

《心理学报》审稿意见与作者回应

题目：中文阅读中无关言语效应的认知机制探究：眼动证据

作者：吴三美 田良苏 陈家侨 陈广耀 王敬欣

第一轮

审稿人 1 意见：

本研究运用眼动技术考察了背景音对阅读理解过程的影响、及其发挥作用的认知机制。研究结果主要发现，背景音对阅读理解的干扰作用主要来自于背景音的语义信息、其作用发生在阅读理解的晚期阶段，从而支持语义干扰假说。本研究的研究问题明确，实验设计较合理，文章撰写清晰。但是尚有以下问题需要作者进一步澄清。

回应：非常感谢审稿专家对文章内容的肯定，也非常感谢专家所提的宝贵意见，这为我们修改论文提供了新的指南和方向。针对专家所提问题，我们已经逐条进行了回复和修改，文中修改的部分用蓝色字体进行了标识。

意见 1：本研究主要探索背景的语义和语音信息对阅读理解的干扰作用，但是不知考察这一问题的理论背景是什么。即，探索这一问题对于我们理解阅读理解的认知机制的新贡献是什么？

回应：非常感谢审稿专家的提问。根据审稿专家建议，我们重点补充和梳理了“探索背景音的语义和语音信息对阅读理解的干扰作用”的理论背景和研究创新点(见正文 P15-19)。具体内容如下：

阅读是人们获取知识的最重要的途径之一，在阅读过程中往往伴随一定的背景音，如新闻广播、同事的聊天声、室友打电话的声音等。研究者将言语类背景声音会对正在进行的认知活动产生干扰的现象称之为无关言语效应(Irrelevant Speech Effect)。虽然有不少研究者关注该现象，也进行了不少研究，但是，关于背景音干扰阅读的内在机制，还有许多重要问题(比如背景音干扰阅读的关键性因素是什么、背景音如何干扰阅读进程等)尚未解决 (Baker & Madell, 1965; Martin et al., 1988; Salamé & Baddeley, 1989; Armstrong et al. 1991; Oswald et al., 2000; Venetjoki et al., 2006; Haka et al., 2009; Haapakangas et al., 2011; Marsh et al., 2014; Hyönä & Eklholm, 2016; Yan et al., 2017; Vasilev et al., 2018; Vasilev, Liversedge, et al., 2019)。关于无关言语效应的产生，研究者从不同的角度提出了不同的解释，其中主要的两种理论是语音干扰假说和语义干扰假说。

语音干扰假说(Phonological-Interference Hypothesis, Salamé & Baddeley, 1982)认为由于

语音会自动进入语音回路，干扰视觉呈现项目的编码和复述，从而产生无关言语效应。而语义干扰假说(Semantic-Interference Hypothesis, Martin, Wogalter, & Forlano, 1988)认为无关言语的语义属性能干扰正在阅读的文本语义表征的建立，从而对阅读产生干扰作用，因而他们认为造成无关言语的干扰效果是言语的语义属性，而不是语音特征。

究竟是语音起干扰作用还是语义起干扰作用，抑或是二者同时起作用？关于这一点还没有定论。另外，已有相关研究采用的语料主要集中在拼音文字，研究结论能否推广至中文阅读尚未可知。汉字和拼音文字属于两种不同的书写系统，其结构单元和结构方式都与拼音文字明显不同，拼音文字属于表音文字，存在形-音的对应规则，语音在词汇通达中具有十分重要的作用，而汉语属于表意文字，字形的表意作用大于拼音文字，它没有明显的形-音对应规则，语音在汉字语义通达中的作用有限(彭聃龄, 2004)，在拼音文字中探讨的无关言语效应在汉语阅读中是怎样的存在，中文阅读的相关研究极少。虽然何立媛等(2015)发现无关言语对中文篇章阅读会产生影响，但是他们的研究并没有进一步探讨究竟是语音成分还是语义成分在无关言语效应中起到关键作用。基于此，本研究进一步探讨中文阅读中无关言语效应的认知加工机制，区分出语音和语义所起到的具体作用。

尤其是，在语音干扰假说和语义干扰假说提出之初，研究者在实验中只记录了被试的整体阅读时间和阅读理解的正确率。这些指标无法反映被试阅读过程中的认知加工过程。因此，研究者只是从整体上来推测背景音对阅读产生了干扰，但是，这种干扰发生的具体机制无从得知。然而，眼动追踪技术能够实时记录被试的阅读过程，可以为深入地分析阅读过程提供更为精确的技术和指标(闫国利等, 2013; Rayner, 2009)。因此，采用眼动追踪技术可以实时记录被试的阅读过程。通过对眼动指标的分析，可以分离出语音或者语义对人类阅读的干扰，进而验证语音干扰理论和语义干扰理论的正确性。因此，本研究结合眼动分析中的指标，探讨无关言语对中文阅读过程的影响。

此外，在中文阅读中，究竟在什么条件下无关言语才会起作用也是一个值得探讨的问题。部分研究得到无关言语对阅读的影响并不显著，如 Boyle 和 Coltheart(1996)发现无论有意义背景音还是无意义背景音，均未对句子判断的正确率和反应时产生影响；而在眼动研究中，Hyönä和 Eklholm (2016)的实验 1 也得到无意义背景音(外语)和有意义背景音(母语)对句子的眼动阅读没有产生明显的干扰作用，目前的听觉干扰理论不能对此进行解释。研究者认为阅读难度可能对阅读产生一定影响(Rayner et al., 2012; 闫国利等, 2008)。Rayner 等人(2012)认为对于熟练阅读者而言，阅读中的许多过程几乎是自动化的，这些过程包括熟悉词汇的识别、新词的解码等。对于已有十几年阅读经验的大学生而言，阅读普通句子极有可能是相对自动化、不怎么耗费认知资源的过程；然而，当阅读难度大的文本时，人们进行较高级的认知加工(例如正确语法结构的建构、词汇语义联系的建立、文本与读者已有经验的融合等)通常会有障碍，阅读速度也会明显下降。闫国利等(2008)发现阅读简单材料的知觉广度显著大于阅

读困难材料；并且，阅读材料的难度会使读者的有效视觉范围缩小，进而降低被试的知觉广度。研究者认为，当阅读较难的材料时，理解者需要花费更长的时间、更多的认知资源来进行难材料的词汇加工，相对的，对该注视词的边缘词汇的注意资源分配就会相应减少，这就导致了阅读者有效视觉范围的缩小，使阅读的知觉广度变小。另外，一项研究背景音乐对词汇记忆的影响的结果显示，任务难度是影响记忆任务中无关言语效应的重要因素，任务难度的变化可以改变视听交互过程中听觉刺激对视觉刺激的影响(高淇, 白学军, 2018), 是否在阅读任务中也存在同样的现象, 目前还不得而知。因此, 在不同难度阅读材料条件下探究无关言语的作用机制就显得十分必要。

综上所述, 本研究对理解中文阅读中无关言语效应的认知机制的研究新贡献在于:

一是从研究选题来看, 本研究探讨了中文(表意文字)阅读中无关言语效应的影响机制。本研究通过比较不同类型的无关言语对句子和篇章阅读的影响, 证实了语义是引发无关言语效应的关键因素, 结果支持语义干扰假说。

二是从研究技术来看, 本研究采用眼动追踪技术, 揭示了在中文阅读中无关言语的语义干扰发生在晚期的词汇加工和语义整合阶段。相比行为实验, 眼动追踪技术能够实时记录被试的即时阅读加工过程, 并且还可以提供多种指标, 可以为具体和深入地探讨无关言语效应的认知机制提供条件。

三是从研究深度来看, 本研究进一步考察了中文阅读中产生无关言语效应的必要条件。本研究不但证实在中文阅读条件下也存在无关言语效应, 并且进一步确定了只有在阅读任务较困难、需要花费较多认知资源时, 无关言语效应才会显现。

意见 2: 本研究设置了三个实验条件: 无声条件、无意义语音、有意义语音。目标任务是中文阅读理解, 无意义语音是西班牙语; 作者旨在通过比较“无声条件”和“无意义语音”条件, 探索背景音的语音信息的干扰作用。但是, 考察背景音的语音干扰作用时, 用西班牙语作为无意义语音条件是否合适? 原因在于, “无意义语音”的理想条件是, 背景音与阅读内容的语音类似、同时又不传达语义信息(或者语义无关)。然而, 在本研究中, 西班牙语与当前阅读的中文内容的语音(如音位、音系等)是否相同、在什么程度上相同, 作者并没有做详细说明和介绍。如果西班牙语的语音和中文的语音无相同之处, 本研究的西班牙背景音可能等同于一般噪声, 因此也不能考察语音干扰的作用。

回应: 非常感谢审稿专家的建议。我们也赞同审稿专家的意见, 西班牙语的语音和汉语的语音确实存在一定的差异, 如: 汉语是声调语言而西班牙语是非声调语言, 汉语是单音节语言, 而西班牙语是多音节语言。但是, 两种语言之间也存在许多共同点, 比如二者都属于拼读型的语音, 部分辅音和元音的发音相似等(赵士钰, 1999; 张燕, 2019)。

另外, 语音干扰假说(Salamé & Baddeley, 1982, 1987)认为干扰的发生是由于无关的语音

进入了语音回路与视觉信息的语音编码产生了混淆,而这与语音本身的特殊属性无关。同时,西班牙语是一种语言,与普通噪音之间存在本质差异,根据 Baddeley 和 Hitch(1974, 1994)的工作记忆模型,作为语言的语音可以进入工作记忆中的语音回路,而一般意义上的噪音,其不具备以上特性。

除此之外,以往的系列研究结果发现,即使无关言语和作为主要任务的视觉刺激之间在语音方面的相似性增大时,其干扰效应并没有相应的增加(Jones & Macken, 1995; LeCompte & Shaibe, 1997; Larsen, Baddeley, & Andrade, 2000)。因而,研究者往往认为“无关言语和主要任务在语音上相似”是非必要的(Vasilev, Liversedge, et al., 2019)。在以往有关无关言语效应的研究中,研究者们使用的语种很多,并没有刻意去使用某一类具体的语种,截止目前,研究者使用过阿拉伯语(Baddeley & Salamé, 1986; Salamé & Baddeley, 1987)、德语(Colle & Welsh, 1976)、俄语(Klatte, Lee, & Hellbruck, 2002; Hyönä & Eklholm, 2016)、日语(Ellermeier & Zimmer, 1997)、汉语(Vasilev, Liversedge, et al., 2019)等在内的一系列不同言语作为无意义背景音来探讨无关言语效应。

不过,我们觉得审稿专家的意见也很有道理,语种间的差异性是否会影响中文阅读中的无关言语效应,这是一个值得进一步研究的问题,再次感谢审稿专家的建议。同时,根据专家意见,我们已在文中实验 1 的材料部分增加了相应的表述(见正文 P20, 第 3 段):

“值得注意的是,西班牙语和汉语的语音虽然存在一定的差异,如:汉语是声调语言而西班牙语是非声调语言,汉语是单音节语言而西班牙语是多音节语言,但是二者也存在一定的共同点,如:二者都属于拼读型的语音,部分辅音和元音的发音相似(赵士钰, 1999; 张燕, 2019)。更为重要的是,研究者往往认为作为无意义背景音条件所使用的实际语言并不是至关重要的(Vasilev, Liversedge, et al., 2019)。因为语音干扰假说(Salamé & Baddeley, 1982, 1987)认为干扰的发生是由于无关的语音进入了语音回路与视觉信息的语音编码产生了混淆,而这与语音本身的特殊属性无关。事实上,无关言语和作为主要任务的视觉刺激之间语音上相似性更大时并不会使干扰增加(Jones & Macken, 1995; Larsen, Baddeley, & Andrade, 2000; LeCompte & Shaibe, 1997)。在以往的研究中,研究者使用了包括阿拉伯语(Baddeley & Salamé, 1986; Salamé & Baddeley, 1987)、德语(Colle & Welsh, 1976)、俄语(Klatte, Lee, & Hellbruck, 2002; Hyönä & Eklholm, 2016)、日语(Ellermeier & Zimmer, 1997)、汉语(Vasilev, Liversedge, et al., 2019)等在内的一系列不同言语作为无意义背景音来探讨无关言语效应。在本研究中,我们采用了西班牙语作为无意义语音条件。”

意见 3: 对于当前阅读任务的“认知负担”的中间调控作用,作者只是通过比较实验 1 和实验 2 的结果,推测而来的。实际上,实验 1 和实验 2 的被试群体不同,如果两个实验得出不同的结果模式,一方面可能是由于两个实验的阅读任务负担不同造成的;另一方面也可能是

由于被试个体差异等无关因素造成的。因此，实验 2 并不能像作者指出的那样揭示阅读任务“认知负担”的中介调控作用。

回应：非常感谢审稿专家的建议。在专家的提醒下，关于“认知负担”的中间调控作用我们进行了进一步的思考，觉得在前言的假设部分直接使用“认知资源”等相关表述的确不是很合适，比较笼统和抽象，相关表述已经修改(详见正文 P18，第 2 段；P19，第 4 段)。

同时，通过查阅文献，可知材料难度对阅读确实会产生一定的影响(Rayner et al., 2012; 闫国利等, 2008)。本研究尝试探讨不同难度的阅读材料中背景音所产生的影响，课题组曾经尝试采用 2×3 的被试内设计，自变量为背景音条件(3 水平)和阅读难度(2 水平)，发现存在两个比较严重的问题：一是被试的阅读材料过多，当简单句和复杂句各 84 句时，总的句子达到了 168 句，尤其是复杂句需要比较多的认知资源加工才能进行理解，在有外在声音干扰的条件下，被试很难坚持去认真理解每一个句子，一定程度上影响了实验的效度；二是实验时间过长，在眼动阅读中，除了阅读句子外，还需要不断进行注视点校正，在校正过程中需要被试集中注意去盯准不断变换位置的黑点，因而实验需要被试高度集中注意力达到至少 70 分钟以上，同时这一过程也使得被试的眼睛非常疲惫，存在较为严重的疲劳效应。

在权衡以后，最终决定采用单因素 3 水平的实验设计，分两个实验进行(研究设计的参考文献：高淇，白学军，2018；何立媛等，2015。高淇和白学军(2018)探讨了无背景音乐、中文流行背景音乐和英文流行背景音乐对大学生中英字词记忆的影响，分两个实验进行，实验 1 使用熟悉中英名词进行即时回忆任务，实验 2 增加了任务难度，使用不熟悉中英字词进行即时再认任务。实验 1 和实验 2 得到了不同的研究结果，他们认为听觉输入语言熟悉度和视觉任务难度都会影响无关言语效应的大小，听觉语言熟悉度的影响受视觉任务难度高低的制约)。与此同时，特别对被试、实验过程进行控制，以保持实验的同质性。具体表现为：一是在被试控制上，选取同一年龄段的学生(18-20 岁)，来自相同的学校，被试的阅读时间、执行功能测试结果并无显著差异；二是在实验过程控制上，采用同一台仪器、同一实验室地点、实验的操作程序相同，以最大程度的控制无关因素。在专家的提醒下，在正文中的相应部分做了更为详细的说明(见正文 P25，第 4 段；P26，第 4 段)。

(正文 P18，第 2 段)“以往研究表明，阅读难度可以对阅读产生一定影响，Rayner 等人(2012)认为对于熟练阅读者而言，阅读中的许多过程几乎是自动化的，这些过程包括熟悉词汇的识别、新词的解码等，普通的句子阅读对于已有十几年阅读经验的大学生而言可能是相对自动化、不费力的过程，然而，当文本难度大、阅读存在困难(很难读懂)时，较高级别的加工(例如正确语法结构的建构、词汇语义联系的建立、文本与读者已有经验的融合等)通常会有障碍，这时阅读速度就会大大降低。闫国利等(2008)发现阅读材料的难度会影响知觉广度，阅读简单材料的知觉广度显著大于阅读困难材料，他们认为占用加工资源较多的难材料

进行词汇通达所需时间更长,在当前注视词的加工过程中由于对难易材料分配的注意资源不同,难材料分配的资源较多,对当前注视词边缘的词的注意资源分配减少,这就导致了阅读者有效视觉范围的缩小,使阅读的知觉广度较小。另外,一项研究背景音乐对词汇记忆影响的结果显示,任务难度是影响记忆任务中无关言语效应的重要因素,任务难度的变化可以改变视听交互过程中听觉刺激对视觉刺激的影响(高淇,白学军,2018),是否在阅读任务中也存在同样的现象,目前还不得而知。因而探讨在不同难度的阅读材料中,无关言语会产生怎样的影响,这一问题值得研究。”

(正文 P19, 第 4 段)“在实验中控制阅读的难度,考察被试在不同背景音下进行不同难度句子阅读的情况,假设在容易的简单句阅读中,无关言语对阅读将不会产生干扰作用,即被试在不同背景音条件下的阅读情况将不会存在差异;而假设在难度大的复杂句阅读中,无关言语对阅读将会产生干扰作用,即被试在不同背景音条件下的阅读情况将会存在差异。”

(正文 P25, 第 4 段)“实验 2 与实验 1 的被试来自相同的学校,两批被试的阅读时间($p=0.87$)、执行功能测试结果(包含 Stroop 任务($p=0.85$)、数字转换任务($p=0.49$)、色点刷新任务($p=0.63$)三种测验)差异不显著。”

(正文 P26, 第 4 段)“与实验 1 同一台仪器、同一实验室地点、实验的操作程序相同等”

相关参考文献:

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. J. (2012). *Psychology of reading* (2nd ed.). New York, USA: Psychology Press.

闫国利, 伏干, & 白学军. (2008). 不同难度阅读材料对阅读知觉广度影响的眼动研究. *心理科学*, 31(06), 1287-1290.

何立媛, 黄有玉, 王梦轩, 孟珠, 闫国利. (2015). 不同背景音对中文篇章阅读影响的眼动研究. *心理科学*, 38(6), 1290-1295.

高淇, 白学军. (2018). 中英流行背景音乐对大学生中英词汇记忆的影响. *心理学报*, 50(1), 1-8.

意见 4: 作者在结果部分呈现了“整体分析”和“局部分析”两种结果。在此之前作者应该清楚说明每种分析具体指的什么, 每种分析的含义、及其能揭示的问题是什么。

回应: 非常感谢审稿专家的建议。根据专家意见, 在实验 1 的结果部分对相关内容进行了补充和修改, 具体表述如下:

(见正文 P21, 第 4 段)“实验指标分析根据目标兴趣区的大小分为两类, 一类是以句子或篇章为兴趣区的整体分析, 一类是以字或词为兴趣区的局部分析。在实验 1 中, 整体分析以整个句子作为目标区, 所分析指标包括总阅读时间(Total reading time, 指阅读者阅读文本的所有注视点的时间的总和)、平均注视时间(mean fixation duration, 指所有注视点的持续时

间的平均值)、注视次数(number of fixations, 指文本被注视的总次数)和回视次数(regression count, 从一个词语由右向左的眼跳次数), 其中, 总阅读时间、平均注视时间反映了句子阅读的整体加工, 注视次数反映了阅读材料的认知加工负荷, 回视次数反映了对句子意义进行整合的过程(闫国利 等, 2013)。局部分析以句子中的双字词作为目标词进行分析(何立媛 等, 2015), 所分析指标包括首次注视时间(First fixation duration, 指首次通过阅读中某兴趣区内的首个注视点的注视时间)、凝视时间(Gaze duration, 指从首次注视点开始到注视点首次离开当前兴趣区之间的持续时间)、总注视时间(Total fixation duration, 指兴趣区内所有注视点的时间的总和)和回视路径时间(Regression path reading time, 指从某个兴趣区的第一次注视开始, 到注视点落到该兴趣区右侧的区域为止, 在这之间的所有注视点持续时间的总和), 其中, 首次注视时间是反映词汇通达早期阶段特征的有效指标, 对语音特性、词汇特征(词频、词长、词的可预测性等)、正字法特征等反应敏感, 凝视时间也是使用广泛的反映词汇通达早期阶段的指标, 对前词汇特征、词汇特征都反应敏感, 总注视时间是反映词汇总体加工的重要指标, 而回视路径时间不仅可以反映词汇通达的加工过程, 而且还能反映后期语义加工和句义整合的过程(闫国利 等, 2013)。”

意见 5: 每个实验是单因素三水平的实验设计, 作者是用什么方法来控制多重比较的?

回应: 非常感谢审稿专家的提问。对于单因素三水平的实验设计, 如果采用传统的方差分析(ANOVA)时, 当自变量水平超过 2 个时, 研究者就需要进行事后多重比较, 比较到底是哪两组之间存在差异, 常用的方法有 LSD 法、Sidak 法、Bonferroni 法、Dunnnett-*t* 检验、Tukey 法、SNK 法、Duncan 法等。然而, 本研究使用的是 R 统计软件(R Core Development Team, 2016)和 lme4 程序包, 采用线性混合效应模型(linear mixed-effects models, LMMs; Baayen et al., 2008)对数据进行统计分析。

R 统计中进行单因素的方差分析和多重比较的方式有两种, 一种是直接使用方差分析, 再采用 LSD 检验、Dunnnett 检验、Tukey 检验等方法进行多重比较(通过运行相应代码实现), 另一种是采用线性回归模型进行分析, 值得一提的是, 不论 *t* 检验、方差分析、还是非参数检验, 其实都是线性回归的特殊形式。本研究采用了拟合线性回归模型的方式, 即使用 sliding contrast 的计算方法, 设计 contrast matrix, 运行拟合的线性回归模型 `depvar.lmer = lmer(depvar ~ Noise + (1+ Noise|Participant) + (1+Noise|Item), datafile)`, 得到有意义语音、无意义语音、无声条件之间两两比对的关系, 实现多重比较的目标。

为什么要采用线性混合效应模型来分析数据? 这是由于相较于传统的方差分析(ANOVA), 采用 LMMs 分析眼动数据具有一定优势。因为在采用 ANOVA 对眼动数据进行分析时, 需要分别进行被试分析(F1 检验)和项目分析(F2 检验), 而在实际眼动数据分析过程中常存在 F1 和 F2 检验结果不一致的情况, 这让研究者难以对实验结果进行合理的解释,

线性混合模型在计算过程中采用最大随机效应结构(maximal random effects structure), 将被试和项目定义为交叉随机效应(crossed random effects)同时纳入模型进行计算, 使结果更为稳定和统一, 同时也避免了当只有 F1 或 F2 检验显著时犯 I 型错误的概率显著增加的问题 (Baayen et al., 2008)。其次, 线性混合效应模型能将所有的原始数据纳入到模型中进行统计, 无需对数据进行平均后再作比较, 其数据利用率更高。采用 R 统计软件进行眼动数据分析的研究诸多, 如: 刘志方等(2020)、张慢慢等(2020)、Vasilev 等(2019)、Zhao 等(2019)人的研究等等。

相关参考文献:

- R Development Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- 刘志方, 全文, 张智君, 赵亚军(2020).语境预测性对阅读中字词加工过程的影响: 眼动证据. *心理学报*, 52(9),1031-1047.
- 张慢慢, 臧传丽, 徐宇峰, 白学军, 闫国利(2020).快速与慢速读者的中央凹加工对副中央凹预视的影响. *心理学报*,52(9), 933-945.
- Vasilev, M. R., Liversedge, S. P., Rowan, D., Kirkby, J. A., & Angele, B. (2019). Reading is disrupted by intelligible background speech: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*. 45(11), 1-93.
- Zhao, S., Li, L., Chang, M., Xu, Q., Zhang, K., Wang, J. X., & Paterson, K. B. (2019). Older adults make greater use of word predictability in Chinese reading. *Psychology and Aging*, 34(6), 780-790.

意见 6: 25 页倒数第二段中“整体分析中有意义语音条件与无声条件相比, 在总阅读时间、回视次数、阅读速度上差异显著, 这与实验 2 及前人的研究(何立媛等, 2015; Vasilev, Liversedge, et al., 2019)结果一致, 总阅读时间、阅读速度是反映篇章整体加工的指标, 而回视次数可以反映对篇章意义进行整合的过程, 有意义背景条件下的整体阅读时间明显更长, 阅读速度更慢、回视次数更多, 说明有意义背景音干扰了篇章阅读的整体加工和意义整合, 一定程度上支持了语音干扰假说”。“一定程度上支持了语音干扰假说”应该是“一定程度上支持了语义干扰假说”。

回应: 非常感谢审稿专家的建议, 的确应该是“一定程度上支持了语义干扰假说”, 已经在文中进行修改(见正文 P33, 第 2 段)。

.....

审稿人 2 意见：

研究问题有意义，研究方法恰当，写作规范，建议修改后酌情发表。意见如下：

回应：非常感谢审稿专家对文章内容的肯定，也非常感谢专家所提的宝贵意见，这为我们修改论文提供了新的指南和方向。针对专家所提问题，我们已经逐条进行了回复和修改，文中修改的部分用蓝色字体进行了标识。

意见 1：有意义背景音条件包含语义属性以及语音属性，并且，背景音的语义信息需通过其语音属性传递给被试。那么，本文的研究结果是否可以完全排除语音干扰的可能性？

回应：非常感谢审稿专家的建议。有意义背景音条件包含语义属性以及语音属性，如何排除有意义背景音条件下语音属性的作用？

本研究主要通过条件间的对比进行推断，我们将有意义背景音条件视为既有语音属性又有语义属性的条件，无意义背景音条件视为只有语音属性但无语义属性的条件(无意义语音条件具有语音属性，但被试不能进行语义通达)，无声条件是既无语音属性也无语义属性的条件。如果将无声条件作为基线条线，无意义背景音条件与无声条件相比所产生的差异可视作为语音属性产生的作用；如果将无声条件作为基线条线，有意义背景音条件与无声条件相比所产生的差异可为语音属性和语义属性产生的作用；如果将无意义背景音作为基线条线，有意义背景音条件与无意义背景音之间存在差别，则可以视为语义属性产生的作用。

在无关言语效应的研究中，研究者并不能选取到纯语义干扰的条件。因此，我们借鉴前人研究，采用间接的方式来说明问题(如 Yan et al., 2017; Hyönä & Ekholm, 2016; 何立媛等, 2015)。

相关参考文献：

Yan, G., Meng, Z., Liu, N., He, L., & Paterson, K. B. (2017). Effects of irrelevant background speech on eye movements during reading. *The Quarterly Journal of Experimental Psychology*, 71(6):1-6.

Hyönä J., & Ekholm, M. (2016). Background speech effects on sentence processing during reading: An eye movement study. *PLOS ONE*, 11(3), Article e0152133.

何立媛, 黄有玉, 王梦轩, 孟珠, 闫国利. (2015). 不同背景音对中文篇章阅读影响的眼动研究. *心理科学*, 38(6), 1290-1295.

意见 2：三个实验使用的被试数不一致，建议对被试数量做出解释说明。

回应：非常感谢审稿专家的建议。根据 Faul 等(2007)所述的方法，用 G*Power 软件计算所需被试数量，达到中等程度的效应量($d=0.25$)和统计检验力($1-\beta=0.8$)所需被试为 29 人。在投稿的时候，3 个实验只有实验 2(有效数据 27 个)未达到预期。因为整体实验是在 2019 年末完成的，原计划在 2020 年 3 月补齐数据后重新分析数据。只是由于疫情缘故无法补数据，

所以在投稿的时候只呈现了 27 名被试的数据。考虑到实验应有的效应量和统计力，依据审稿专家的意见，我们现将实验 2 增加到了和实验 1 相同的被试数量，结果发现，实验 2 目前的研究结果与增加被试前的结果模式相一致。具体情况详见正文实验 2 被试(正文 P25，第 4 段)和结果分析(正文 P26-28)部分。

相关参考文献：

Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

意见 3：文中 LMM 部分尚未清楚报告随机效应，建议明确报告具体的 random-effects structure，或者报告完整的 model。

回应：非常感谢审稿专家的建议。本研究中 LMM 的完整 model 是：depvar.lmer = lmer(depvar ~ Noise + (1+ Noise|Participant) + (1+Noise|Item), datafile)，random-effects structure 是 (1+Noise|Participant) + (1+Noise|Item)，已经在文中增加相应的文字表述：

(见正文 P22，第 2 段)“在本研究的模型中，声音为固定因素，同时考虑被试和项目两个随机因子，在每个随机因子上，既考虑了随机截距，又考虑了固定效应的随机斜率。”

意见 4：文末作者提出类似的应用实例，即背景音对驾驶员的分心想象。本文背景音对阅读的干扰效应(语义干扰效应更大)是否可推广到驾驶员驾驶？即有意义的背景音比无意义的背景音对驾驶造成更大的干扰？

回应：非常感谢审稿专家的建议。背景音下的阅读与背景音下的驾驶行为存在一定的相似性，两者都是较为常见的“双任务”。杨萌等人(2011)认为在驾驶行为中，背景音乐的歌词可能通过增加认知负荷占用更多的视觉注意资源。当驾驶期间出现言语干扰时，驾驶员会不自觉地形成了一个“双任务”，这会增加其认知负荷(Blanco et al., 2006)，并直接影响驾驶行为(Lee, Lee, et al., 2007)，表现为减弱驾驶员对汽车的控制能力(Rakauskas et al., 2004)、减慢行驶的速度(Strayer & Drews, 2004)、延缓或打断对道路信息的加工(Lee, Caven, et al., 2001)。但是从背景音对驾驶行为的相关研究来看，不能推断出有意义的背景音比无意义的背景音对驾驶造成更大的干扰，考虑到研究在表达上的严谨性，修改稿中去掉了由背景音对阅读的干扰效应(语义干扰效应更大)推及到背景音下的驾驶行为的相关表述，并将相关文字进行了修改，具体见文中蓝色字体(正文 P37，第 3 段；P38，第 1 段)。

相关参考文献：

Blanco, M., Biever, W. J., Gallagher, J. P., & Dingus, T. A. (2006). The impact of secondary task cognitive processing demand on driving performance. *Accident Analysis & Prevention*, 38(5), 895-906.

- Lee, Y. C., Lee, J. D., & Boyle, L. N. (2007). Visual attention in driving: the effects of cognitive load and visual disruption. *Human Factors*, 49(4), 721-733.
- Rakauskas, M. E., Gugerty, L. J., & Ward, N. J. . (2004). Effects of naturalistic cell phone conversations on driving performance. *Journal of Safety Research*, 35(4), 453-464.
- Strayer, David, L. , Drews, & Frank, A. (2004). Profiles in driver distraction: effects of cell phone conversations on younger and older drivers. *Human Factors*, 46, 640-649.
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.
- Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159-177.
- Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, 17(3), 134-141.
-

第二轮

审稿人 1 意见:

在第一轮提出的大部分评审意见，作者都进行了较好地回应。但是，尚有以下问题需进一步澄清。

回应: 非常感谢审稿专家提出的宝贵意见。针对专家所提问题，我们已经逐条进行了回复和修改，文中修改的部分用蓝色字体进行了标识。

意见 1: 还是上次提到的问题，用“西班牙语”作为无意义语音条件不太合适。与汉语相比，西班牙语在声调、辅音、元音等方面存在很大差异，听者可能只把其作为非语音信息来处理，不进行进一步的语音编码和识别，因此其语音干扰作用有限。因此，本研究并不能排除语音干扰假设。

回应: 非常感谢审稿专家的意见。很抱歉关于此问题我们之前的论述可能没有阐明得很清楚，对此我们进行了深入反思，并补充了新的数据和文献对该问题进行论证。

在以往的无关言语效应研究中，研究者常使用无意义的无关言语(如，听不懂的外语、顺序被打乱的母语)作为载体探讨语音的作用(Baddeley & Salamé, 1986; Salamé & Baddeley, 1987; Oswald et al., 2000; Hyönä & Eklholm, 2016; Yan et al., 2017; Vasilev, Liversedge, et al., 2019; Meng et al., 2020)，这些研究者认为无意义的无关言语具有言语的语音特征但不能进行语义通达，因而其对阅读所产生的影响，可以认为是语音属性产生的作用。作为无意义的无关言语至少需要满足三个必要条件：一是必须具有言语的语音特征；二是不能进行语义通达；三是与主要任务无关(Vasilev et al., 2018)。

本研究中使用西班牙语作为无意义的无关言语，我们需要论证其是否满足以上三个条

件。对于没有学习过西班牙语的汉语母语被试而言，其不能理解西班牙语所表达的意义(第二个条件满足)，同时西班牙语和主要的阅读任务无关(第三个条件满足)，现在关键是需要论证西班牙语是否满足第一个条件，即在实验过程中被试是否会对其进行语音层面的加工，还是只将其视为非语音信息来处理。针对此问题，我们收集了两个方面的数据。

首先，通过追溯实验过程中被试的表现，发现被试将西班牙语视为了一种“听不懂的外语”。事实上，为了保证西班牙语能够满足本实验的要求，在做实验的过程中我们进行了初步的操纵检验。比如，在每个实验结束后，主试会随机抽取部分被试(3个实验每个实验随机抽取了20人，共60人)询问一些有关实验实施情况的问题，其中有一个问题是“在阅读过程中您听到了几种声音条件”，结果有58人回答听到了“无声、中文和听不懂的外语”三种类型的声音，只有2人的答案是“不记得了”。绝大部分被试(除了2个回答不记得的被试)将西班牙语视为了一种“听不懂的外语”。也就是说，被试能够意识到它是一种语言，但是无法听懂具体的内容，即被试可加工其语音，但是无法通达语义。由于在完成初稿时限于篇幅，且考虑到沿用的是以往研究者操控材料所采用的方法(Yan et al., 2017; Vasilev, Liversedge, et al., 2019; 何立媛等, 2015)，因而没有在文中进行详细描述。

同时，为了进一步确保实验材料的有效性，我们重新对实验用的声音刺激进行了评定，结果发现：对于中文母语者而言，西班牙语被认为是一种不能进行语义加工的外语，与中文(有意义语音)和噪音(非语音)存在显著差异。具体评定如下：首先，我们选取了三类声音材料：中文音频(10个，从中文实验材料中随机抽取)、西班牙语音频(33个，所有西班牙语的实验材料)、噪音(10个，包括白噪音和粉红噪音，如语音频谱噪声(Vasilev, Liversedge, et al., 2019)、瀑布声(何立媛等, 2015)、雨声(何立媛等, 2015)、飞机噪音(Sörqvist, 2010)、火车汽笛声(Hygge, 2003)、交通噪音(Hygge, 2003)等)。然后，将所有声音刺激按照随机的顺序进行打乱。要求评定者评定“所听音频是否为一种语言”以及“能否对该音频进行语义加工”，他们需要依据自己的确认程度进行7点评分，1表示评定者非常确认所听音频“不是语言/不能进行语义加工”，7表示非常确认所听音频“是语言/能进行语义加工”。整个评定过程持续30分钟左右。

共有25名大学生参加了此次的评定实验。结果发现：(1)在判断声音刺激是否为语言方面：所有被试判断中文语音为语言，中文条件下的评定得分： $M=6.80$ ； $SD=0.24$ ；所有被试判断西班牙语为语言，西班牙语条件下的评定得分： $M=6.65$ ； $SD=0.34$ ；所有被试判断噪音为非语言，噪音条件下的评定得分： $M=1.32$ ； $SD=0.49$ 。通过方差分析发现三种声音的得分之间存在显著差异($F=1754.32$, $p<0.001$)，事后检验得到，被试在判断西班牙语和中文为语言时的得分没有显著差异($p=0.16$)，西班牙文条件和噪音条件相比差异显著($p<0.001$)，中文条件和噪音条件相比差异显著($p<0.001$)。结果说明，被试认为西班牙语和中文一样是一种语言，但噪音不是一种语言。(2)在判断能否对声音刺激进行语义加工的方面：中文条件下的评定

得分为： $M=6.83$ ； $SD=0.39$ ，西班牙文条件下的评定得分为 $M=1.11$ ； $SD=0.22$ ，噪音条件下的 $M=1.18$ ； $SD=0.17$ 。通过方差分析发现三种声音的得分之间存在显著差异($F=3568.52$, $p<0.001$)，事后检验得到，西班牙文条件和中文条件的差异显著($p<0.001$)，西班牙文条件和噪音条件相比差异不显著($p=0.39$)，中文条件和噪音条件相比差异显著($p<0.001$)。结果说明，被试认为西班牙语和噪音一样都不能进行语义加工，而中文可以进行语义加工。材料评定的结果和实验询问被试的结果相一致。通过评定可知，被试认为西班牙语是一种语言，但不能进行语义通达，它和噪音(非语音信息)有显著差别。中文和西班牙文在语音上存在一定的共同点，如：二者都属于拼读型的语音，部分辅音和元音的发音相似等(赵士钰，1999；张燕，2019)。因而我们认为被试有对西班牙语进行语音层面的加工，其满足作为无意义无关言语的第一个条件。

不可否认的是，西班牙语和汉语的语音确实存在一定的差异，如：汉语是声调语言而西班牙语是非声调语言、部分辅音和元音的发音存在差异等，这种差异是否能显著影响无关言语效应中语音所产生的干扰作用？从以往的系列研究结果发现，当无关言语和作为主要任务的视觉刺激在语音上的相似性增加时，无关言语的干扰效应并不会相应增加(Jones & Macken, 1995; LeCompte & Shaibe, 1997; Larsen, Baddeley, & Andrade, 2000)，这可能是由于无关言语作为任务完成中的一种背景音，是被试需要忽略的部分，因而其与主要任务之间在语音上的相似性所能产生的作用十分有限。同时，根据 Baddeley 和 Hitch(1974, 1994)的工作记忆模型，语音成分能够进入工作记忆中的语音回路，而普通的噪声则不能进入。语音干扰假说(Salamé & Baddeley, 1982, 1987)认为干扰的产生是因为无关言语可以进入语音回路，而不是由言语本身的特殊属性所引起。这些研究结果和理论学说削弱了考虑保证“无关言语和作为主要任务的视觉刺激之间语音上相似”的必要性。因而有研究者认为作为无意义背景音条件所实际使用的具体语言并不是至关重要的(Vasilev, Liversedge, et al., 2019)。

在以往的研究中，研究者曾使用过包括阿拉伯语(Baddeley & Salamé, 1986; Salamé & Baddeley, 1987)、德语(Colle & Welsh, 1976)、俄语(Klatte, Lee, & Hellbruck, 2002; Hyönä & Eklholm, 2016)、日语(Ellermeier & Zimmer, 1997)、汉语(Vasilev, Liversedge, et al., 2019)、维语(Meng et al., 2020)等在内的不同语言(外语)作为无意义语音条件来探讨无关言语效应。在最近的一项研究中，Meng 等人(2020)在探讨文本加工中认知任务对无关言语效应的调节作用时，采用的是以中文为母语的被试，无意义无关言语条件用的是维语(该研究发表于 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2020 年第 10 期)；而 Vasilev, Liversedge 等人(2019)探讨不同类型的背景音对英文文本阅读的影响时，采用的是以英文为母语的被试，无意义的无关言语条件采用的是中文普通话(该研究发表于 *Journal of Experimental Psychology: Human Perception and Performance*, 2019 年第 11 期)。当然，以往也有研究者采用“顺序被打乱的母语”作为“无意义的无关言语条件”(Oswald et al., 2000; Yan

et al., 2017), 这种情况下, 背景音与阅读内容的语音类似且在句子层面不能通达语义信息, 然而, 它也存在不足之处, 一是在词的层面上还是存在语义通达(孟珠, 闫国利, 2018), 二是不按正常顺序呈现的母语不符合被试的使用习惯, 可能会存在新异性等无关因素的混淆。

总的来说, 目前还未发现较完美的无意义语音条件(即背景音与阅读内容的语音类似、同时又不传达语义信息), 这是在无关言语效应研究中需要去进一步思考和解决的问题, 我们也会根据专家的宝贵意见, 接下来对该问题进行更为深入的探讨, 再次感谢专家的提问。同时根据专家意见, 我们也意识到了应该更为详细地阐明“被试在听到西班牙语时有感知其为一种不能进行语义加工的语言”, 因而文中实验 1 的实验材料和实验程序部分增加了相应的表述:

(见正文 P24, 第 2 段) “25 名未参加正式实验的大学生对音频材料进行了评定, 在所评定的音频中, 除了中文和西班牙语外, 还增加了一类声音, 即噪音(如, 飞机噪音(Sörqvist, 2010)、交通噪音(Hygge, 2003)等), 通过三种声音之间的对比以检验本研究所选取的材料是否符合实验的要求。要求评定者评定“所听音频是否为一种语言”以及“能否对该音频进行语义加工”, 他们需要依据自己的确认程度进行 7 点评分, 1 表示评定者非常确认所听音频“不是语言/不能进行语义加工”, 7 表示非常确认所听音频“是语言/能进行语义加工”。结果得到: (1)关于语言的判断, 中文条件下的评定得分: $M=6.80$; $SD=0.24$, 西班牙文条件下的评定得分: $M=6.65$; $SD=0.34$, 噪音条件下的评定得分: $M=1.32$; $SD=0.49$, 通过方差分析事后检验得到西班牙文条件和中文条件的差异不显著($p=0.16$), 西班牙文条件和噪音条件相比差异显著($p<0.001$), 中文条件和噪音条件相比差异显著($p<0.001$), 说明被试认为西班牙语和中文一样是一种语言, 与噪音有显著区别。(2)关于语义加工的判断, 中文条件下的评定得分: $M=6.83$; $SD=0.39$, 西班牙文条件下的评定得分: $M=1.11$; $SD=0.22$, 噪音条件下的评定得分: $M=1.18$; $SD=0.17$, 通过方差分析事后检验得到西班牙文条件和中文条件的差异显著($p<0.001$), 西班牙文条件和噪音条件相比差异不显著($p=0.39$), 中文条件和噪音条件相比差异显著($p<0.001$), 说明被试认为西班牙语和噪音一样都不能进行语义加工, 而中文可以进行语义加工。通过评定可知, 被试认为中文和西班牙语都是一种语言, 中文可以进行语义加工, 但西班牙文不能进行语义通达, 它们和噪音(非语音信息)有显著差别。”

(见正文 P25, 第 2 段) “为确保数据的有效性, 在实验结束后, 要求被试回答关于实验情况的几个问题, 如“以前是否学习过西班牙语”“在实验过程中有听到几种声音条件”等, 以排除被试能听懂西班牙语或实验过程中声音未出现等无关因素。”

意见 2: 对于单因素被试内设计(单因素三水平), 最少被试量是 29 人。但是, 作者把实验 1 和实验 2 进行了实验间比较, 这相当于两因素混合实验设计(3×2), 29 人的被试量是否足够大?

回应：非常感谢审稿专家的提问。关于被试量的问题，可能是在回复审稿专家意见时没有说出具体的被试数，只是说了用 G*power 软件算出的最小被试量是 29 人，同时将实验 2 增加到了和实验 1 相同的被试数量，从而给专家造成了一种误解，即认为实验 1 和实验 2 的被试量是 29 人。需要说明的是，29 人只是关于实验被试量的最低参考标准，根据前人的相关研究(Yan et al., 2017; Hyönä & Ekholm, 2016)，我们在最低被试量的基础上还增加了一定量的被试，第一次投稿时实验 1 的被试量是 51 人，在第二轮投稿时将实验 2 增加到了和实验 1 相同的被试数量，实际上目前实验 1 和实验 2 的被试数分别都是 51 人。同时，实验 1 和实验 2 所采用的阅读材料分别是 84 个和 81 个句子，每种背景音条件下的句子阅读数量分别为 28 句和 27 句，以保证实验有良好的统计功效。

第三轮

审稿人 1 意见：

感谢作者再次认真详细地回答并澄清了我的问题。我没有进一步的问题。本研究通过几个实验层层递进、检验一个核心假说，有一定的理论贡献，建议发表。另外，作者在二审修改稿 P24 页增加的第二段的内容，没有必要，因此建议删除。原因在于，这段内容测验的目的在于证明“被试认为中文和西班牙语都是一种语言，中文可以进行语义加工，但西班牙语不能进行语义通达”，而这是显而易见不需要证明的。

回应：非常感谢审稿专家对文章内容的肯定。根据审稿专家的建议，删除了二审修改稿 P24 页第二段中增加的内容，并对文章的其他内容进行了检查，敬请编委斧正。

编委复审意见：

该论文经两轮审稿，作者目前已经较好地完成审稿人提出的问题，两审稿人均认为文章可以发表。我也通读了全文，该文有理论的支撑，假设清楚，研究数据也能够说明所关注的问题。一处小的修改建议：建议作者在前言中对研究假设的描述用文字叙述（而非以纲要性的方式，即 H0、H1 等）的方式来进行，以增强文章的可读性。

回应：非常感谢编委专家对文章的肯定以及所提的宝贵意见，我们已经依据建议去除了 H0、H1 等表述，对研究假设的描述改用于文字叙述的形式，同时对文章内容进行了再次检查，敬请主编斧正。

主编意见：同意外审和编委意见，建议录用。