

《心理学报》审稿意见与作者回应

题目：汉语复合词视觉识别的时间进程：基于同形语素的行为与 ERP 证据

作者：吴建设 常嘉宝 邱寅晨

第一轮

审稿人 1 意见：该研究采用行为和 ERP 实验考察汉语复合词视觉识别的时间进程，研究操纵了同形异音歧义词，同形同音歧义词、相同语素以及控制条件，考察汉语复合词形音义的激活过程，思路清晰，设计合理，数据分析和结果描述规范，讨论较为充分，具有一定的贡献，研究还存在以下几个小问题，供作者思考。

意见 1：前言部分，研究问题的提出较明确，建议文献综述的写作与问题的提出对应起来。

回应：谢谢审稿专家的宝贵建议，已遵照建议对文献综述部分进行了修改与调整，并将文献综述部分归为“正字法与语义”、“音位与语义”两大部分。参见稿件中红色修改部分。

意见 2：启动范式以及启动时间的长短对同形词干/语素抑制效应应有不同的影响，本研究为何考虑采用掩蔽启动范式？SOA 的确立有何依据？

回应：感谢审稿专家的宝贵意见。已将相关理由补充到稿件中。

本研究为了探究亚词汇/语素的早期加工过程而使用了掩蔽启动范式。按照 Forster et al. (1987) 的建议，在被试无法有意识感知启动词的情况下，他们就无法借助它进行有意识的策略加工。Feldman et al. (2012) 在总结掩蔽范式时也提到“学者们通常认为大于 100ms 的 SOA 就无法揭示视觉识别过程中的早期加工阶段”。而为了探究亚词汇/语素的加工过程，先前研究也多数采用了掩蔽启动范式。例如 Rastle et al. (2004) 采用掩蔽启动范式探究了英语复杂单词的语素效应 (cleaner-CLEAN, corner-CORN, brothel-BROTH)，并将 SOA 设置为 42ms；而为了考察汉语复合词的语素效应，Zhou et al. (1999)、Tsang et al. (2014)、Wong et al. (2014)、Wu et al. (2016) 等研究也都采用了掩蔽启动范式，并分别将 SOA 设置为 57ms、40ms、47ms、53ms。上述四项实验中所涉及汉语复合词具体示例如下：

Zhou et al. (1999)：目标词为“华贵”，启动词为“华丽”/“华侨”/“滑翔”/“完整”

Tsang et al. (2014)：目标词为“公厕”，启动词为“公园”/“公雞”/“皮带”

Wong et al. (2014)：目标词为“机场”，启动词为“机会”/“基础”/“跑道”/“政策”

Wu et al. (2016)：目标词为“公园”，启动词为“公眾”/“公雞”/“草地”/“嗅覺”

此外，在文献阅读中，我们发现启动时间的长短对同形词干/语素抑制效应有着不同影响，而在掩蔽启动范式下同形语素的抑制效应是本研究的关注焦点。

综合以上情况，本研究确定使用掩蔽启动范式。而 SOA 的确立则基于以上学者的建议与研究。同时为了确保被试无法观测到启动词的存在，本研究在前期的预实验中进行了反复测试与调整，最终确定为 47ms。

以下仅列出本稿件中未出现的参考文献：

Feldman, L. B., Kostić, A. Gvozdenović, V., O'Connor, P. A., and Moscoso del Prado Martín, F. (2012). Semantic similarity influences early morphological priming in Serbian: a challenge to form-then-meaning accounts of word recognition. *Psychon. Bull. Rev.* 19, 668–676.

Forster, K. I., Davis, C., Schoknecht, C., and Carter, R. (1987). Masked priming with graphemically related forms: repetition or partial activation?. *Quarterly Journal of Experimental Psychology*, 39(2), 211–251.

Wong, A. W. K., Wu, Y., & Chen, H. C. (2014). Limited role of phonology in reading Chinese two-character compounds: Evidence from an ERP study. *Neuroscience*, 256, 342-351.

意见 3: 本研究所采用的材料举例存在词性的差异，如“作坊”为名词，其他条件符合词为动词，这是否会影响复合词的加工。此外，46 对材料仅在文内列举了一对，材料或者材料举例未附在文末，请补充。

回应: 感谢审稿专家的宝贵意见。本研究在设计之初，也曾考虑到关于材料词类的控制问题。但在实际材料编制中，由于实验材料对音位、语义的特殊要求，最终初步可以获得的材料仅有 60 余组，在控制、调整好首字、尾字字频、笔画后，无法进而控制词类问题。而在其他利用汉语复合词的独特语言特征的类似实验中，多数研究也由于类似原因而未对材料的词类进行严格控制。比如，Tan & Perfetti (1999)考察汉语复合词的音位激活情况的研究中，使用了诸如“设备”、“遗弃”等不同词性的实验材料；Zhou et al.(1999)通过五项实验来探究汉语复合词加工过程中形音义的作用，也因实验材料编制难度大，而未再考虑控制词类（以“华贵”为目标词为例，启动词分别为“华丽”、“华侨”、“滑翔”和“完整”，以“简易”为目标词为例，启动词分别为“轻易”、“贸易”、“压抑”和“诬陷”）。本研究继而通过采用相同的目标词这样“项目内”设计来减少词类的可能影响。而无论行为数据抑或是 ERP 数据，虽然启动词不同，但都是基于相同目标词而完成数据收集的。

感谢审稿专家的提醒。已将本研究的实验材料作为附录附在稿件后。

意见 4: 本文未按照自检报告要求提供统计分析的 95%CI。

回应: 感谢审稿专家的提醒，结合审稿专家二的建议，我们已将包含统计结果、“95%CI”等相关信息作为“表 3”增加到稿件中。

表 3: P200、早期 N400、晚期 N400 各成分统计结果汇总表

成分	位置	(I) 基线	(J) 启动类型	差值 (I-J)	P 值	95% 置信区间	
						下限	上限
P200	中线：额中央区	+O+P-S	+O-P-S	-0.686	0.065	-1.4	0.029
			+O+P+S	-.685*	0.014	-1.264	-0.107
		-O-P-S	+O-P-S	-.704*	0.038	-1.381	-0.028
			+O+P+S	-.704*	0.049	-1.406	-0.002
	中线：中央区	+O+P-S	+O-P-S	-1.041*	0.000	-1.619	-0.462
			+O+P+S	-.589*	0.022	-1.116	-0.063
-O-P-S		+O-P-S	-1.078*	0.002	-1.831	-0.325	
早期 N400	中线：前额区	-O-P-S	+O-P-S	-.923*	0.027	-1.772	-0.075
			+O+P-S	-.973*	0.049	-1.943	-0.003
			+O+P+S	-1.156*	0.002	-1.959	-0.353
	中线：额中央区	-O-P-S	+O-P-S	-1.066*	0.025	-2.035	-0.098
			+O+P-S	-0.952	0.05	-1.905	0.001
			+O+P+S	-1.338*	0.003	-2.292	-0.383
	中线：中央区	+O+P-S	+O-P-S	-.581*	0.027	-1.113	-0.048
			+O-P-S	-1.474*	0.002	-2.483	-0.464
		-O-P-S	+O+P-S	-.893*	0.039	-1.753	-0.033
			+O+P+S	-1.129*	0.013	-2.075	-0.183
右侧：额中央区	+O+P-S	+O-P-S	-.588*	0.049	-1.174	-0.001	
右侧：中央区	+O+P-S	+O-P-S	-.617*	0.02	-1.162	-0.072	
晚期 N400	中线：额中央区	+O+P-S	+O+P+S	-1.073*	0.003	-1.832	-0.314
	中线：中央区	+O+P-S	+O-P-S	-.948*	0.002	-1.609	-0.287
			+O+P+S	-1.125*	0.000	-1.774	-0.476
		+O+P-S	+O-P-S	-.712*	0.047	-1.418	-0.006
	中线：中顶区	+O+P+S	+O+P+S	-1.010*	0.001	-1.645	-0.375
			+O-P-S	-1.015*	0.049	-2.028	-0.003
		+O+P+S	-1.313*	0.014	-2.421	-0.205	
	中线：顶区	+O+P-S	+O-P-S	-1.067*	0.02	-2.007	-0.127
			+O+P+S	-1.178*	0.000	-1.888	-0.468
	右侧：中央区	+O+P-S	+O-P-S	-.692*	0.048	-1.381	-0.003

注：* 表明均值差值显著， $p < 0.05$ 。

意见 5：请按照自检报告要求，告知如何确定被试量。

回应：谢谢审稿专家的宝贵意见，本研究使用了 G*power (3.1)确定了本研究的被试量。所考虑主要情况以及计算步骤说明如下：

鉴于前期 Wu et al. (2016)关于复合词中的语素效应的 ERP 研究中报告了 η^2 约为 0.154-0.420。且在实验中他们的被试人数为 24 人；而 Tsang et al. (2014)在关于复合词的语素效应的行为研究中报告了 η^2 约为 0.134-0.586，在实验中被试人数为每项实验 24 人。因而在 G*power (3.1)的计算中我们选择“A priori”方式，选择“F tests-ANOVA Repeated Measures”，参考上述研究将 η^2 谨慎地设定为 0.10【按照 Cohen (1988)属于中等以上效果量。由于 G*power 主窗口的“Effect Size”与 F 检验中的 η^2 并不一致，我们在“Dertermine”模块中进行了相应转化】，然后分别设定 $\alpha=0.05$ ，Power=0.95，由此计算出所需样本数为 21 人。这与上述 Tsang et al. (2014)、Wu et al. (2016)的研究人数较为接近。而考虑到 ERP 数据收集中所产生的眼电伪迹等问题，我们在 Tsang et al. (2014)、Wu et al. (2016)实验的基础上增加 4 人，并最终确

定两项实验的人数分别都为 28 人。而在完成数据分析与处理后，有效被试人数最终各为 25 人。

相关软件下载地址：G*power (3.1)下载地址：<http://www.gpower.hhu.de/>

意见 6：图 4 的比较图中，请用灰色条标记出分析的时段。P10，脑电数据分析部分，删除了错误率较高的两个刺激，请报告平均错误率。P18,P19 方差分析部分，“偏 η^2 都具有较大的效果量”，请告知效果量大小的评价标准。

回应：谢谢审稿专家的宝贵建议，已按要求标出时段。我们已按要求在稿件中补充报告了具体的错误率。

因实验一为真假词判断任务，被试在正确率这一项指标上并无显著差异，四种实验条件的正确率都在 95% 以上。但在被试实验后的跟踪问卷统计结果显示，被试普遍对“委蛇”（读音为：wei1/yi2）和“上声”（读音为：shang3/sheng1）”的读音判断错误率较高，分别为 47% 和 39%。而其他 46 组材料的读音的判断准确率均在 88% 以上。因而在 ERP 实验中，我们将此两组试次予以删除。所删除的两组试次具体为：

目标词：委任。 启动词：委蛇 委婉 委派 黑夜

目标词：上楼。 启动词：上声 上司 上车 南极

谢谢审稿专家的提醒，我们已经将具体标准添加入稿件中，并进行了相应修改。

按照 Cohen (1988)的建议，在 $\eta^2=0.01$ 或以上为较小的效果量，而 $\eta^2 = 0.06$ 或以上则为中等程度的效果量，而 $\eta^2 = 0.14$ 或以上则为较大的效果量。上述标准也在心理学界获得普遍认可，如剑桥大学认知与脑科学实验室亦推荐使用上述判断标准（参见：<http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize>）。

以下仅列出本稿件中未出现的参考文献：

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

意见 7：建议在总讨论部分分成几个小部分逐一回答问题提出时的的问题，以加强对问题回答的针对性。

回应：谢谢审稿专家的宝贵建议，已遵照建议并结合审稿专家二的意见对总讨论部分进行了修改与调整，参见稿件中红色部分。

.....
审稿人 2 意见：

研究设计比较严谨，细致，为相关领域提供了新证据。有几点建议或意见，供作者参考。

意见 1：词位和词素（语素）在概念上如何区分？有些含糊。

回应：谢谢审稿专家的宝贵建议，已遵照建议对稿件进行了部分补充与说明。

在普通语言学中，语素（morpheme）被定义为最小但自由的意义结合体，主要功能是作为构成词语的组成成分（参见 Cruse, 1986）。如“夏日”是由“夏”和“日”两个语素组成，但“玻璃”拆分后没有意义，因而只有一个语素。而根据 Crystal（1995）的定义，词位（lexeme）是词汇意义的单元，是抽象意义的基本单位，即具有相同意义的所有不同形式。它不管单词的曲折变化如何，也不管其中有多少个单词。比如英语中的 run, runs, ran and running 都是同一个词位的不同形式，而 kick the bucket, pass away, die 也都是同一个词位（DIE）的不同形式。而上述第一个例子中的词条（lemma）则是 run，它代表了所有其他的不同形式并成为词典中约定俗成的词条。

但在心理学中,词位(lexeme)/词条(lemma)的定义与普通语言学中不尽相同。Garrett (1975)首先在词汇通达模型中区分了形态与句法加工,但在其理论中他并未使用lexeme/lemma。而Kempen等学者(Kempen & Huijbers, 1983; Kempen & Hoenkamp, 1987)引入了lexeme/lemma来区分形态与句法加工。按照其理论,一个词汇的词位(lexeme)专指其形态音位特征,而词条(lemma)则专指其语义句法特征。而类似词位与词条的区分在此后的Levelt et al., 1989, 1998; Roelofs et al., 1996的理论中也发挥了重要作用,lexeme/lemma的区分也由此逐渐在心理学的研究中得以确立。而本研究文献综述中所提到的词位(lexeme)/词条(lemma)的区分正来自这种心理学传统。

以下仅列出本稿件中未出现的参考文献:

- Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- Crystal, D. (Ed.) (1995). *The Cambridge Encyclopedia of The English Language*. Cambridge: Cambridge University Press, p. 118.
- Garrett, M.F. (1975). The analysis of sentence production. In: Bower, G.H. (Ed.), *The psychology of learning and motivation* (Vol. 9) (pp. 133–177). Academic Press, New York.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201–258.
- Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14, 185–209.
- Levelt, W.J.M., (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1998). *A theory of lexical access in speech production*. Behavioral and Brain Sciences.
- Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, 35, 854–876.

意见 2: 研究的第一个问题是“复合词的视觉识别中是否存在语素分解过程?”以往在汉语复合词加工研究中,不知道是否有支持不存在语素分解过程的观点和证据。在研究的早期,基于拼音文字的多词素词存在整词加工和分解加工的争论,但对于汉语复合词的加工似乎并没有这种对立观点的争论,建议作者可以再斟酌该问题的必要性。

回应: 谢谢审稿专家的宝贵建议。同时我们参考审稿专家一的建议,在删除研究中原来的第一个问题后,结合文献与分析讨论,将其提炼、修改为“同形语素中的歧义是否一定会触发抑制效应?”,借此探讨了本文所聚焦的两类同形歧义语素在启动效应上的不同及其原因。请参见稿件中红色修改部分。之前包括第一个研究问题的主要考虑也补充简述于下,并保留了部分相关的分析讨论,统一结合到研究问题一中。

虽然支持汉语复合词分解加工的研究较多(如:陈宝国、彭聃龄, 2001; 陈宝国等, 2003; 张清芳等, 2004),但现有关于汉语复合词的研究仍存在关于“词提取”与“字提取”的讨论以及语义在单词分解加工中所发挥作用的讨论。支持“词提取”的研究主要有 Peng et al. (1999)、丁国盛、彭聃龄 (2006)、张珊珊 (2006)。Peng et al. (1999)、丁国盛、彭聃龄 (2006) 所提出层间-层内联结模型 (inter-and intra-connection model, 简称 IIC 模型) 认为整词表征和语素的激活是平行的, 刺激输入可分别激活整词表征和语素表征。IIC 模型还提出语素分解非唯一的通达复合词语义的途径的假设。而张珊珊、赵仑等 (2006) 基于 6 项 ERP 实验的研究结果而得出的结论为: “词更有可能成为提取和存储的基本语言单位, 而不是字、语素和短语”。而支持“语义在单词分解加工中发挥决定作用”的研究主要来自王春茂、彭聃龄 (1999)、王文斌 (2001) 以及玛依拉·亚克甫、周晓林 (2004) 关于维吾尔语的研究。王春茂、彭聃龄 (1999)、王文斌 (2001) 的研究均认为在汉语复合词的识别过程中语义透明度

具有决定性的作用。而玛依拉·亚克甫、周晓林（2004）的研究发现：启动词和目标词之间只有在具有词素加语义关系时或者仅仅具有语义关系时，才有显著的启动效应。如果启动词与目标词之间仅仅具有词素（词形）关系，而没有语义关系，或者仅仅具有词形关系，它们之间的启动效应微乎其微。

而本研究的同形异音歧义语素（仅具有字形）、同形同音歧义语素（兼具字形与音位）以及相同语素条件（具备相同形音义）这三种不同的启动条件的对比将让我们可以一窥语义在汉语复合词分解加工中的作用。此外，本研究在考察汉语复合词分解加工中语义作用的同时，利用汉语复合词独特的语言特征同时考察了音位与语义的互动，其研究结果也对现有复合词分解加工的文献起着一定的补充与借鉴作用。

以下仅列出本稿件中未出现的参考文献：

- Peng, D. L., Liu, Y., & Wang, C. (1999). How is access representation organized? The relation of polymorphemic words and their morphemes in Chinese. In J. Wang, A.W. Inhoff & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 65–89). Mahwah, NJ: Lawrence Erlbaum.
- 丁国盛、彭聃龄. (2006). 汉语逆序词识别中整词与词素的关系. 当代语言学, 8(1), 36–45.
- 张珊珊、赵仑、刘涛、顾介鑫、杨亦鸣. (2006). 大脑中的基本语言单位——来自汉语单音节语言单位加工的 ERPs 证据. 语言科学, 3, 3-13.
- 王春茂、彭聃龄. (1999). 合成词加工中的词频、词素频率及语义透明度. 心理学报, 3, 266-273.
- 王文斌. (2001). 汉语并列式合成词的词汇通达. 心理学报, 2, 117-122.
- 玛依拉·亚克甫、周晓林. (2004). 词素效应到底是什么？——来自维吾尔语的证据. 心理学报, 36 (5), 515-524.

意见 3：研究的第二个问题及其假设的陈述也可以再斟酌。本研究的目的是探讨 P200 所反映的认知成分？还是以 P200 为研究指标，通过 P200 揭示不同认知成分的加工进程？

回应：谢谢审稿专家的宝贵建议。结合审稿专家一的建议，我们已对研究问题二进行了提炼与修改，以便在表达上更加清晰。参见稿件中红色修改部分。

意见 4：建议在 ERP 结果图上标出进行统计分析的时间窗口。

回应：谢谢审稿专家的提醒，我们已按要求采用不同灰度标出了时段。

意见 5：ERP 不同成分的统计分析结果单用文字描述不易理解，建议通过一个表进行显示。

回应：感谢审稿专家的宝贵建议，结合审稿专家一的建议，我们已将包含统计结果、“95%CI”等相关信息作为“**表 3: P200、早期 N400、晚期 N400 各成分统计结果汇总表**”增加到稿件中。

第二轮

审稿人 1 意见：修改稿有明显提升，特别问题提出的针对性有增强，前言的叙述也更为集中，有以下修改建议。

意见 1：文字再精炼些。有些表达，如果能统一起来更好，以便理解。比如有的地方是“同形同义语素”和“同形异义语素”，有的地方是“相同语素”和“同形语素”，叫法不统一。

回应：谢谢审稿专家的宝贵建议。我们按照专家建议，三位作者对稿件全文文字再次进行了独自审读，并各自划出认为需要进行文字精简与修改的部分。在此基础上三位作者进行了集中讨论，最后由第一作者基于讨论结果完成了相应文字的修改。

之所以在稿件中存在“同形异义语素”、“同形同义语素”、“相同语素”、“同形语素”以及“同

形异音歧义语素”、“同形同音歧义语素”这些不同名称的使用，主要是因为不同文献所涉及的具体语言刺激材料有所不同，且多数文献并未如本研究中那样区分“同形异音歧义语素”、“同形同音歧义语素”这两类条件。因而在总体介绍不同文献时，为了行文简洁需要，常常使用了不同条件的上义词。为了避免混淆，我们遵照专家意见对上述所有术语针对性地逐一核查，并完成了简化与统一命名处理，同时对部分不合规的命名进行了相应修改，最终在稿件中将以上命名简化为 4 种。具体处理所遵循原则如下表所示：

表 1：本稿件中不同语素条件的命名规则

	统称	具体名称	本研究实验条件	示例（标红）
1.	同形语素	异音同形语素	实验条件 1	作坊 vs 作画
2.		同形同音语素	实验条件 2	作息 vs 作画
3.	相同语素	相同语素	实验条件 3	作诗 vs 作画

意见 2：实验一的设计，感觉针对性不强。从前言中，得出的结论是启动（或抑制）效应与 1）是否掩蔽、2）SOA 长短，3）实验任务有关。但实验一直接选择了掩蔽条件和短 SOA 条件，为何没有考虑其他条件，不太清楚。

回应：谢谢审稿专家的宝贵建议。关于不同启动范式对语素加工的影响在 Zhou et al. (1999)、Feldman et al. (2012)、Tsang et al. (2014) 的研究中已经分别得以验证，且 Zhou et al. (1999)、Tsang et al. (2014) 的研究皆涉及汉语复合词的同形语素加工。上述学者的具体研究结论简要补充如下。而参考审稿专家建议，我们业已将 Zhou et al. (1999)、Feldman et al. (2012)、Tsang et al. (2014) 关于不同启动范式对语素加工启动效应影响的研究结论进而补充到实验一的研究程序中，具体参见研究程序中的红色修改部分。

- 在 Zhou et al. (1999) 的研究中，基于 5 项实验，他们同时考察了不同启动范式【掩蔽、非掩蔽启动范式以及长短 SOA (57ms 和 200ms)】对汉语复合词同形语素加工启动效应的影响。研究总体发现：在 SOA 为 57ms 时，掩蔽启动范式与非掩蔽启动范式的启动效应没有本质上的差别，但在非掩蔽启动且 SOA 为 200ms 时语义抑制效应显著。在 Tsang et al. (2014) 的研究中，他们亦考察了不同启动范式【掩蔽启动范式 (SOA=40ms)、无掩蔽启动范式 (SOA=40ms) 以及长时延迟范式 (SOA=2000ms 或按键反应)】对汉语复合词同形语素加工启动效应的影响。研究发现相同语素条件与同形语素条件在掩蔽启动范式下表现相同，但在无掩蔽启动范式中同形语素条件表现则较为异常。由上述两项研究结果可见，单词视觉识别的加工过程在非掩蔽启动且 SOA 为 200ms 时、在无掩蔽启动范式中较为容易受到启动范式的影响。
- 而关于掩蔽启动范式下的 SOA 时长，Feldman, Kostić, Gvozdenović, O'Connor & Martín (2012) 曾提到：若 SOA 大于 100 ms 时，启动研究的结果很难用于解释单一的、孤立的词汇加工过程。因为被试可以使用启动词来预测目标词，因而情景知识、其他策略知识可能会污染词汇加工结果。一般掩蔽启动研究者认为大于 100ms 的 SOA 并不能揭示视觉识别的早期加工过程。只有采用较短 SOA 所产生的启动效应才反映了加工的早期阶段。因而，他们推荐采用较短 SOA 掩蔽启动范式来研究单词的早期视觉识别加工过程。

在前言中，通过分析现有同形词干/语素研究所采用的不同启动范式，我们旨在澄清不同研究结论相互冲突的一个重要原因：长 SOA 会引发同形语素抑制效应，而短 SOA 会引发同形语素促进效应。此外还发现部分实验任务以及不同控制条件（与无关控制条件相比，或是与相同语素条件相比）也会引发不同结果。由于本研究更多关注不同类型（同音或异音）的同形语素是否会触发类似的相对于无关控制条件的促进效应以及相对于相同语素条件的抑

制效应，并由此借助同形语素效应考察正字法加工与语义加工之间、音位加工与语义之间的互动效应。因而，基于 Zhou et al. (1999)、Feldman et al. (2012)、Tsang et al. (2014) 等学者的上述研究结论以及其他类似视觉识别早期阶段研究所采用范式【如 Wong et al. (2014)、Wu et al. (2016)】，本研究直接选择采用短 SOA 的掩蔽启动范式，未再同时考察不同启动范式对不同类型汉语同形语素加工启动效应的可能影响，以避免本研究所涉及因素过多而带来的数据庞杂或失去研究焦点。但显然这是一个值得进一步思考的研究问题，我们将按照审稿专家的建议，在将来下一步的研究中仔细分析、考虑。

以下仅列出本稿件中未出现的参考文献：

- Feldman, L. B., Kostić, A. Gvozdenović, V., O'Connor, P. A., and Moscoso del Prado Martín, F. (2012). Semantic similarity influences early morphological priming in Serbian: a challenge to form-then-meaning accounts of word recognition. *Psychon. Bull. Rev.* 19, 668–676.
- Wong, A. W. K., Wu, Y., & Chen, H. C. (2014). Limited role of phonology in reading Chinese two-character compounds: Evidence from an ERP study. *Neuroscience*, 256, 342-351.

第三轮

审稿人 3 意见：本研究修改稿比初稿逻辑明显清晰了，有很好的提升，但仍有以下问题需作者进一步思考并修改。

意见 1：引言部分的理论描述还是有点弱，到底是由哪一个理论驱动研究者去研究基于汉语复合词的同形语素的互动呢？引言部分需将“正字法与语义的互动”以及“语音与语义的互动”进行梳理，使脉络清晰，而不是仅仅分为这两个部分来罗列文献。

回应：谢谢审稿专家的宝贵建议。我们按照专家第 1 条、第 2 条、第 10 条建议，在重新检索、补充与本研究相关文献（截止今年 5 月）的基础上，对稿件的理论框架与文献综述进行了再次梳理（参见文中红色部分）。文中主要的修改部分说明如下：

（1）在文献综述梳理后，在引言结束部分凸显、补充了与单词识别/形态加工相关的 5 个模型；具体修改如下：

总而言之，针对同形词干/语素的研究为了解单词的视觉识别加工提供了一个特别的窗口。在识别单词的过程中，同形词干/语素的相同字形可能会同时激活与之对应的不同语义信息或音位信息。它不仅被用来验证不同的单词识别或形态加工模型，如：词汇分解-词条模型（Taft, 1994, 2004; Taft & Nguyen-Hoan, 2010; Xu, & Taft, 2014）、超词汇模型（Giraudo & Grainger, 2001, 2003）、字形-语义模型（Rastle et al., 2004; Meunier & Longtin, 2007; Rastle & Davis, 2008; Crepaldi et al., 2010）、形态-正字法/形态-语义混合模型（Diependaele et al., 2005, 2009; Beyersmann et al., 2012, 2014）或平行分布加工模型（Parallel Distributed Processing, PDP）（McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989; Gonnerman, Seidenberg & Andersen, 2007; Plaut & Booth, 2000），还可以与 ERP 等神经成像技术结合以考察单词识别的时间进程。

（2）补充了 Zhang et al. (2017)、Zou et al. (2012, 2019)等在听觉识别范式或视听跨通道范式下完成的近期研究，以及 Simon et al. (2012)采用 MEG 完成的相关研究（皆已在文中使用红色标出）。此外，还补充了一篇基于芬兰语同形词干的研究（Jarvikivi et al., 2009）。关于还对 Taft & Nguyen-Hoan (2010)、Xu & Taft (2014) 等学者提出的基于词条的形态加工模型进行了补充介绍。

（3）将所有引用文献按专家第 4 条、第 5 条建议进行了再次审读与细节核对，并进行

了适当的补充（见“引言”中标红部分）。

（4）在讨论部分增加了关于上述模型的相应讨论与分析（见“总讨论”中标红部分）。

意见 2：“正字法相同”与“同形”还是有区别的，需作者进一步厘清，不能一会儿正字法一会儿同形，用法需一致。

回应：谢谢审稿专家的宝贵建议。为避免错误或混淆，我们按照专家建议，对稿件全文进行了逐一核对。并遵照语言学的具体定义以及 Zhou et al. (1999) 文中的使用标准统一了两者的用法。具体规则如下：

（1）在与音位、语义相对或相关时，全部采用“正字法”；

（2）在本文具体实验条件以及指明不同语素类型时使用“同形”；

（3）两者的英语对等翻译分别采用“orthograph”（正字法）或“orthographic”（正字法的）（具体指代单词的拼写系统）以及“homograph”（同形）或“homographic”（同形的）【具体指代拼写相同但意义不同的单词或语素（发音不一定相同）】。

意见 3：P15-16. “Tan & Perfetti （1999）进而操控了汉语复合词的位置和读音一致性（读音不一致条件为多音字，见以下例子中划线部分）两个因素，将其分为首字一致条件（如：形成）、首字不一致条件（如：见识）、尾字一致条件（如：明净）和尾字不一致条件（如：体重）。研究表明，语音不一致条件下的反应时更长，他们由此认为汉语复合词的识别过程经历了两个阶段：语境无关阶段会激活一个字的所有语音项，但在语境敏感阶段，符合语境的语音项才会被筛选出来。”对这一研究的描述不清楚。

回应：谢谢审稿专家的宝贵建议。我们按照专家建议，对 Tan & Perfetti（1999）的研究再次进行了审读，并进一步补充了实验细节。然后在此基础上将其重新梳理以符合整体文献综述的需要。具体修改如下：

Tan & Perfetti（1999，实验二）采用词汇判断范式，通过操控汉语复合词中组成汉字的位置和读音一致性（读音不一致条件为多音字，见以下例子中划线部分）两个因素，将其分为首字一致条件（如：形成）、首字不一致条件（如：见识，首字读音分别为 jian4/xian4）、尾字一致条件（如：明净）和尾字不一致条件（如：体重，尾字读音分别为 zhong4/chong4）。研究表明，音位不一致条件下的反应时相比控制条件显著更长，首字与尾字读音不一致条件分别慢 22ms、28ms，但两种不一致条件之间没有显著差异。他们由此认为被试在汉语复合词识别过程中对组成汉字的音位信息高度敏感，但对处于首字或尾字语境并不敏感。因而复合词的识别过程包括了其组成汉字的整合过程，且涉及了音位信息。首字的加工应该与语境（复合词）无关，它会激活该复合词首字的所有对应语音项。

意见 4：“而 Liu et al.（2017）采用视听双通道探究句子加工语境下汉语同形语素的认知加工。通过语音干扰条件的设置，他们发现语音干扰条件在 P200 波幅上更负，而同时由于语音干扰而导致语义获取难度增加，在 N400 波幅上也更负。上述研究总体表明，被试对英语同音词中的不同语义信息、对汉语复合词中同形语素中的不同语音信息敏感。但 Zhou et al.（1999）实验四也提供了相反的证据。”这一段描述有误，需再仔细校对。另外，用 Zhou et al.（1999）的实验结果来对应 Liu et al.（2017）的结果说提供了相反的证据十分不妥。

回应：谢谢审稿专家的宝贵建议。我们按照专家建议，我们按照专家建议，对 Liu et al.（2017）的研究再次进行了审读，并进一步补充了实验细节，并增加了同一研究组基于功能连接分析的另一篇论文（Zhang et al., 2017）。为尽量精简引言中的文献综述内容，对该研究组关注的脑区信息（与 P200、N400、LPS 所出现脑区较为一致）未再具体介绍。然后在此基础上将研究内容重新梳理以符合整体文献综述的需要。

关于将 Liu et al. (2017) 与 Zhou et al. (1999) 的对比, 由于研究范式、使用材料、启动时间等多处不同, 我们完全同意审稿专家的建议: 比较确实不合适。在修改稿中我们因而区分了基于多音字的研究以及基于异音同形语素的研究。具体修改如下:

而 Liu et al. (2017) 采用视听双通道范式, 在句子加工语境(视听通道同时逐字呈现 500ms, ISI=500ms) 下对比了音位相同/语义相同条件(如: 校长/zhang3)、音位相同/语义不同条件(如: 家长/chang2)、音位不同/语义相同条件(如: 船长/yuan2)、音位不同/语义不相同条件(如: 班长/mu4), 研究发现音位相同/语义相同条件较其他条件出现 P200 (280–380ms) 增强效应。而上述四种条件的 N400 (400–550ms) 效应也依次增强, 音位不同/语义不相同条件与其他三种条件相比在 N400 波幅上显著更负。而晚期正波(LPS) 在 600–700ms 时窗上音位相同/语义不同条件相比音位相同/语义相同条件出现显著增强效应, 但在 750–800ms 时窗上, 各条件之间不再有显著差异。他们由此推断 P200 效应可能与音位加工有关, N400 效应可能与视听信息整合过程中的语义不匹配有关(且音位信息可能调节了语义的提取), 而晚期正波的第一时窗则可能表明视听信息整合过程中的冲突检测过程。上述结果在 Zhang et al. (2017) 基于功能连接分校的研究中也得到进一步的证实。

意见 5: P16. “考虑到启动时间长短对同形词干/语素抑制效应的不同影响, 我们认为有必要进一步考察在掩蔽启动范式下同形语素的促进效应是如何产生的。”这里的问题提出不明确, 前面引言中并未涉及多少掩蔽启动的问题, 怎么就突然出现有必要研究了昵? 需修改。

回应: 谢谢审稿专家的宝贵建议。我们按照专家建议, 在“引言”的文献综述中补充说明了各研究所具体使用的范式(包括掩蔽启动范式的使用及其 SOA) 及其对比结果。关于同形语素的总体结果是: 同形同音语素条件在长时启动范式下会产生抑制效应, 但在掩蔽范式下会产生促进效应; 而异音同形语素条件在长时启动范式下同样产生抑制效应, 但在掩蔽范式下无抑制效应(Zhou et al., 1999, 实验 4b), 而 Zou et al. (2012) 在听觉识别范式下(SOA=150ms) 的 ERP 研究也同样未能发现它的抑制效应(即使与相同语素条件相比)。

在单词识别/形态加工研究中, 自从 Forster & Davis (1984)、Grainger & Segui (1990) 首次采用掩蔽范式以来, 无论是词缀研究(如: hunter-HUNT)、假派生词研究(如: brother-BROTH) 还是诸如本研究的同形词干研究, 绝大多数类似研究都采取掩蔽范式(但 SOA 不尽相同) 以试图揭示单词视觉识别早期阶段是否存在词汇形态分解、基于字形的早期自动加工、或分离形态-正字法与形态-语义过程(可参见 Xu & Taft, 2014)。有鉴于此, 本研究为了分离同音/异音同形语素条件以及相同语素条件的早期自动化加工阶段(该阶段被多数学者认为是无语义参与的) 与晚期词汇语义加工过程, 也为了与早期采用长时启动范式的 ERP 研究(如 Barber et al., 2002; Dominguez et al., 2004) 相区别, 由此采用了掩蔽范式。这也在 Wu et al. (2016) 等学者在 ERP 中所采用的范式。Forster (1998) 在对掩蔽范式进行详细考察后也认为“尽管有种种不足, 它仍然可以被作为完全自动化过程的一个指标”(p.229)。

基于上述说明, 我们对研究范式与 SOA 选择的理由进行了补充说明。见以下以及文中的红色部分:

在研究范式的选择上, 鉴于先前考察单词形态-正字法加工、形态-语义加工的研究采用掩蔽范式以分离单词识别的早期过程与晚期过程(见前言中的相关文献综述), 本研究同样采用了掩蔽范式【根据 Zhou et al. (1999)、Feldman et al. (2012)、Tsang et al. (2014) 等学者的研究结果, 非掩蔽范式且 SOA 大于 100ms 时, 单词视觉识别的早期过程容易受到策略加工的影响; 而 Forster (1998) 在对掩蔽范式进行详细考察后也认为“尽管有种种不足, 它仍然可以被作为完全自动化过程的一个指标”(p.229)】。而在 SOA 的选择上, 为了确保被试无法观测到启动词的存在, 本研究在前期的预实验中进行了反复测试与调整, 最终确定

SOA 为 47ms，并严格地设置了前掩蔽与后掩蔽。

意见 6：表 1 中出现许多“首字词频”和“尾字词频”，有误，应该为字频。

回应：谢谢审稿专家的宝贵提醒。已全部查找、审读、改正。

意见 7：实验二中任务改为“真假人名判断”意义何在？人名的语义比较复杂，是否经过了评定？

回应：谢谢审稿专家的宝贵建议。Zhou et al. (1999) 在其研究中提到，词汇判断任务本身可能引发了音位加工。为考察单词识别加工的早期阶段与同形语素识别的自动化过程，我们因而在本研究中特意遵循 Laszlo & Federmeier (2007, 2011) 的建议与实践将行为实验中的词汇判断任务改为人名判断任务。在此任务下，被试被要求仅仅对“陆毅”这样的人名目标词做出反应，对其他目标词无需做出按键反应，由此尽量避免了任务以及按键反应的直接影响以及被试可能的策略加工。由于人名仅作为本实验的填充材料，并未计入本研究的数据分析中。人名的语义经过其他未参加被试的删选，以确保删除任何与汉语复合词较为接近的人名。

遵照专家的提醒，为避免混淆，我们已经将上述简化说明补充到“方法”中的相应部分（见稿件中对应标红部分）。本研究中所使用具体人名材料示例如下：

陆毅/屈原/姜潮/费翔/汪皓/嵇磊/宋茜/曹帆/罗晋/李白/邬兰/阮玲/严鹏/
邓雯/倪妮/薛钗/梁虹/邵逸/梅婷/魏征/贺龙/黎姿/陆源/刘涛/乔宇/董卿/
吕沁/姚笛/闵聪/于彤/唐强/戴娇/舒畅/何炅/穆晖

意见 8：总讨论中“汉语双字复合词的视觉识别过程存在对环境变量的依赖，其视觉识别过程更有可能是“分解+组合”两者融为一体的过程。”该结论存在过度推论，需提供更加有力的佐证。

回应：谢谢审稿专家的宝贵建议。我们完全同意专家的意见。由于我们已经按照审稿专家第 1 条建议在稿件中补充了关于相关单词识别/形态加工模型的说明，我们进而也对此处“总讨论”中的上述分析内容进行了修改。我们整体删除了上述关于环境变量的讨论，并基于本研究结果增加了关于不同单词识别模型的分析讨论（见文中标红部分）。

意见 9：全文所引文献太陈旧，仅一篇 17 年和 16 年的文献无法有力支撑本文最终的结论，需作者补充该领域的最新文献。

回应：谢谢审稿专家的宝贵建议。如我们在回复专家第 1 条建议中所言，我们业已按照专家建议，对与本研究相关的文献进行了再次检索，并同时完成所有本研究的引用文献与新增文献进行了再次审读。在此基础上，我们完成了文献综述的再次梳理以及下数文献内容的添加与补充（详见稿件中、参考文献中红色标红部分）。以下为在本次修改中所增加的具体文献：

（1）对文献综述中文献的更新与补充，共 5 篇；

Jarvikivi, J., Pykkonen, P., & Niemi, J. (2009). Exploiting degrees of inflectional ambiguity: stem form and the time course of morphological processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 221-237.

Simon, D. A., Lewis, G., & Marantz, A. (2012). Disambiguating form and lexical frequency effects in MEG responses using homonyms. *Language and Cognitive Processes*, 27, 275-287.

Zhang, Z., Zhang, G., Zhang, Y., Liu, H., Xu, J., & Liu, B. (2017). Cross-modal integration of polyphonic characters in chinese audio-visual sentences: a mvpa study based on functional connectivity. *Experimental Brain Research*, 235(12), 3743-3755.

Zou, L., Desroches, A. S., Liu, Y., Xia, Z., and Shu, H. (2012). Orthographic facilitation in Chinese spoken word recognition: an ERP study. *Brain Lang.*, 123, 164–173.

Zou, L., Packard, J.L., Xia, Z., Liu, Y., & Shu, H. (2019). Morphological and whole-word semantic processing are distinct: Event related potentials evidence from spoken word recognition in Chinese. *Front. Hum. Neurosci.*, 13, 133.

(2) 关于单词识别/语素加工模型文献的补充, 共 11 篇;

Beyersmann, E. , Coltheart, M. , & Castles, A. . (2012). Parallel processing of whole words and morphemes in visual word recognition. *The Quarterly Journal of Experimental Psychology*, 65(9), 1798-1819.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, 20(1–2), 75–114.

Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition*, 37(6), 895–908.

Girardo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review*, 8, 127–131.

Girardo, H., & Grainger, J. (2003). A supralexical model for French derivational morphology. In D. Sandra & H. Assink (Eds.), *Reading complex words* (pp. 139–157). Amsterdam, The Netherlands: Kluwer.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Pt. 1. An account of basic findings. *Psychological Review*, 88, 375–407.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, 107, 786–823

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language & Cognitive Processes*, 23(7/8), 942–971.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.

Taft, M., & Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes*, 25, 277–296.

Xu, J., & Taft, M. (2014). Solely soles: Inter-lemma competition in inflected word recognition. *Journal of Memory and Language*, 76, 127–140.

(3) 针对审稿专家各条建议在研究方法上的补充说明相关文献, 共 3 篇;

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research*, 27(2), 203-233.

Kiefer, M., & Brendel, D. (2006). Attentional modulation of unconscious “automatic” processes: evidence from event-related potentials in a masked priming paradigm. *Journal of Cognitive Neuroscience*, 18(2), 184-198.

Laszlo, S., & Federmeier, K. D. (2007). Better the dvl you know: acronyms reveal the contribution of familiarity to single-word reading. *Psychological Science*, 18(2), 122-126.

(4) 其他在文献梳理、分析讨论过程中的说明补充文献, 共 2 篇;

Laudanna, A., Voghera, M., & Gazzellini, S. (2002). Lexical representations of written nouns and verbs in Italian. *Brain and Language*, 81, 250–263.

Meng, X. , Jian, J. , Shu, H. , Tian, X. , & Zhou, X. . (2008). ERP correlates of the development of orthographical and phonological processing during Chinese sentence reading. *Brain Research*, 1219, 91-102.

第四轮

审稿人 3 意见:

作者针对审稿人提出的问题逐条做了回答和修改,补充了新的文献,补充了研究假设,清晰了所研究的问题,建议修后发表。但以下问题仍需作者认真修改。

意见 1: 由于补充了很多新的文献,其中有一些明显作者还没有能够清楚驾驭其方法和结论,导致引言中有些字句含混不清。如:(1) P21:“因而认为在掩蔽范式下同形词干可能在识别前激活了多重词条从而促进了加工过程(p.132),因而同形词干抑制效应源于与意识相关的选择机制。”这句话不明白作者要表达什么,需修改。(2) P24:“他们由此认为被试在汉语复合词识别过程中对组成汉字的音位信息高度敏感,但对处于首字或尾字语境并不敏感。”所引文献中并没有提及语境的问题,且已有相关研究发现阅读过程中复合词的首字是对语境非常敏感的,所以该结论需要作者认真回读文献看是否正确,以免引起误导。(3) P24:“上述结果在 Zhang et al. (2017) 基于功能连接分校的研究中也得到进一步的证实。这句话表达有误。”总之,类似以上小错误文中还有不少,需作者认真查正并修改,建议有些文献可以适当删减。

回应: 谢谢审稿专家的宝贵建议。我们遵照专家建议,在重新逐条查证文献的基础上,按要求完成了稿件中相关表述的核对与修改(参见稿件中红色部分)。各条在具体修改后如下:

(1) 因而认为同形词干促进效应的可能机制为:启动词所激活了多重词条促进了目标词的识别加工过程(p.132),由此同形词干效应涉及了某种有意识的选择机制。

(2) 他们由此认为被试在汉语复合词识别过程中对组成汉字的音位信息高度敏感,但对处于首字或尾字位置并不敏感。因而复合词的识别过程包括了其组成汉字的整合过程,且涉及了音位信息。首字的加工会激活该复合词首字的所有对应语音项。

(3) 而 Zhang et al. (2017) 采用 Liu et al. (2017) 相同的设计,基于功能连接分析(采用多体素模式分析, MVPA),发现在 P200、N400、LPS 不同时窗下功能分析的分类模式产生了改变,同时它们在脑区空间分布上也明显不同。

与之同时,遵照审稿专家要求,我们在通读稿件的基础上,调整了部分文字表述;此外我们还部分删减或简化了关于 Simon et al. (2012) (研究采用 MEG 而非 ERP 技术)、Zou et al. (2019) (研究采用了听觉识别范式而非视觉识别范式) 等部分并非直接相关文献的内容(见文中对应红色字体部分)。

意见 2: ERP 实验中任务换为人名判断,是否有研究者采用这种方式? 需附上参考依据。因是人名则按键不是人名则不按键的任务有可能涉及到被试的注意控制过程,从而影响 ERP 早期成分的结果。

回应: 谢谢审稿专家的宝贵提醒。伊利诺伊大学香槟分校的 Federmeier 等学者在其单词识别的系列研究中多次采用此任务(如: Laszlo & Federmeier, 2007, 2008, 2011)。而在此之前,北京大学周晓林教授在其 Zhou et al. (1999) 关于汉语复合词的视觉识别研究中也对词汇判断任务是否可能强制激活了语义加工有过提醒。有鉴于此,我们在本研究中采用了人名判断任务,被试仅仅被要求对人名进行强制反应,因而避免了对目标词的强制按键反应。基于上述学者们的研究实践与建议,与词汇判断任务相比,人名判断任务可能更符合掩蔽启动范式的内在要求。遵照审稿专家的提醒并基于以上理解,我们对相关部分具体修改如下(参见文中红色部分):

鉴于 Zhou et al. (1999) 曾提醒词汇判断任务中的强制性按键反应可能会影响单词视觉识别过程,并有可能激活了语义加工。为避免此种影响,我们参考 Federmeier 等学者的系

列研究（如：Laszlo & Federmeier, 2007），将行为研究中的词汇判断任务改为真假人名判断任务。被试仅被要求对作为填充材料的人名进行按键反应，而对于实验各条件的目标词都无需按键反应。我们因而去掉了行为实验中的假词填充材料而对应增加了 46 组由两字构成的人名（如：陆毅）。

第五轮

编委专家意见：经过几轮审稿，文章有了很大提高，但还存在以下问题。

意见 1：文章的文字太晦涩，许多句子不通顺，如“认为”之后一般应该有逗号，还有如“启动词所激活了多重词条促进了目标词的识别加工过程”、“结合实验结果他们认为形态加工的早期阶段为自下而上的”、“一个明显的遗憾”（应该为缺陷或缺点）。建议作者找一位中文比较好的人审读，将文字搞通顺。文章也太长，适当压缩一下篇幅。

回应：谢谢编委专家的宝贵建议。遵照专家建议，我们逐条完成稿件中上述文字表述上的修改（包括与“认为”相似的“发现”、“表明”、“推断”等）。同时，我们按专家要求对稿件尽量压缩了篇幅，将正文字数压缩减少约 4000 字。在此基础上，我们还按要求邀请了一位中文专业的老师帮助通读修改稿，对所有较为晦涩的文字表述进行了适当调整。

意见 2：文章中不要出现“我们”之类的第一人称的表达，如“我们的假设是：”应该改为“研究假设是”。

回应：谢谢编委专家的宝贵建议。遵照专家要求，我们将稿件中出现的所有“第一人称表达”进行了相应处理，根据语境分别修改为“研究”、“本研究”或“研究者”。

意见 3：研究假设的提出用了好几段话，请简化。

回应：谢谢审稿专家的宝贵建议。遵照专家建议，我们对稿件中三条“研究假设”进行了简化（参见稿件中红色部分）。各条在具体修改后如下：

（1）音位信息是否参与了汉语复合词的视觉识别加工？如果同形语素中的音位信息未被激活，同形同音语素条件（如：作息）与同形异音语素条件（如：作坊）都会产生类似 Dominguez et al. (2004) 研究中的同形语素抑制效应；而如果同形语素中的音位信息被激活，那么仅仅同形同音语素条件仍然会产生同形语素抑制效应，而同形异音语素条件则不会产生类似抑制效应。

（2）音位信息何时参与了单词视觉识别加工？如果 P200 成分表征了音位加工，则音位不一致条件（同形异音语素条件、控制条件）相对于音位一致条件（同形同音语素条件、相同语素条件）在 P200 成分上将会更负。反之则无法断定音位加工出现在 200ms。而判断音位加工是否参与的另一个较为可靠的指标则是：同形语素抑制效应在何时产生。

（3）形态-正字法加工与形态-语义加工是否在 N400 这个时窗发生分离？鉴于同形同音、同形异音、相同语素条件在正字法上完全相同，若正字法与语义加工在 N400 发生分离，三者早期 N400 成分上应无显著差异，但在晚期 N400 成分上会有所不同；而如果三者未发生分离，则它们在 N400 成分上应无显著差异。鉴于五种单词识别模型对正字法与语义加工激活的时间有不同的预测，借助上述研究结论亦可同时对其进行验证。

意见 4：掩码启动应该改为掩蔽启动，一些概念在提出时应该解释，如同形语素与相同语素。

回应：谢谢编委专家的宝贵建议。我们完全同意专家建议，并已遵照专家建议，将所有“掩

码启动”替换为“掩蔽启动”。同时，在稿件中适当增加了对“同形语素”、“相同语素”、“语素变体”等语言学概念的具体解释（参见稿件中红色部分）。

意见 5：符号与文字存在不对应的情况，如异音同形语素条件与其简化符号不对应，改为同形异音语素条件更好。

回应：谢谢编委专家的宝贵建议。之前译文主要考虑的是，译文与英文原文尽量对应。我们完全同意专家意见，如此修改后，它们不仅与稿件中的简化符号对应，且更符合汉语表达习惯。我们业已遵照专家建议在稿件中完成了相应修改。

意见 6：表 2 的数据应该有标准差，应该有效果量与 95% 置信区间，统计量应该为斜体，表 3 应为三线表，图 4 线太粗，不够清晰。

回应：谢谢编委专家的宝贵建议。遵照专家建议，我们在上述表格中重新添加了相应数据，也同时对表格格式按要求进行了重新调整（表 3 由于数据较多，因而仍保留了三条辅助线）。

意见 7：结论应该改写，现有的结论不像结论，请参考学报已经发表的文章。

回应：谢谢编委专家的宝贵建议。我们遵照专家建议并参考学报所发表最新论文（王娟、马雪梅、李兵兵、张积家. 2019. 汉字形声字识别中义符和声符的家族效应,《心理学报》第 8 期），对结论进行了重新改写（参见稿件中红色部分）。