



人类听觉特点



语音的产生及数字模型



听觉系统和语音感知



语音的质量评价



听觉系统和语音感知

人类说话和收听的过程

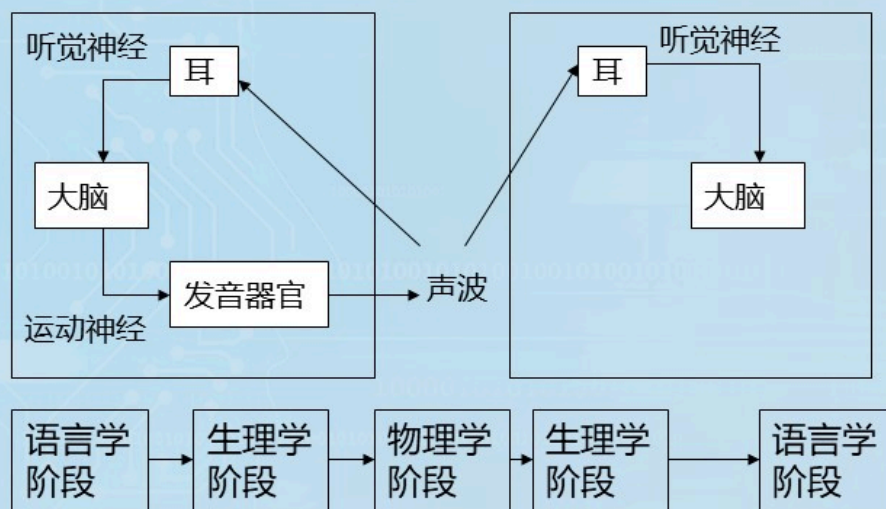


图6 语言通道



影响语音感知的主要因素



听觉范围

频率范围

正常人可听声音的频率范围为0.016~16kHz，年轻人可听到20kHz的声音，而老年人可听到的最高频率为10kHz左右

感觉域

容忍的最高声压。当声压高到一定程度时，耳朵会出现不适感。对正常人而言，一般取120dB为不适阈，140dB为痛阈，且与频率无关

强度差阈

正常人对频率固定的声音所能辨别的最小强度差值



影响语音感知的主要因素



音调

对于频率低的声音，听起来感觉它的音调“低”，而频率高的声音，听起来感觉它的音调“高”

音调与声音的频率并不成严格的正比关系，它还与声音的强度及波形有关



影响语音感知的主要因素



掩蔽效应

当人耳听到两个强度不同的声音时，强的声音的频率成分会影响人耳对弱的声音的频率成分的收听

频率掩蔽

通常，低音容易掩蔽高音，而高音掩蔽低音较难

时间掩蔽

强声音后面的弱音容易被掩蔽；强声音前面的弱音也容易被掩蔽



影响语音感知的主要因素



掩蔽效应

噪音对单音的掩蔽

一个单音可以被以它为中心频率，具有一定频带宽度的连续噪音所掩蔽

噪音掩蔽的临界带宽

如果在这一频带内噪声功率等于该纯音的功率，这时该纯音处于刚能被听到的临界状态，则称这一带宽为临界带宽

填空题 3分

掩蔽效应分为频域掩蔽和 [填空1] , 或 [填空2] 和异时掩蔽, 后者又分为超前掩蔽和 [填空3] 。

单选题 1分

掩蔽效应分为同时掩蔽和()。

- ☐ A 频域掩蔽
- ☐ B 超前掩蔽
- ☐ C 滞后掩蔽
- ☒ D 异时掩蔽

单选题 1分

异时掩蔽可分为()和滞后掩蔽。

- ☐ A 同时掩蔽
- ☐ B 时域掩蔽
- ☐ C 频域掩蔽
- ☒ D 超前掩蔽

人耳掩蔽效应

一个较弱的声音(被掩蔽音)的听觉感受被另一个较强的声音(掩蔽音)影响的现象称为人耳的“掩蔽效应”。人耳的掩蔽效应。被掩蔽音单独存在时的听阈分贝值,或者说在安静环境中能被人耳听到的纯音的最小值称为绝对闻阈。实验表明,3kHz—5kHz绝对闻阈最小,即人耳对它的微弱声音最敏感;而在低频和高频区绝对闻阈要大得多。在800Hz--1500Hz范围内闻阈随频率变化最不显著,即在这个范围内语言可懂度最高。在掩蔽情况下,提高被掩蔽弱音的强度,使人耳能够听见时的闻阈称为掩蔽闻阈(或称掩蔽门限),被掩蔽弱音必须提高的分贝值称为掩蔽量(或称阈移)。

频域掩蔽效应

一个强纯音会掩蔽在其附近同时发声的弱纯音,这种特性称为频域掩蔽,也称同时掩蔽(simultaneous masking)。如,一个声强为60dB、频率为1000Hz的纯音,另外还有一个1100Hz的纯音,前者比后者高18dB,在这种情况下我们的耳朵就只能听到那个1000Hz的强音。如果有一个1000Hz的纯音和一个声强比它低18dB的2000Hz的纯音,那么我们的耳朵将会同时听到这两个声音。要想让2000Hz的纯音也听不到,则需要把它降到比1000Hz的纯音低45dB。一般来说,弱纯音离强纯音越近就越容易被掩蔽;低频纯音可以有效地掩蔽高频纯音,但高频纯音对低频纯音的掩蔽作用则不明显。

由于声音频率与掩蔽曲线不是线性关系,为从感知上来统一度量声音频率,引入了“临界频带(criticalband)”的概念。通常认为,在20Hz到16kHz范围内有24个临界频带。

时域掩蔽效应

除了同时发出的声音之间有掩蔽现象之外,在时间上相邻的声音之间也有掩蔽现象,并且称为时域掩蔽。时域掩蔽又分为超前掩蔽(pre-masking)和滞后掩蔽(post-masking),如图12-05所示。产生时域掩蔽的主要原因是人的大脑处理信息需要花费一定的时间。一般来说,超前掩蔽很短,只有大约5~20ms,而滞后掩蔽可以持续50~200ms。这个区别也是很容易理解的。



影响语音感知的主要因素



语音感知和理解

语音感知和理解是一个复杂的过程，它包含自下而上和自上而下的过程，前者在于收集语音信号中所含有的信息，但光靠这些信息还不足以进行语言理解，还要由收听者根据语法和句法知识对语音信息进行理解。



影响语音清晰度的主要因素



语音强度

语音强度，影响对语音的正确辨别率



对语音的掩蔽作用

对于纯音掩蔽而言,低频纯音对语音的掩蔽要大于高频纯音。



影响语音清晰度的主要因素



频率选择性

虽然语音信号的大部分功率包含在低频分量之中，但是它们对清晰度的贡献并不是很大

去掉高频成分
对于辅音清晰度影响大

去掉低频成分
对于元音清晰度影响大



影响语音清晰度的主要因素

频率选择性



用低通滤波器

保留5KHz以下的频率成分，清晰度不受影响

保留1.5KHz以下的频率成分，清晰度约下降一半

保留200Hz以下的成分时，清晰度降为零



影响语音清晰度的主要因素

频率选择性



用高通滤波器

保留400Hz以上的频率成分清晰度基本不受影响

保留1000Hz以上的部分，则语音信号的功率可能损失了约80%，
但清晰度却仅下降了10%

保留2300Hz以上的频率成分，清晰度下降一半左右

保留6kHz以上的频率成分时，清晰度降为零



影响语音清晰度的主要因素



限幅

峰值削波

将幅度超过某一门限
的值限制在门限上

中心削波

将幅度小于某一门限
的值置为零



影响语音清晰度的主要因素



限幅的影响

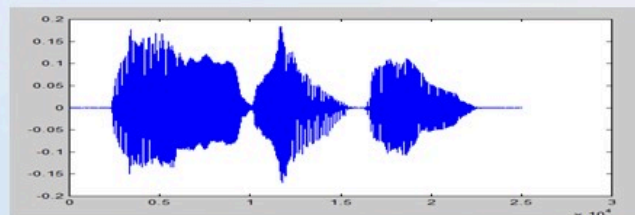
在峰值无限削波的情况下，仍然相当好地保留了单词的清晰度
削去声波幅度的一半，清晰度几乎降为零

结论：语音信号中的大部分信息都保存在其低幅值的部分

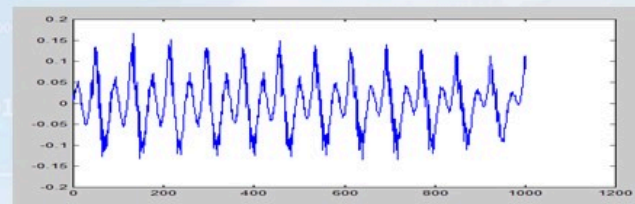


语音信号的特征波形

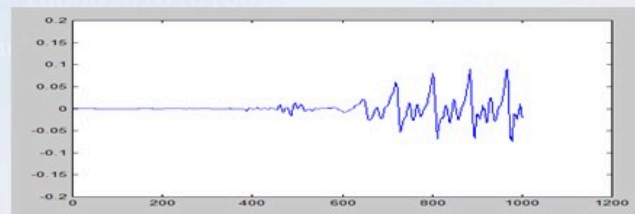
静息波



准周期波



噪声波



脉冲波



语音信号的特征波形



元音的共振峰

元音的产生是通过声带的准周期振动，经声道调制，由口鼻辐射出来。不同的元音，其频谱特性是不同的。各个元音的差异，可以用元音的前三个共振峰频率 f_1 、 f_2 、 f_3 来表示。 F_1 分布在290 Hz ~ 1 kHz范围内， F_2 分布在500 Hz ~ 2.5 kHz范围内， F_3 分布在2.5 ~ 4 kHz范围内。