

《心理学报》审稿意见与作者回应

题目：自然情境下舌尖效应的认知老化——日记研究

作者：赵瑞琪 娄昊 欧阳明昆 张清芳

第一轮

审稿人 1 意见：

意见 1：该研究采用日记法对自然情境下舌尖效应发生的机制进行探究，研究采用日记研究法，对自然情境下的舌尖效应老化现象进行四周的追踪研究，比较了汉语情境下青年人和老年人舌尖效应日常发生的情况，并考察了舌尖效应认知和元认知两方面的特点。研究发现，自然情境舌尖效应是言语产生认知过程老化的结果，同时也是元认知过程的调节双重作用的结果。研究具有一定的理论和实践意义，具有发表的价值。有如下问题需要修改：

本文的引言和逻辑是“舌尖效应老化的意义——舌尖效应老化的理论解释——舌尖效应实验室研究方法的缺陷——自然情境研究方法的优点——汉语背景下舌尖效应——研究方案和假设”，引言逻辑清晰，但是，在问题的提出部分，在文字表述方面，如何更好地体现“自然情境”、“追踪研究”、“汉语情境”以及“认知和元认知特点”，在这几个方面，哪些是主要的创新点？创新的理论意义是什么？建议作者考虑。

回应：非常感谢审稿专家的肯定和意见，使得我们重新思考了本文的研究特色和创新点，重新构建了前言的结构。第一，我们介绍了 TOT 认知老化的研究视角和影响因素，有关言语产生过程的 TOT 研究分别采用了认知视角和元认知视角，在认知视角下，研究者提出了传输缺陷假说和阻塞假说，研究中强调了语言学属性和个体特点对 TOT 认知加工过程的影响。元认知视角下的研究则关注的是 TOT 发生时个体的体验和状态，强调体验对言语产生过程的积极作用，包括了个体的情绪状态以及对线索的元认知体验等。目前，尚未有研究综合考察认知视角下的影响因素与元认知视角下的影响因素是如何共同影响了 TOT 的解决时间。第二，比较了自然情境和实验室研究中的不同结果，且自然情境下人们可以利用的线索更多，更有利于 TOT 的解决。通过这一部分内容的阐述，表明了自然情境研究的必要性。第三，比较了汉语和字母语言在口语产生过程中语义和音韵加工的模式，强调了汉语口语产生过程的独特特点。

综上，本研究采用生态效度较高的日记研究法，对自然情境下的 TOT 老化现象进行为期四周的追踪研究，考察比较汉语情境下青年人和老年人 TOT 日常发生的情况；所记录的情况包括 TOT 认知和元认知两个方面的特点，通过方差分析和回归分析考察 TOT 的产生机制和影响因素；采用多水平数据分析方法，探究 TOT 记录对后续发生情况的影响，并考察元认知状态对认知过程的影响。

基于已有研究现状和本研究内容，我们认为本研究的创新点和理论意义在于：第一，从认知视角和元认知视角两个方面同时考察 TOT 的认知老化及其影响因素，并且在研究中采用了多种数据分析方法。第二，日记研究得到的是多水平数据，被试每次的记录嵌套在被试个体中，通过多水平数据分析，我们考察了在 TOT 的记录过程中，前面所发生的 TOT 的记录、评价和解决时间对于后续所发生 TOT 的状态和解决时间的影响，这是实验室情境中无法考察的问题。根据本研究的发现，我们认为 TOT 的认知状态和元认知状态下的因素都对 TOT 的解决时间产生了影响。

具体的阐述见前言（p.11-15）。

意见 2: 引言第三行“TOT 是发生在言语产生领域的一种特有现象, 小到儿童, 大到成人都会产生(Hanly & Vandenberg, 2010), 特别是老年人…”表述欠妥当, 建议修改为“TOT 是发生在言语产生领域的一种特有现象, 儿童和成人都会产生这种现象(Hanly & Vandenberg, 2010), 特别是老年人…”

同样, 引言第四行:“这在某种程度暗示了 TOT 存在认知老化现象, 然而对于该现象发生机制的研究, 尤其是自然情境下 TOT 产生原因及认知意义的研究目前尚且缺乏。”建议修改为“这在某种程度暗示了 TOT 存在认知老化现象, 然而对于该现象发生机制的研究, 尤其是自然情境下 TOT 产生原因及认知意义的研究目前尚且缺乏。”

回应: 谢谢审稿专家的意见, 已对相应部分进行修改。

意见 3: 在研究方法部分, 建议说明被试的招募形式, 另外, 被试学历的计分方式是否正确? 请作者考虑?

回应: 我们通过在高校中张贴招募启事、相关学生网站上发帖、微信公众号发文等形式招募青年被试; 通过在高校中张贴招募启事和走访高校社区、公园招募老年被试。

根据审稿专家建议, 修改稿中我们参照已有研究, 把被试学历修改为被试受教育年限(何洁莹, 张清芳, 2017; 彭华茂, 毛晓飞, 2018)。通过对青年人和老年人的受教育年限进行统计分析, 我们发现两组被试的受教育年限差异不显著($t(60) = 1.19, p = 0.24$)。也就是说, 本研究对青年人和老年人的受教育年限进行了严格匹配。

相关信息已经增加至文章的方法部分(见 p.15-16)。

参考文献:

何洁莹, 张清芳. (2017). 老年人书写产生中词汇频率和音节频率效应的时间进程: ERP 研究. *心理学报*, 49(12), 1483-1493.

彭华茂, 毛晓飞. (2018). 抑制对老年人舌尖现象的影响. *心理学报*, 50(10), 1142-1150.

意见 4: 在结果部分, 作者使用了未矫正过的自由度“结果显示, 词类主效应显著($F(3.02, 175.23) = 44.74, p < .001, \eta_p^2 = .44$); 词类和年龄的交互作用显著($F(3.02, 175.23) = 2.72, p = .046, \eta_p^2 = .05$)。”建议作者作出解释和说明。

回应: 本文重复测量方差分析球形假设不成立时, 使用 Greenhouse-Geiss 法校正。我们已经修改了文中的自由度。

意见 5: 在讨论部分, 应结合研究结果对“自然情境”研究方法, 进行深入的讨论。

回应: 感谢审稿专家的建议。在前言的第二部分, 我们对比了自然情境与实验室研究的结果, 已有研究表明自然情境下 TOT 的解决概率要高于实验室研究, 而且自然情境的研究方法具有较高的生态效度, 能够收集被试日常生活情境中的数据, 如工作场所、在家中、交流过程中的数据, 可以“将生活再现。”(Bolger et al., 2003)。日记研究是一种更为开放的形式, 被试可以用自己的语句记录事件, 想法, 感受和行为(Poppleton, Briner, & Kiefer, 2008)。被试可以在 TOT 发生后及时记录与之相关的问题, 降低了回溯偏差(Reis & Gable, 2000)。日记研究所收集的数据为多水平嵌套数据, 被试每次的记录嵌套在被试个体中, 既能考察每个被试记录间的影响(within-person), 也能考察被试之间的差异和影响(between-person), 且存在时间滞后效应(Bolger, Davis, & Rafaeli, 2003), 即将时间因素考虑在内, 考察被试前一次的认知和行为状态对后续记录的影响。通过日记研究方法, 我们可以考察变量随时间的变化, 短时状态体验和行为间的关系以及人或情境特征与短时状态体验和行为间的关系。

针对自然情境的研究特点, 我们分析了 TOT 状态发生时 TOT 的特点, 个体的身心状态, 具体采取的解决策略, 影响 TOT 解决时间的因素, 以及 TOT 日记研究过程中前期的记录和

评价对后续 TOT 的影响，我们在相应部分对此做了深入讨论，重点关注了自然情境与实验情境下所得到的不同结果，强调了日记研究的特色（见 p.25-29）

.....

审稿人 2 意见：

意见 1：舌尖效应是老年人主述的言语产生问题之一，该研究采用日记法研究老年人的日常生活中的舌尖效应特点，很有生态特点，在方法上与以往大多舌尖效应研究的不同，值得鼓励！研究虽然报告了丰富的老年人和青年人的舌尖效应的现象特征，但在论文写作上还存在很多问题甚至疏漏，请作者作出重大修改，提升论文质量，更好展示自己的研究构想和结果。具体问题如下：

文献综述部分主要阐述了有关 TOT 产生的几种理论以及实验室研究的缺陷，而对本研究要探讨的问题涉及不足。本研究主要是描述了自然情境下老年人舌尖效应的特征以及 TOT 产生的影响因素，但在文献综述部分，对于后面数据分析涉及到的各个指标或者方面都没有阐述，这样造成了文献综述和具体研究设计和结果的脱节。同时，也让本研究的研究问题十分不明确。比如为什么要研究自然情境下的老年人舌尖效应的现象？以前的实验室研究和当前研究结果比，有什么特征没有体现出来的？作者是认为元认知保持良好，是老年人 TOT 增多的机制吗？请作者重新构思文献综述和问题提出部分，要讲述一个逻辑完整的故事，而不是一些文献的罗列而已。

回应：非常感谢审稿专家的肯定和意见，使得我们思考了本文的研究特色和创新点，重新构建了前言的结构。第一，我们介绍了 TOT 认知年老化的研究视角和影响因素，有关言语产生过程的 TOT 研究分别采用了认知视角和元认知视角，在认知视角下，研究者提出了传输缺陷假说和阻塞假说，研究中强调了语言学属性和个体特点对 TOT 认知加工过程的影响。元认知视角下的研究则关注的是 TOT 发生时个体的体验和状态，强调体验对言语产生过程的积极作用，包括了个体的情绪状态以及对线索的元认知体验等。目前，尚未有研究综合考察认知视角下的影响因素与元认知视角下的影响因素是如何共同影响了 TOT 的解决时间。第二，比较了自然情境和实验室研究中所得到的不同结果，且自然情境下人们可以利用的线索更多，更有利于 TOT 的解决。通过这一部分内容的阐述，表明了自然情境研究的必要性。第三，比较了汉语和字母语言在口语产生过程中语义和音韵加工的模式，强调了汉语口语产生过程的独特特点。

综上，本研究采用生态效度较高的日记研究法，对自然情境下的 TOT 老化现象进行为期四周的追踪研究，考察比较汉语情境下青年人和老年人 TOT 日常发生的情况；所记录的情况包括 TOT 认知和元认知两个方面的特点，通过方差分析和回归分析考察 TOT 的产生机制和影响因素；采用多水平数据分析方法，探究 TOT 记录对后续发生情况的影响，并考察元认知状态对认知过程的影响。

具体的阐述见前言（p.11-15）。

意见 2：文献综述部分，有些表述不是特别清楚或者准确。如“汉语 TOT 发生可能主要以语义信息提取为主”，一句，在其所在段落中，似乎不能由上文阐述推导而出。又如“我们推测 TOT 老化机制可能既反映言语产生时具体信息提取的老化，也反映老年人对自己的只是量良好监测能力，以及对言语产生过程良好的控制能力，可能这些能力并不会随年龄而衰退。”这句话，所谓 TOT 老化，是不是就是指老年人 TOT 现象比青年人多？如果是，那么按照这句话的字面意思，其老化机制应该是指老年人出现 TOT 增多的原因或者内在心理过程，也就是可能是老年人具体信息提取能衰弱造成 TOT 增多，也可能是老年人对言语产生过程有良好控制而产生的。这后半句是否非常矛盾？为什么有良好监控反而会增多 TOT？

回应：谢谢审稿专家的问题。我们已经对前言做了根本性的修改，您所提出的问题在前言的第一部分“TOT 认知老龄化的研究视角和影响因素”中进行了阐述。

TOT 是口语产生过程中所出现的一种单词提取失败的现象，儿童和成人都会产生这种现象(Hanly & Vandenberg, 2010)，特别是老年人，TOT 几乎每天都会发生(Brown, 2012)，这表明口语产生过程存在认知老龄化现象。研究发现 TOT 发生时人们一般可以提取目标词的语义信息(Fieder, Nickels, & Biedermann, 2014)，而不能或只能部分提取音韵信息(Cleary, Konkel, Nomi, & McCabe, 2010)。第二，TOT 发生时通常伴随有意识的元认知体验，包括紧张、焦虑、兴奋等情绪，及目标词即将出现的紧迫感，并且这些元认知体验可以有效地预测随后目标词的成功提取(Schwartz, 2002)。可以看出，TOT 发生时人们的言语产生过程出现困难，同时伴随有强烈的元认知体验。研究者对此现象从认知和元认知两种视角进行解释。

从认知视角而言，TOT 反映了言语产生中信息提取的失败，而老化会进一步加剧这种失败，通常以传输缺陷假说和阻塞假说进行解释（见 p.12），研究中强调语言学属性和个体的一般认知能力（如抑制无关信息的能力）对 TOT 过程的影响。元认知视角下的研究则关注的是 TOT 发生时个体的体验和状态，强调体验对言语产生过程的积极作用。根据 TOT 的元认知理论，言语产生时元认知系统会对个体所感知的各种线索进行目标词的可提取性(retrievability)的启发式推理判断(Schwartz & Metcalfe, 2011)，即对记忆系统中目标词的可提取性进行监测（见 p.12-13）。

综上，TOT 的认知老龄化机制是一个非常复杂的问题，同时受到了认知视角的因素和元认知视角因素的影响，我们在文章的前言和讨论部分都对此进行了讨论。

意见 3：被试样本量是否合适，请进行 G Power 分析。

回应：谢谢您的建议。我们采用 G Power 3 计算了最小样本含量，针对文中两因素被试间方差分析，取 $\alpha = 0.05$, $\beta = 0.2$ ，效应量取“中”(0.25)计算，在 F 检验中所需样本量为 45。针对文中三因素重复测量方差分析，取 $\alpha = 0.05$, $\beta = 0.2$ ，效应量取“中”(0.25)计算，在 F 检验中所需样本量为 34。本文数据分析中的样本量为 62，符合样本量的要求。对此，我们增加了脚注进行说明（见 p.17）。

意见 4：表 1 中，被试学历为分类数据，报告平均数是没有意义的。另外，由于涉及到言语产生，被试受教育水平和词汇量或者言语能力应进行测量并进行差异检验，必要的话，还应在后续分析中作为协变量控制。

回应：感谢审稿专家的意见。修改稿中我们参照已有研究，把被试学历修改为被试受教育水平，其操作性定义是受教育年限（何洁莹，张清芳，2017；彭华茂，毛晓飞，2018）。通过对青年人和老年人的受教育年限进行统计分析，我们发现两组被试的受教育年限差异不显著($t(60) = 1.19$, $p = 0.24$)，表明本研究中青年人和老年人的受教育水平是匹配的。

关于认知老龄化的研究，在匹配青年组和老年组时，一般都是利用量表评估老年人的认知能力，只要得分大于特定的分数，则认为老年人的认知能力没有损害。研究中同时匹配老年人和青年人的受教育年限（何洁莹，张清芳，2017；杨群，张清芳，2015；Blackett et al., 2017；Macdonald, 2013；Segaert et al., 2018）。本研究采用 MoCA 量表对老年人一般认知能力是否属于正常范围进行甄别，包括非言语能力（如视空间与执行功能、短时记忆等）和言语能力（如图片命名、言语流畅性和词汇理解）两部分。MoCA 量表属于标准参照测验，量表总分为 30 分，个体的量表得分大于或等于 26 分被认为属于认知正常(Castro & James, 2014；Nasreddine et al., 2005)。本研究中结果分析部分的数据均来自一般认知能力正常的老年被试（MoCA 平均得分 27.59，标准差 1.27），且匹配了教育年限，同时考虑到使用这种方法时认知能力正常的老年人与青年人的得分比较接近，因此相关研究在统计分析时不考虑 MoCA 得分的个体差异对结果的影响，我们在数据分析中也采用了该领域内得到研究者认可的方法。

参考文献：

杨群，张清芳. (2015). 口语产生中词频效应、音节频率效应和语音促进效应的认知老龄化. *心理科学*, 38(6), 1303-1310.

何洁莹, 张清芳. (2017). 老年人书写产生中词汇频率和音节频率效应的时间进程: ERP 研究. *心理学报*, 49(12), 1483-1493.

彭华茂, 毛晓飞. (2018). 抑制对老年人舌尖现象的影响. *心理学报*, 50(10), 1142-1150.

Blackett, D. S., Harnish, S. M., Lundine, J. P., Zezinka, A., & Healy, E. W. (2017). The Effect of Stimulus Valence on Lexical Retrieval in Younger and Older Adults. *Journal of Speech, Language, and Hearing Research*, 60(7), 2081-2089.

Castro, N., & James, L. E. (2014). Differences between young and older adults' spoken language production in descriptions of negative versus neutral pictures. *Aging Neuropsychology & Cognition*, 21(2), 222.

Macdonald, G. (2013). *Aging and semantic processing*. Windsor, Ontario, Canada: Unpublished doctoral thesis, University of Windsor.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699.

Segaert, K., Lucas, S. J., Burley, C. V., Segaert, P., Milner, A. E., Ryan, M., & Wheeldon, L. (2018). Higher physical fitness levels are associated with less language decline in healthy ageing. *Scientific reports*, 8.

意见 5: 研究工具介绍太过简单, 不应所有信息都见附录。应报告各指标的含义、计分方法等。由于文献综述未涉及研究工具中的各项指标, 所以也令人无法明白, 为什么要从这些方面去记录和测量。另外, 从附录题目看, 元认知的条目没有任何解释和描述, 被试是否能够理解其含义?

回应: 感谢您的建议。已在修改稿“研究工具”部分给出具体的认知和元认知指标(见 p.16), 在结果部分增加了编码和计分方式(见 p.17-23)。

对于元认知的条目, 在附录中的题目上没有给出详细描述和解释, 但是, 在正式调查前, 主试会向被试介绍 TOT 记录本记录的基本要求, 并对被试进行填写训练。给被试详细解释最后的元认知条目是指发生了 TOT 后, 当时所体验到的不同的身体心理状态, 要求其按照实际情况填写。而且, 在被试填写的一个月中, 主试每周对被试的记录情况进行提醒, 通过电话访谈的方式, 询问被试的填写情况, 及时解答被试的疑问, 确保了被试是理解其具体含义的(见 p.16-17“研究流程”)。

意见 6: 所有的指标都应报告描述统计结果, 且建议合在 1-2 个表呈现, 而不是现在这样, 哪个显著了, 就报告一下描述统计, 写到哪里了, 就随手呈现一个表格。图也如此。

回应: 感谢您的建议。修改稿中将所有结果整理成了 3 个表格(表 2-4), 分别为表 2: 不同年龄和不同性别下 TOT 发生时目标词的基本特征, 表 3: 不同年龄和不同性别下 TOT 发生时提取的各类信息的特点, 表 4: 青年组和老年组中不同 TOT 解决策略所占比例(%)平均值和标准差。文中共包括了四个图, 分别呈现了年龄与词类之间的交互作用(图 1 不同词类的目标词产生 TOT 的年龄差异), 年龄与身心状态之间的交互作用(图 2 青年组和老年组发生 TOT 时的身心状态), 年龄与解决 TOT 策略之间的交互作用(图 3 青年组和老年组解决 TOT 所用策略比较)。

意见 7: 讨论 4.2 中, 有关“青年人体验的疲惫感和兴奋性情绪均比老年人更强。”的结果解释比较牵强, 引用 SST 的观点显得更加牵强。

回应: 感谢审稿专家的建议。修改稿中我们引用已有研究的发现, 从青年人和老年人采取的不同解决策略, 以及不同的元认知过程的角度对此结果进行了解释。具体的修改见 p.27-28。

意见 8: 作者在研究局限中指出, 日记法研究得到的数据是多水平数据, 那么为什么本研究不采用多水平数据分析方法剥离个体层面的变异带来的影响? 这是可以做到的事情, 完全不必成为局限。

回应: 感谢审稿专家的建议。本研究中被试在连续 28 天记录所发生的 TOT 情况, 数据为多水平嵌套数据, 被试每次的记录嵌套在被试个体中。考虑到日记研究中的时间滞后效应

(time-lagged effect), 即将时间因素考虑在内, 考察被试前一次的认知和行为状态对后续认知和行为的影响(Bolger, Davis, & Rafaeli, 2003; Ohly, 2010; Eldahanet al., 2016), 本研究中为被试对前一次的认知和行为状态的评价和解决可能会对后续 TOT 的发生和解决产生影响。在多水平数据分析中, 我们以 TOT 发生后的解决时间为结果变量, 分别以熟悉度(认知因素)、被试的兴奋度, 疲惫感, 舒适度, 紧张度(元认知状态)为预测变量, 考察被试的认知状态、元认知状态对 TOT 发生后解决时间的影响, 并进一步考察每次发生 TOT 后被试的认知和元认知状态对后续 TOT 发生的认知和元认知状态的影响, 一共形成五个模型(1 熟悉度, 2 兴奋度, 3 疲惫感, 4 舒适度, 5 紧张度)。

具体的模型构建及结果见“3.2.5 认知视角因素和元认知视角因素对后续 TOT 的影响: 多水平数据分析”(p.23-25)。

第二轮

审稿人 1 意见:

意见 1: 作者很好的回答了审稿人的问题, 建议作出小的修改后发表。问题提出部分: “采用多水平数据分析方法, 探究 TOT 记录对后续发生情况的影响, 并考察元认知状态对认知过程的影响。我们推测 TOT 老化机制可能既反映了老年人言语产生时在认知层面具体信息(如音韵信息)提取能力的衰弱, 也反映了老年人的元认知监测和控制能力依然保存良好, 并不会随年龄而出现衰退。”为什么会有这样的假设和推测? 希望论文在此处进行阐述。

回应: 非常感谢审稿专家的肯定和意见。我们的推测是基于已有理论观点和多种数据分析方法提出的。根据 TOT 认知视角中的两个理论, 传输缺陷假说(Burke et al., 1991)认为, 老化会减弱目标词语义和音韵节点之间的联结, 语义节点的激活不能有效地传输至音韵节点, 从而导致目标词音韵信息提取的失败产生 TOT。阻塞假说(Jones, 1989)则认为, 由于老年人对非目标词的抑制能力衰退, 更多非目标词的激活阻塞目标词提取, 产生了更多 TOT(Hasher & Zacks, 1988; Stoltzfus, Hasher, & Zacks, 1996)。而根据 TOT 的元认知理论, TOT 是一种元认知监测判断, 会影响元认知的控制系统, 控制系统会根据监测结果对言语产生的具体认知过程进行调节, 从而实现 TOT 发生后目标词的成功提取。TOT 发生后, 对于目标词的提取, 老年人甚至比青年人提取率略高(Schwartz, 2002), 这反映出老年人对言语产生监测和控制的调节能力并不会因年龄增长而受到影响, 对言语产生过程仍具有较高的监测和控制能力。这一部分我们在“1.1 TOT 认知老化的研究视角和影响因素”中做了详细阐述。

结合两类视角中的理论观点以及研究发现, 我们推测在汉语 TOT 的研究中, 老年人 TOT 现象的增多既反映了老年人口语产生过程中音韵信息激活的减弱, 也与元认知视角下老年人对 TOT 的监测和控制有关。在研究中我们通过多种分析方法, 包括方差分析、回归分析以及多水平数据分析来验证上述研究假设。我们已经对问题提出部分的阐述做了修改(见 P.19)。

意见 2: 引研究在工具部分, 建议介绍“回溯问卷和 TOT 记录本”的信、效度。

回应: 本文中对 TOT 的测量采用了回溯问卷和 TOT 记录本, 其内容是根据 TOT 的认知视角和元认知视角构建的。已有研究者采用此问卷和记录本对英语和西班牙语中的 TOT 发生状态进行记录和评估, 研究者认为其内容效度很好地反映了 TOT 的认知特点和元认知特点(Burke, Mackay, Worthley, & Wade, 1991), 实验室研究也证实了采用回溯问卷和 TOT 记录本所测量结果的可靠性(Burke et al., 1991; Gollan, Montoya, & Bonanni, 2005; D' Angelo & Humphreys, 2015; Warriner & Humphreys, 2008)。我们的发现与英语研究中的结果一致, 证实了关于 TOT 状态的记录存在跨语言的一致性, 同时表明此问卷和记录本的设计有很好的内容效度。

已有的研究均未报告此问卷和记录本的信、效度。问卷的一部分是有关每次发生 TOT

的内容，每次发生 TOT 的内容和状态是不同的，不同的内容之间不能进行比较（详细内容见方法部分的介绍）。问卷的其它部分是有关发生 TOT 状态时个体的主观体验，我们利用本研究中所得到的数据对此做了重测信度分析。重测信度是指用同一个量表对同一组被试重复施测所得结果的一致性程度，其大小等于同一组被试在两次测验所得分数的皮尔逊积差相关系数，考察跨时间的一致性。TOT 记录本中最后被试身心状态的评分是连续变量，被试每次发生 TOT 后对此进行记录，我们计算了每次身心状态评估之间的重测信度，其信度系数为 0.62 ($p < 0.01$)，具有较好的跨时间一致性（见 P.24 结果部分脚注 2）。

参考文献：

Burke, D. M., Mackay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: what causes word finding failures in young and older adults? *Journal of Memory & Language*, 30(5), 542-579.

Gollan, T. H., Montoya, R. I., & Bonanni, M. P. (2005). Proper names get stuck on bilingual and monolingual speakers' tip of the tongue equally often. *Neuropsychology*, 19(3), 278.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, 98(1), 194-198.

D' Angelo, M. C., & Humphreys, K. R. (2015). Tip-of-the-tongue states reoccur because of implicit learning, but resolving them helps. *Cognition*, 142, 166-190.

Warriner, A. B., & Humphreys, K. R. (2008). Learning to fail: Reoccurring tip-of-the-tongue states. *The Quarterly Journal of Experimental Psychology*, 61(4), 535 - 542.

意见 3: 在讨论部分，建议结合本研究的研究结论，对本研究研究方法和工具的信、效度以及在该领域研究中的推广和应用做出阐述。

回应: 非常感谢审稿专家的建议。与已有研究结果的比较以及对身心状态的重测信度分析表明我们所使用的问卷和记录本具有较好的信度和效度。在后续 TOT 的实验室研究中可以使用这样的测量和评定考察 TOT 的认知和元认知状态，在严格的实验室控制情境下，考察这些因素如何影响了 TOT 的解决以及使用的策略。例如，在实验室情境中可能语义信息更容易诱发被试的汉语 TOT，被试发生 TOT 后提供相应的语义关联的替代词能够缩短被试的 TOT 解决时间等。自然情境下 TOT 的研究所得到的描述性结果为实验室研究提供了真实的 TOT 材料，研究者可以根据具体的研究目的选择合适的 TOT 语料进行考察。我们对此在讨论部分进行了阐述（见 P.33）。

.....

审稿人 2 意见:

意见 1: 3.2.1 中，报告了 TOT 解决的总次数，但在结果表中没有看到该数据结果，请补充。

回应: 谢谢审稿专家的建议。我们已在表 2 中补充了 30 名青年人和 32 名老年人 TOT 的“发生总数”。

意见 2: 结果报告了 TOT 的性别差异，但性别差异并没有在前言部分提出。鉴于本文篇幅较长，结果很多，建议作者删去对性别差异的分析。

回应: 非常感谢审稿专家的建议。我们对本文结果部分重新进行了数据分析，删除了性别这个被试特征变量，并修改了相应图表内容（详见结果呈现的各个部分）。

意见 3: 总体结果显示，老年人的 TOT 比年轻人多，但元认知体验不比年轻人高。这和前言中提出的观点是不一致的。如何解释这一结果？是否说明元认知观点下的 TOT 现象

其实还有待进一步证实？请在讨论中补充对这一结果的思考。

回应：非常感谢审稿专家的建议。本研究结果发现，老年人发生的 TOT 数量多于青年人，在元认知体验(身心状态)的评分上青年人体验到更多的疲惫感和兴奋感，而在其它体验上两组之间无差异，这表明 TOT 的发生数量与元认知体验无关。TOT 的发生并不是连续多次的，并不是元认知体验导致个体产生更多的 TOT，而是个体在产生 TOT 后伴随有特定的元认知体验。青年组和老年组在元认知体验上的不同和相同提示两组被试对元认知体验的不同方面的敏感性不同。在回收问卷访谈的过程中，我们发现青年人对身心状态的关注更多，老年人由于遗忘等原因，可能会产生一定的记录偏差。已有研究发现，老年人可以对目标词的提取状态进行监测和控制，且目标词成功提取的概率与青年人相当 (Burke et al., 1991)，我们的研究中虽然老年人身心状态评分比青年人低，但是老年人 TOT 成功提取率高达 94%，甚至略高于青年人(92%)，说明老年人对语言产生过程中的监测和控制可能与青年人相当，并可能通过元认知状态的评估帮助他们进行 TOT 状态的解决。

这提示我们在后续研究中需要关注 TOT 的认知因素和元认知因素之间的关系，及其可能的交互作用对 TOT 产生后解决时间及其解决策略的影响。本研究中通过回归分析和多水平分析做了尝试，但仍然需要控制严格的实验室研究对此问题进行深入研究。

我们在讨论部分对此进行了详细阐述（见 P. 31）。

第三轮

主编意见：

研究从认知和元认知的角度，通过回溯问卷和日记调查法，探究了自然情境下青年人和老年人汉语舌尖效应的发生特点及机制，生态效度高，研究具有高水平的理论水平和创新性。经过多轮修稿，作者在行文结构、逻辑框架及结果呈现、解释方面进行了进一步修改和完善，符合学报发表文章的要求，建议发表。