

《心理学报》审稿意见与作者回应

题目：汉语音节累积词频对同音字听觉词汇表征的激活作用

作者：方杰, 李小健, 罗畏畏

第一轮

审稿人 1 意见：

意见 1：作者提到“不能通达的词汇却可以受词汇自身信息影响而得到不同程度的表征激活，提示我们可以通过考察词汇表征的激活(activation of lexical representation)来分离词汇通达过程和分析词汇表征。”如何分离？请作者进一步阐述。

回应：感谢审稿人的建议。在“1.2 听觉词汇通达中的同音词表征激活”部分的第二段，作者进行了进一步的阐述。

意见 2：作者提到“Levelt 等人(1999)提出言语产生词汇通达的两阶段分离激活模型，中间环节分离出先后进行的词条(lemma)和词位(lexeme)两个节点，假定同音词在词位节点共享语音表征，激活语音表征可以反馈到早一阶段的词条节点，激活其它同音词，形成干扰。”如果是分离两阶段模型，其观点是词条选择和音韵编码两个阶段之间不存在交互作用，音韵编码阶段的激活不会反馈回到词条编码。与作者在这里提到的言语产生中“激活语音可以反馈到早一阶段的词条阶段”的说法是矛盾的。

回应：感谢审稿人的建议。原文中的内容“假定同音词在词位节点共享语音表征，激活语音表征可以反馈到早一阶段的词条节点，激活其它同音词，形成干扰。”已被删去，不影响所要引出的表征共享模型。

意见 3：作者提到“这些不一致的观点说明研究方法、材料、语言和知觉反应通道等的差异会带来结果和解释上的混淆，因此采用直接的而不是间接的途径来考察听觉的词汇通达是非常必要的。”在文中作者并未对提到的研究做出进一步的阐述，就提出了上述结论，不能令读者信服。请作者做进一步的阐述。

回应：感谢审稿人的建议。文章对这段进行了修改，成为一个需要探讨的问题：“这些不一致说明累积词频作用不稳定，还是实验范式不合适，也有待探讨。”并且在“4.综合讨论”的末尾第二段回答了这个问题。

意见 4：作者基于言语产生研究领域中的争论提出了“同音字听觉词汇通达的表征激活是否可以被音节的累计词频作用解释”这一研究问题。这里涉及到了两个研究领域：言语产生和言语理解，这是两个不同的认知加工过程，当然，也存在一些联系。作者应进一步阐述其研究问题是如何基于言语产生领域的争论提出的？其研究是否能解决言语产生领域的争论？

回应：感谢审稿人的建议。这个问题提得好。是上一稿把言语产生领域的模型与争论放在过于重要的位置了。经过全篇文章的梳理和修改，文章更多从听觉言语理解已有的研究出发(例如周晓林等, 1994; 李小健等, 2011)，把言语产生与其它非听觉理解研究领域(包括视觉刺激判断)提出的问题一起作为研究问题的部分背景。

然而，需要说明的是，同音字(词)的表征问题，只有言语产生研究提出较明确的理论和模型，我们还不知道同音字表征是否与加工(言语理解和言语产生)有关。所以本文必须对那些模型进行听觉理解实验的检验。在“引言 1.3”及“4 综合讨论”里都对此进行了阐述。目

前的结论是，言语理解和言语产生领域所得的结论很不同，一个领域的模型不太能够适用于另一领域。

意见 5：所有实验的数据分析中只进行了被试分析，未做项目分析。这一类的研究都需要同时做被试分析和项目分析，需要增加项目分析的结果。

回应：感谢审稿人的建议。关于项目分析的问题，作者做如下说明和解释：

只做被试分析是根据统计原理采取的做法。只有当一个因素（变量）符合随机抽样的条件（如同被试是大学生里随机抽样那样）才能采用含随机因素的方差分析模型，把该因素设为随机因素来计算。重复测量中的被试就是随机因素。每个被试在每个变量水平上有一个项目变量的均值（例如平均反应时），所有被试对这个变量水平就组成了一个随机分布，当样本足够大时，就符合中心极限定理，可以进行均值 Z-检验，有多个水平和多个变量就可以进行方差分析。类似，当实验刺激（项目）是随机抽样得来，样本足够大，也可以进行项目的方差分析。如果是被试内设计，在多个项目条件下重复测量，项目条件之间就组成项目间设计，要使用不同的方差分析模型（随机或相关区组的组间分析）。**对我们的实验来说，刺激需要同时控制不少条件，等同许多非实验变量，找到的刺激几乎就是所有符合条件的同音字了，并没有随机抽样性质，失去推广到更多项目和总体上去的意义。**

意见 6：实验一的标题为“音节累积词频如何受同音字具体词频的影响”，标题未表达出所要研究的问题。实验一目的是考察音节累积词频和具体词频对同音字加工过程的影响，并非具体词频如何影响累积词频。

回应：感谢审稿人的建议。实验 1 的标题改为“累积词频与具体词频”。

意见 7：实验二的标题为“音节累积词频如何受同音字数的影响”，标题未表达出所要研究的问题。实验二目的是考察具体词频不变的条件下，音节累积词频和同音字数这两个因素对同音词加工过程的影响。

回应：感谢审稿人的建议。实验 2 的标题改为改为“累积词频与同音字数”。

意见 8：实验 1a 为“音一字同音判断任务”，实验 1b 为“词汇判断任务”。同样性质的实验材料，有时表达为“字”，有时表达为“词”。请作者统一表达。

回应：感谢审稿人的建议。本文研究汉语单音节词。大量使用“同音字”说法是为了突出中文特点，采用“字”来表示《词频词典》里的“单字词”。本文作者还专门在引言开始第一段作了如下说明：单音节“词”和“字”，都是指中文信息处理词切分意义上的单字词，以帮助读者理解。

但是，当涉及其他语言，如英语的同音词，常用的术语(如词汇通达，词汇判断任务等)，我们只能采用“词”的说法来兼容。我们也尝试把“词汇判断任务”改为“单字词判断任务”，似乎变成一个大家不熟悉的范式了，反而增加了读者的阅读负担

意见 9：作者指出“条件‘累积词频高-同音字数多’的反应时最慢，这不能归结为刺激里含有较多的同音高频字(该条件有最多的 10 个同音高频字，见表 2)，因为高频字能帮助加快反应时间”，但未对其进行解释。

回应：感谢审稿人的建议。本文作者已经在“3.1.5 实验结果”部分的最后一段话中，增加了说明。

意见 10：分析实验 2a 和 2b 的错误率数据。

回应：感谢审稿人的建议。本文作者已经增加了对实验 2a 和 2b 的错误率数据分析。

意见 11：视觉词汇判断是一个判断所呈现字词是否为真字和假字的任务，与视觉同音字词汇通达过程有一定联系，但并不能与同音字词汇通达的加工过程等同。作者多次提到“视觉词汇判断不是考察同音字词汇通达和表征激活的可靠范式”，建议作者将弱化对这一结论的阐述。

回应：感谢审稿人的建议。正如审稿人所言，“视觉词汇判断”是一个实验任务或范式，而“词汇通达”是从词的形式获取词义（Jay 的定义，见引言 1.1 第二段），确有不同。所提到的“视觉词汇判断不是考察同音字词汇通达和表征激活的可靠范式”，比较含混，已经删去。

意见 12：实验结果表明视觉词汇判断过程中也发现了具体词频效应，这是否表明同音字通达中的词频效应可能发生在词汇判断过程？这一发现与同音字通达过程中的词频效应有何不同？

回应：感谢审稿人的建议。本研究采用的音-字同音判断任务，其中含有“看字”环节，所以可能本身就存在视觉词频效应。实验是为了考察听觉词汇通达过程，所以要在音-字同音判断任务所记录的反应时里排除视觉环节的词频效应，才采用了相同刺激的视觉词汇判断任务，所得的反应时作为音-字同音判断反应时的协变量，经过协方差分析被排除其影响。这个做法类似 Levelt 等人（1999）在英-荷语翻译任务后做一个英语词汇判断任务，反应时相减，用以排除识别英语所产生的词频效应。本文在引言 1.4 说明了这个做法的逻辑。

意见 13：综合讨论需要与前言中所提到问题的联系起来进行深入讨论，阐明其研究的理论意义。

回应：感谢审稿人的建议。这次修改，作者对文章的前言和综合讨论部分做了大幅度的增删，将综合讨论与前言中所提到问题的联系起来进行深入讨论，阐明了研究的理论意义。

意见 14：视觉词汇判断的词频效应是自动的和稳健的。什么是“稳健的”？

回应：感谢审稿人的建议。“稳健的”来自英文文献常用的词“robust”（意思是强健的、稳固的），表示一个效应不仅稳定地出现，而且效应比较强。鉴于这个词和翻译的不熟悉性，我们尽量改用其它词。

审稿人 2 意见：

意见 1：听觉词汇通达中，具体词频（书面词频）的作用可能与任务有关，本研究中的实验任务是音-字同音判断任务，实验程序是“字音播放完后间隔 50ms，屏幕中心呈现一个汉字，被试的任务是尽快又准地按键判断听到的字音和看到的汉字是否同音。”因此听觉“字音”加工有更充足的时间，判断反应时中包含更多的成分是汉字识别的时间，因而反应时较少受到累积词频对通达“音节表征”的影响，更多受到具体词频（书面词频）的影响。如果任务采用纯听觉任务，结果是否可能会有不同？Li, X. J., Wang, W. N., & Li, X. Q. (2011)和 Zhou, X., & Marslen-Wilson, W. (1994)研究结果的不同，是否可能与两个实验中采用的实验任务不同有关？

回应：感谢审稿人的建议。这是关于听觉词汇通达实验方法的大问题，已经论述在相关文献中(李小健, 方杰, 楼婧, 2011; 李小健, 王文娜, 李晓倩, 2011)，本文在修改中改进了论述，这里只作一些具体的简要解释。

（1）问“Li, X. J., Wang, W. N., & Li, X. Q. (2011)和 Zhou, X., & Marslen-Wilson, W. (1994)研究结果的不同，是否可能与两个实验中采用的实验任务不同有关？”是的。Zhou (1994) 研

究的对象是双音节词,而词素在双字词里的语义是确定的(假定 Zhou 没有选用双字同音词),所以没有同音词汇歧义问题。听者对语义清楚,就没有判断的障碍,只是有词频影响(反应时)的大小。这与 Li (2011) 研究的对象是汉语单音节同音字还是很不同的,因为单音节的听觉词汇通达会产生同音词汇歧义,不能通达词义。于是研究的是词汇表征的激活,探测某些表征激活带来的促进或抑制,而不是词汇完全通达。本文的修改稿在综合讨论里用实验 2 的结果解释了 Zhou (1994) 的结论(为何累积高频产生抑制作用)。

(2) 对于音-字同音判断任务过程中涉及了什么加工,本文在引言(1.4)介绍了实验范式,其中提到任务过程中含有“看字”环节,所以需要在音-字同音判断任务所记录的反应时里排除视觉环节的词频效应,采用了相同刺激的视觉词汇判断任务,所得的反应时作为音-字同音判断反应时的协变量,用于协方差分析,把视觉的汉字识别产生的差异抵消。这个做法类似 Levelt 等人(1999)在英-荷语翻译任务后做一个英语词汇判断任务,反应时相减,用以排除识别英语所产生的词频效应。协方差分析抵消了视觉的效应后,方差就集中到“与预期相符”(或“与预期违背”)的听觉激活效应上。其逻辑是,听一个音节,当看到的同音字与听字音所自动激活的字义相符时,反应会较快、较准确,从而知道所激活的同音字表征是与预期同类的。

(3) 采用纯听觉任务是可以的,但是在听觉下输入的同音单字词会产生同音词汇歧义,导致通达障碍,在不能通达词义的情况下,实验者要确定听者的听觉激活了什么样的词汇表征,还是需要采用一些办法。我们的实验采用了“与预期相符”(或“与预期违背”)的探测方法。用视觉汉字做心理预期的探测(音-字同音匹配),利用了汉字在视觉下语义明确。如果要用语音的声响做探测,可以采用含目标字的双音节词(音-音汉字匹配)。这两者都能使听者做明确的判断,但音-字同音匹配需要处理视觉的外加效应,音-音汉字匹配要处理双字词的整词效应。有的实验(周海燕和舒华,2008)让被试听双音节词,再用视觉呈现其尾字。总之都需要有被试的明确判断。注意到,不做判断的纯听觉任务,只能用脑科学探测方法(脑电、脑成像等)来测量心理表征的激活。

意见 2: 第 5 页,“按照词汇通达的一些传统理论(如 McClelland & Elman, 1986; Marslen-Wilson, 1987),一个音节的同音字由于共享一个语音表征,该音节的所有同音字都能被该音节激活。如果是这样,都激活起来的同音字表征会如何累积起来发生作用呢?是合力起促进作用还是竞争起抑制作用?其结果与累积词频又是什么关系?”表述不够清楚。“同音字表征”、“同音字听觉词汇通达的表征激活”的含义是什么?心理词典中,语音表征和正字法表征应该是分离的。根据本研究的结果,作者认为同音字的语音表征是单一的,其正字法表征是多个,在听觉词汇通达上是一对多的关系?还是同音字的语音表征也是分离的?

回应: 感谢审稿人的建议。修改稿已经把意思较含混的话修改了,使用较明确的说法。“同音字表征”、“同音字听觉词汇通达的表征激活”的含义,在修改后的引言 1.2 里给予了更清楚的介绍。词汇表征的激活可以是不同类型的表征的激活,而词汇通达需要获得词义。同音字的表征激活之所以特别,是因为听觉输入的单音节同音字会产生词汇歧义,难以获得通达,但是存在某些表征的激活,而且影响了后期获得词义。

意见 3: 第 11 页,“结果首先指出,音节的同音族内高频字在听觉词汇通达的速度和准确性上都显示了促进作用。”表述不够清楚,听觉词汇通达应该指在听觉通道获得语音的意义。为什么“音节的同音族内高频字”促进了“听觉词汇通达的速度和准确性”?是跨通道的促进吗?本研究中采用的“音-字同音判断任务”中视觉呈现的作用是什么?如果探测纯听觉通道的“听觉词汇通达”,结果会如何?

回应：感谢审稿人的建议。审稿人所提的问题的确需要更清楚地表述。词汇通达与词汇表征的激活不同。我们在修改后，更明确指出这点（如引言 1.2），在全文里尽可能分清这两个术语的使用。例如，在实验 1 的讨论开始时指出，“引言 1.4 说明了，使用协变分析，从音-字同音判断的观察量中去除视觉词汇判断中相同变量的效应，所得的结果能够比较集中反映听觉词汇通达过程中的表征激活效应。”然后，把原文的“音节的同音族内高频字在听觉词汇通达的速度和准确性上都显示了促进作用”改为“具体词频（这里其实就是音节的同音族内字频）的高频在听觉词汇通达中的表征激活速度和准确性上都显示了相对于低频的促进作用。”

至于所问的“音-字同音判断任务”和纯听觉通道的“听觉词汇通达”，在问题 2 里已经给予了回答。

第二轮

审稿人 1 意见：

意见 1：作者根据评审者的意见，对“汉语音节累积词频对同音字听觉词汇通达的表征激活作用”一文做了认真的改进，同意发表。但在文章写作上还需要对词句进一步斟酌、改进，使文章的流畅性、可读性更好。

回应：感谢审稿人的建议。作者已经再一次对全文进行了词句斟酌、修改，确保文章有较好的流畅性和可读性。

意见 2：另外，在修改的文章中，有时对评审者提出的问题直接回答，会使文章的连贯性下降。例如，在“4 综合讨论”“根据以上讨论，我们再来解释前面提出的一些有待探讨的问题。Zhou 等人（1994）……，而观察同音字的各种表征激活及其影响关系恰恰是在同音词汇歧义下最合适。”可以简化解释，加强文章的连贯性。其它部分请作者自查一下。

回应：感谢审稿人的建议。作者已经就此对综合讨论修改，并且增加了小标题，提高所述内容的逻辑清晰度，改善连贯性。

审稿人 2 意见：

意见 1：视觉言语理解同样存在词义加工的问题，比如同样字形的词，可能具有多重意义。

回应：感谢审稿人的建议。在“1.1 同音词的听觉词汇通达”部分的第二段，作者专门谈论了这个问题，虽然视觉言语理解同样存在词义加工的问题，但“同音异形词的通达障碍只是在听觉理解中发生的，区别于同音同形词的视觉词汇通达障碍。”

意见 2：“由于在通达词汇形式(word form access)阶段发现词频效应，也发现“slips of the tongue”的说话错误更多表现为将同音低频词说成同音高频词”。这里指的是同音词错误吗？不论高频或低频的发音都是相同的，那么如何知道被试说错了？

回应：感谢审稿人的建议。一些拼音语言（意大利语、法语等）的同音高频词义与低频词义对应不同的性别（gender），研究者可从看图说词任务被试说出的词性知道通达了同音的高频还是低频词义。作者将文章此处修改为：发现“slips of the tongue”的说话错误更多是将同音词的低频词义误以高频词义的性别形式说出（意大利语等）。

意见 3：语言产生领域内有关同音词的争论是如何与言语理解领域内的有关同音词的争论联系起来的？这似乎是两个不同的问题。不清楚作者在这里占用大量篇幅论述言语产生领域内有关同音词的理论争论对于本研究的意义何在？而且，作者在最后指出：“仅就同音字累积词频的表征激活作用看，言语理解与言语产生的同音字通达过程及同音字表征激活很可能是不同的”。这一结论是如何得出的？如果是这样，那也就没有必要在前言中叙述言语产生领

域对同音词的争论了。

回应：感谢审稿人的建议。言语理解和言语产生是两个不太相同的问题，但研究文献使用了不少看似相同的概念，如“词汇通达”、“词汇表征”等。概念混用的原因之一，是言语理解和言语产生具有一定相似的和相互作用的加工过程。尤其是“同音词表征”，是先在言语产生领域由 Levelt (1994, 1999) 等人提出了加工模型和表征模型后，才从词汇通达过程分离出来，给予专门研究。然而，“词汇通达”、“词汇表征”等概念在言语理解和言语产生两者之间混用容易给、也已经给读者造成相当大的混淆。文献还没有见到能够区别两者的同音词表征理论和模型。由于这个历史和逻辑关系，本文需要以一定篇幅介绍言语产生领域的前期发现，在文章的最后给出我们的（两者不同的）结论，回答争论。上一稿没有在这个问题上交代得足够清楚，现在已经修改。

意见 4：关于项目分析，第一，心理语言学的研究都需要同时做项目分析和被试分析。不做项目分析不知道所选项目是否能代表总体样本。第二，不能理解作者的解释，为什么项目的总体不是正态分布？如果所选择的项目不是从一个总体呈正态分布的样本中选择出来的，那么是否应该将项目作为一个协变量来分析。第三，这样选择的样本会对结果产生什么影响？

回应：感谢审稿人的建议。我们对实验数据进行了项目分析，修改稿中，符号 F_2 表示以测试项目为随机变量进行的项目分析结果。

对于审稿人提出的第二个问题“为什么项目的总体不是正态分布？……”和第三个问题。我们的数据是反应时，是项目（单字词）上记录的反应时。反应时数据在通常情况下不是正态分布的。为了符合统计模型，项目反应时的样本要足够大（每个条件的项目足够多，至少 30 个），其均值做成的数据形成一个均值空间（总体），就可以认为这个均值空间是具有正态分布的（中心极限定理）。

意见 5：排除被试对视觉呈现汉字加工过程可能引起的效应，作者介绍了两种方法，为什么用了第一种方法？如果用第二种方法，是否会得到不同的结果？

回应：感谢审稿人的建议。本文之所以会用第一种方法，原因如下：

- 1、方法一是有探测听觉表征激活的研究先例的，本文只是沿用前人的方法。
- 2、本文采用第一种方法，与这项研究要与 Levelt 等人(1999)在言语产生研究的结果进行比较有关，所以采用类似 Levelt 等人(1999)采用的方法（我们用协变量分析比 Levelt 等人用反应时相减更有统计学理据）。
- 3、李小健、方杰和楼婧(2011)采用方法一，控制累积频率，同时变化具体频率和同音字数，考察同音字听觉词汇表征，本文的研究结果正好和这个研究进行结果比较和分析，详见综合讨论 4.2。
- 4、获得协变量的视觉词汇判断实验也有其自身意义，可以为更多比较提供证据（见综合讨论 4.3）。

关于“如果用第二种方法，是否会得到不同的结果？”，已有研究先例（李小健、方杰，楼婧，2011；李小健、王文娜，李晓倩，2011）看出两种方法都有探测听觉表征激活的能力，都是可以采用的，并且第二种方法在分离听觉表征激活上的道理更彻底、清晰，因此，第一种方法能够得出的结论，第二种方法也能得出。但是，基于使用第一种方法的四个原因的综合考虑，本文没有使用第二种方法。为了避免读者产生误解，在修改稿中，作者删除了关于第二种方法的介绍，改为：“其他去除看字环节的视觉识别效应的具体方法可参见李小健、王文娜和李晓倩(2011)研究”，让有兴趣的读者选择进阶阅读。

意见 6：什么是“声母、韵母、声调尽可能均匀分布”？

回应:感谢审稿人的建议。对文章“2.1.3 实验材料”做了以下叙述上的修改:在选择音节时,声母、韵母、声调在各条件的音节里出现的广度(尽可能都使用)和均匀度(各自出现的次数)尽可能接近。

意见 7:实验 1 的结果发现,累积频率的主效应显著:即无论具体频率的高低,累积高频率下的反应时(621ms)显著慢于累积低频率下的反应时(611ms)。但从图 1 来看,实验 1 的结果主要表现为具体低频下,累积高频和累积低频存在显著差异(20 ms 的差异)。这是不是也表明具体频率的高低也对反应时产生了影响?

回应:感谢审稿人的建议。实验 1 的音-字同音判断反应时分析结果是“具体词频和累积词频的交互作用不显著,累积词频的主效应不显著,仅有**具体词频的主效应显著**。”由于累积词频和具体词频没有交互作用,只需分别看累积词频的主效应和具体词频的主效应,不需再分开一个变量的两个水平来看另一个变量的差异(简单主效应)。因此,具体数字的差异(20 ms 或别的数量差异)不说明本质问题,都要从统计上是否有显著的准则来看。

意见 8:真假字判断是一个正字法判断任务,而音一字判断任务是所呈现词语的声音与视觉呈现单字的发音是否一致。这其中所涉及到的认知过程不同,语音所激活的程度也不同。那采用真假字判断的数据作为协变量分析,能否解决作者所提到的音一字判断任务中视觉呈现单字所引起的混淆?

回应:感谢审稿人的建议。视觉词汇判断的反应时作为协变量,能否消除音-字同音判断的反应时中的视觉效应,正是在设计这个实验的方法时需要深入思考的问题,是难点。作者认为采用真假字判断的数据作为协变量分析的方法能够解决音一字判断任务中视觉呈现单字所引起的混淆。原因有以下几点:

- 1、采用真假字判断的数据作为协变量分析的方法来解决音一字判断任务中视觉呈现单字所引起的混淆是有探测听觉表征激活的实验先例,例如李小健、方杰和楼婧(2011)研究,因此可以采用。
- 2、Levelt 等人(1999)在英-荷语翻译任务后做一个英语词汇判断任务,反应时相减,用以排除识别英语所产生的词频效应,分离出荷语产生的词频效应。本文采用的方法(采用汉字词汇判断任务的反应时和错误率作为协变量分析,以去除看字环节的视觉识别效应)是在言语产生的同音词表征研究(如 Levelt 等人(1999))的基础上发展而来的。

采用真假字判断的数据作为协变量分析的方法其实是利用了认知心理学普遍存在和使用的认知“减法原理”,大部分认知心理学的变量分离都多多少少依靠这个减法准则。从认知操作来讲,真假字判断的数据作为协变量分析的方法可以大体剥离视觉词频效应,但是否“恰好”从音-字同音判断反应时里排除视觉词汇判断反应时的变差,一点不多,一点不少,很难这么说。相对而言,协变分析方法比著名学者 Levelt 等许多人在词汇通达研究上经常采用的反应时相减方法要更好。

意见 9: wong & chen, 2008: w 应大写。

回应:感谢审稿人的建议。wong & chen, 2008 的 w 和 c 都已经改为大写 Wong & Chen, 2008。

意见 10: 词频没有标注单位。

回应:感谢审稿人的建议。词频单位已经补上(见表 1 的注)。本文中的单字词词频都来自宇航出版社出版的《现代汉语常用词词频词典》(1990 年版),是七百五十万个进入统计的一字词中出现 5 次或 5 次以上该字的频度(即次数)。

意见 11: 数据表示 $F(1, 31) = 3.14$, $p = 0.086$, $\eta_p^2 = 0.092$, 最后一个指标的意义是什么? 一般 APA 规定的都是要表示出 MSE。

回应: 感谢审稿人的建议。 η_p^2 是方差分析中使用的一个效应量指标。由于统计显著性实际上是效应量和样本量共同作用的结果, 因此, 当统计检验显著后, 效应量就是研究者最关心的, 需要报告独立于样本量的效应量大小。效应量 η_p^2 的计算公式如下:

$$\text{效应量} = \eta_{\text{partial}}^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}}$$

η_p^2 表示要研究的实验因素的方差(分子)在排除了其他实验因素后所占总方差(分母)的比例为多大, 比例越大, 说明实验因素的效果量越大, 可以比较其他变量看哪个的作用较大。 η_p^2 值可以通过 SPSS 软件的 General Linear Model 运行方差分析, 勾选 Options 框中的

Estimates of effect size 即可得到, 方便易行。更详细的内容可参见如下文献:

American Psychological Association. (2010). *Publication manual of the APA* (6th ed.). Washington, DC: Author.

郑昊敏, 温忠麟, 吴艳. (2011). 心理学常用效应量的选用与分析. *心理科学进展*, 19(12), 1868–1878.

卢谢峰, 唐源鸿, 曾凡梅. (2011). 效应量: 估计、报告和解释. *心理科学探新*, 31(3), 260–264.

胡竹菁, 戴海琦. (2011). 方差分析的统计检验力和效果大小的常用方法比较. *心理科学探新*, 31(3), 254–259.

APA 手册(第六版)才明确要求在方差分析中报告效果量 η_p^2 的大小, 在 APA 之前版本的手册中没有这个要求。目前, APA 手册(第六版)明确要求在方差分析中报告效果量 η_p^2 大小的这一举措, 已经被国内有关学者在方差分析中广泛使用, 据笔者统计, 从 2012 年 9 月至今, 《心理学报》发表的文章中, 至少有 20 篇以上的文章在方差分析中使用了效果量 η_p^2 或 η^2 , 例如[李轩, 刘思耘. (2012). 汉语短序列回忆中的语音相似性和视觉相似性效应. *心理学报*, 44(12), 1571–1582]。

意见 12: 在图上标注出各个条件下的 SD。

回应: 感谢审稿人的建议。已经在所有数据分析图形(图 1 到图 3)上增加了各个条件下的 SD, 图中括号内数字即表示 SD。

意见 13: 19 页 15 行: 研究指出, 同音族内的高频字激活最快, 低频字则收到抑制(李小健等人, 2011)。“收到”应该为“受到”。

回应: 感谢审稿人的建议。已经将“收到”改为“受到”, 即研究指出, 同音族内的高频字激活最快, 低频字则受到抑制(李小健等人, 2011)