

《心理科学进展》审稿意见与作者回应

题目：动态视觉-动作控制中的信息整合机制——基于卡尔曼滤波模型

作者：陈中廷 张子扬 连玉净 高天泽

第一轮

审稿人意见：

该文章从传统心理学研究的“刺激-反应”模型切入，试图以贝叶斯框架拆解以往的静态黑箱假设，从更具生态效度的视角揭示外部输入与内在先验的动态整合和迭代更新机制，对感知觉机制研究及心理学实验设计具有重要的前沿理论意义。研究结构清晰，阶段划分合理，前期数据也显示出一定的可行性。我并非计算建模领域专家，仅从大同行的视角提出阅读过程中的一些疑问。

回应：

感谢审稿专家的肯定！下文对每条意见进行了回应和修改。正文修改部分使用蓝色字体标明；在正文部分增加了行号，以方便定位具体修改位置。

意见 1：

作者首先是从对以往“刺激-反应”模型的探讨来切入的。按我的理解，作者在后文中的研究实际关心的是动态视动任务中感觉加工、反馈、信息整合等实时动态过程；但同时，作者又在自检报告、摘要、“理论构建”章节等处多次提到“科学问题：动态任务中‘刺激-反应’的映射关系与机制”“本构想立足于‘刺激-反应’关系这一核心问题”“关注信息整合成分对‘刺激-反应’映射关系的影响”“从而对‘刺激-反应’关系提出一种具备动态调节机制的解释视角”，这让人有些困惑。

我认为，“刺激-反应”模型作为一个较为抽象、笼统的宏观概念，并不适合写在具体科学研究问题中——因为任何行为任务范式本质上都可以视为一个“刺激-反应”映射过程。作者完全可以仅将此概念作为讲故事的引子，而不让其涉及具体问题。若采用此路径，建议修改摘要及正文、弱化这一概念，使其仅出现在问题提出部分，同时将核心科学问题及研究落脚点明确表述为“动态视觉-动作控制中的信息整合机制”之类，这样便已经构成了一个较为完备的科学问题与研究方案。

如果作者坚持将“刺激-反应”作为本研究的核心问题，则需将其具象化为可操纵或观测的因素或变量。从直觉而言，“刺激-反应”概念本身隐含了单次、离散的前提，是一种脉冲式的映射；那么在作者所采用的动态连续视动追踪任务中，“刺激”和“反应”将分别对应哪些具体因素或变量？作者虽然希望在研究模块一中分离出动态任务中的核心参数和测量指标，但其中很可能包含实时后验反馈、整合权重、知觉决策等相关参数，若再考虑运动控制系统的输出，到底以哪个变量来衡量“反应”很难定论。这与其说解释了连续动态任务中的“刺激-反应”关系，不如说拓展了这一关系的概念，用一系列动态整合和反馈过程来取代简单的“刺激-反应”映射。否则，如果要用同一套变量，既用来模拟动态过程，又用来定义“刺激-反应”映射过程，那就很难说是研究意义的递进，反而有同义反复之嫌。

回应：

感谢审稿专家对这一核心问题提出的关键性意见。我们十分同意，目前的正文中缺少对于动态追踪任务中“刺激-反应”关系具体操作化定义的描述；若将其作为本研究的核心问

题，必须对其进行明确的操作化界定，避免停留在抽象的理论表述。为更好地阐明相关概念，我们在正文部分做以下补充修改：

a.在 2.1 部分增加了对动态追踪任务数据格式及其理论意义的阐述（第 15 页 11 行至 16 行）

修改：

在此类任务中，行为反应不再以单次试次中的独立输出形式出现，而是表现为与时间变化对应的轨迹信号，例如手部位置、眼动速度或注视点位置等。因此，反应的记录方式也由传统的“刺激呈现—按键反应”转变为对连续行为状态的高时间分辨率采样，在采集得到的数据层面仍然表现为时间维度的“刺激-反应”对应。此时，每一时刻的行为输出均可视为对当前信息整合结果的即时表达，既包含来自外部刺激的驱动，也反映了先前状态与反馈信息的累积影响。这些研究共同指出，在连续交互、实时反馈及任务目标稳定存在的背景下，个体的信息加工过程表现出明显不同于静态任务的结构特征。……

b. 在 3 研究构想开头部分修改以明确对“刺激-反应”关系操作化定义及其模型化表达的说明（第 19 页 5 行至 10 行）

修改：

本研究旨在围绕视觉-动作控制中信息加工机制的统一建模问题，系统探讨不同任务条件下感知输入、反馈调节与先验预测之间的动态整合关系。前期研究表明，静态试次任务与动态连续任务在加工结构上存在显著差异，而现有理论框架尚不足以解释任务结构如何影响信息利用策略。在本研究中，我们将“刺激-反应”关系操作化定义为时间序列上的条件映射，即在任一时刻，行为输出是刺激状态及其历史信息的函数，而非对当前刺激的独立瞬时响应。在连续视觉-动作任务中，该映射由多源信息整合生成，这种映射的生成（process）机制可以数学化表达为卡尔曼滤波模型，其参数作为这种机制的刻画指标。因此，本研究将“刺激-反应”关系定义为一组可估计的动态整合参数及其变化规律，期望构建一个具备参数化解释力的加工模型，并在多个层面展开实证检验。

意见 2：

作者在“2.3 卡尔曼滤波模型的应用”和“3 研究构想”章节中，提出了很多变量符号，但对其意义的解释似乎不甚明确。例如研究一提到的“参数 R_p ”首次出现时并未定义，虽然后文提到了“视觉加工噪音 R_p ”，但放在前面的语境中还是不明确，比如前期数据中其与加工阈限呈现出相关，阈限确实是一个具体指标，但“加工噪音”并不是，像我一样的外行读者可能很难理解它究竟对应哪种可观测的行为表现。此外，我认为目前图 1、图 2 中的图示也过于分散和抽象，并不足以阐明研究的计算基础。

与基金申请书不同，作为公开发表的学术论文，若篇幅允许，不妨插入理论模型的一个或几个核心公式，并对相关的变量做统一说明；在公式中未出现的变量，则上下文中也不必专门单独命名，以免增加阅读负担。另外，当具体的实验范式确定后，这些抽象的变量最终体现在哪些测量指标上，例如是反应时，误差，准确率，还是某项眼动指标？也有必要明确。比如前面说的“加工噪音 R_p ”，在视动控制任务中到底如何定义和测量。即使实际开展研究时的方案可能会有调整，但在研究构想中也应有初步设定或举例说明。

回应：

感谢审稿专家的细致建议。我们十分同意原稿中对关键变量（如 R_p 、 K_t 等）的定义与行为意义说明不够充分，可能增加读者理解负担。针对这一问题，我们在“研究模块一”中补充了我们研究中使用的卡尔曼滤波模型核心公式，对变量之间的关系进行了明确表达，

在其中对关键参数（尤其是 R_p 和 K_t ）的含义进行了集中说明，我们希望以此提升整体表述的清晰性与可读性。以下是具体修改内容（第 21 页 12 行至第 22 页 7 行）。

修改：

对于被试在辨别任务中的反应，我们通过拟合累积正态分布曲线获得其参数 σ^2 ，将其作为辨别阈限的指标。对于追踪任务中采集得到的被试反应轨迹数据，我们基于卡尔曼滤波框架对连续反应过程进行建模。具体而言，被试在时刻 t 的反应更新可表示为：

$$\hat{S}_t = \hat{S}_{t-1} + K_t(I_{t-\Delta t} - \hat{S}_{t-1})$$

即被试在时刻 t 的反应 $\hat{S}_t$ 由当前的物理刺激 $I_{t-\Delta t}$ 和前一时刻的追踪反应 $\hat{S}_{t-1}$ 共同加权形成。其中的权重即卡尔曼系数 K_t ，可写为：

$$K_t = \frac{P_{t-1} + Q}{P_{t-1} + Q + R_p}$$

该参数由三个不确定性共同决定：上一时刻的预测不确定性 P_{t-1} 、刺激本身的变化程度 Q （在每个实验处理条件下为固定值），以及感知噪声方差 R_p 。其中， R_p 表示被试在感知当前刺激 $I_{t-\Delta t}$ 时所引入的内部误差大小（Bonnen 等，2015），而 P_{t-1} 表示基于上一时刻反应对当前状态进行预测时的不确定性。这两个参数共同决定系统在更新反应时应更依赖过去还是当前信息。需要注意的是， R_p 描述的是个体相对稳定的感知噪声水平，因此在同一实验条件下通常假设其为恒定值。 R_p 与 K_t 均通过拟合得到的后验估计得出。

意见 3：

有些研究的描述过于笼统，一些关键性的任务和信息没有解释。对于贯穿各研究的视动追踪任务，却一次都没有指明该任务到底要测量被试的什么指标，只说“分析被试追踪过程”“被试的追踪反应”“观察被试的视动追踪”，过于笼统。实验 6 中什么叫“产生与传统任务一致的学习效果”？实验 7 中什么叫“反馈结构”“优化动作反馈的方式”“反馈信息的冗余性”？每个实验至少应说清实验任务以及要操纵和观测的变量；不同实验采用同一观测指标的，至少首次提到时要说清。不然读者很难理解研究的设计和目的。

回应：

感谢审稿专家的细致审读与重要建议。我们同意，原文中对于各实验任务的具体内容、操纵方式及观测指标的说明不够明确，可能影响读者对研究设计与目的的理解。遵循审稿专家意见，我们在 3 研究构想部分对相关内容进行了补充与细化，明确了各实验的任务流程、操纵变量及对应的测量指标。具体修改如下：

a. 在 3.1 部分增加了研究一的示意图（第 21 页 1 行）。

修改：

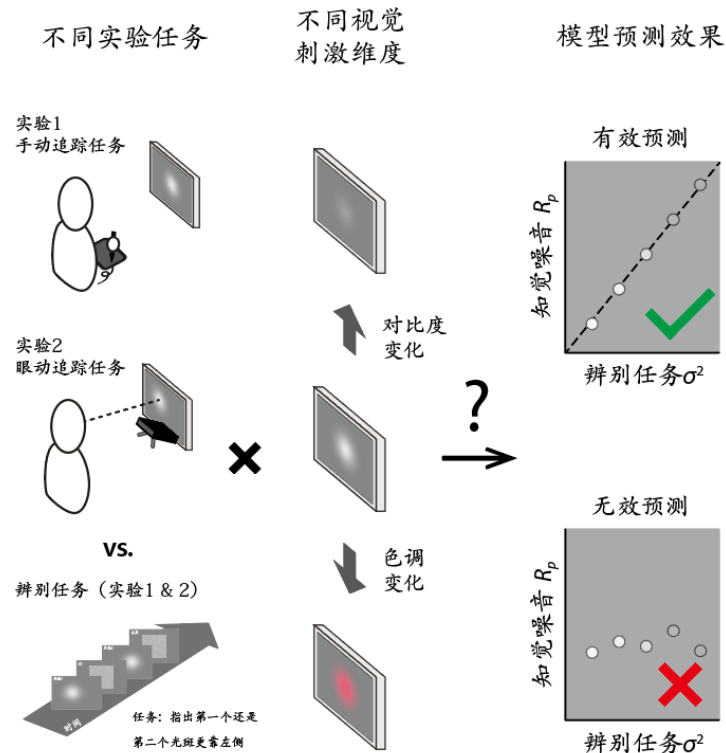


图 3：研究一（实验 1、2）实验设计及预期结果

b.在 3.2 部分补充了研究三与研究四的相关内容图示（第 23 页 4 行至第 24 页 7 行）。

修改：

其中，实验 5 计划通过在实验 1 的辨别任务（见图 3）基础上提升被试的亮度对比度敏感性，并检验知觉学习后，视动追踪任务反应中视觉加工噪音 R_p 是否会下降，以及视觉信息的权重 K_i 是否会提升。在实验 6 中，我们计划采用视动追踪任务作为知觉学习的训练任务，探究该训练是否能改变视动过程的信息整合，并检验其是否在行为与模型层面产生与传统辨别任务一致的学习效应。具体而言，本实验将比较 a) 训练前后追踪误差（tracking error）与反应延迟（lag）是否显著下降；b) 参数 R_p 是否下降（表征知觉编码精度提升），以及 K_i 是否发生系统性变化；c) 上述变化模式是否与实验 5 中辨别任务诱发的学习效应在 R_p 与 K_i 上的变化方向一致，从而判断两种训练是否涉及相同的视觉加工机制。

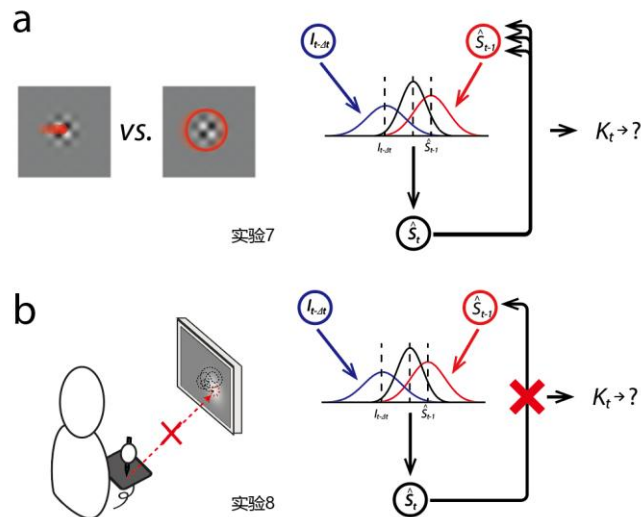


图 4 研究四（实验 7、8）实验设计

研究四则着重于反馈信号的调整对于视动过程信息整合的影响。其中，实验 7 将通过改变视觉反馈形式来操纵反馈信息的结构，使反馈信息具有更大的冗余性。具体而言，在标准条件下，被试通过单一位置点（红点）获得自身位置反馈；而在高冗余条件下，反馈将以与目标轮廓匹配的空心圆形式呈现，使被试能够同时获得位置与空间结构的一致性信息（见图 4a），从而提高单位时间内可用于误差评估的信息维度。实验 8 则通过使反馈信号在试次中随机短暂消失来操纵反馈信息的不确定性（见图 4b），其不确定性将通过反馈消失的时间比例进行量化。

在研究五中，我们计划操纵刺激运动轨迹的可预测性，从而观察先验知识是否会系统性改变视动控制过程中的信息整合。我们期望，可以从内隐（实验 9）和外显（实验 10）的两个角度同时探索先验的轨迹知识对于视动加工的影响。

我们计划在实验 9 中比较随机高斯噪音生成的轨迹和两种正弦波叠加生成的轨迹（相位随机与相位固定）下被试的手动追踪任务表现，从而探索无法外显习得的轨迹可预测性对于信息整合的影响。包含两项实验，分别探索横向一维运动轨迹（实验 9a）及二维运动轨迹（实验 9b）中被试的反应。在实验 10 中，我们计划通过以不同比例（0%~100%）加权叠加一维随机高斯噪音生成的轨迹和频率为 0.5 Hz 正弦波轨迹，同样生成具有不同规律性的轨迹。不同于实验 9，如此生成的轨迹可以令被试对轨迹的规律性有外显的意识。该实验将比较 a) 不同条件下追踪误差与反应延迟的变化趋势；b) 模型参数 R_p 是否随可预测性变化保持相对稳定；c) 整合权重 K_t 是否随正弦成分比例增加而下降，以刻画先验信息在信息整合中的调节作用。

意见 4:

“3.1 研究模块一”第二段中提到“相关任务流程与上述研究一中的流程一致”，但这里是文中首次提到研究一的内容，不知“上述”何在？而且这里本身就是在描述研究一的内容，那究竟是想说谁和谁相一致？需要澄清。如果是让读者看图，应直接引用图片。

回应:

感谢审稿专家指出的这一问题。“相关任务流程与上述研究一中的流程一致”之后提到的实验材料等内容属于研究一已实施实验的一部分，这一句为笔误，已作删除处理。现在“被试将参与三项任务，分别为……”可以与“在每一项任务中，……”合理衔接。

另有几个小问题:

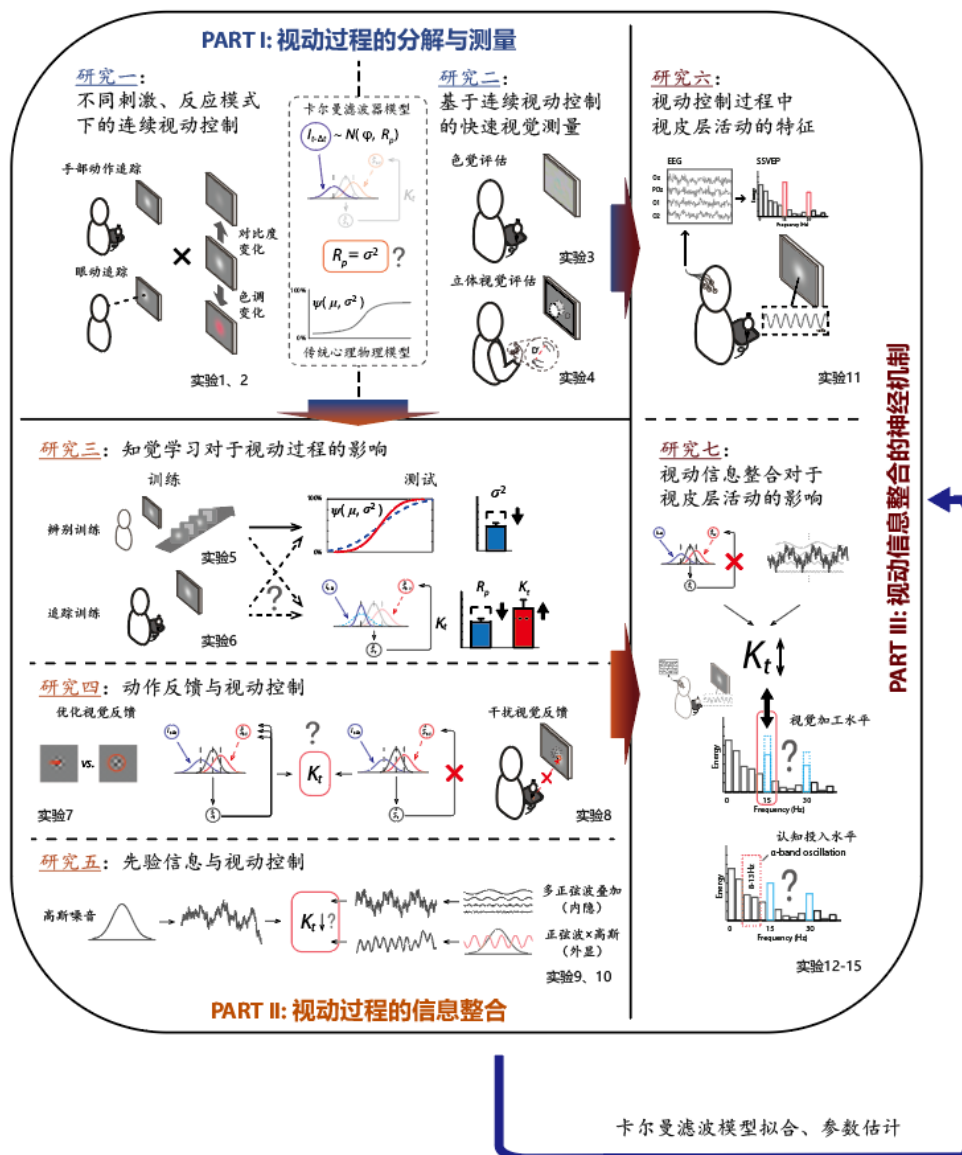
意见 5:

图 2 大框下面的小图太小了，文字看不清。其中一部分好像是跟图 1 一样，那可以去掉文字只保留图示，或是干脆使用文字描述。右边的小框则是完全不知道在干嘛。“JAGS”似乎是引用的一个算法程序，那不妨直接改用简要文字描述其作用，以免读者困惑。

回应:

感谢审稿专家指出的这一问题，我们删去了图 2 大框下的模型示意图和 JAGS 程序示意图，改为简要的文字描述。

修改:



意见 6:

“3.1 研究模块一”末段“(AUC = 1)”，其他地方用的都是全角括号，应统一。

回应:

感谢审稿专家指出的这一格式问题，已作相应修改。

意见 7:

“3.2 研究模块二”第三段中“我们计划在实验 9 中比较随机高斯噪音生成的轨迹（G）和两种正弦波叠加生成的轨迹（相位随机，S-F；相位固定，S-L；）下被试的手动追踪任务表现”，里面的“G”“S-F”“S-L”等代号在上下文并未出现，可以删去，以免给读者增加无用信息。

回应:

感谢审稿专家指出的这一问题，已删去本句的缩写代号。

意见 8:

“3.2 研究模块二”第四段“卡尔曼系数 K”，上文用的都是 K_t ，若是同一指标，则符号应保持一致。

回应:

感谢审稿专家指出的这一问题，已作相应修改。

意见 9:

“4 理论构建”第一段，“尽管已有地模型……”，应为“已有的”。

回应:

感谢审稿专家指出的这一语法问题，已作相应修改。

第二轮

审稿人意见:

作者对文章做了较为完善的修改，让文章的可读性有了很大提升；不过也让我注意到了上次没有发现的一个问题，虽然不影响整个研究的意义和有效性，但也属于关键的方法学问题，需要做出说明或修改。

回应: 感谢审稿专家的肯定！下文对每条意见进行了回应和修改。正文修改部分使用红色字体标明。

意见 1:

“3.1 研究模块一”次页第 12 行，“对于被试在辨别任务中的反应，我们通过拟合累积正态分布曲线获得其参数 σ^2 ，将其作为辨别阈限的指标。”这里有一个问题。用累积正态分布曲线拟合心理物理曲线时，辨别阈限一般为 σ （或与 σ 成正比），而非 σ^2 。虽说在数值上二者显然成正相关，但是二者量纲不同，在统计或物理意义上也并不能混为一谈： σ 与刺激强度的单位相同，其倒数 $1/\sigma$ 可以直接描述辨别的敏感度（sensitivity）；而 σ^2 更偏向于表示系统的噪声或不确定性，类似于本文模型中 R_p 的概念，其倒数衡量辨别的精度（precision）。除非有足够充分的理由（需要特别强调），否则将 σ^2 作为辨别阈限实际是一个方法学错误。如果作者仍希望检验 R_p 对辨别阈限的预测功能，那就需要修改分析方法，将其与 σ 而非 σ^2 进行相关分析，同时也需要对前期数据进行重新分析和报告；如果作者本意是希望检验 R_p 与 σ^2 的相关关系，那么就需要修改相关表述，将“辨别阈限”改为相应的其他指标。

回应:

感谢审稿专家指出这一关键问题。我们同意，原稿中将累积正态分布曲线拟合所得的 σ^2 表述为“辨别阈限”并不准确，容易混淆阈限与噪声/不确定性这两个不同层面的概念。结合

本研究的实际目的，我们在此处更希望检验的是连续追踪任务中提取的视觉加工噪声参数 R_p 与传统辨别任务中噪声成分之间的关系，而非将 σ^2 直接作为辨别阈限进行解释。

基于这一考虑，我们对相关内容作以下相应修改（第 24 页 12 行至 13 行）。

修改：

对于被试在辨别任务中的反应，我们通过拟合累积正态分布曲线获得其参数 σ ，并取 σ^2 作为辨别任务中信息加工噪声的指标。……

与此同时，后续段落中的相关表述也已统一由“加工阈限”修改为“加工噪声”，以避免概念混用，并使其与 R_p 的含义保持一致（第 24 页 15 行至 17 行）。

修改：

根据我们前期测得的行为数据（ $N = 28$ ）可以看出，手动追踪任务中提取的 R_p 可以较好地关联视觉加工噪声（ $r(145) = .655, p < .001$ ），而眼动追踪任务尽管与视觉加工噪声之间也存在联系，但似不足以有效评估加工噪声（ $r(145) = .163, p = .049$ ）。这可能与眼动过程中的外周视野加工以及眼跳现象有关。

另有 3 个小建议：

意见 2：

“3.1 研究模块一”次页第 17 行，“其中的权重即卡尔曼系数 K_t ”，“权重”补充为“刺激权重”会更明确。

回应：

感谢审稿专家指出的这一问题，已作相应修改，该段落跟随意见 3 迁移至 2.3 节相应位置（第 20 页 19 行）。

意见 3：

以上提到的本次新增的公式及变量解释部分，目前位于 3.1 节中，是否挪到 2.3 节中更合适？毕竟该公式在图 1 中就已出现，其中的变量也在 2.3 节中首次提到。而且这算是基本原理，放在目前位置反而打断 3.1 节对具体实验设计的阐述。在 3.1 节中将涉及到的变量再提一下即可。

回应：

感谢审稿专家指出这一问题。我们同意，原稿中将新增的公式及变量解释部分放在 3.1 节中，确实会在一定程度上打断该部分对具体实验设计的展开；同时，相关公式与变量本身属于卡尔曼滤波模型的基本原理，与 2.3 节“卡尔曼滤波模型的应用”的内容联系更为紧密。根据审稿专家建议，我们已将该部分内容调整至 2.3 节中（第 20 页 16 行至 26 行）。

同时，对 3.1 节的原先对应位置作相应的衔接表述（第 24 页 14 行）。

修改：

……对于追踪任务中采集得到的被试反应轨迹数据，我们基于卡尔曼滤波框架对连续反应过程进行建模，通过后验估计得到 R_p 与 K_t 的估计值。

意见 4：

“3 研究构想”第一段新增部分，“在本研究中，我们将‘刺激 - 反应’关系操作化定义为时间序列上的条件映射……瞬时响应”，其中的“条件映射”是什么意思？之前在别处好像没有见过这个术语，也没有查到一个准确的定义。如果确认是一个既有术语的话，可以

附上英文，否则建议换一个更准确的词。另外，“行为输出是刺激状态及其历史信息的函数”，按我对后文卡尔曼滤波模型的理解，这里的“历史信息”应该是指“行为输出的历史”而不是“刺激状态的历史”吧？如果是这样的话，当前写法会让读者以为是后者。建议改为“行为输出是刺激状态及历史输出信息的函数”或“行为输出是刺激状态及上一时刻输出的函数”。

回应：

感谢审稿专家指出这一表述问题。我们同意，“条件映射”这一说法不够准确，也并非广泛使用的既有术语，可能增加读者理解负担。我们已将该表述细化为“刺激联合时间序列上递归更新（recursive updating）的效应与反应之间的映射关系”，以更准确地表达本研究所关注的并非当前刺激与当前反应之间的瞬时对应，而是连续时间过程中当前行为输出由当前刺激状态与前一时刻行为输出共同决定的动态关系。同时，我们也将“行为输出是刺激状态及其历史信息的函数”修改为“行为输出是刺激状态及上一时刻输出的函数”，以避免“历史信息”被误解为刺激状态的历史。我们希望通过这一修改，使相关概念的操作化定义与卡尔曼滤波模型的逻辑保持一致，并进一步提升表述的可读性（第 22 页 5 行至 6 行）。

修改：

.....在本研究中，我们将“刺激-反应”关系操作化定义为刺激联合时间序列上递归更新（recursive updating）的效应与反应之间的映射关系，即在任一时刻，行为输出是刺激状态联合上一时刻输出的函数，而非对当前刺激的独立瞬时响应。

第三轮

审稿人意见：

作者已对我的全部意见做了充分的回应和修订，不过图 3 下面第 4-6 行，“以分析被试的.....视知觉加工阈限”，该句末的“加工阈限”是否也应该随后文一起改为“加工噪声”？除此以外没有其他问题了。

回应：

感谢审稿专家指出这一问题。我们已经对相应文本作出修（第 25 页 4 行至 6 行），修改部分使用绿色字体标明。

修改：

其中，研究一计划采用视觉追踪任务，以分析被试的手部动作追踪过程（实验 1）或眼动追踪过程（实验 2）是否可以反映被试对于刺激不同维度（对比度、色调）的视知觉加工噪声。...

编委复审意见：

稿件已达到发表水平，同意接收。