# 基本语音学知识

1. syllabic音节 syllabic音素 logographic音标 alphabetic字母的 consonantal辅音的
2. syllabic——Japanese（音节语言） alphabetic——Roman（字母语言）
3. 日语的音节是由一组汉字的草书形式修改而成，这些汉字用来表示声音。
4. 口语是由更小的语言单元组成的，是构成我们所有现代音位学理论基础的Ur-theory。
5. 语音学研究——语言的声音如何由声带产生，如何在声学上实现，如何数字化和处理。
6. 在不同上下文中音素发音是不同的
7. 单词的发音分为音素phones和语段segments
8. International Phonetic Alphabet(IPA)国际音标致力于表示世界所有语言的发音，根据IPA的原则，同一句话可以有不同的方法翻译

# 发音

1. 空气通过声带时，两块声带合在一起会震动，声带震动产生的语音为浊音，不震动产生的语音为清音。
2. 语音可分为consonants(辅音)和vowels(元音)两大类，consonants在产生时受到的阻挡较小，一般为浊音，比辅音响亮且持续时间长。
3. 辅音中，发音部位图把语音分为等价类：唇音（labial）、齿音（dental）、齿龈音（alveolar）、上颚音（palatal）、软腭音（velar）、喉音（glottal）
4. 元音分类(元音在语音识别中较为重要)：
   1. 塞音(stop/plosives)：closure->release
   2. 鼻音(nasal)：软腭下降，气流通过鼻腔流出
   3. 擦音(fricative)：擦塞音(affricates)
   4. 半元音(approximant)：边音(lateral)
   5. 颤音(tap或flap)

# 音节

1. ﻿辅音和元音结合成音节，音节由一个元音（或响音sonorant）和周围一些联系非常紧密的辅音结合而成。
2. 音节组成——音节核（处于音节核心部位的元音）、音节头（onset，如果音节头有一个以上的辅音则称为complex onset）、音节尾（coda）。音节核+音节尾=韵（rime）或韵脚（rhyme）
3. phonotactics是指音素彼此间跟随关系的约束条件
4. 重读音节（accented syllable），**对应语言标志是音高重音（pitch accent）**，被凸显的单词或音节为音高重音的负荷者（bear）。（sentence stress）
5. 重读中的一个重要的因素——词重读需要在语音字典中表示出来
6. 非重读元音可能被弱化以至于进一步成为弱化元音 （发音动作没有全部完成）。常见的弱化元音是非重读央元音（schwa）——[ax]
7. 任何元音特别是双元音尽管在非重读位置上也能保住完整的特性
8. 凸显度有5个潜在级别：重读(accented)、重音(stressed)、第二重音(secondary stress)、实足元音(full vowel)、弱化元音(reduced vowel)
9. Phoneme音位，用户表示不同上下文中[t]的相似点。音位在不同上下文的表层实现称为音位变体(allophones)。/t/用来表示音位，在ARPAbet中用因素集合表示抽象的音位单元。
10. 相邻的音会彼此汇合影响，称为弱化（reduction）或省音（hypoarticulation）
11. 语音同化就是改变某一个语音片段使之与周围的语音片段的发音更加相似，其中一种常见的同化是palatalization
12. /t//d/容易发生脱落
13. 当一个音受到了相邻音素的发生器官动作的影响而发声的时候，我们就说这样的发音受到了协同发音（coarticulation）的影响
14. 区别特征（distinctive features）特征为二元变量，特征总体可以分为两种：发音部位（place）和发音方法（manner）
15. 发音部位：唇音性，舌尖性，舌面性
16. 发音方法：辅音性，连续性，响音性
17. 元音特征：高音性，低音性，后音性，圆唇性
18. 音位规则（phonological rule）
19. 影响语音变异的因素：语速（每秒的音节数），单词的频率和可预见性，心态，单词的语音学音系学形态学

# 信号和声学语音

1. frequency--cycles per second(Hz)
2. amplitude（振幅）
3. period（周期）T = 1/f
4. We represent sound waves by plotting the change in air pressure over time. 使用空气压力相对于时间的变化来代表声波
5. 声波波形图可以测度空气分子的compression（压缩量），rarefaction（吸入量）
6. 声波的数字化表示
7. 空气压强->模拟电信号
8. 模拟电信号->数值型号（analog-to-digital conversion）：抽样（sampling）和量化（ quantization ），对于每个信号周期应该至少有2个采样（一个测量声波正侧部分，另一个测量声波负侧部分），少于两个可能会完全地遗漏声波频率
9. 能测量的最大频率的波就是频率等于抽样率一半的波，对于给定的抽样频率的最大频率为Nyquist频率。
10. 由于大多数人类语言的频率低于10kHz，因此必须有20kHz的抽样率。（电话的语音经过开关网络过滤，频率低于4kHz，对于电话带宽的语音8kHz的采样率），对于麦克风通常采用16kHz的抽样率。
11. 对振幅的实数表示为整数的过程称为量化，由于整数有最小颗粒度，因此在量化范围的值都相同的表示
12. 对于振幅存储形式的影响参数是采样频率和采样大小，电话语音8kHz——8bit，麦克风16kHz——16bit。另一个影响参数是通道，对于立体声可以将所有的通道存到一个文件中也可以存到不同文件中。最后一个参数是每个采样是线性存储还是压缩存储。通常的对电话语音的压缩方法是μ-law。对于对数压缩算法的直观理解是：人的听觉在音高较低时更加敏感。非log压缩的线性值通常指linear PCM。
13. wav文档结构：有44个字节的标题以及数据块
14. 声波的主峰有声带开启而形成，声带振动的频率（复杂波的频率）称为基音频率（基频F0）。
15. 声波波形的纵轴为空气压强大小，零值表示标准大气压强
16. 在计算一段时间内的平均振幅时使用均方根
17. 信号的强度（power）与振幅的平方相关是振幅的均方
18. intensity音强是信号强度对于听觉阈值的归一化
19. 语音的音高——基音频率，一般来说基音频率越高，音高越高。但是由于人的感知在不同频率的敏锐性不同，所以音高和频率并不完全线性相关
20. 语音的响度——信号强度，也成非线性关系因此采用μ-law压缩音高
21. mel是音高单位，是在感官上的音高单位
22. 抽取F0的算法称为基频抽取，常用自相关函数法
23. 元音的波形发音时间较长，波峰较高
24. 塞辅音是一个沉静的间隙后跟随一个振幅上的爆破
25. 擦音为不规则的噪声波形。
26. 一个信号的声谱（spectrum）可以表示信号的频率成分和频率成分的振幅，
27. 对声波信号采用傅里叶变换可以得到平滑的声谱
28. 元音的语谱图（spectrogram），频谱上x轴表示时间，y轴表示频率的赫兹数。频谱一个点上的暗度表示频率成分的振幅大小。暗点表示有较高的振幅，亮点表示有较低的振幅
29. 每个元音有一系列暗色条纹在多个频率带上，每个暗色条纹称为formant
30. 共振峰的不同位置可以把不同的元音区分来开，还可以区分鼻音音素和边音音素。
31. formant是由口腔共鸣引起的。
32. 声源滤波器模型，声源产生脉冲，声腔（滤波器）使脉冲成型，声腔像一个滤波器或放大器，通过改变声腔的形状可以使不同频率的声音得到放大，声门在产生脉冲时会产生基波以及与基波频率成整数倍的谐波，通过改变声腔的形状是的不同的谐波得到放大，有同样基频的波在通过不同声腔位置的时候会引起不同的谐波得到放大