

《心理学报》审稿意见与作者回应

题目：视听时、空一致性对 Pip-and-Pop 效应的影响

作者：唐晓雨，崔鑫忠，高敏，袁梦莹

第一轮

审稿人 1 意见：

本研究使用动态视觉搜索范式，并使用眼动技术自动记录被试搜索过程中的眼球运动，通过分析眼动指标可以考察搜索集合，视听空间一致性，视听时间一致性三个因素对搜索中注意转移的影响，以探究其如何影响 Pip-and-Pop 效应，进一步求证是多感觉通道整合还是 oddball 在起作用。实验具有一定的创新意义，一些意见供参考：

意见 1：摘要只是现象结果的描述，理论意义在哪里？摘要中并没有明确体现出来。

回应：感谢审稿专家的宝贵意见。根据审稿专家的建议，连同第二条审稿意见，我们在摘要中补充了本研究的理论意义以及自变量引入的理论意义。(详见文中第 13 页摘要红色字体部分)

“摘要 多感觉整合遵循空间原则和时间原则，有研究表明 Pip-and-Pop 效应产生的原因是多感觉整合，那么 Pip-and-Pop 效应是否同样遵循空间原则和时间原则呢？本研究采用动态视觉搜索范式，通过两个眼动实验考察空间一致性(实验 1)和时间一致性(实验 2)对 Pip-and-Pop 效应的影响。结果发现：(1)当视觉目标颜色变化伴随一个同侧的听觉刺激时(同侧条件)所产生的 Pip-and-Pop 效应最大，其次为双侧条件，对侧条件没有发现 Pip-and-Pop 效应。(2)视听刺激同时呈现所产生的 Pip-and-Pop 效应最大，随视听刺激呈现时间间隔的增加 Pip-and-Pop 效应减弱。结果表明，视听刺激在空间、时间上越一致，Pip-and-Pop 效应越大，即视听时、空一致性对 Pip-and-Pop 效应具有调节作用，这为多感觉整合在产生 Pip-and-Pop 效应中的作用提供了证据。”

意见 2：同问题一，摘要中为什么要做这样的操控，这样的自变量的引入的意义在哪里？与理论意义之间的联结在哪里并没有体现出来。

回应：感谢审稿专家的宝贵意见。根据审稿专家的建议，连同第一条审稿意见，我们在摘要中补充了本研究的理论意义以及自变量引入的理论意义。(详见文中第 13 页摘要红色字体部分)

“摘要 多感觉整合遵循空间原则和时间原则，有研究表明 Pip-and-Pop 效应产生的原因是多感觉整合，那么 Pip-and-Pop 效应是否同样遵循空间原则和时间原则呢？本研究采用动态视觉搜索范式，通过两个眼动实验考察空间一致性(实验 1)和时间一致性(实验 2)对 Pip-and-Pop 效应的影响。结果发现：(1)当视觉目标颜色变化伴随一个同侧的听觉刺激时(同侧条件)所产生的 Pip-and-Pop 效应最大，其次为双侧条件，对侧条件没有发现 Pip-and-Pop 效应。(2)视听刺激同时呈现所产生的 Pip-and-Pop 效应最大，随视听刺激呈现时间间隔的增加 Pip-and-Pop 效应减弱。结果表明，视听刺激在空间、时间上越一致，Pip-and-Pop 效应越大，即视听时、空一致性对 Pip-and-Pop 效应具有调节作用，这为多感觉整合在产生 Pip-and-Pop 效应中的作用提供了证据。”

意见 3：为什么之前的文献表明-100ms 最快，当前实验结果发现是 0ms 最快？

回应：抱歉在文中有关不同时间条件下结果的不当表述。前人文献还是当前的实验结果，均发现 0ms 是最快的，Pip-and-Pop 效应是最大的。相关内容正文中已修改。(详见文中第 16 页红色字体)

“另一方面，以往有关视听刺激时间一致性对 Pip-and-Pop 效应影响的研究结果也是不一致的。Van der Burg 等(2008)使用搜索集为 48(1 个目标刺激和 47 个干扰刺激)的动态视觉搜索范式，在研究中设置了 8 种声音目标间隔(Tone-Target-Interval, TTI)条件，结果发现 TTI 为-100ms、-50ms、-25ms、0ms、25ms、100ms 时，听觉刺激才会促进对视觉目标的搜索，视听刺激同时呈现时(0ms 条件)的促进作用最大，且声音目标间隔与反应时之间呈 U 型函数关系，即视听刺激呈现的时间一致性越高，搜索越快。”

意见 4：当前的研究是否可以进一步解决多感觉整合还是 oddball 刺激吸引注意这个争论呢？作者在讨论中做了一些阐述，但是在摘要和前言中并没有十分清晰的阐述和相应的预期。

回应：感谢审稿专家的宝贵意见。当前的研究结果表明 Pip-and-Pop 效应至少有部分是由多感觉整合引起的，但也不能完全排除 oddball 刺激吸引注意在其中的作用。正文前言中空间

一致性和时间一致性变量引入的部分已加入相关理论阐述，前言最后也加入了相应的预期。
(详见文中第 15~17 页红色字体部分)

“一方面，前人有关视听刺激空间一致性对 Pip-and-Pop 效应影响的研究结果并不一致。Ngo 和 Spence 考察了视听刺激空间一致性对 Pip-and-Pop 效应的影响，他们发现与无空间信息线索或无效空间信息线索相比，具有空间信息的听觉或触觉线索会提高视觉搜索的效率(Ngo & Spence, 2010a)，即视听刺激空间一致性对 Pip-and-Pop 效应产生了促进作用。所以研究者认为 Pip-and-Pop 效应是视听刺激整合引起的，空间位置越一致，视听刺激越容易进行整合，Pip-and-Pop 效应越大。但 Fleming 等(2020)研究发现，视听刺激的空间位置是否一致，都会提高搜索效率。行为与脑电结果均显示空间位置一致与否对 Pip-and-Pop 效应的影响差异不显著，表明视听刺激空间一致性并没有对 Pip-and-Pop 效应产生影响(Fleming et al., 2020)。这说明 Pip-and-Pop 效应可能并不是视听刺激整合引起的。此外，上述两项研究结果不一致的原因还可能是与二者视觉刺激在空间排列上的差异有关，Ngo 和 Spence (2010)在动态视觉搜索范式中将视觉刺激分成左右两区域呈现；而后者 Fleming 等(2020)在此基础上缩小了搜索区域且增加了干扰线段的方向，使得动态视觉搜索任务的空间关联性变得不如前者那么密切。本研究使用动态视觉搜索范式，将视觉搜索区域分为左右两侧(与中心点有明显距离)，线段整齐地排列在网格中(以搜索集 36 为例，左右分别为 3×6 的网格)，并使用眼动技术自动记录被试搜索过程中的眼球运动，同时眼动数据可以揭示实时信息处理过程以及搜索过程中的眼跳计划，通过分析眼动指标可以考察视听空间一致性对搜索中注意转移的影响，以探究视听空间一致性如何影响 Pip-and-Pop 效应。

另一方面，以往有关视听刺激时间一致性对 Pip-and-Pop 效应影响的研究结果也是不一致的。Van der Burg 等(2008)使用搜索集为 48(1 个目标刺激和 47 个干扰刺激)的动态视觉搜索范式，在研究中设置了 8 种声音目标间隔(Tone-Target-Interval, TTI)条件，结果发现 TTI 为-100ms、-50ms、-25ms、0ms、25ms、100ms 时，听觉刺激才会促进对视觉目标的搜索，视听刺激同时呈现时(0ms 条件)的促进作用最大，且声音目标间隔与反应时之间呈 U 型函数关系，即视听刺激呈现的时间一致性越高，搜索越快。Van der Burg 等认为，同时呈现的听觉刺激以及在整合时间窗内(-100ms~100ms)呈现的听觉刺激引发的搜索效率提高是由于跨通道整合增强了视觉目标的凸显性(Van der Burg et al., 2008)。但也有研究发现视听刺激时间一致性并不会影响 Pip-and-Pop 效应，Zou 等(2012)研究者在搜索集为 36 的条件下设置了听觉刺激与视觉目标颜色变化同时呈现、听觉刺激先于视觉目标颜色变化 100ms 呈现以及听觉刺激随机呈现三种条件，研究结果发现这三种条件都会促进视觉目标的搜索，但它们之间

并没有显著差异。Zou 等认为这三种条件都提高搜索效率且没有差异的原因是无空间信息的听觉刺激引发的“冰冻效应”所致(Zou et al., 2012), 即当听觉刺激与视觉目标同步呈现时, 被试主观上感知到的视觉目标呈现的时间变长了(Ngo & Spence, 2012; Tsai & Yeh, 2013; Vroomen & De Gelder, 2000)。所以 Zou 提出为 Pip-and-Pop 效应并不是因为多感觉整合引起的, 也不遵循时间原则。”

“基于上述视听刺激空间、时间一致性对 Pip-and-Pop 效应的影响研究结论并不一致, 本研究将眼动技术运用到动态视觉搜索范式中, 通过两个实验考察两个问题: (1)调控搜索集大小和听觉刺激呈现位置, 探究视听刺激空间一致性如何影响 Pip-and-Pop 效应。(2)调控搜索集大小和听觉刺激与视觉目标的间隔时间, 探究视听刺激时间一致性如何影响 Pip-and-Pop 效应。如果 Pip-and-Pop 效应产生的原因是多感觉整合, 那么 Pip-and-Pop 效应遵循空间一致性和时间一致性原则, 即视听刺激呈现的空间位置越一致, 产生的 Pip-and-Pop 效应越大; 视听刺激呈现的时间越一致, 产生的 Pip-and-Pop 效应越大。”

意见 5: 结果部分, “所有被试的正确率均达到 98%”, 应改为“所有被试的正确率均达到 98% 以上”

回应: 感谢审稿专家的宝贵意见。已在文中将“所有被试的正确率均达到 98%”改为“所有被试的正确率均达到 98% 以上”。(详见文中第 20 页、26 页红色字体部分)

意见 6: 结果部分: “并剔除反应过快(在目标变化前反应)和反应时间过长的试次”。这里最好写出具体的删除标准, 比如是快于 150ms, 慢于 1000ms 的试次删除。

回应: 感谢审稿专家的宝贵意见。首先解释为什么没有给出例如“快于 150ms, 慢于 1000ms 的试次”这样具体的数据, 因为每个试次目标颜色变化的时间不同, 目标颜色变化随机出现在 3、4、5、6、7、8 间隔之后, 所以目标第一次变化的时间不是一个固定值, 无法给出“比如快于 150ms”这样的固定值。其次, 我们之前的描述确实过于笼统, 反应过快指目标首次变化前被试就已经做出反应的试次, 反应时间过长试次指被试在目标变化了 10 次及以上依旧没有做出反应的试次。已在文中做出详细的修改。(见文中结果与分析部分, 第 20 页、26 页红色字体)

“剔除反应错误以及反应过快和过慢的试次, 反应过快指目标首次变化前被试就已经做出反应的试次, 反应过慢试次指被试在目标变化了 10 次及以上依旧没有做出反应的试次 (Van der Burg et al., 2008)。最终剔除试次占总试次的 5.5%。”

意见 7：结果呈现，最好将分析方法和结果分成两个部分进行描述，分析方法单独一个部分进行描述。

回应：感谢审稿专家的宝贵意见。已将分析方法和结果分析分开进行描述。(详见文中 19~20 页数据分析、结果与分析红色字体部分)

“2.1.4 数据分析

正确率(Accuracy)指被试正确完成试次的概率。反应时(Reaction time)指搜索屏呈现到被试做出反应之间的时间间隔。搜索效率指在正确反应的基础上，反应的快慢。反应时越长，说明被试找到目标的速度越慢，搜索效率越低；反应时越短，说明被试找到目标的速度越快，搜索效率越高。

反应时和正确率只能说明听觉刺激能否提高搜索效率，为了进一步探究听觉刺激具体如何影响搜索效率，根据以往研究选择了三个眼动指标(Zou et al., 2012)：①平均注视次数(Mean fixation number)指落入当前兴趣区注视点的次数。该指标可以有效地反映知觉负荷大小，知觉负荷越大，平均注视次数越多，搜索效率越低；②平均眼跳幅度(Mean saccade amplitude)指从当前注视到下一次注视之间的距离。眼跳幅度越大，说明被试在眼跳前获取的信息越多，搜索效率越高；③平均注视时间(Mean fixation duration)指落入当前兴趣区所有注视点的注视时间的平均。该指标反应整体的加工过程，Zou 等(2012)用该指标衡量冰冻效应，探究呈现的听觉刺激是否拓展了平均注视时间。

为了考察视听空间一致性和搜索集对 Pip-and-Pop 效应的影响，对所有指标都进行了 3(搜索集大小：36, 48, 60)×4(视听空间一致性：无声音，同侧，对侧，双侧)的重复测量方差分析。

2.2 结果与分析

剔除反应错误以及反应过快和过慢的试次，反应过快指目标首次变化前被试就已经做出反应的试次，反应过慢试次指被试在目标变化了 10 次及以上依旧没有做出反应的试次(Van der Burg et al., 2008)。最终剔除试次占总试次的 5.5%。

2.2.1 正确率

.....”

意见 8：实验二开头最好加入一段实验二的目的简介。

回应：感谢审稿专家的宝贵意见。已在实验 2 开头加入实验目的简介，同时给实验 1 和实验 2 加上题目，以达到简介各个实验的目的。(详见文中 17 页、25 页标题红色字体部分)

“实验 1：视听空间一致性对 Pip-and-Pop 效应的影响

实验 2：视听时间一致性对 Pip-and-Pop 效应的影响

实验 2 的目的是考察 Pip-and-Pop 效应是否遵循时间原则，即时间一致性对 Pip-and-Pop 效应是否具有调节作用。”

.....

审稿人 2 意见：

研究通过两个实验试图揭示视听时、空一致性对 Pip-and-Pop 效应的影响，通过在不同空间，不同时间呈现听觉刺激，两个实验共同揭示了视听刺激在空间、时间上越一致，Pip-and-Pop 效应越大。本研究试图通过多感觉整合和的 oddball 刺激吸引注意探究 Pip-and-Pop 效应的内在机制，对揭示 Pip-and-Pop 效应的产生原因具有比较重要的理论意义。①然而，本研究也存在一些不足之处，如全文有部分多字或少字现象，建议全文阅读并进行改正；②对一些学术性概念没有进行精细的定义，比如搜索效率；③同时在结果的分析 and 讨论中逻辑串联性不强，只是单纯的把结果罗列，应该讲清楚为何这些指标(反应时，平均注视次数，眼跳幅度)共同推出了我们的结果(搜索效率)。总体来说，建议进行进一步修改完善。

回应：感谢审稿专家的宝贵意见。①已通读全文进行修改。②、③在实验结果前加入了数据分析部分，解释了每个指标的意义，以及每个指标的变化对应的搜索效率变化，说明了这些指标的变化如何体现搜索效率的改变。(详见文中第 19 页数据分析部分)

“2.1.4 数据分析

正确率(Accuracy)指被试正确完成试次的概率。反应时(Reaction time)指搜索屏呈现到被试做出反应之间的时间间隔。搜索效率指在正确反应的基础上，反应的快慢。反应时越长，说明被试找到目标的速度越慢，搜索效率越低；反应时越短，说明被试找到目标的速度越快，搜索效率越高。

反应时和正确率只能说明听觉刺激能否提高搜索效率，为了进一步探究听觉刺激具体如何影响搜索效率，根据以往研究选择了三个眼动指标(Zou et al., 2012)：①平均注视次数(Mean fixation number)指落入当前兴趣区注视点的次数。该指标可以有效地反映知觉负荷大小，知觉负荷越大，平均注视次数越多，搜索效率越低；②平均眼跳幅度(Mean saccade

amplitude)指从当前注视到下一次注视之间的距离。眼跳幅度越大,说明被试在眼跳前获取的信息越多,搜索效率越高;③平均注视时间(Mean fixation duration)指落入当前兴趣区所有注视点的注视时间的平均。该指标反应整体的加工过程,Zou 等(2012)用该指标衡量冰冻效应,探究呈现的听觉刺激是否拓展了平均注视时间。

为了考察视听空间一致性和搜索集对 Pip-and-Pop 效应的影响,对所有指标都进行了 3(搜索集大小: 36, 48, 60)×4(视听空间一致性: 无声音, 同侧, 对侧, 双侧)的重复测量方差分析。”

意见 1: 实验 1 设置了搜索集大小(36, 48, 60)、变化间隔为 50、100 或者 150ms, 共有九个变化间隔(一个变化周期), 三种变化间隔随机出现、以及视听空间一致性(无声音, 同侧, 对侧, 双侧)。

1、三个变量中变化间隔的设置意义不明确, 并且也没有对其设置的目的和意义进行解释, 在分析过程中也没有运用到这个变量。如果有目的的设置, 请进行进一步的解释。

回应: 感谢审稿专家的宝贵意见。实验 1 中的自变量是搜索集大小和空间一致性, 变化间隔并不是实验的自变量, 所以在分析时没有考虑。视觉刺激的变化间隔(50、100 或者 150ms)参考了 Van der Burg 等(2008)、Ngo(2010a)、Zou 等(2012)的研究, 如此设置的原因是避免视觉刺激在进行连续变化时有很高的规律性。

2、搜索集大小的设置通过阅读全文可以知道是为了检验知觉负荷对 Pip-and-Pop 效应的调节作用, 但是在实验 1, 以及前言实验 1 的铺垫(空间一致性)中并未提及到知觉负荷, 只是前言中在讲时间一致性(实验 2 的铺垫)中提到了知觉负荷, 而在这里直接强调“本研究调控两种知觉负荷差距较大的条件(搜索集为 36 和 60)”, 所以这个变量设置的逻辑有点混乱, 建议再梳理一下。

回应: 感谢审稿专家的宝贵意见。前言空间一致性部分确实没有提到知觉负荷, 对此进行修改, 在时间一致性介绍结束后引出知觉负荷, 并且简要的概括了以往研究中知觉负荷在时间、空间一致性中对 Pip-and-Pop 效应的影响, 来说明引入搜索集大小这个自变量的原因。其次在实验 2 中解释了为什么实验 1 的搜索集大小是(36、48、60), 而实验 2 改为(36、60)。(见文中第 16~17 页、25 页实验与仪器材料红色字体部分)

“上述研究发现视听刺激时间一致性对 Pip-and-Pop 效应影响的结果不一致, 还可能是实验任务中的搜索集大小不同。Kösem 和 Van(2012)在搜索集大小为 60 的研究中发现视听时间一致性在听觉刺激影响视觉目标搜索的过程中起重要作用(Kösem & van Wassenhove,

2012)。相比 Van der Burg 等(2008)以及 Kösem 和 Van(2012)的实验任务, Zou 等(2012)实验可能由于搜索任务过于简单, 导致三种时间条件没有产生差异。此外, 在空间一致性对 Pip-and-Pop 效应影响的相关研究中, Ngo 等(2010)没有发现搜索集大小(24, 36, 48)的调节作用, 而 Fleming 等(2020)使用相同的搜索集大小却发现了知觉负荷对空间一致性的调节作用。综上所述, 知觉负荷在空间一致性和时间一致性中的调节作用并不一致, 且上述研究中知觉负荷的调节作用仅反映在反应时这个指标上。所以本研究引入搜索集大小这一变量, 并通过行为与眼动两方面数据分析以探究在不同知觉负荷下视听刺激时间一致性和空间一致性对 Pip-and-Pop 效应的影响。”

“实验仪器同实验 1。视觉刺激材料有两点不同于实验 1: 第一, 本实验中搜索集大小为 36 和 60, 由于实验 1 中只在平均眼跳幅度上发现了搜索集 48 与 36 和 60 有显著差异, 而反应时、平均注视次数以及平均注视时间上均无显著差异, 所以取消了搜索集为 48 的条件。”

意见 2: 实验 1 结果发现在同侧条件下呈现听觉刺激, 反应时最短, 平均注视次数最少, 眼跳幅度最大, 因此认为在空间一致条件下搜索效率最高, 产生的 Pip-and-Pop 效应最强, 但是还可以存在一种可能这种搜索效率并不是由于听觉刺激增强了视觉目标变化的凸显性导致的, 而是听觉刺激在同侧出现, 被试把听觉刺激单纯的当作了一个提示线索, 提示目标刺激会在线索一侧出现, 导致的搜索效率的提升。也需要在实验 1 的结论部分解释一下。

回应: 感谢审稿专家的宝贵意见。审稿人提出的“提示线索”这一可能性与实验 1 讨论中“有效空间线索”所表达的意思相同, 所以我们将“提示线索”这一可能性在文中相应地方加入作为补充解释。听觉刺激确实可能作为一个提示线索, 从而促进搜索效率的提高。但听觉刺激的提示作用是把被试引导到同侧, 并没有直接定位到目标, 被试要继续进行搜索, 那么被试仍然会受到听觉刺激的作用, 此时听觉刺激会起到增强视觉目标变化凸显性的作用, 因为在双侧条件下, 即使没有提示线索, 仍然发现了 Pip-and-Pop 效应。所以可能是听觉刺激本身的提示及其与视觉目标整合的共同作用导致搜索效率的提高, 产生 Pip-and-Pop 效应。(详见文中第 23~24 页红色字体部分)

“同侧条件下, 当听觉刺激从目标侧声道发出, 在听觉刺激的有效空间信息的作用下, 被试可以直接快速地定位到目标侧进行搜索, 在目标侧通过少量注视, 大幅度眼跳快速的搜索到目标线段, 搜索时间短, 效率高, 所以产生的 Pip-and-Pop 效应最大。这说明听觉刺激可能被当作提示线索, 从而导致搜索效率的提高。但听觉刺激引导被试到同侧后并不代表可

以立即定位到目标的位置，接下来的搜索过程中，听觉刺激仍然起到增强视觉目标变化凸显性的作用，因为在双侧条件下，即使没有提示线索的作用，也发现了 Pip-and-Pop 效应。所以可能是听觉刺激的提示及其与视觉目标整合的共同作用导致搜索效率的提高，产生 Pip-and-Pop 效应。而对侧条件下，当听觉刺激从非目标侧声道发出，在听觉刺激的无效空间信息作用下，被试先在非目标侧进行大量搜索后再回到目标侧搜索，在非目标侧进行多次注视，眼跳幅度不断缩小进行细致搜索，没有发现目标线段后再回到目标侧进行搜索，相比于同侧条件，对侧条件在非目标侧的无效搜索产生了搜索成本，通过两个条件的注视次数差异热点图也可以明显地发现对侧条件在非目标侧上的搜索成本(见图 3)，所以同侧条件的搜索效率显著高于对侧条件。以往研究发现与不提供空间信息的听觉刺激相比，提供空间信息的听觉刺激能够减少被试辨别和检测目标的潜伏期(Perrott et al., 1990)。因此，听觉线索和视觉目标在空间位置上一致可以有效的将被试的空间注意指向视觉目标，更加利于被试搜索到视觉目标。此外，也有研究发现来自不同通道的刺激位置接近时，产生的多感觉相互作用最强；不同通道的刺激之间空间距离增加时多感觉相互作用减弱(Lewald et al., 2001; Slutsky & Recanzone, 2001; Soto-Faraco et al., 2003; Stein & Stanford, 2008)。综上，本实验中的同侧条件可能是听觉刺激的提示作用和视、听刺激方位一致产生了较强的多感觉相互作用，二者共同促进了搜索效率，所以产生的 Pip-and-Pop 效应最大；而对侧条件可能是听觉刺激的无效提示作用和两个通道刺激方位不一致，距离增大，多感觉相互作用减弱，所以不足以产生 Pip-and-Pop 效应。”

意见 3:

1、实验 2 的实验设计和程序中，并没有讲清楚实验的具体流程和被试的任务，虽然我们可以了解到实验 2 的具体流程和被试任务与实验 1 一致，但是这里也应该明确实验 2 与实验 1 一致之处。

回应：感谢审稿专家的宝贵意见。实验 2 与实验 1 的实验流程是完全一致的，已在文中加入说明。(详见文中第 25 页实验设计和程序红色字体部分)

“实验采用 2(搜索集大小：36, 60)×6(视听时间一致性：无声音条件，-200ms, -100ms, 0ms, 100ms, 200ms)两因素被试内实验设计。实验具体流程如图 4a 所示，与实验 1 的实验流程完全一致。正式实验包括 12 个 block.....”

2、①实验 2 的结果与分析部分，很多句子多次反复出现，过于冗杂，建议简略。②并且通过实验 2 的结果图可以发现，在搜索集大小为 60，听觉刺激呈现时间为(-200ms、200ms)

的条件下反应时和平均注视次数都有较大的差异,但是在实验结果和分析中并没有提到这个结果,是否也存在一个听觉刺激先于视觉刺激出现和听觉刺激后于视觉刺激出现的视觉搜索效率的差异呢?

回应: ①感谢审稿人的宝贵意见, 已对本文的实验结果部分进行简略。②在搜索集大小为 60 时, -200ms 与 200ms 条件下的反应时($p = 0.09$)和平均注视次数($p = 0.058$)无显著差异, 不能说明存在一个听觉刺激先于视觉刺激出现或听觉刺激后于视觉刺激出现的搜索效率差异。其次这个结果不能用来说明时间一致性对 Pip-and-Pop 效应的影响, 所以我们没有把这个结果放入文章进行详细解释。

3、实验 2 的讨论部分提到了“冰冻效应”是指当目标刺激与听觉刺激同时呈现时, 被试报告对视觉目标的主观“冰冻”感受, 被试主观上认为视觉目标呈现的时间延长了, 这里的时间延长是指被试主观认为延长, 并不是客观上的延长, 而在接下来的讨论中“冰冻效应”会导致平均注视时间变长。这里的逻辑需要重新梳理。

回应: 感谢审稿专家的宝贵意见。“冰冻效应”最初是 Vroomen 和 De Gelder (2000)提出的, 是被试报告的一个主观现象, Zou 等(2012)提出可以用平均注视时间这个眼动指标来客观衡量冰冻效应。我们参考了 Zou 等(2012)的研究, 用平均注视时间来衡量冰冻效应。已在文中对此部分的逻辑重新进行了梳理。(详见文中第 30 页红色字体部分)

“在平均注视次数眼动指标中发现了与反应时相同的结果, 平均注视次数与 TTI 之间的关系呈 U 型曲线: TTI 的绝对值越小, 平均注视次数越少; 平均眼跳幅度与平均注视时间结果也表明只有 TTI 在 0ms, -100ms, 100ms 条件下, 才会优化搜索的眼动行为。这种在眼动行为上的优化在以往研究中被认为是冰冻效应(freezing effect)所致, Zou 等(2012)使用平均注视时间来量化冰冻效应。即无空间信息的听觉刺激“冰冻”了眼动, 突然出现的听觉刺激抑制了眼跳, 使眼跳延迟, 从而搜索过程中的平均注视时间变长, 那么被试就可以在更长的时间和更广阔的空间内通过增大眼跳幅度、减少注视次数对视觉目标进行搜索, 从而更精确地引导接下来的注视指向目标, 加快对目标的定位。此外, 由这种“冰冻”所引起的时间变长不仅可以提高目标在当前区域的搜索效率, 也可以快速拒绝目标不在的区域, 从而提高整体的搜索效率(Zou et al., 2012)。与以往研究结果不同的是, 本研究发现 0ms 条件时的平均注视时间最长, 这说明听觉刺激对视觉搜索的促进作用并不全是冰冻效应引起的。综上表明 Pip-and-Pop 效应遵循时间一致性原则, 视听刺激呈现时间越一致, 产生的 Pip-and-Pop 效应越大。”

意见 4： 在两个实验的结果与分析部分均把正确率的结果放在这一部分的总括段，建议作为一个小节即可，因为这个正确率的结果，并没有起到一个总结的效果。

回应：感谢审稿专家的宝贵意见。已在文中结果部分将正确率作为一个小节呈现。(详见文中 20 页、26 页红色字体部分)

第二轮

审稿人 1 意见：

意见 1：这个听觉刺激和当前任务是无关的，为什么会提升搜索效率？提高了警觉状态？可以增加一些视听整合提升视觉效率这方面的内容简单介绍。

回应：感谢审稿专家的宝贵意见。虽然这个听觉刺激和当前任务是无关的，但听觉刺激与视觉目标变化在时间上绑定在一起，即听觉刺激与目标线段颜色变化同时呈现，这种时间上的同步呈现使听觉刺激增强了视觉目标变化的凸显性，从而让视觉目标在众多不断变化的刺激中凸显出来。根据上述解释，这种搜索效率的提升是多感觉整合所导致的。对于提高警觉状态的解释，Van der Burg 等人(2008)通过将同步的听觉线索换成视觉闪光的视觉线索做对比，在视觉线索条件下没有发现 Pip-and-Pop 效应，所以这种搜索效率的提升并不是提高了警觉状态。根据审稿专家的意见，已补充“视听整合提升视觉搜索效率”这方面的内容。(详见文章 20-21 页蓝色字体)

“有研究认为 Pip-and-Pop 效应产生的原因是多感觉整合(Chamberland et al., 2016; Van der Burg et al., 2008; Van der Burg et al., 2011)。多感觉整合(multisensory integration, MSI)是指来自不同感觉通道的信息相互作用，并整合为统一、连贯且有意义的知觉过程(Talsma et al., 2010; Tang et al., 2016; Van der Stoep et al., 2017)。Van der Burg 团队的研究发现，同步呈现的听觉刺激与视觉目标颜色变化后 50ms 在顶枕区发生了早期的多感觉整合，并且这一早期多感觉整合与行为学上的 Pip-and-Pop 效应显著相关，即早期多感觉整合越大，搜索效率的提高越大。Van der Burg 等人认为，通过早期整合将听觉与视觉目标绑定，使得视觉目标在众多不断变化的刺激中凸显出来，更好地捕获注意，与只有视觉目标条件相比搜索效率显著提高，所以多感觉整合在 Pip-and-Pop 效应中有着关键作用(Van der Burg et al., 2008; Van der Burg et al., 2011)”

意见 2：①当前任务相比于前面两个研究的优势在哪里？两个矛盾结果的实验都是采用动态

视觉搜索范式，当前研究同样也是采用动态视觉搜索范式，将区域划分为左右两侧，线段整齐排列的意义在哪里？需要明确说明出来？是相比于之前的研究，更好地建立了空间关联性，操控空间一致性吗？②以及为什么要用眼动指标，采用眼动指标的优势在哪里？也需要明确指出来。

回应：感谢审稿专家的宝贵意见。①当前任务相比于前面两个研究具有更好的空间关联性，从而更好地操控空间一致性，根据审稿专家的意见，已补充研究任务优势和改动意义。②采用眼动指标平均注视次数、平均眼跳幅度可以表明听觉刺激具体如何促进搜索效率，眼动指标平均注视时间可以探究 Pip-and-Pop 效应的产生机制。(详见文章 21-22 页、23 页蓝色字体)

① “一方面，前人有关视听刺激空间一致性对 Pip-and-Pop 效应影响的研究结果并不一致。Ngo 和 Spence(2010)研究发现与无效空间信息线索相比，具有空间信息的听觉刺激会提高视觉搜索的效率，即视听刺激空间一致性影响了 Pip-and-Pop 效应。因此，Ngo 和 Spence 认为 Pip-and-Pop 效应是视听刺激整合引起的，空间位置越一致，视听刺激越容易进行整合，Pip-and-Pop 效应越大(Ngo & Spence, 2010a)。但 Fleming 等人(2020)研究表明，行为与脑电结果均显示视听刺激空间位置一致与否对 Pip-and-Pop 效应的影响差异不显著，他们由此提出空间一致性并不是 Pip-and-Pop 效应产生的必要条件，视听刺激整合并没有起到重要作用(Fleming et al., 2020)。上述两个研究结果不一致的原因可能与刺激在空间排列上的差异有关。Ngo 和 Spence 将刺激随机排列在 10×10 的隐形网格中，按照搜索集大小在该隐形网格上随机位置呈现视觉刺激。Fleming 等虽也如此在随机位置上呈现视觉刺激，但他们缩小了搜索区域且增加了干扰线段的方向，同时还增加了人机距离，降低了视觉上的空间关联性。所以本研究为了更好地建立空间关联性，将视觉搜索区域分为左右两侧(与中心点有明显距离)，线段整齐地排列在网格中(以搜索集 36 为例，左右分别为 3×6 的网格)，以此更明确地操纵空间一致性从而探究视听空间一致性如何影响 Pip-and-Pop 效应。”

② “基于上述视听刺激空间、时间一致性对 Pip-and-Pop 效应的影响研究结论并不一致，视觉搜索任务中的行为反应时只能反映听觉刺激是否提高了搜索效率，而眼动数据可以根据眼跳和扫视等反映搜索中的实时信息处理过程。具体来说，通过平均注视次数反映知觉负荷大小，平均眼跳幅度反映在眼跳前被试获取的信息量，从而揭示听觉刺激具体如何提高搜索效率。此外，平均注视时间可以衡量是否有冰冻效应产生，从而说明 Pip-and-Pop 效应的产生是否有 oddball 刺激吸引注意的作用，进而可以探究 Pip-and-Pop 效应的产生机制。所以本研究将眼动技术运用到动态视觉搜索范式中，通过两个实验考察以下两个问题：(1)调控搜索集大小和听觉刺激呈现位置，探究视听刺激空间一致性如何影响 Pip-and-Pop 效应。(2)调

控搜索集大小和听觉刺激与视觉目标的间隔时间，探究视听刺激时间一致性如何影响 Pip-and-Pop 效应。”

意见 3：预期中，反向的预期也需要写出来，反之如果视听整合没在起作用，则表现出什么结果？

回应：感谢审稿专家的宝贵意见。如果不是视听整合起作用，作用的原因是 oddball 刺激吸引注意，那么无论空间和时间一致与否，突然出现的听觉刺激都会促进视觉搜索效率，条件之间没有显著差异。根据审稿专家的意见，对反向预期的内容加以补充。(详见文章 23 页蓝色字体)

“如果 Pip-and-Pop 效应产生的原因是多感觉整合，那么 Pip-and-Pop 效应遵循空间一致性和时间一致性原则，即视听刺激呈现的空间位置越一致，产生的 Pip-and-Pop 效应越大；视听刺激呈现的时间越一致，产生的 Pip-and-Pop 效应越大。如果 Pip-and-Pop 效应产生的原因是 oddball 刺激吸引注意，那么无论空间和时间一致与否，突然出现的听觉刺激都应影响视觉搜索效率。”

意见 4：结论部分只是实验结果的描述，需要添加一个得出什么推断，验证什么假设？

回应：感谢审稿专家的宝贵意见。根据审稿专家的意见，在结论部分重新梳理了实验结果的描述，加入了本研究能说明的机制问题以及未来对相关研究问题的展望。(详见文章 38 页蓝色字体)

“结论

视听刺激空间一致性、时间一致性对 Pip-and-Pop 效应具有调节作用。视听刺激空间位置越一致，Pip-and-Pop 效应越大；视听刺激呈现时间越一致，Pip-and-Pop 效应越大。当前研究结果为 Pip-and-Pop 效应产生的原因是多感觉整合提供了证据。此外，在未来的研究中应基于动态视觉搜索范式，设置明确的视听整合条件与 oddball 刺激吸引注意条件进行对比，进一步实验探究 Pip-and-Pop 效应的产生机制。”

.....

审稿人 2 意见：作者针对一审意见进行了修改，对学术性概念进行精细的定义，结果分析中的逻辑性增强。总体来说，文章有一定改进。但是，通读全文发现文章还有几处地方可以进一步改善。

意见 1: 前言部分在阐述视听刺激空间一致性对 Pip-and-Pop 效应影响的研究结果不一致时语言有些混乱, 逻辑不自洽。①例如正文第 3 页中“此外, 上述两项研究结果不一致的原因还可能是与二者视觉刺激在空间排列上的差异有关”, 这个“还可能是”是承接的哪个可能性呢? ②是怎么通过上文的描述了解到是研究结论不一致的原因与二者视觉刺激在空间排列上的差异有关呢? ③视听刺激空间一致性对 Pip-and-Pop 效应影响也是如此。建议作者对这个问题进行重新梳理。

回应: 感谢审稿专家的宝贵意见。①对于“还可能是”的承接确实存在逻辑上的问题, 根据审稿专家的意见, 对这两部分内容的逻辑进行重新梳理。②视听空间一致性中, 两个研究都是基于动态视觉搜索范式的研究, 不一致的地方是两个研究在刺激呈现上的差异(文中已经加入详细差异点), 所以二者结论不一致的原因可能是与刺激在空间上呈现的差异有关。③视听时间一致性中, 以往研究使用的搜索集不同, 使用较大搜索集的研究发现了时间一致性对 Pip-and-Pop 效应的影响, 使用较小搜索集的研究没有发现时间一致性对 Pip-and-Pop 效应的影响, 所以以往研究结论不一致的原因可能是搜索集大小的作用。(详见文章 21-22 页蓝色字体)

① “一方面, 前人有关视听刺激空间一致性对 Pip-and-Pop 效应影响的研究结果并不一致。Ngo 和 Spence(2010)研究发现与无效空间信息线索相比, 具有空间信息的听觉刺激会提高视觉搜索的效率, 即视听刺激空间一致性影响了 Pip-and-Pop 效应。因此, Ngo 和 Spence 认为 Pip-and-Pop 效应是视听刺激整合引起的, 空间位置越一致, 视听刺激越容易进行整合, Pip-and-Pop 效应越大(Ngo & Spence, 2010a)。但 Fleming 等人(2020)研究表明, 行为与脑电结果均显示视听刺激空间位置一致与否对 Pip-and-Pop 效应的影响差异不显著, 他们由此提出空间一致性并不是 Pip-and-Pop 效应产生的必要条件, 视听刺激整合并没有起到重要作用(Fleming et al., 2020)。上述两个研究结果不一致的原因可能与刺激在空间排列上的差异有关。Ngo 和 Spence 将刺激随机排列在 10×10 的隐形网格中, 按照搜索集大小在该隐形网格上随机位置呈现视觉刺激。Fleming 等虽也如此在随机位置上呈现视觉刺激, 但他们缩小了搜索区域且增加了干扰线段的方向, 同时还增加了人机距离, 降低了视觉上的空间关联性。所以本研究为了更好地建立空间关联性, 将视觉搜索区域分为左右两侧(与中心点有明显距离), 线段整齐地排列在网格中(以搜索集 36 为例, 左右分别为 3×6 的网格), 以此更明确地操纵空间一致性从而探究视听空间一致性如何影响 Pip-and-Pop 效应。”

② “另一方面, 以往有关视听刺激时间一致性对 Pip-and-Pop 效应影响的研究结果也是不一致的。Van der Burg 等人(2008)使用动态视觉搜索范式, 在研究中设置了 8 种声音目标

间隔(Tone-Target-Interval, TTI)条件, 结果发现 TTI 为-100ms、-50ms、-25ms、0ms、25ms、100ms 时(0ms 为同时呈现, 负数为提前呈现, 正数为滞后呈现), 听觉刺激才会促进对视觉目标的搜索, 视听刺激同时呈现时(0ms 条件)的促进作用最大, 且声音目标间隔与反应时之间呈 U 型函数关系, 即视听刺激呈现的时间一致性越高, 搜索越快。Van der Burg 等人认为, 同时呈现的听觉刺激以及在整合时间窗内(-100ms~100ms)呈现的听觉刺激引发的搜索效率提高是由于跨通道整合增强了视觉目标的凸显性(Van der Burg et al., 2008)。但也有研究发现视听刺激时间一致性并不会影响 Pip-and-Pop 效应, Zou 等人(2012)比较了 0ms、-100ms 以及听觉刺激随机时间呈现三种时间条件, 发现这三种条件都会促进视觉目标的搜索, 但它们之间没有显著差异, 因此 Zou 等人(2012)认为 Pip-and-Pop 效应是无空间信息的听觉刺激引发的“冰冻效应”所致, 即 oddball 刺激吸引注意的作用, 无论听觉刺激何时呈现都会促进视觉搜索, 而不是视听刺激整合作用的结果(Zou et al., 2012)。

上述研究使用的搜索集大小不同, 在搜索集为 48 和 60 的研究中发现了视听时间一致性对 Pip-and-Pop 效应的影响(Van der Burg et al., 2008; Kösem & van Wassenhove, 2012), 在搜索集 36 的研究中没有发现视听时间一致性对 Pip-and-Pop 效应的影响(Zou et al., 2012)。所以搜索集大小可能在视听时间一致性影响 Pip-and-Pop 效应的过程中起着调节作用。此外, 在视听空间一致性影响 Pip-and-Pop 效应的研究中, 搜索集大小在其中的调节作用也是不一致的(Ngo & Spence, 2010a; Fleming et al., 2020)。可能的原因是在搜索集较小的条件下知觉负荷较低, 只要听觉刺激存在, 无论其是否与视觉目标空间或时间上一致都可以提高视觉搜索效率。为了进一步确定这个问题, 本研究引入搜索集大小这一变量, 通过行为与眼动两方面数据探究在不同知觉负荷下视听刺激时间一致性和空间一致性对 Pip-and-Pop 效应的影响。”

意见 2: 在先前审稿意见中提到的三个变量中变化间隔的设置意义不明确。作者在修改中已经在实验仪器和材料中进行了详细阐述变化间隔的设置, 并且在意见回复中表明了三个变量变化间隔的设置的目的。①建议作者最好在文章中也有体现出来这个设置的目的和意义, 方便读者更好的理解。②以及需要在文中阐释清楚在实验 2 中为何把设置三个变化间隔变为设置一个变化间隔。

回应: 感谢审稿专家的宝贵意见。根据审稿专家的意见, 已补充变化间隔的设置目的与意义以及实验 2 中为何三个变化间隔变为一个两部分内容。(详见文章 24 页、31 页蓝色字体)

① “变化间隔为 50、100 或者 150ms, 三种变化间隔随机出现, 共有九个变化间隔(一个变化周期), 以此避免视觉刺激在进行连续变化时有很高的规律性。目标线段颜色变化遵

循以下 3 种规则：(1)目标线段颜色单独变化，并且每个周期只能改变一次，目标线段颜色变化的平均频率为 1.11Hz(即每 900ms 变化一次)；(2)目标颜色变化之前是 150ms 的变化间隔，目标颜色变化之后是 100ms 的变化间隔。由于视听整合的时间窗大约为 $\pm 100\text{ms}$ ，在目标变化前后的时间间隔如此设置以确保明确的视听整合；(3)每个周期的前三个间隔目标线段颜色不会变化。”

② “第二，在听觉刺激与视觉目标变化不一致条件下，为了使听觉刺激与视觉干扰刺激的变化同步呈现，将本实验中的线段变化间隔由实验 1 中的 50ms、100ms、150ms 三种改为 100ms 一种变化间隔。”

意见 3：结论部分写的比较像教科书式的写法，建议参考一下论文结论的写法。不仅要写出得出的结论，更要体现出结论的意义，甚至提出未来的研究展望。现在的结论内容深度略显不足。

回应：感谢审稿专家的宝贵意见。根据审稿专家的意见，在结论部分重新梳理了实验结果的描述，加入了本研究能说明的机制问题以及未来对相关研究问题的展望。(详见文章 38 页蓝色字体)

“结论

视听刺激空间一致性、时间一致性对 Pip-and-Pop 效应具有调节作用。视听刺激空间位置越一致，Pip-and-Pop 效应越大；视听刺激呈现时间越一致，Pip-and-Pop 效应越大。当前研究结果为 Pip-and-Pop 效应产生的原因是多感觉整合提供了证据。此外，在未来的研究中应基于动态视觉搜索范式，设置明确的视听整合条件与 oddball 刺激吸引注意条件进行对比，进一步实验探究 Pip-and-Pop 效应的产生机制。”

第三轮

审稿人 2 意见：

作者针对二审意见进行了修改，总体来说，文章质量提升很大，但是前言部分存在一点小问题。

意见 1：Pip-and-Pop 效应产生的原因是多感觉整合或是 oddball 刺激(奇异性刺激)吸引注意。两个原因的提出，以及推出“空间原则(the spatial rule)和时间原则”，到最后提出“Pip-and-Pop 效应是否同样受视听刺激呈现的空间一致性和时间一致性的影响？”这个讨论点。这其中的

逻辑并不顺畅，希望作者可以加以梳理调整。

回应：感谢审稿专家的宝贵意见。上述讨论点确实存在逻辑不顺畅的问题，我们对这部分内容重新进行了梳理：Pip-and-Pop 效应产生的原因存在争议，以往研究从多感觉整合的空间和时间两个原则出发，试图通过考察 Pip-and-Pop 效应是否也遵循空间原则和时间原则，来探究 Pip-and-Pop 效应产生原因是否是多感觉整合。(详见文章 22 页紫色字体)

“Pip-and-Pop 效应产生的原因是多感觉整合还是 oddball 刺激吸引注意仍然存在争议。以往研究表明多感觉整合遵循两个原则——空间原则(the spatial rule)和时间原则(the temporal rule)，即当来自不同感觉通道的信息在大致相近的空间位置呈现时，多感觉整合效应最大 (Spence, 2013; Van der Stoep et al., 2017)；当来自不同感觉通道的信息在大致接近的时间呈现时，多感觉整合效应最大(Fister et al., 2016; Stevenson et al., 2012)。所以有研究通过调控空间一致性与时间一致性来探究 Pip-and-Pop 效应产生的原因是否是多感觉整合(Van der Burg et al., 2008; Ngo & Spence, 2010a; Zou et al., 2012)。如果 Pip-and-Pop 效应遵循空间和时间原则，那么 Pip-and-Pop 效应产生的原因存在多感觉整合的作用；如果 Pip-and-Pop 效应不遵循空间和时间原则，那么 Pip-and-Pop 效应产生的原因可能是 oddball 刺激吸引注意，即突然出现的听觉刺激影响了视觉搜索效率。”