人机交互技术：
BCI自适应性及特征提取

未知参数

在大部分BCI问题中，最佳参数设置是未知的：

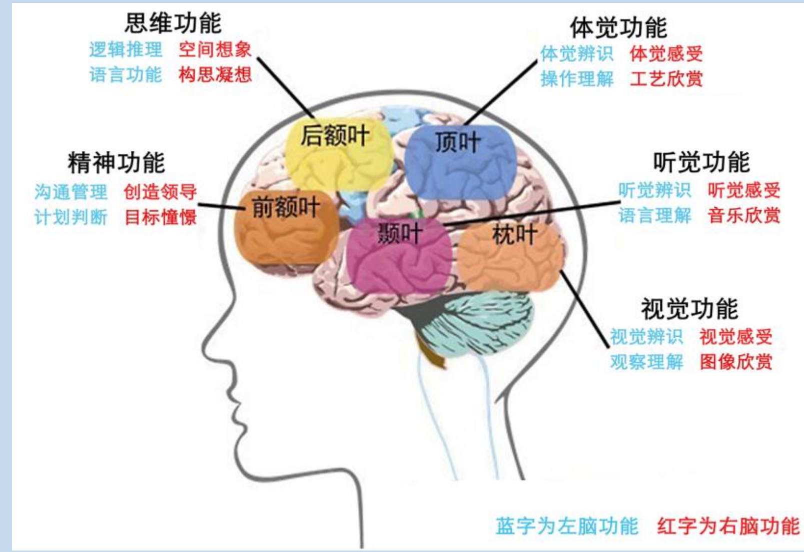
- 依赖于某些可变参数：电极位置，认知状态
- 随实验对象、任务而改变
- 依赖于一些不易测量的因素：大脑功能地图，大脑沟回等

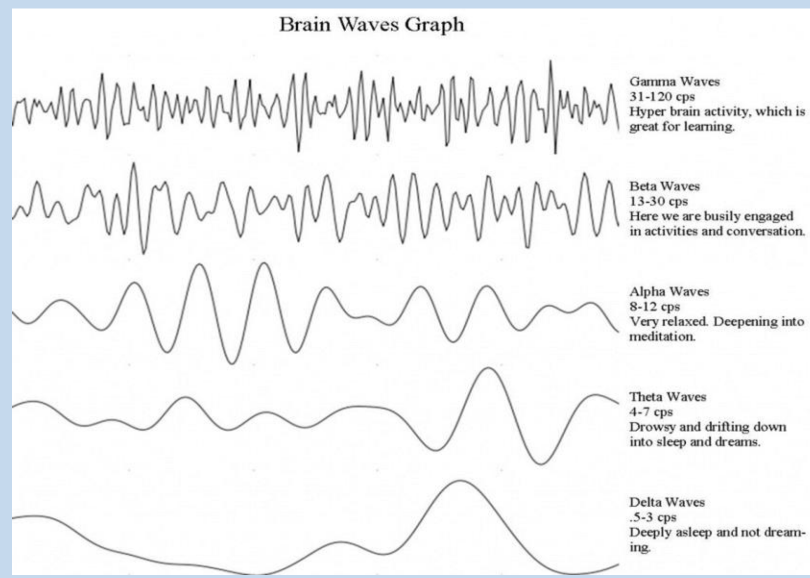
如何优化这些未知参数 ？

先验知识

先验知识通常跟神经科学相关：

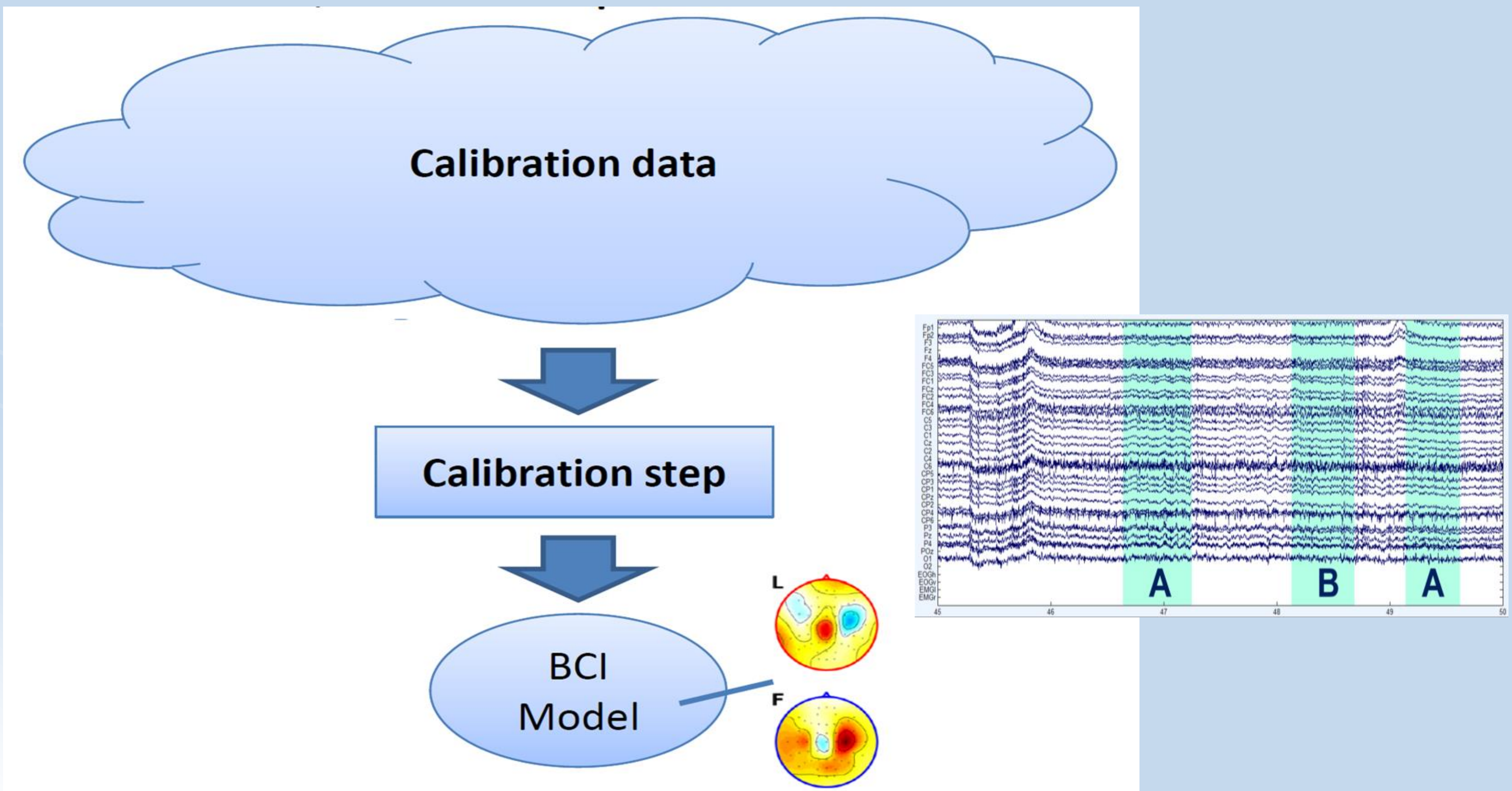
- 大脑解剖图、功能地图
- 时间信息：神经响应延迟P300，反应时间
- 振荡过程的频带信息：alpha，beta，gamma



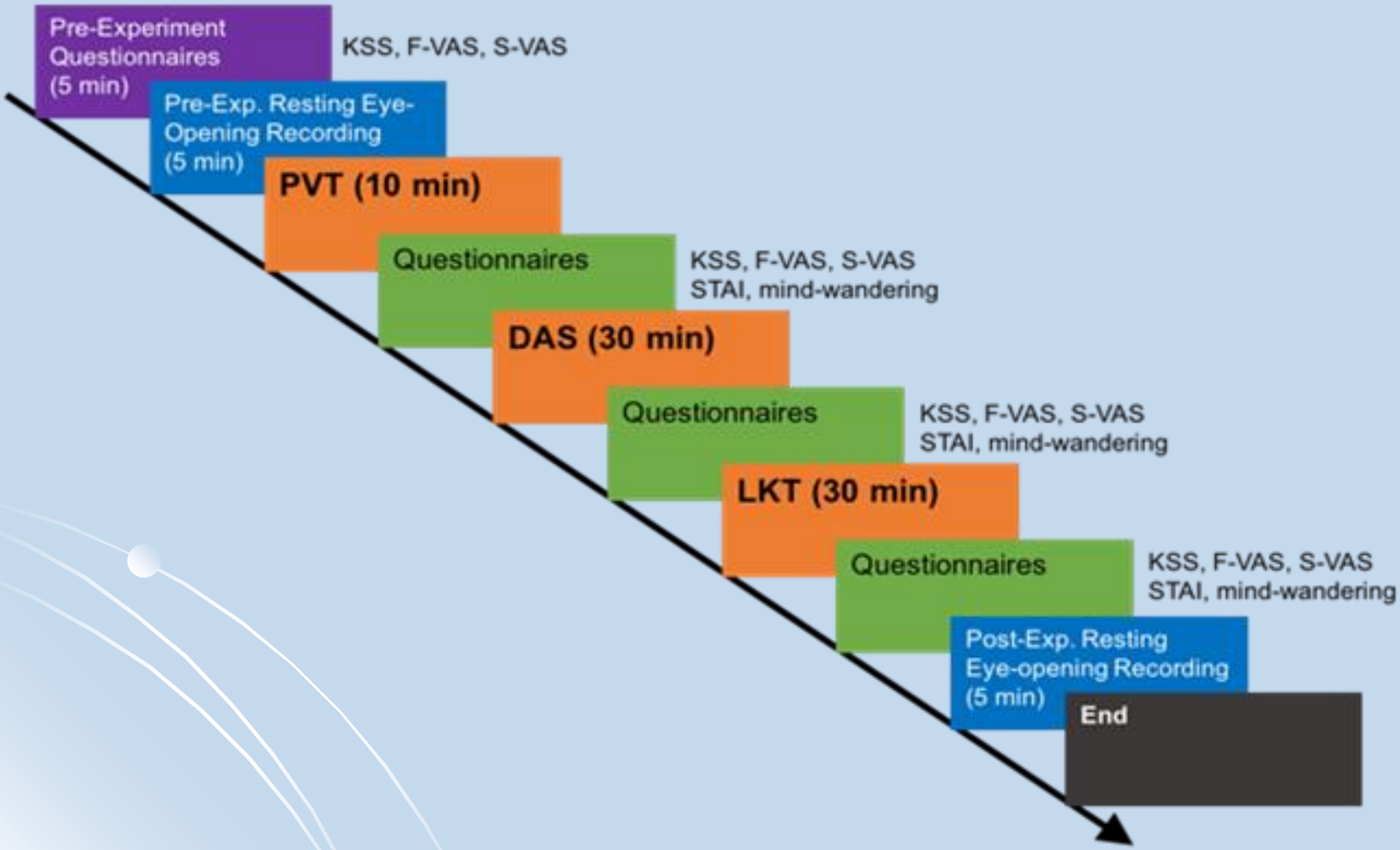


模型校准

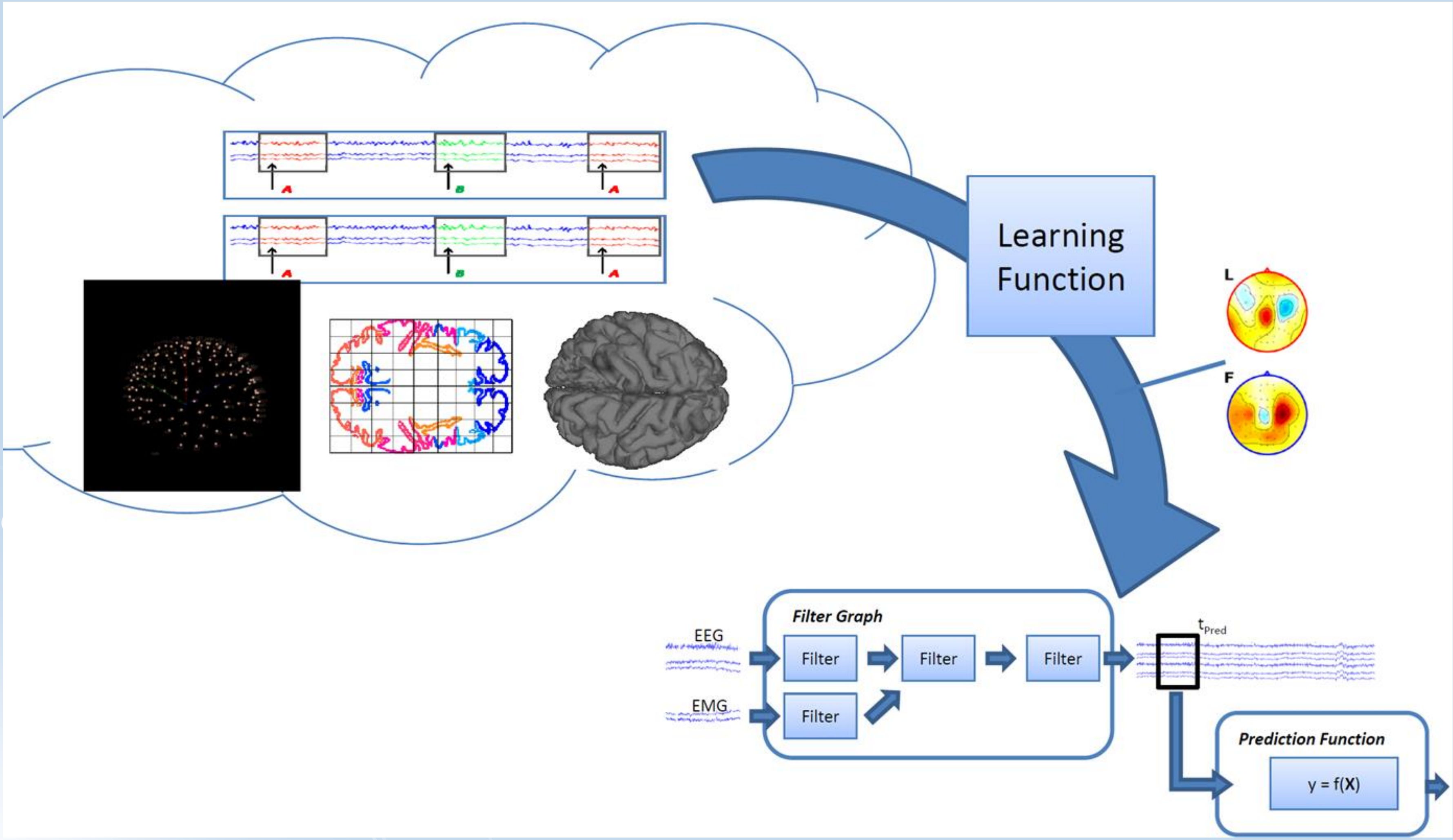
- 设计一个校准试验，采集数据来校准模型
- 使用同样的场景和设备
- 包含多次重复试验以尽量考虑变异性



试验设计


Pre-Experiment Questionnaires (5 min)
KSS, F-VAS, S-VAS
Pre-Exp. Resting Eye-Opening Recording (5 min)
PVT (10 min)
Questionnaires
KSS, F-VAS, S-VAS
STAI, mind-wandering
DAS (30 min)
Questionnaires
KSS, F-VAS, S-VAS
STAI, mind-wandering
LKT (30 min)
Questionnaires
KSS, F-VAS, S-VAS
STAI, mind-wandering
Post-Exp. Resting Eye-opening Recording (5 min)
End

完整过程



Learning Function
L
F
Filter Graph
EEG
EMG
Filter
Filter
Filter
Filter
t_{Pred}
Prediction Function
y = f(X)

特征提取

- 现成的机器学习算法通常不能直接用于原始EEG信号：
  - ❖ 原始信号维数太大
  - ❖ 原始信号信息含量低
- 特征提取：
  - ❖ 降维
  - ❖ 提高特征信息量

k近邻学习

*k*近邻(*k*-Nearest Neighbor, *k*NN)学习是一种常用的监督学习方法:

- 确定训练样本，以及某种距离度量。
- 对于某个给定的测试样本，找到训练集中距离最近的*k*个样本，对于分类问题使用“投票法”获得预测结果，对于回归问题使用“平均法”获得预测结果。还可基于距离远近进行加权平均或加权投票，距离越近的样本权重越大。
  - 投票法：选择这*k*个样本中出现最多的类别标记作为预测结果。
  - 平均法：将这*k*个样本的实值输出标记的平均值作为预测结果。

维数灾难 (Curse of Dimensionality)

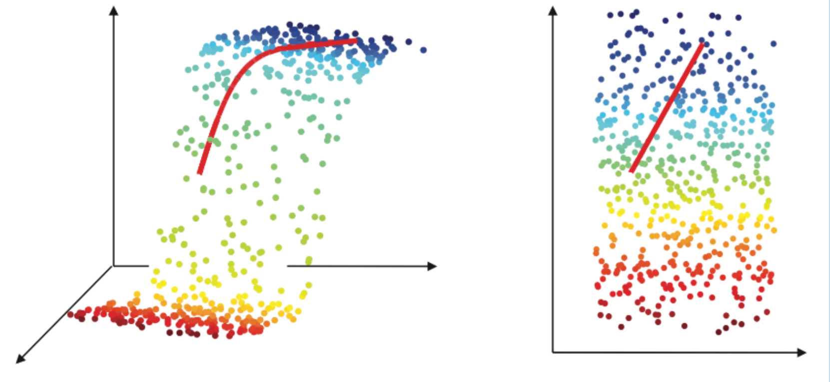
K近邻分析基于一个重要的假设：任意测试样本$\boldsymbol{x}$附近的任意小的δ距离范围内总能找到一个训练样本，即训练样本的采样密度足够大，或称为“密采样”。然而，这个假设在现实任务中通常很难满足：

- 若属性维数为1，当δ=0.001，仅考虑单个属性，则仅需1000个样本点平均分布在归一化后的属性取值范围内，即可使得任意测试样本在其附近0.001距离范围内总能找到一个训练样本，此时最近邻分类器的错误率不超过贝叶斯最优分类器的错误率的两倍。若属性维数为20，若样本满足密采样条件，则至少需要$(10^3)^{20}=10^{60}$个样本。
- 现实应用中属性维数经常成千上万，要满足密采样条件所需的样本数目是无法达到的天文数字。许多学习方法都涉及距离计算，而高维空间会给距离计算带来很大的麻烦，例如当维数很高时甚至连计算内积都不再容易。
- 在高维情形下出现的数据样本稀疏、距离计算困难等问题，是所有机器学习方法共同面临的严重障碍，被称为“维数灾难”。

降维

- 缓解维数灾难的一个重要途径是降维(dimension reduction)。
- 通过某种数学变换，将原始高维属性空间转变为一个低维“子空间”(subspace)，在这个子空间中样本密度大幅度提高，距离计算也变得更为容易。

- 为什么能进行降维？
- 数据样本虽然是高维的，但与学习任务密切相关的也许仅是某个低维分布，即高维空间中的一个低维“嵌入”(embedding)，因而可以对数据进行有效的降维。



(a) 三维空间中观察到的样本点　　(b) 二维空间中的曲面

图 10.2　低维嵌入示意图

线性降维方法

- 欲获得低维子空间，最简单的是对原始高维空间进行线性变换。给定d维空间中的样本$\mathbf{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m)\in\mathbb{R}^{d\times m}$，变换之后得到$d'\leq d$维空间中的样本

$$\mathbf{Z}=\mathbf{W}^{\mathrm{T}}\mathbf{X},$$

其中$\mathbf{W}\in\mathbb{R}^{d\times d'}$是变换矩阵，$\mathbf{Z}\in\mathbb{R}^{d'\times m}$是样本在新空间中的表达。

- 变换矩阵**W**可视为d'个d维属性向量。换言之，$\boldsymbol{z}_i$是原属性向量$\boldsymbol{x}_i$在新坐标系$\{\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_{d'}\}$中的坐标向量。若$\boldsymbol{w}_i$与$\boldsymbol{w}_i$

$(i\neq j)$正交，则新坐标系是一个正交坐标系，此时**W**为正交变换。显然，新空间中的属性是原空间中的属性的线性组合。

- 基于线性变换来进行降维的方法称为线性降维方法，对低维子空间性质的不同要求可通过对**W**施加不同的约束来实现。

主成分分析 (Principal Component Analysis, PCA)

- 对于正交属性空间中的样本点，如何用一个超平面对所有样本进行恰当的表达？

- 容易想到，若存在这样的超平面，那么它大概应具有这样的性质：
  - 最近重构性：样本点到这个超平面的距离都足够近；
  - 最大可分性：样本点在这个超平面上的投影能尽可能分开。

主成分分析

$$\begin{aligned} \max_{\mathbf{W}} \quad & \mathrm{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{W}) \\ \text{s.t.} \quad & \mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{I}. \end{aligned}$$

对优化式使用拉格朗日乘子法可得

$$\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{W} = \lambda\mathbf{W}.$$

只需对协方差矩阵$\mathbf{X}\mathbf{X}^{\mathrm{T}}$进行特征值分解，并将求得的特征值排序：$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$，再取前$d'$个特征值对应的特征向量构成$\mathbf{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{d'})$，这就是主成分分析的解。

主成分分析

- 降维后低维空间的维数d'通常是由用户事先指定，或通过在d'值不同的低维空间中对k近邻分类器（或其它开销较小的学习器）进行交叉验证来选取较好的d'值。

- 对PCA，还可从重构的角度设置一个重构阈值，例如$t=95\%$，然后选取使下式成立的最小d'值：
$$\frac{\sum_{i=1}^{d'}\lambda_i}{\sum_{i=1}^{d}\lambda_i}\geq t.$$

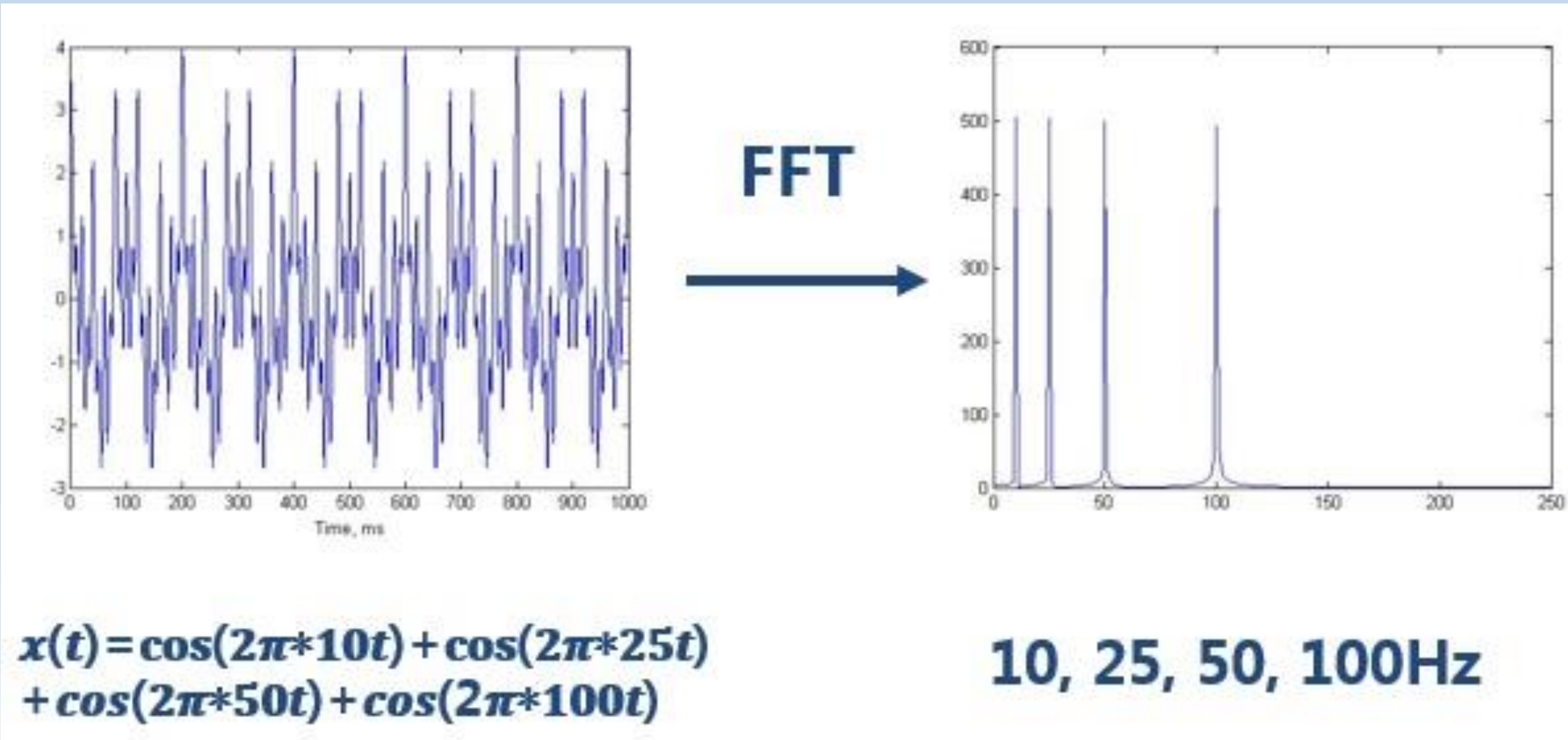
- PCA仅需保留 $\mathbf{W}$ 与样本的均值向量即可通过简单的向量减法和矩阵-向量乘法将新样本投影至低维空间中。

- 降维虽然会导致信息的损失，但1)舍弃这些信息后能使得样本的采样密度增大，2) 当数据受到噪声影响时，最小的特征值所对应的特征向量往往与噪声有关，舍弃可以起到去噪效果。

特征提取方法

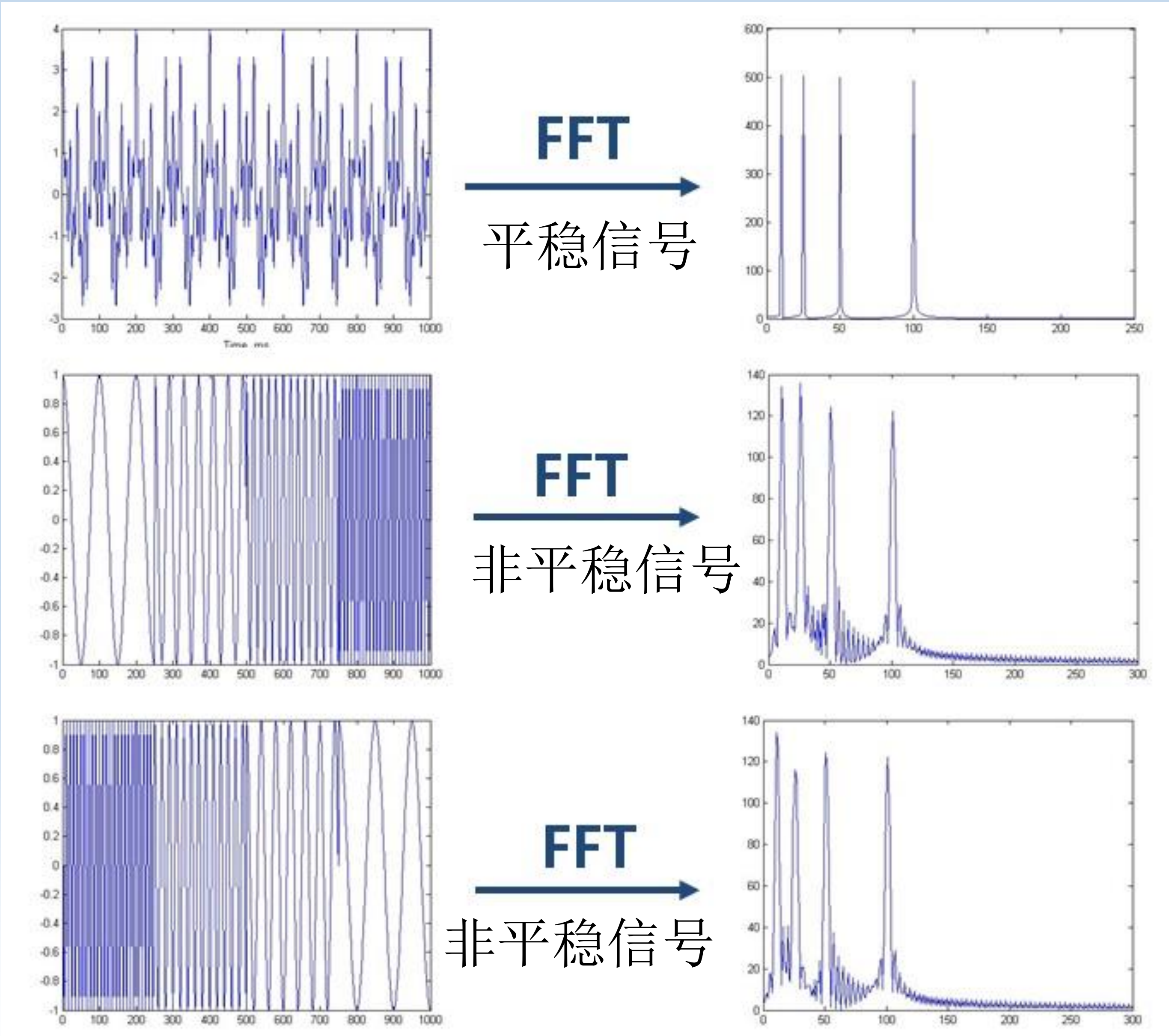
- 依赖于具体的脑电过程
- 振荡过程：
  - ❖ 对数方差：方差取对数后分布更加合理
  - ❖ 傅里叶频谱
- 事件相关电位：
  - ❖ 波峰的时延、高度、宽度
  - ❖ 信号分段的均值
  - ❖ 小波变换的系数

傅里叶变换



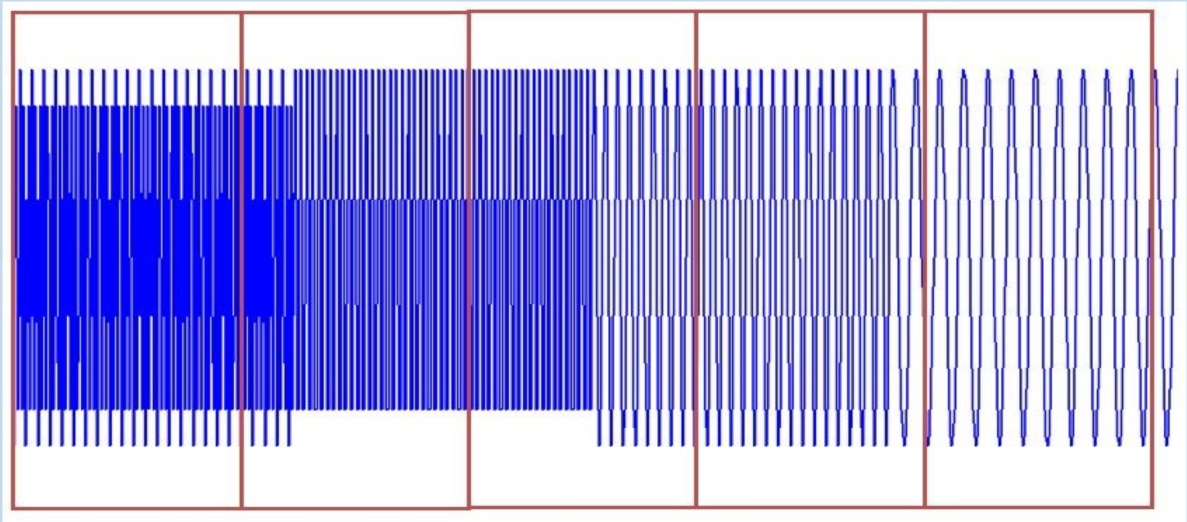
Time, ms
FFT
x(t)=cos(2π*10t)+cos(2π*25t)
+cos(2π*50t)+cos(2π*100t)
10, 25, 50, 100Hz

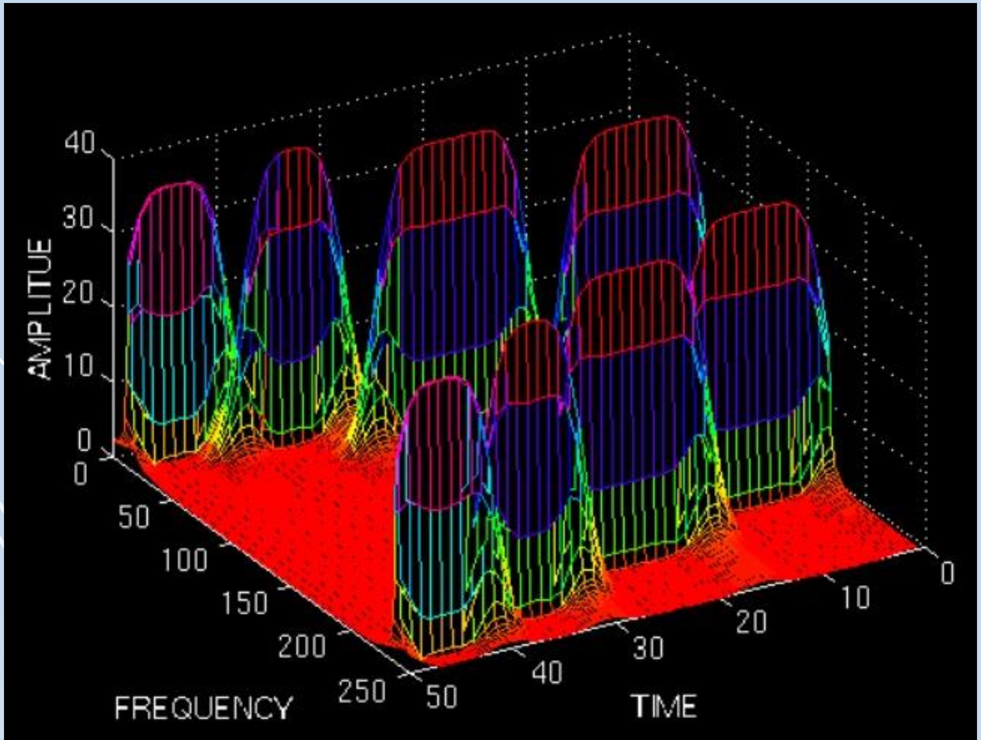
傅里叶变换



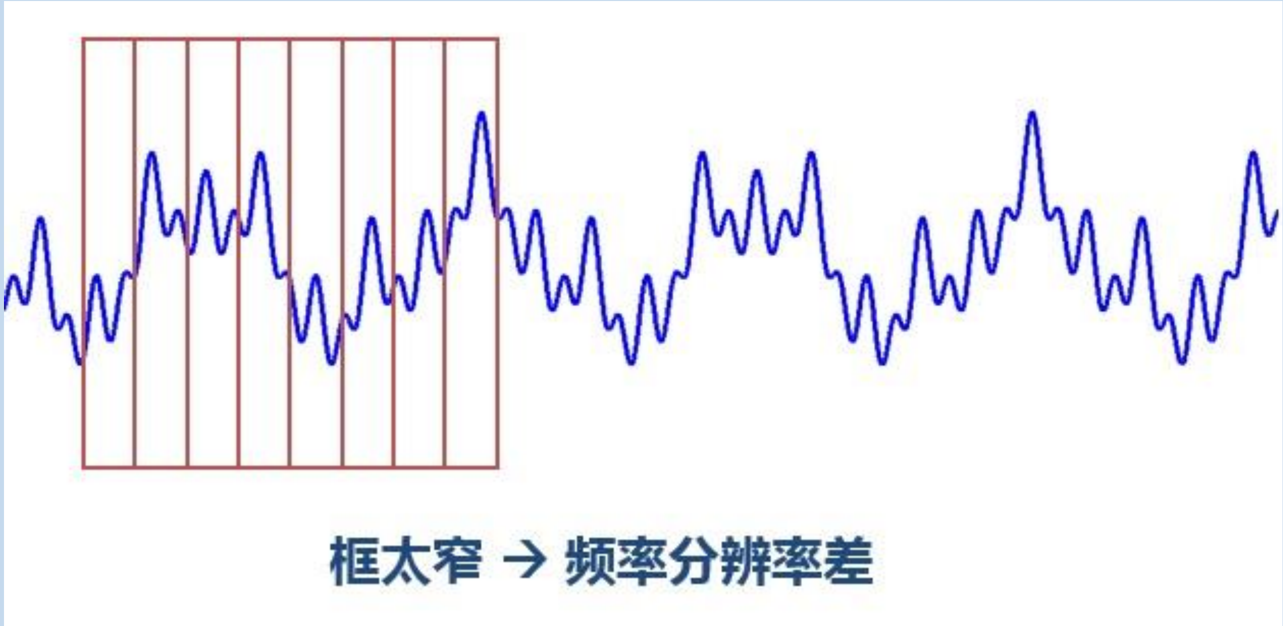
- 傅里叶变换处理非平稳信号有天生缺陷。它只能获取一段信号总体上包含哪些频率的成分，但是对各成分出现的时刻并无所知。因此时域相差很大的两个信号，可能频谱图一样。
- 对于非平稳信号，还需要知道各个成分出现的时间。

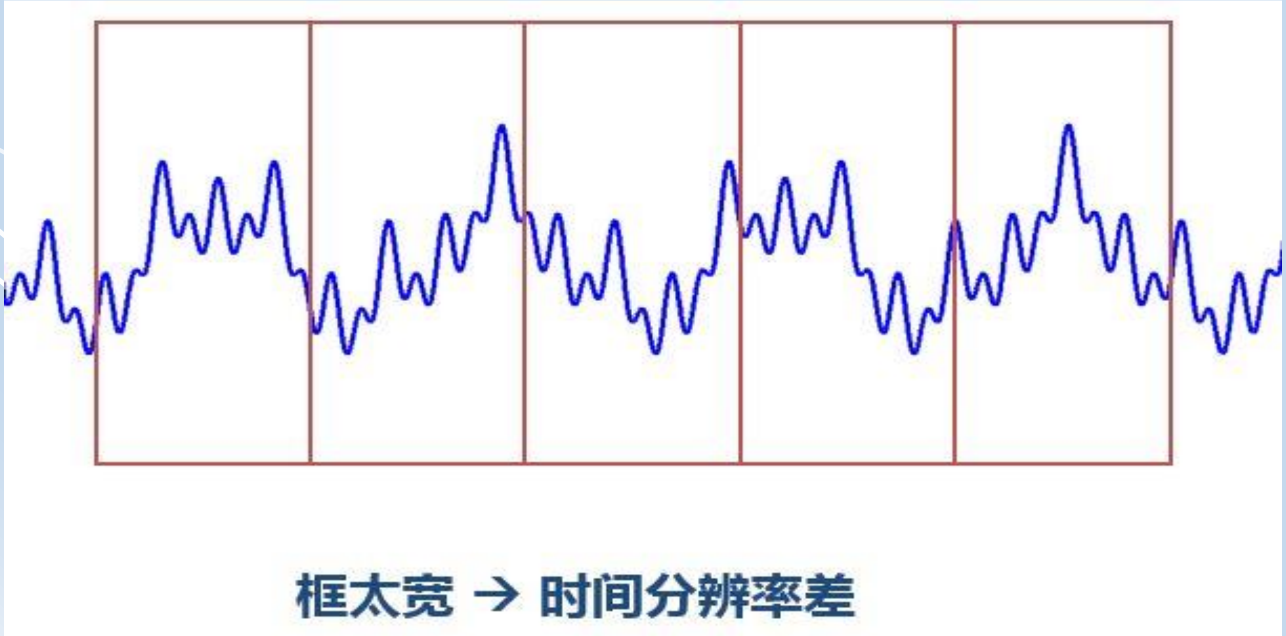
短时傅里叶变换



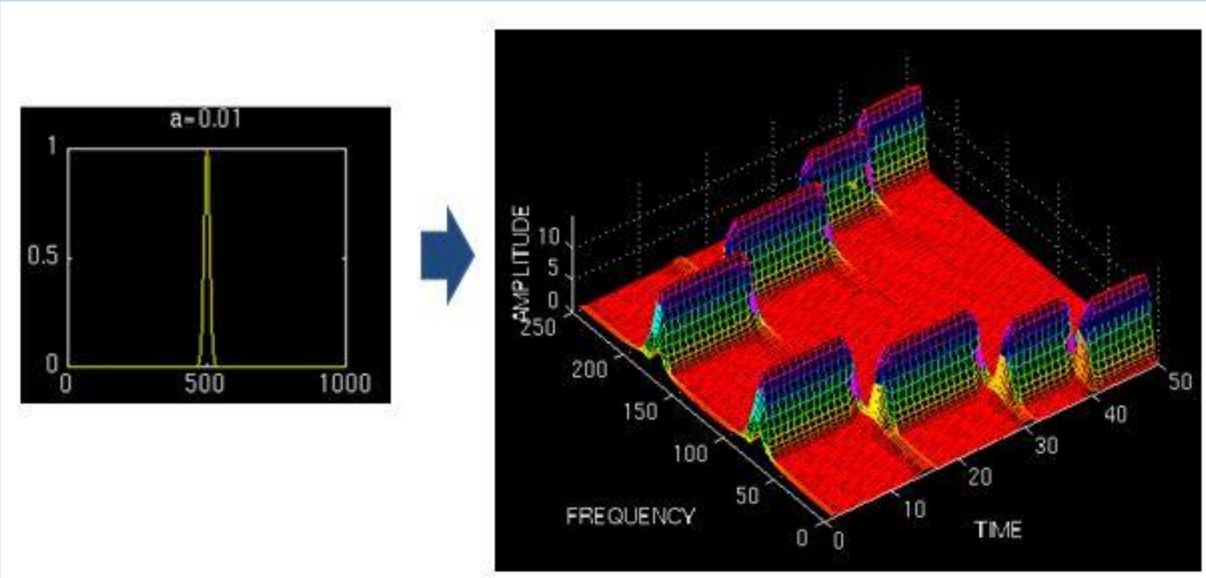

AMPLITUE
40
30
20
10
0
0
50
100
150
200
250
FREQUENCY
50
40
30
20
10
0
TIME

短时傅里叶变换问题：窗函数



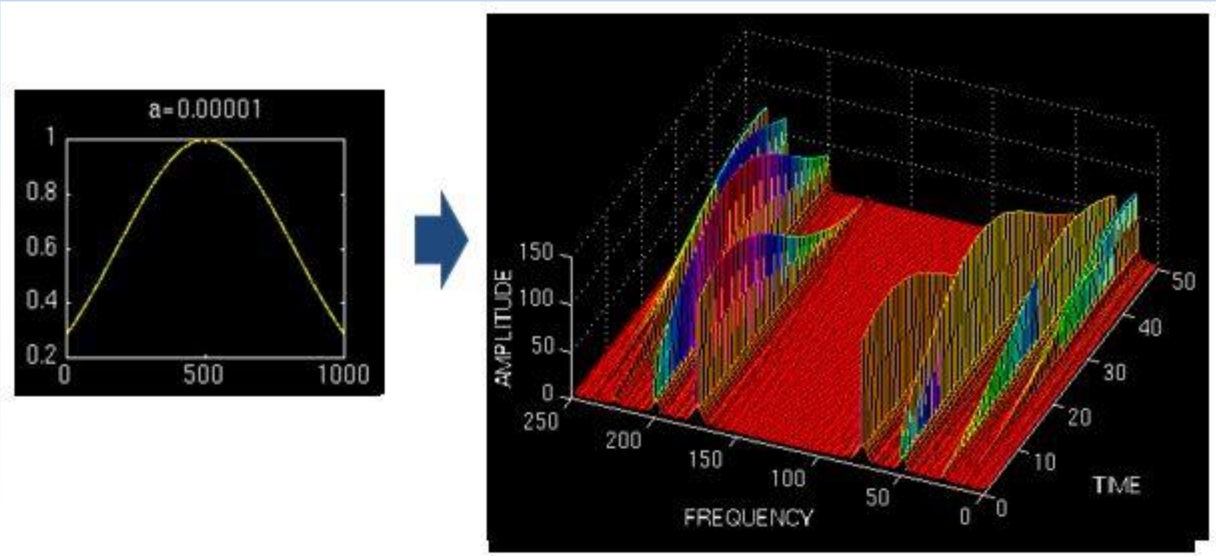


短时傅里叶变换问题：窗函数



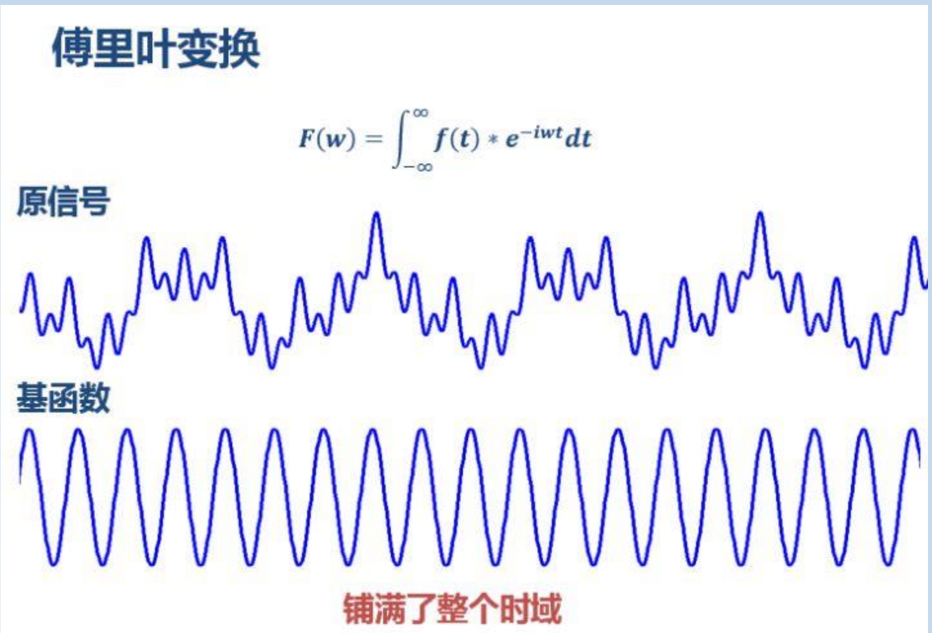
窄窗口时间分辨率高，
频率分辨率低。

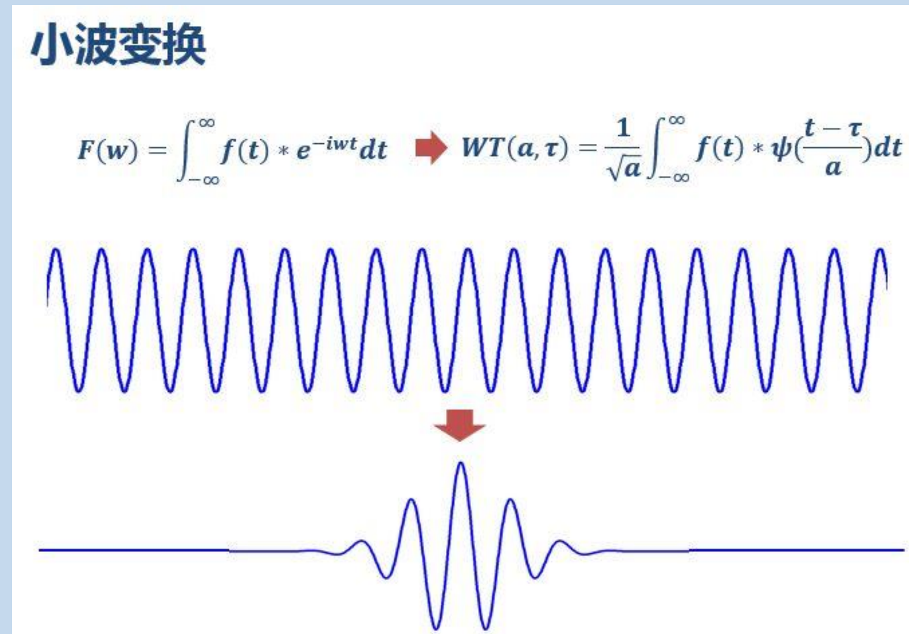
高频用窄窗口，
低频用宽窗口。



宽窗口时间分辨率低，
频率分辨率高。

傅里叶变换与小波变换

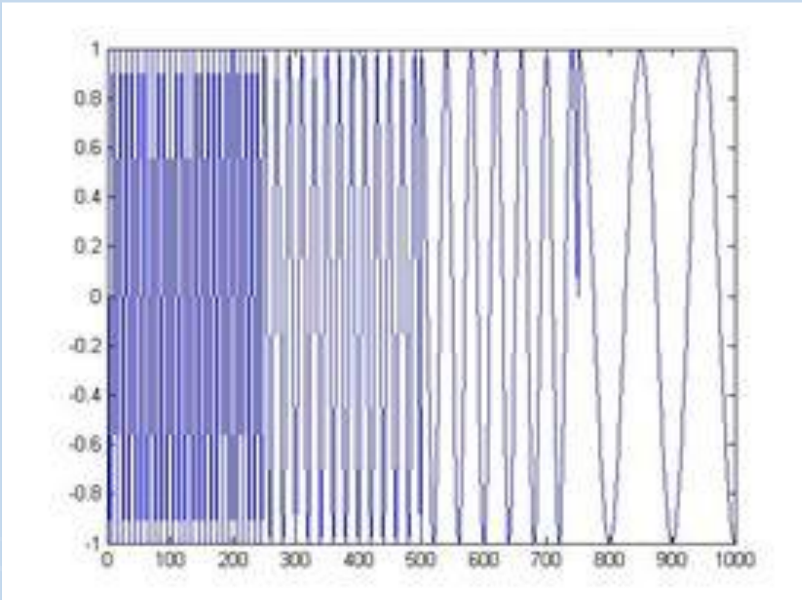




$$WT(a,\tau)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t) * \psi\left(\frac{t-\tau}{a}\right) dt$$

尺度：频率（反比）
平移量：时间

傅里叶变换与小波变换



时域信号