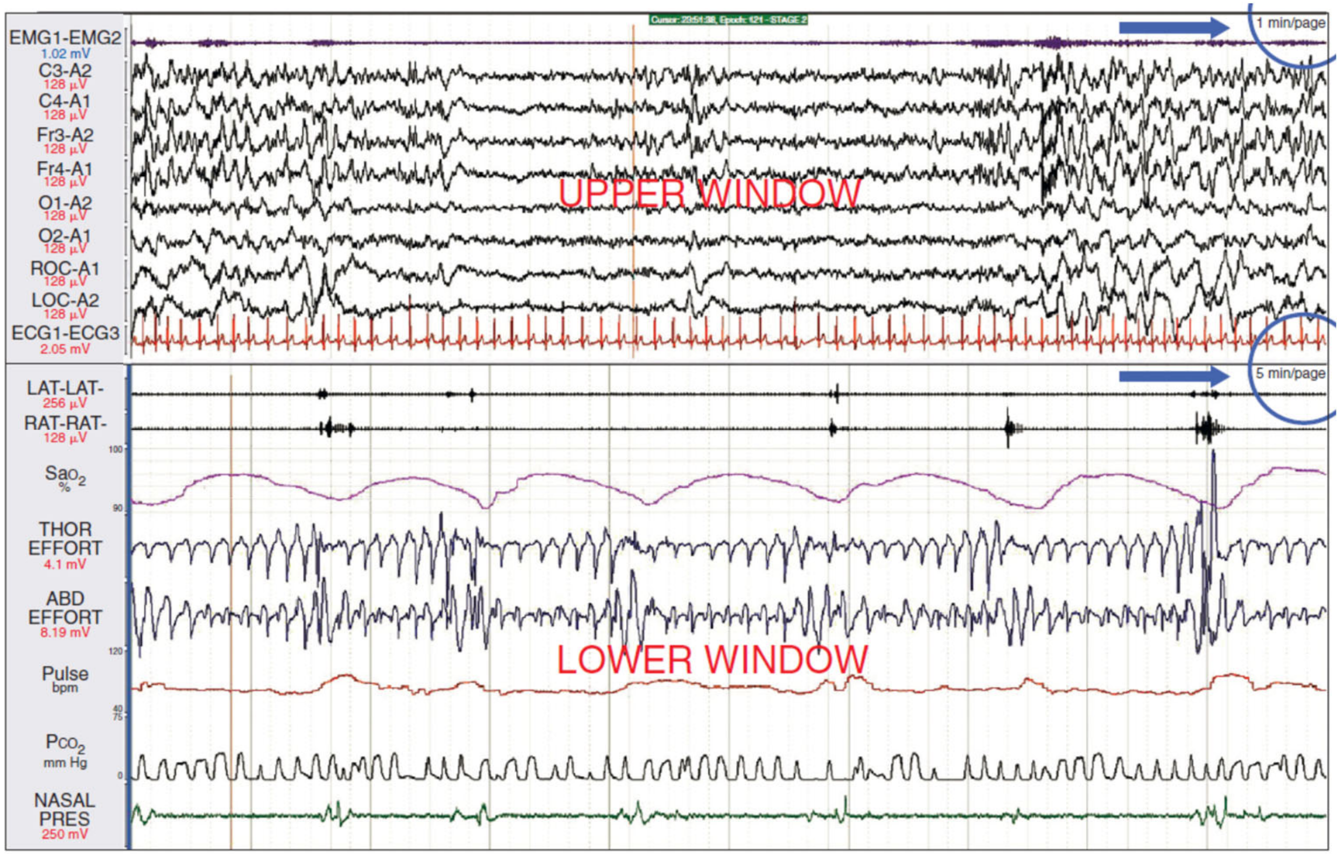
随时间演变的不确定信号

- 例如多导睡眠图



- 一个呼吸周期内通过鼻腔的气流速度随着时间变化
- 不同的呼吸周期的气流速度的变化并不完全一致
- 这些随机现象无法作为经典的随机变量来研究

第四讲 随机信号基础2

随机信号作为一个随机过程

学习要点

- 随机过程的概念
- 随机信号的概率分布函数与数字特征
- 随机信号平稳性，遍历性的定义
- 弱平稳的检验方法
- 随机信号的置信区间
- 对生物医学信号采集与处理的启示

随机过程

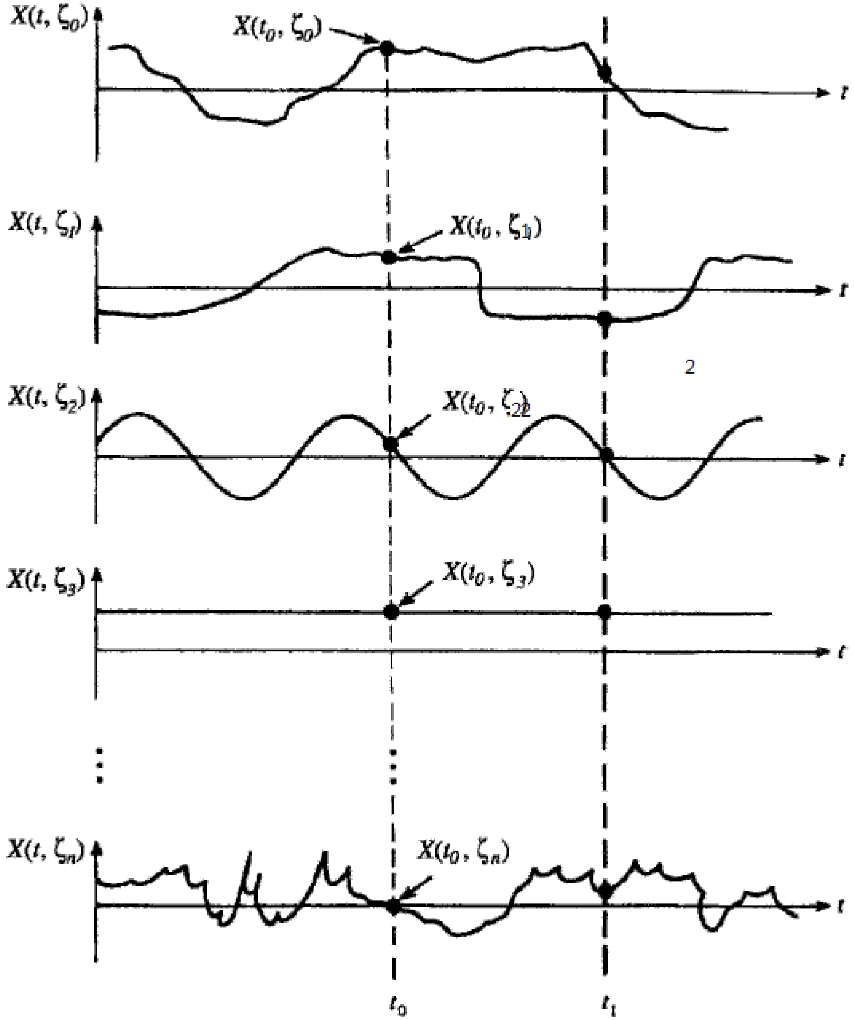
- 随机过程作为一门学问
  - 是概率论的“动力学”部分，研究随时间演变的随机现象
  - 以概率论（研究随机变量的理论）为基础，但是形成为一门相对独立的课程和研究领域

- 随机过程作为一类现象
  - 指代随时间演变的随机现象本身

随机过程的基本描述

- 指代随时间演变的随机现象本身
  - 是依赖于某个参数的一簇（可无穷多个）随机变量，用随机函数描述
    - 这个参数可以是时间，也可以是空间，还可以是频率等
  - 数学表达：设T 是一个实数集，依赖于参数$t \in T$的一簇随机变量$X(t)$称为随机过程，记为$\{X(t), t \in T\}$
    - 对于任意时刻$t = t_i$，$X(t_i)$是一个随机变量，可用概率论研究
    - $X(t)$是随机过程在t时刻的状态
    - 对于一切$t \in T$，$X(t)$所有可能取值的全体是随机过程的状态空间
  - 随机过程需要用“随机过程”这门学科来研究

随机过程的n个样本



- n个样本，在随机过程中，可以称为样本函数
- 这些样本函数仅仅是这个随机过程的具体实例
- 随机过程包含这样样本在内的全体“曲线”样本函数，$X(t, \zeta)$
- 选取特定时刻，例如t_0，对应一个随机变量$X(t_0)$

概率论与随机过程的比较

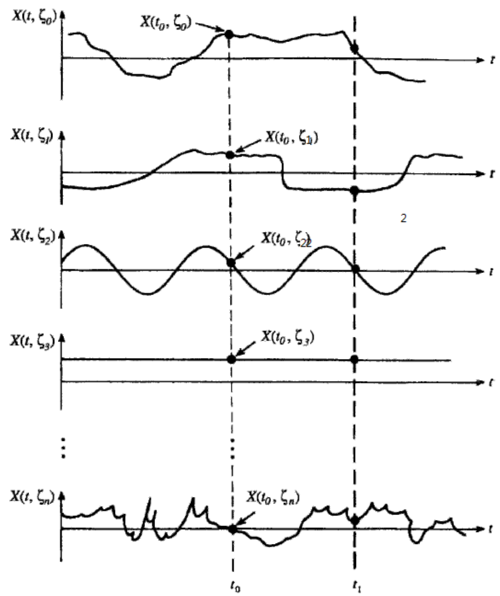
	概率论	随机过程
研究对象	随机变量	引入了时间维度的随机变量，即随时间变化的随机现象
举例	心率	心率随时间的变化
研究内容	概率分布	?
	数字特征（矩）	?

概率论vs随机过程

	概率论	随机过程
研究对象	随机变量	随时间变化的随机现象，引入了时间维度的随机变量
举例	心率	心率随时间的变化
研究内容	概率分布 【与时间无关】	概率分布 【是时间的函数】
	数字特征（矩） 【如均值、方差、偏度等，与时间无关】	数字特征（矩函数） 【是时间（或其它变量）的函数】

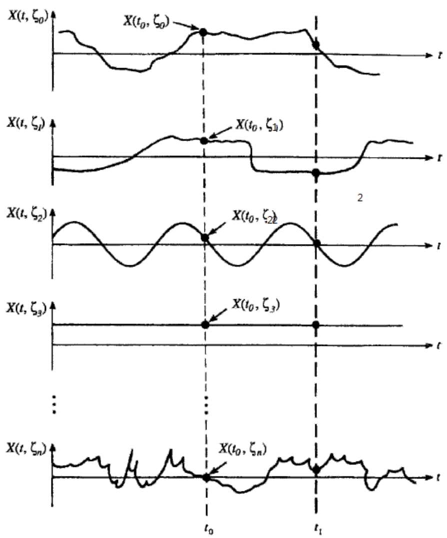
随机信号的概率分布

- 随机过程$X(t)$的一维概率分布函数
  - $F_X(x_1, t_1) = P(X(t_1) <= x_1)$
  - $f_X(x_1, t_1) = dF_X/dx_1$
  - 概率分布是时间t_1的函数
    - 概率分布可能随时间变化



随机信号的概率分布函数

- 随机信号在不同时刻$t_1, t_2, \ldots, t_n, \ldots$的取值
  - 对应随机变量$X(t_1), X(t_2) \ldots$
  - $(X(t_1), X(t_2), \ldots)$构成高维随机变量



*n*维分布函数簇

$F_X(x_1, x_2, \ldots, xn; t_1, t_2, \ldots, tn) =$

$P(X(t_1) \le x_1, X(t_2) \le x_2, \ldots, X(tn) \le xn)$

当*n*足够大， *n*维分布函数簇可以接近完全刻画随机过程的特性

随机信号的数字特征

- 随机过程$\{X(t), t \in T\}$，在任意时刻$t_1 \in T$，$X(t_1)$对应一个随机变量
- 随机过程的均值函数

$$\mu_{X(t)} = E[X(t)]$$

  - 是所有样本函数的平均值，是时间的函数
  - 是随机过程$X(t)$在各个时刻的摆动中心

- 随机过程的二阶中心矩

$$\sigma_X^2(t) = E\{[X(t) - \mu_X(t)]^2\}$$

  - 是随机变量围绕“摆动中心”变化范围的一个度量

- 本课程仅以高斯信号为例理解医学信号处理，则我们仅需考虑一阶矩和二阶矩

自相关函数

- 定义
  - 设 $x(t_1)$和$x(t_2)$是随机过程X(t)在t_1和t_2时刻的随机变量，$x(t_1)$和$x(t_2)$的乘积的统计均值为自相关函数

$$\varphi_x(t_1, t_2) = E[x(t_1)x(t_2)]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x(t_1)x(t_2)]\ f(x(t_1), x(t_2))\ dx(t_1)\ dx(t_2)$$

自协方差函数

- 定义
  - 设 $x(t_1)$和$x(t_2)$是随机过程X(t)在t_1和t_2时刻的随机变量，称$x(t_1)$和$x(t_2)$的二阶联合中心矩为X(t)的自协方差函数

$$\mathrm{Cov}[x(t_1),x(t_2)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\alpha_1-m_x(t_1))(\alpha_2-m_x(t_2))\,f_{x(t_1)x(t_2)}(\alpha_1,\alpha_2)\,d\alpha_1 d\alpha_2$$

其中$f_{x(t_1)x(t_2)}(\alpha_1,a_2)$ 是二阶概率密度函数，$m_x(t_1)$ 和$m_x(t_2)$分别是t_1和 t_2的均值

  - 自协方差函数是随机过程去均值（直流分量）后剩余部分的相关函数
  - 若对于任意的t_1和t_2, 都有 $Cov[x(t_1),x(t_2)]=0$，则随机过程的任意两个时刻状态间不相关

随机过程的平稳性

- 随机信号的概率密度分布是时间的函数
- 一阶平稳过程

$$f_{x(t_1)}(\alpha)=f_{x(t_2)}(\alpha)$$

- 二阶平稳过程

$$f_{x(t_1+\tau)x(t_2+\tau)}(\alpha_1,\alpha_2)=f_{x(t_1)x(t_2)}(\alpha_1,\alpha_2)$$

- 强平稳，又称严平稳或者狭义平稳过程

$$f_{x(t_1+\tau)\ldots x(t_n+\tau)\ldots}(\alpha_1,\ldots,\alpha_n,\ldots)=f_{x(t_1)\ldots x(t_n)\ldots}(\alpha_1,\ldots,\alpha_n,\ldots)$$

应用中的平稳性

- 概率分布函数描述随机信号的困难
  - 很难获得概率分布函数

- 具体应用中，矩估计和矩描述更实际

- 宽平稳过程 – 均值和协方差不随时间变化

$$E[x(t_1)] = E[x(t_2)]$$

$$\mathrm{Cov}[x(t_1), x(t_2)] = \mathrm{Cov}[x(t_1), x(t_1 + \tau)]$$

平稳性的评价

- 很多随机信号分析方法的基本要求是随机信号具有弱平稳性
  - 例如将要讨论的功率谱分析
- 平稳性判定的最简单方法 – 考察产生信号的物理机制和环境
  - 例1：不同单词的发声需要声强和频率的变化，而且单词较短，信号很难在较长时间内实现平稳
  - 例2：自发脑电在警觉、困倦、智力思考等状态之间的切换，脑电波特征会发声变化，如果人的状态保持稳定，则脑电波的特征可能在这段时间保持平稳性
- 统计检验方法
  - 考察随机信号均值和方差等统计矩是否随时间变化，评价的弱平稳性

平稳性的统计检验方法

- 步骤
  - 将记录信号分成多个片段
    - 片段要足够长，对于包含振荡分量的随机信号，每个片段都必须包含多个振荡周期
    - 例如心电包含振荡分量，过短则不能反映信号的真正特性
  - 估计各个片段的均值和方差
  - 用标准统计检验方法检验各个片段的均值和方差的异同
    - 分两段的均值检验 (T统计量) 和方差检验 (F统计量)
    - 分多段比较的方差分析法 (ANOVA，此课程不要求)
    - 对于非高斯信号*，则需要建立各个信号分段的概率密度函数或者累计分布函数模型，比较这些模型的稳定性 (非参，K-S检验，本课程不要求)

例：两段信号的平稳性检验

- 信号分为两段
  - S1: 1至35个数据点；S2: 36-70数据点；样本均值分别为 53.17，49.11；方差分别为 176.78，96.17
  - 置信水平为95%，即假设检验的$\alpha = 0.05$
  - 检验均值是否相等的T 统计量服从t分布
    - 自由度$v = N_1+N_2 - 2 = 35+ 35 - 2 = 68$
    - $$T = \frac{\hat{m}_1 - \hat{m}_2}{\left(\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\left(\frac{N_1\hat{\sigma}_1^2+N_2\hat{\sigma}_2^2}{N_1+N_2-2}\right)\right)^{\frac{1}{2}}} = \frac{53.17-49.11}{\left(\frac{2}{35}\frac{35(176.78+96.17)}{68}\right)^{\frac{1}{2}}} = 1.43$$
    - 显著区域临界$T_c|_{v=68,\ a=0.05} = 2$，即满足 $P(|T|>=T_c)\ |_{v=68} = 0.05$
    - 计算的T值小于显著区域的临界值T_c，即T值在置信区，不显著，所以可以认为信号的均值是平稳的

例：两段信号的平稳性检验

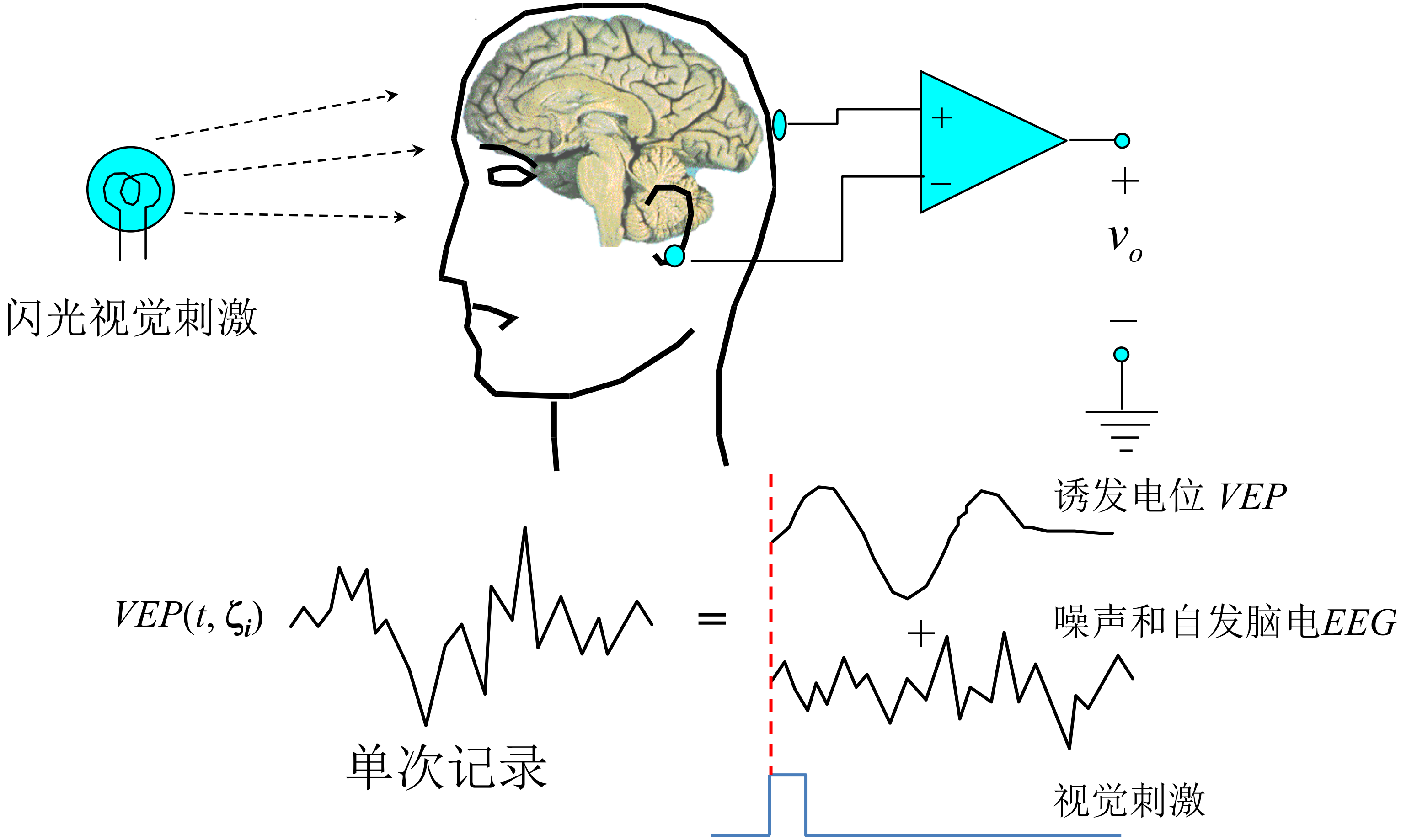
- 检验方差是否相等的F统计量
  - 自由度 $v_1 = N_1-1 = 34$， $v_2 = N_2-1 = 34$
  - $$F = \frac{\frac{N_1}{v_1}\hat{\sigma}_1^2}{\frac{N_2}{v_2}\hat{\sigma}_2^2} = 1.84$$
  - 双侧检测的置信区间 $F_{34,34;0.975} \le F \le F_{34,34;0.025}$ $0.505 \le F \le 1.97$
  - 上面计算的F =1.84在置信区间内，因此可判断方差具有平稳性
- 因为均值和方差都可以认为具有平稳性，在95%的置信水平下，原信号可以看做弱平稳随机信号

各态遍历性

- 定义：如果所有样本函数在**某一特定时刻**的统计特性与**单一样本函数在长时间**内的统计特性一致，则称该随机过程各态遍历
  - 如果只能观测一个样本函数，则无法用实测数据检验各态遍历性

- 应用场景
  - 在应用中同时获取大量样本函数非常困难，常常只能观测一个样本函数
  - 我们测量的生物医学信号，都假设其具有各态遍历性，通过多次测量估计出随机信号的特征和参数

视觉诱发电位信号的测量



实际随机信号采集

- 理想情况，完全观测数据
  - 观测N个放大器的输出，结果记为 $x_i(t)$, $i = 1, 2, \ldots, N$。当N趋近无穷大时，这些样本函数的集合就构成了放大器输出可能经历的整个过程，即随机过程$X(t)$
- 现实情况
  - N有限
  - 大多数实验数据采集情况下，$N = 1$。例如脑电信号采集中每个传感器测量对象仅有一个对应的放大器
  - 假设随机遍历性
  - 分时间先后顺序，多次重复测量M次。例如重复进行某种特定的视觉刺激，并记录对应的视觉诱发脑电

生物医学信号

- 可以试图用随机过程描述，进行理论上的探讨
- 医学信号处理书籍中，随机信号==随机过程
- 实际被处理的信号是采集记录下来的结果，只能是随机信号的样本
  - 例如我们记录的脑电信号
    - 时间有限的采样，大脑一直在活动，仅采取一段时间的信号
    - 空间有限的采样，不可能将所有位置的信号都记录下来
- 生物医学信号处理的目标：从样本信号推测总体的特征，根据样本的统计量和抽样分布进行统计推断

随机信号的统计推断

- 点估计
- 区间估计
- 假设检验

诱发脑电原始数据

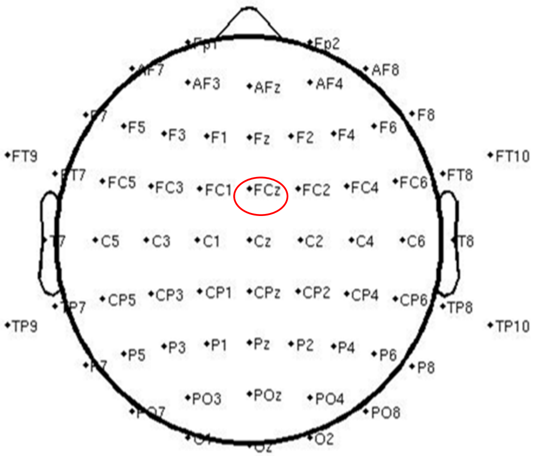
frontiers in
PSYCHOLOGY

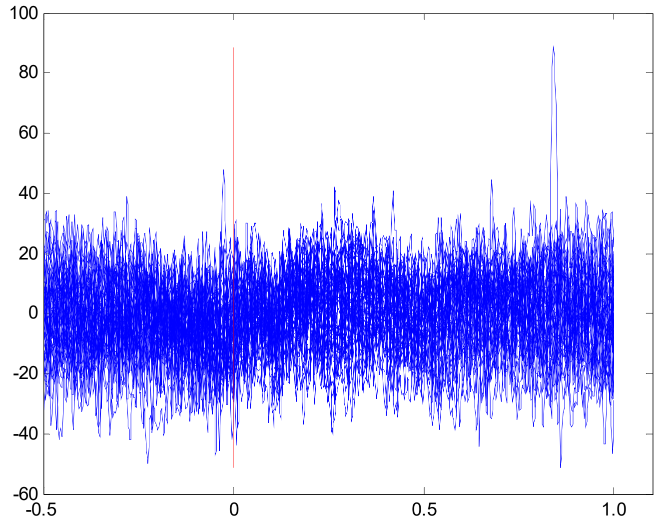
ORIGINAL RESEARCH ARTICLE
published: 09 November 2011
doi: 10.3389/fpsyg.2011.00331

Larger error signals in major depression are associated with better avoidance learning

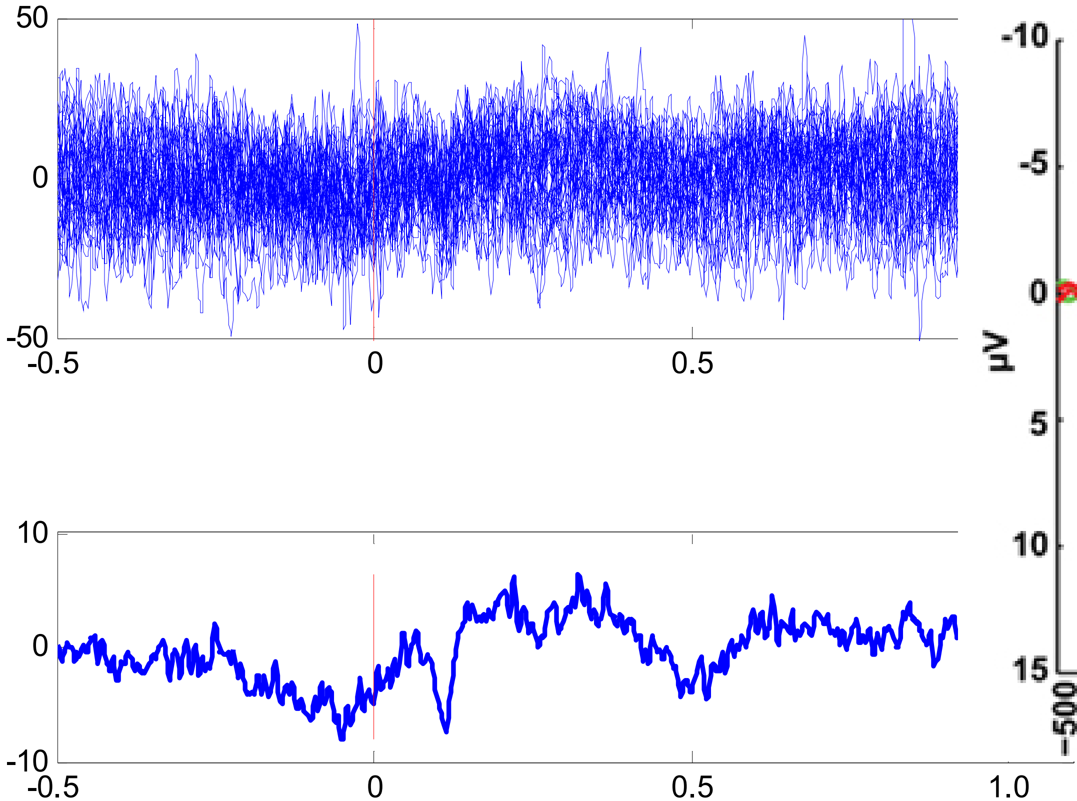
James F. Cavanagh[1,2]*, *Andrew J. Bismark*[2], *Michael J. Frank*[1,3,4] *and John J. B. Allen*[2]

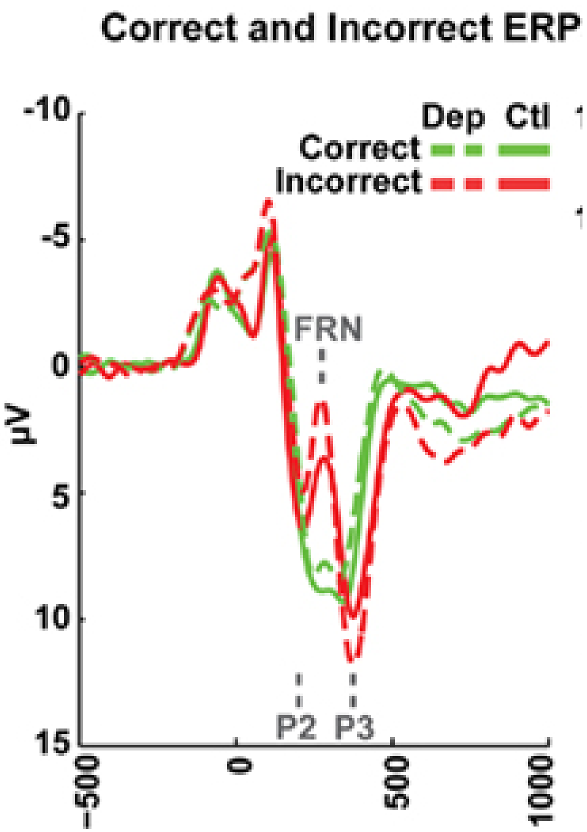
[1] *Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA*
[2] *Psychology Department, University of Arizona, Tucson, AZ, USA*
[3] *Department of Psychiatry, Brown University, Providence, RI, USA*
[4] *Brown Institute for Brain Sciences, Brown University, Providence, RI, USA*





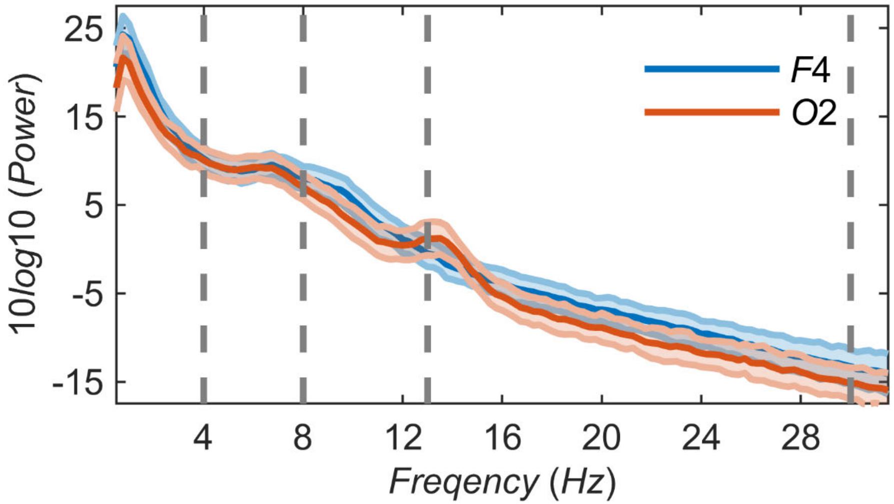
样本的平均－总体均值的点估计





每个时间点上的均值——随机信号的点估计

区间估计



每个频率点的功率的均值及其置信区间（Fig. 2A in Cui et al., 2023）

假设检验

NeuroImage 214 (2020) 116767

ELSEVIER

Contents lists available at ScienceDirect

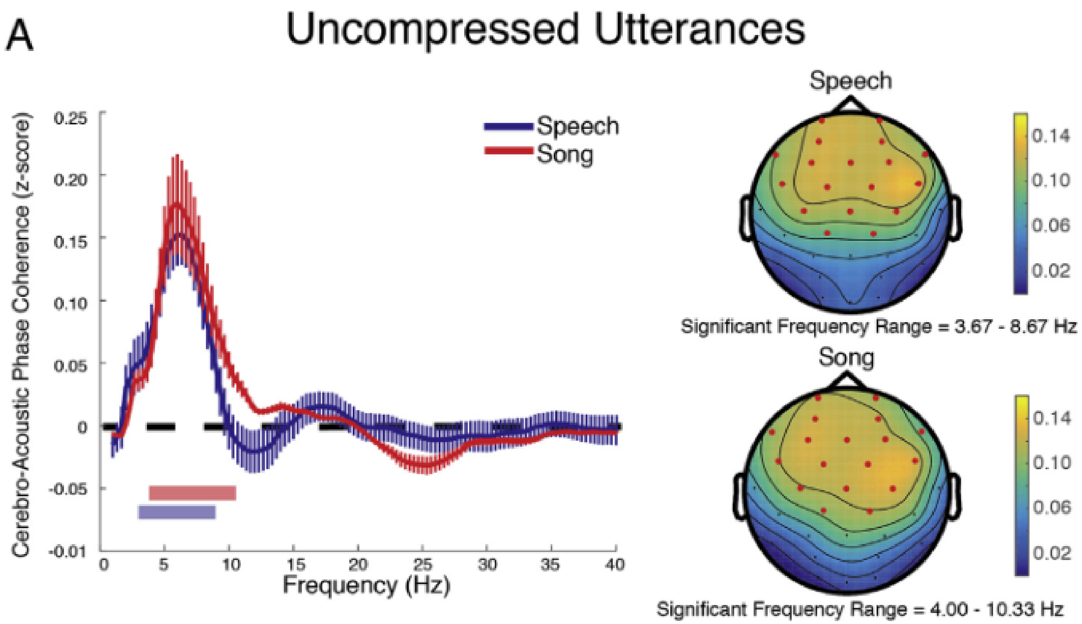
NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Music as a scaffold for listening to speech: Better neural phase-locking to song than speech

Christina M. Vanden Bosch der Nederlanden [*], Marc F. Joanisse, Jessica A. Grahn

Department of Psychology & Brain and Mind Institute, The University of We



随机信号处理的三项数学基础

	概率论	随机过程[1]	统计学	随机信号
研究对象	随机变量：无时间维度	随机函数：随某时间[2]变化的随机现象	随机样本（以及样本矩）	随机信号的样本（以及随机信号的样本矩）
举例	心率作为随机变量	视觉诱发电位	样本：对50人的一次心率检测，或者一段时间的即时心率数据	测量50次的视觉诱发电位
研究内容	概率分布 【与时间无关】	概率分布【是时间的函数】	抽样分布—统计量的概率分布	随机信号统计量的概率分布(抽样分布)
	数字特征（矩） 【特定数值，如均值】	数字特征（矩函数） 【是时间的函数】	基于统计量的推断（参数估计-点估计和区间估计，假设检验）	基于统计量的推断（参数估计-点估计和区间估计，假设检验）

[1] 将随机过程看作是随机信号的理论部分，随机信号才考虑信号的样本;
[2] 时间或者其他维度，例如空间位置。

小结与讨论

- 随机过程的概念和基本理论
  - 作为学科的随机过程，作为研究对象的随机过程
  - 随机过程与概率论的关系
  - 随机过程的基本描述
    - 概率分布函数与数字特征
    - 随机信号平稳性，遍历性的定义
    - 弱平稳的检验方法
- 对生物医学信号采集与处理的启示
  - 信号采集通常需要假设随机信号具有遍历性，重复测量，获得随机信号的多个样本
  - 基于样本不同时间段内的均值和方差判定弱平稳性
    - 平稳信号：功率谱估计，自相关函数估计等
    - 非平稳信号：短时傅里叶变换，小波变换等