

《心理学报》审稿意见与作者回应

题目：汉语口语产生中音节和音段的促进和抑制效应

作者：岳源 张清芳

第一轮

审稿人 1 意见：

意见 1：文章 3 个实验的证明逻辑较弱。假设音节与音段均是音韵编码过程的单元（而非作者所推论的——仅仅音节是音韵编码单元），似乎也可以解释三个实验的结果。例如，实验 1 的结果可以解释为音韵编码阶段，首音段与韵均得到激活，因而产生促进效应。而在实验 2 和/或 3 中，首音段和/或韵受到了抑制——这种抑制发生在音韵编码阶段，导致部分条件下出现抑制效应。这意味着，基于目前 3 个实验的结果，似乎不能确切地得出文章所推论的结论，尽管实验结果本身是可靠的。

回应：感谢审稿专家给出的可能性解释。首先，需要澄清的一点是：实验 2 与 3 所发现的抑制效应不可能发生在音韵编码阶段。这是因为实验 2 采用的是延迟命名任务，在命名信号出现时，被试已经完成了音韵编码和语音编码两个阶段的加工，反应时中仅包括了发音阶段。实验 3 采用的时延迟命名和发音抑制两种任务的结合，在此任务中被试不能完成语音编码阶段的加工，因此反应时中包括了语音编码和发音阶段。实验 2 和 3 中所采用的任务均不包含音韵编码阶段的加工，因此两个实验中所发现的抑制效应不可能发生在此阶段。由于我们对实验逻辑描述的不够清晰使得审稿人产生了误解，在文中的前言部分，我们对此做了进一步阐述。

根据 O'Seaghdha 等（2010）提出的汉语口语产生中的“合适单元假设”，汉语的口语产生中最先选择的单元是音节，之后音节进一步分解为音素或音段信息（音韵编码阶段），准备发音运动程序（语音编码过程），最后进行发音输出口语产生的结果（发音阶段）。基于这一模型，我们认为：实验 1 中发现的音节和音段的促进效应发生在音韵编码阶段，而实验 2 和 3 中发现的抑制效应发生在语音编码阶段或发音阶段。同时，结合已有的研究结果，我们进行了效果量 Cohen *d* 值大小的比较和分析，音节的促进效应在已有的研究结果中均达到了较高程度，而音段促进效应的效果量较弱。综上，基于本研究 and 已有研究的结果，我们认为音节是音韵编码的单元之一，对于音段在汉语口语产生中的作用机制，需要采取更为敏感的任务或实验技术（如事件相关电位技术）进行研究。

我们接受审稿专家的建议，在总讨论中对 3 个实验的综合结果及其认知机制做了更进一步的阐述。在结论和摘要部分弱化了研究结论，指出音节是音韵编码的单元之一，音段的促进效应较小。根据本研究结果并不能排除其在音韵编码阶段所起的作用，对其的作用机制需要更进一步的探索。

意见 2：实验 1 使用的即时命名任务，比较自然，因而生态学效度较高。相比之下，实验 2 和 3 的任务非常特殊，日常语言情境中很少涉及。这意味着，实验 1 与实验 2/3 性质或情境可能截然不同。这种情况下，将 3 个实验的结果放到一起比较/讨论，是否合适？

回应：我们同意审稿人的看法，实验 2 和 3 的确是生态效度较低不自然的口语产生任务。尽管如此，在口语产生研究中，延迟命名以及延迟命名与发音抑制任务的结合是一种较为常见的实验任务，一般用来探索某个具体的效应是否发生在发音阶段或者语音编码阶段（Laganaro & Alario, 2006; Zhang & Damian, 2010）。例如，在本文中提到过的 Laganaro 和 Alario (2006)

发表在 *Journal of Memory and Language* 上的研究，就采用了延迟命名与发声抑制相结合的方法，用来探测音节频率效应究竟发生在口语产生过程中的哪个阶段，并得出了该效应位于语音编码阶段的结果，与研究假设一致，因此实验操纵及其结果是可靠的。本研究中，实验 1 的平均反应时最长，实验 3 次之，实验 2 最短，与实验 1 包括了口语产生的所有阶段，实验 3 的反应时包括了语音编码和发音阶段，实验 2 的反应时仅包括了发音阶段，反应时的长短也表明了实验操纵是有效的。

本研究中，实验 1 采取的图画命名任务，实验 2 为图画命名和延迟命名的结合任务，实验 3 为图画命名与延迟命名和发音抑制结合的任务。因为每种任务下所包含的认知加工过程不同（见前言），通过不同任务之间的比较能够考察音节和音段效应在图画命名过程中的发生阶段。所采取的任务与我们的研究目的相契合。因此我们认为实验 1 与 2、3 的比较是有意义的、恰当的。在前言以及各个实验部分我们也强调了各个实验之间比较的目的及意义。

意见 3: 目前的三个实验中，图画和干扰词同时呈现，即 $SOA=0$ 。与操纵上述 3 个不同的实验任务相比，系统变化/操纵 SOA ，似乎更有助于讨论音节和音段等不同表征激活的时间进程问题。

回应: 根据已有的实验结果，当 SOA 为 $0ms$ ，即图画和干扰字同时呈现时，无论干扰字与目标字音节重叠还是音段重叠，所得到的语音促进效应最强（Zhang & Weekes, 2009; Zhang, Chen, Weekes, & Yang, 2009; Zhao, La Heij, & Schiller, 2012; Wong & Chen, 2008; 2009）。本研究主要目的是为了考察汉语口语产生过程中音节效应和音段效应的认知机制，首先我们需要重复出该效应， SOA 为 $0ms$ 时效应最强，因此我们采取了这一 SOA 水平。第二，在 Wong 和 H.-C. Chen (2008) 的研究中，他们操纵了 SOA 这一变量，发现 SOA 和语音相关变量没有交互作用，因而无法作为操纵变量来探测语音相关变量引发的被试反应时差异的机制。这可能是由于从音节分解为音段的时间非常短，采用行为反应时间无法探查。在之后 Wong 和 Chen 所做的一系列相关实验中，他们所采用的 SOA 也都为 0 。因此在本研究中选取了图画和干扰字同时呈现这一实验条件，并进一步采用各种实验任务的结合和比较考察音节和音段效应的认知机制。

意见 4: 在“1 引言”一节中，并没有包含与实验证明逻辑有关的、任何有帮助的信息。更好的写法是，在“引言”一节的最后，阐述三个实验放在一起的、总体的证明逻辑，并预览三个实验的主要发现。目前的写法，文章的可读性较弱。

回应: 感谢审稿专家的建议。在修改稿中，我们在引言的最后对总体的实验逻辑做了详细阐述，并在介绍每个实验的方法之前进一步阐述实验 2 和 3 的实验逻辑，在实验 2 和 3 的结果讨论部分结合实验逻辑进行了较为深入的讨论。

意见 5: “4.3 讨论”一节的内容完全是实验 3 结果总结，而不是针对实验 3 结果进行的讨论，应该重写。

回应: 已按照建议对实验 3 结果进行了讨论。

意见 6: 实验 2 的讨论中，作者说“韵相关产生的抑制效应量大于音节相关，但两者之间的差异未达到显著差异的水平”（见 3.3 一节）。然而，在“5 总讨论”一节中，文章说“相比而言，在语音编码和发音阶段，音段的效应量大于音节。这表明……在语音编码和发音阶段，音段的作用大于音节”。统计检验差异不显著的情况下，这样的讨论是错误的。建议根据统计检验结果，重新进行相应的讨论。

回应：感谢审稿专家指出上述不一致之处。在实验 2 的讨论中，我们所指的是：干扰字和目标词之间分别存在韵和音节相关时，所产生促进效应的大小之间的差异。在总讨论中，审稿专家所指出的部分我们比较的是：根据具体的促进效应以及对应的标准差计算出来的效果量 Cohen *d* 值的大小。确实是由于我们表达的不明确，造成了混乱。在修改稿中，我们已经对具体的促进效应量和效果量（Cohen *d*）做了明确区分（见实验 2 讨论和总讨论以及对审稿专家 2 第 6 个问题的回答）。

审稿人 2 意见：

意见 1：延迟命名结合发音抑制的任务，反应时为什么包括了语音编码过程，排除了音韵编码过程？其中的逻辑是什么，没有交代清楚。排除了音韵编码，如何保证反应正确？

回应：感谢审稿专家的问题。在修改稿中，我们已经在引言最后一段对研究逻辑进行了详细阐述（同时请参考对审稿专家 1 第 1 个问题的回答）。同时，文中所提到的“排除”是指：在延迟命名（实验 2）及其与发音抑制结合的任务（实验 3）中，由于被试已经完成了音韵编码过程，因此在提示可以进行发声反应的线索出现后被试对目标图片进行命名，这两种情况下所测得的反应时中均未包含音韵编码过程这一加工阶段。实验 2 和 3 的正确率都高达 98% 以上，这也表明被试确实进行了音韵编码过程，保证了反应的正确率。我们在阐述实验逻辑时删除了“排除”这一词语，以免引起读者误解。

意见 2：“四种条件下干扰字的字频（《现代汉语频率词典》，1986）和笔画数匹配，统计结果无显著差异。”建议给出具体数值和统计检验值。

回应：已在材料介绍部分增加了每组平均值和统计检验值，并在附录 1 中增加了每组实验材料的字频和笔画数。

意见 3：实验 2 的讨论中，“实验 2 发现了韵相关产生的抑制效应量大于音节相关，但两者之间的差异未达到显著差异的水平”，既然不显著，说“大于”就不够严谨。

回应：已修改。

意见 4：实验 2 的讨论中，“我们的实验发现与已有结果（Wong et al., 2008）一致。”Wong 等发现的是促进效应，本实验 2 发现的是抑制效应，如何“一致”，还需要进一步解释。

回应：这里的叙述存在不当。我们的本意是指：因为两个实验中都发现了韵的效应，无论其效应是促进还是抑制，这表明两个实验中被试对韵的信息都比对首音的信息更为敏感。在这一点上两个实验一致，文中已经做了相应修改。

意见 5：实验 3 的讨论中，“另一方面，尽管音节相关条件与无关条件相比仅有 12ms 的差异，其效应量小于音段”，12ms 与表 1 中的结果不一致。

回应：在实验 3 中，音节相关条件与无关条件相比的反应时应该为 10 ms，我们已改正此错误。

意见 6：总讨论中有如下说法“在音韵编码阶段，音节的效应量大于音段；相比而言，在语音编码和发音阶段，音段的效应量大于音节”。效应量是否经过统计检验？

回应：这里的“效应量”指的是效果量 Cohen *d* 值，表示的是实验组条件与无关条件之间的差异量的效果量大小。效果量是表示实验效应强度或者变量关联强度的指标（Snyder & Lawson, 1993），它不受样本容量大小的影响（或者影响较小）（郑昊敏，温忠麟，吴艳，2011）。

计算和报告效果量可以区分统计显著性和实际显著性 (Kirk, 1996), 而且可以通过元分析方法比较同一效应在不同研究中的实验结果 (Hunter & Schmidt, 1990) 来估计统计检验力 (Cohen, 1988)。一般来说不会比较效果大小之间的差异是否显著, 只是根据效果大小值确定其可靠性的程度。0.2 < Cohen d < 0.5 表示效果量大小弱, 0.5 < Cohen d < 0.8 表示效果量大小中等, Cohen d > 0.8 表示效果量大小强 (Cohen, 1988)。因此, 根据计算出的 Cohen d 值, 我们对音节效应和音段效应的效果量大小进行了定性判断。如表 2 所示, 音节效应的 Cohen d 均大于 0.8, 效果量大, 而韵促进效应的 Cohen d 值均较小, 效果量很弱。

由于我们表达的不明确, 造成了混乱。在修改稿中, 我们已经对具体的促进效应和效果量 Cohen d 值做了明确区分, 并修改量上述审稿专家所指出部分的错误。

意见 7: 总讨论中应强调本研究的创新性发现, 如何解答了现有研究中的分歧等, 并与问题提出部分紧密呼应。

回应: 感谢审稿专家的建议。我们修改了前言和总讨论部分, 并强调了本研究的新发现及其理论意义。

意见 8: 文章还存在一些笔误和表述不规范的问题, 如: 1) “音节和音段的促进效应发生汉语口语词汇产生中的音韵编码阶段”, 2) “首音段相关条件下显著反应时显著慢于无关条件”, 3) “采用图画-词汇干扰范式中在粤语中发现了韵母和音调结合条件下的促进效应”。请作者举一反三, 认真修改。

回应: 我们已经通读了全文, 尽力修改了文中描述不统一或者错误之处。

第二轮

审稿人 2 意见

意见 1: 总讨论中, 在本文的主要发现讨论完后, 又增加了一大段关于效果量 Cohen d 的说明和解释, 并增加了表 2, 这个安排似乎不太妥当, 导致总讨论的文脉中断、可读性下降。可考虑把对 Cohen d 的相关解释说明放在方法中, 表 2 可考虑放在实验 1 的结果或讨论中。

回应: 我们感谢并接受审稿人的第一个建议, 将对 Cohen d 值的涵义解释移到了实验 1 的结果部分, 这样便于读者在第一次出现 Cohen d 值就明确地知道其涵义, 更容易理解实验结果。

表 2 主要是想将本实验 1-3 的总结果与已有的研究结果进行跨实验间的比较, 因此我们倾向于仍然将表 2 放在总讨论部分。

意见 2: 最后一段的结论应该精简, 表述主要的发现而不是再次重复主要结果。

回应: 感谢审稿专家仔细阅读, 我们删除了最后一段中的结果重复部分。

意见 3: 图 1 应该放在更前面一些, 离相关表述近点。

回应: 感谢审稿专家的建议, 在修改稿中已将图 1 放在了“合适单元假说”所在段落的下方。

意见 4: “另一方面, 尽管音节相关条件与无关条件相比仅有 10ms 的差异, 其效果量 Cohen d 值小于音段相关条件。”前后表述之间的逻辑关系是什么, 为什么用“尽管”? **回**

应: 这是一处笔误, 修改稿中已将“尽管”一词删去。

意见 5:

一些表达不通顺或不够规范, 标点使用不当, 如:

(1) “Wong 和 H.-C. Chen (2008, 2009) 采用图画-词汇干扰范式中在粤语中发现……”;

回应：已修改为：“Wong 和 H. -C. Chen（2008，2009）采用图画-词汇干扰范式，在粤语中发现……”。

（2）“可以这些效应发生在口语产生中单词形式编码（音韵编码或者语音编码）还是最后的发音输出阶段”；

回应：已修改为：通过比较三个实验中是否出现音节效应或音段效应，可以推测这些效应发生在口语产生中单词形式编码（音韵编码或者语音编码）还是最后的发音输出阶段。

（3）“为了进一步确定实验 1 中发现的音段促进效应是否发生在口语产生中的语音编码阶段。”此处不应用“。”；

回应：已修改。

（4）“使得被试不能进行语音编码阶段的加工（Baddeley, Lewis, & Vallar, 1984）提示线索出现后，要求被试说出目标图画的名称。”中间缺一个“，”或“。”；

回应：已修改。

（5）“O’ Seaghdha 等（2010）提出的汉语口语产生的单词形式编码模型（如图 1 所示）的合适单元假设。”这句似乎没完；

回应：已修改。

（6）有些地方应该用中文“，。”但使用了英文“，.”；

回应：已通读全文并修改，

请作者举一反三，认真通读。

回应：非常感谢审稿专家的仔细审阅，除了上述修改之外，我们仔细阅读了全文，对所发现的不通顺之处进行了逐一修改。

第三轮

主编意见：

意见 1：建议在摘要中的结果部分给出一些具体结果，目前比较多的是结论表述。

回应：已在摘要中增加了结果的描述。

意见 2：表格的规范：表 1、表 2 改为三线表。

回应：已将表 1 和表 2 改为三线表。

意见 3：关于 Cohen d 的一段解释冗余“Cohen d 值是衡量统计检验效果大小（效应量）的指标之一，效果量是表示实验效应强度或者变量关联强度的指标（Snyder & Lawson, 1993），它不受样本容量大小的影响（或者影响较小）（郑昊敏，温忠麟，吴艳，2011）。计算和报告效果量可以区分统计显著性和实际显著性（Kirk, 1996），而且可以通过元分析方法比较同一效应在不同研究中的实验结果（Hunter & Schmidt, 1990）来估计统计检验力（Cohen, 1988）。 $0.2 < \text{Cohen } d < 0.5$ 表示效果量大小弱， $0.5 < \text{Cohen } d < 0.8$ 表示效果量大小中等， $\text{Cohen } d > 0.8$ 表示效果量大小强（Cohen, 1988）。”建议简化或者删除。

回应：感谢主编的建议。我们删除了解释冗余的部分，保留了对 Cohen d 强弱范围的界定部分，以使读者了解本研究中 Cohen d 值效果量大小的范围。修改后的部分如下：

Cohen d 值是衡量统计检验效果大小（效应量）的指标之一，效果量是表示实验效应强度或者变量关联强度的指标（Snyder & Lawson, 1993），它不受样本容量大小的影响（或者

影响较小) (郑昊敏, 温忠麟, 吴艳, 2011)。0.2 < Cohen d < 0.5 表示效果量大小弱, 0.5 < Cohen d < 0.8 表示效果量大小中等, Cohen d > 0.8 表示效果量大小强 (Cohen, 1988)。

我们将删除的部分补充在了总讨论中对已有研究结果进行综合比较的元分析部分。

意见 4: 实验 2 的问题提出部分和实验 3 的讨论部分建议补充文献依据。

回应: 感谢主编的细心阅读, 我们在这两部分增加了相应的参考文献, 在参考文献列表中也增添了相应的文献。

意见 5: 建议斟酌如下表述: “最近, 一些研究者提出了前运动 (Pre-motor) 和运动加工过程的分离假设 (Eichhoff, Heim, Zilles, & Amunts, 2009; Hickok & Poeppel, 2007) ……” 2009 年已经是 5 年前, 很难说是最近。

回应: 感谢您指出文中描述不确切的部分, 已修改为如下表述:

在口语产生研究领域, 一些研究者提出了前运动 (Pre-motor) 和运动加工过程的分离假设 (Eichhoff, Heim, Zilles, & Amunts, 2009; Hickok & Poeppel, 2007) ……

另外, 我们增加了 3 篇文献 (Hickok, 2012, 2014; Roelofs, 2014) 指出研究者对这一问题的关注, 并在参考文献列表中增加了相应文献。这一问题并非本研究主题, 因此未做深入阐述。