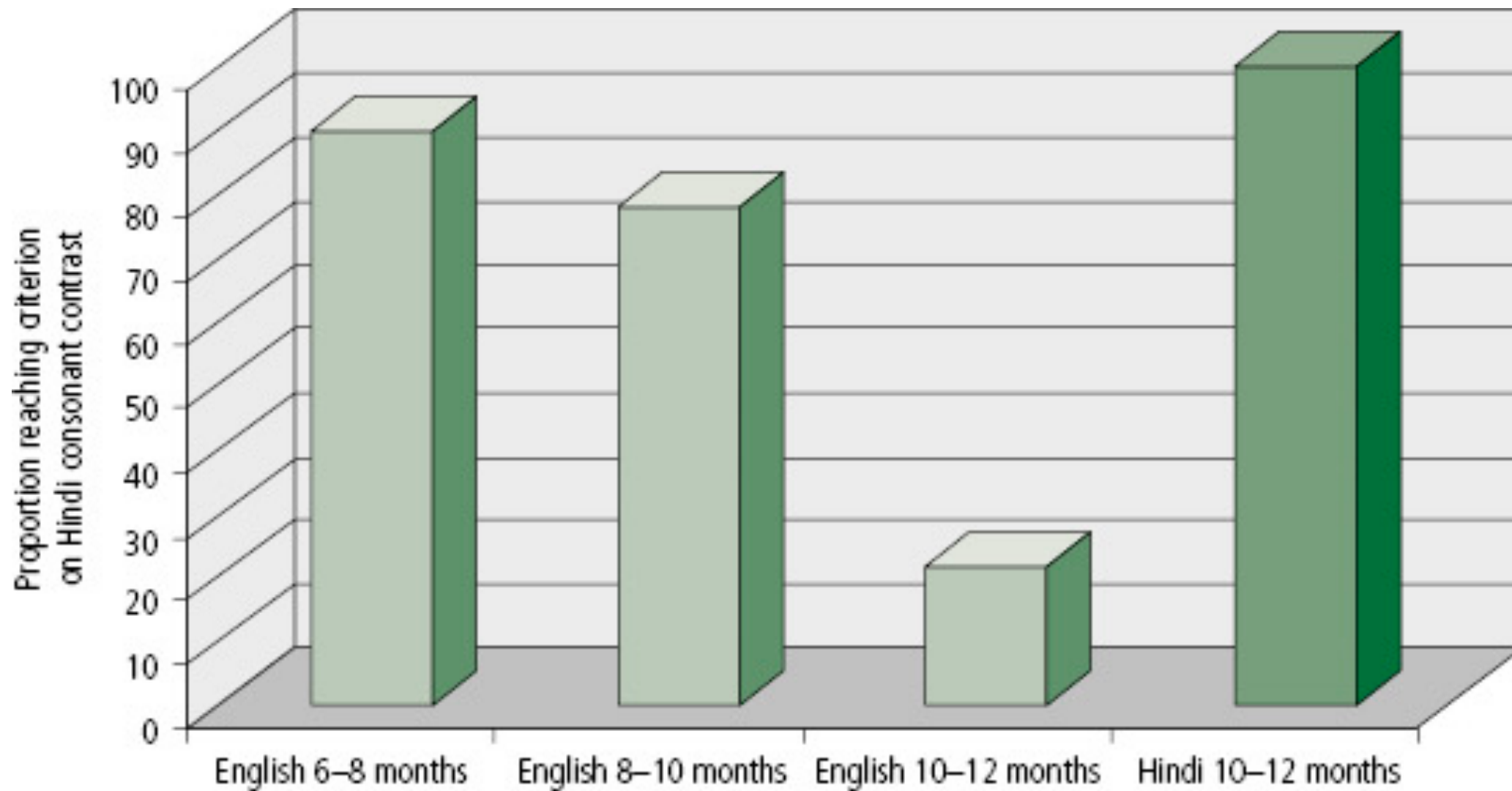
研究者指出：

- 婴儿以一种接近范畴感知（categorical perception）的方式辨别浊声与清声的闭塞辅音
- 因为婴儿接触到的语音有限，缺乏发音的经验，所以这种符合语言学模式的范畴感知可能是与生俱来的

依赖语言的感知的发展

- 孩童什么时候失去对非母语的音位的辨别能力？

转头偏好法 (Preferential Head-Turning Procedure)



Werker and Tees, 1984

小结：母语的范畴感知的发展

- 婴儿能辨别许多跨语言使用的语音区别
- 这种能力随着他们对特定语言的经验积累而下降
- 至一周岁前，他们会被调适到（tuned to）自己的母语




处理可变性

- 我们知道，无论是单个说话人的语音流，还是不同说话人之间的语音流，都是高度可变的
 - 可变性（variability）如何影响我们对音位范畴的感知？

范畴感知的语境效应

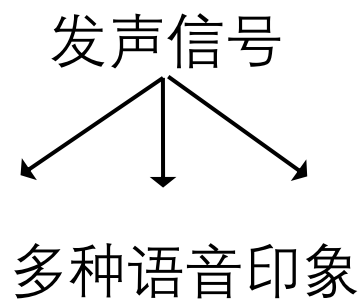
Ladefoged and Broadbent (1957)

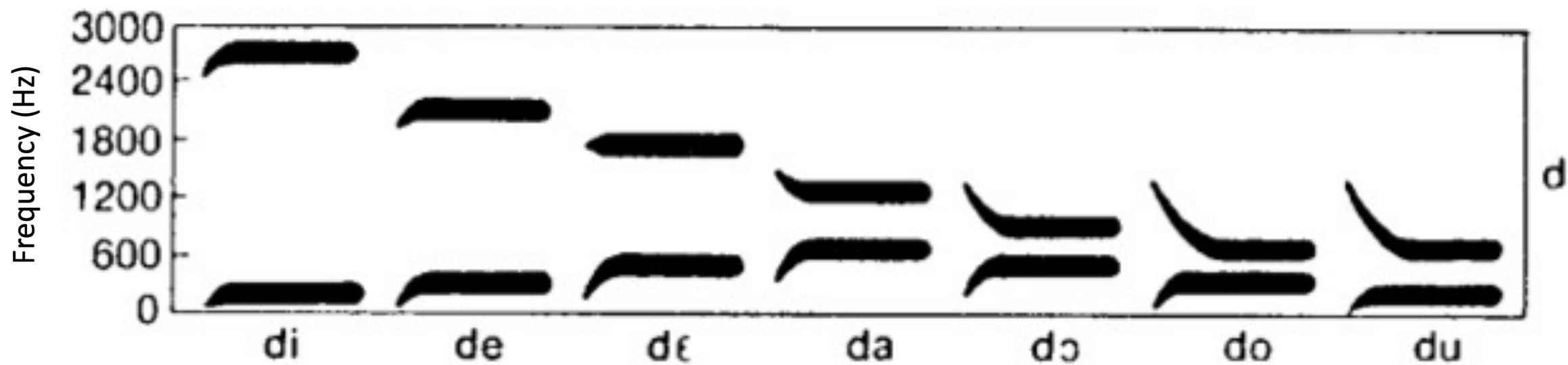
- “Please say what this word is: bit bet bat but”

		F1 of CARRIER
bet		200-380 Hz
bit		380-660 Hz
		

范畴感知的语境效应

同一种发声信号（acoustic signal）并不总能唤起相同的语音印象（percept）。

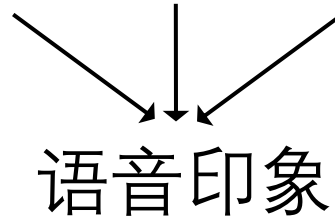




音节/di/ 与 /du/的程式化的共振峰过渡 (stylized formant transitions)

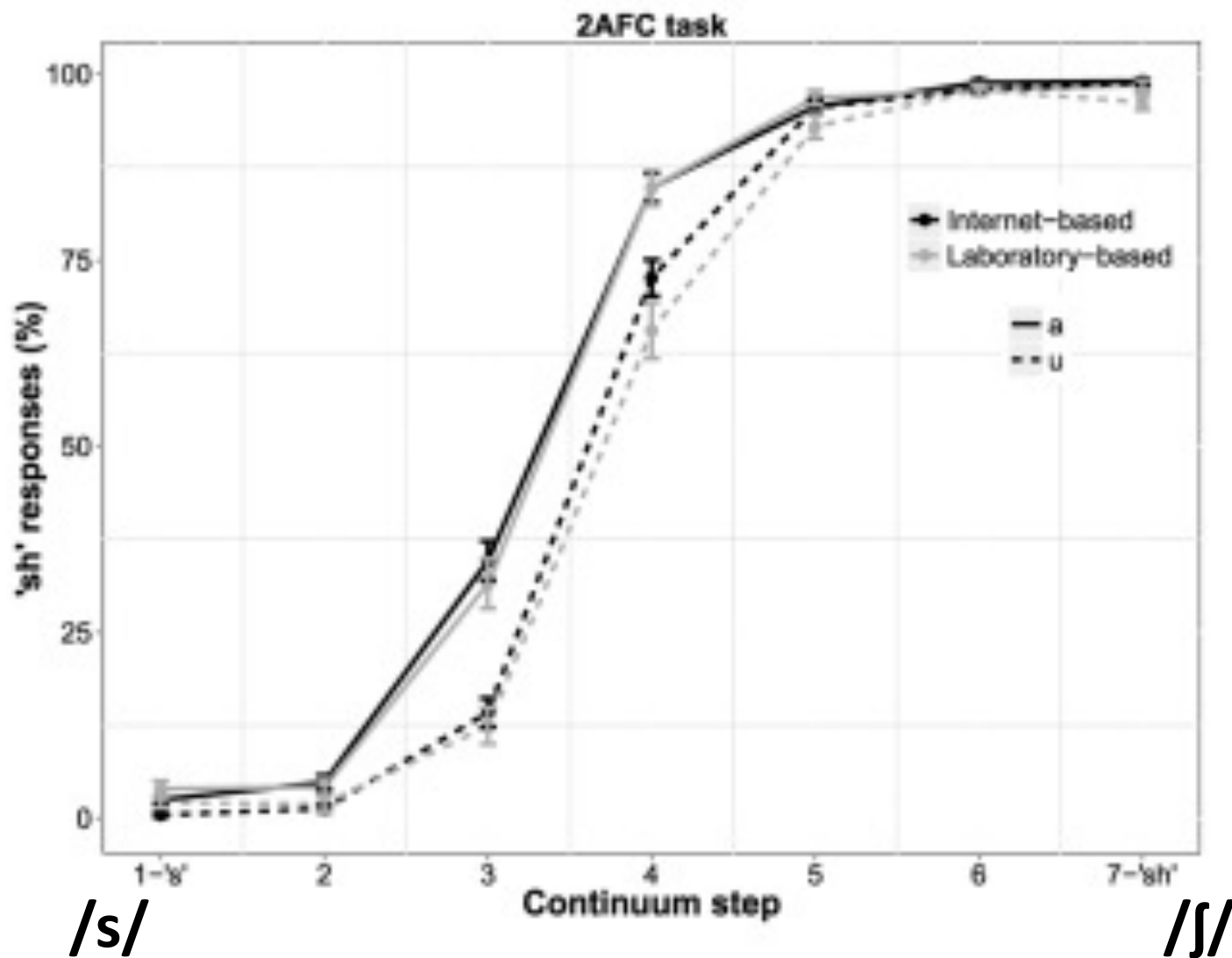
不同的发声信号可能唤起相同的语音印象

发声信号



引起同一辅音印象（percept of the same consonant）的共振峰过渡的方向与程度（direction and extent of formant transitions）因辅音出现的元音语境（vowel context）的不同而不同

协同发音的感知补偿 (Perceptual Compensation)



小结：范畴感知的语境效应

- 范畴感知是离散的（discrete），也是**相关的（relative）**
- 听音者对**感知的语境高度敏感（highly sensitive to context）**
- 说话人在某种程度上会**补偿（compensate）**发声信号的变化
 - 与说话人相关的可变性（Speaker-related variability）
 - 协同发音的可变性（Coarticulatory variability）
 - ？

影响言语感知的其他因素

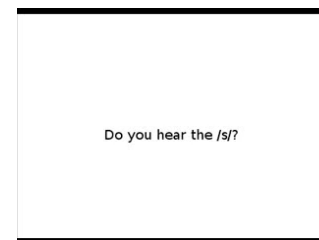
- 我们还可以在言语感知中运用哪些知识？

范畴感知的多模态效应 (Multimodal Effect)

- Information from the auditory stream is not the only information we use to form an auditory percept. Information from other modality is integrated very rapidly.
- 要想形成听觉印象 (auditory percept)，我们并不仅仅从听觉流中获取信息。来自其他模态 (modality, 感觉模式) 的信息也会被迅速整合。



音位恢复效应 (Phoneme restoration effect) (Warren & Warren 1970)



音位恢复效应

(Warren & Warren 1970)

- 参与者听一段口语，其中有一个音位被一声咳嗽代替了
“The state governors met with their respective legi*latures convening in the capital city.”
- 参与者或是没有注意到丢失的语音，声称不存在丢失的语音；或是误以为丢失的是其他语音

Phoneme restoration (cond.)

模糊的词语片段

...the *eel was on the orange. (peel)

...the *eel was on the axle. (wheel)

...the *eel was on the shoe. (heel)

...the *eel was on the table. (meal)

参与者能够理解整个句子，并且报告说“听到”的是对应的正确的句子。

总结

- 我们可以把听到的辅音归类为**离散的范畴（音位）**
- 对于音位范畴的感知**在幼年**就开始发展，并被调适到其母语
- 言语感知既依赖于**自下而上的（声学）**线索和**自上而下的知识**
- 听音者在很大程度上会把交际渠道（communication channel）中的变化**规范化处理（normalize）**