

《心理学报》审稿意见与作者回应

题目：《转换概率和词长期期待对语音统计学习的影响》

作者：于文勃；王璐；瞿邢芳；王天琳；张晶晶；梁丹丹

第一轮

审稿人 1 意见：

该文章考察了转换概率和词长因素对汉语读者进行人工语音统计学习的影响。研究有一定的创新和特色，为理解汉语语音学习的机制提供了实验证据。有以下意见或建议。

意见 1：摘要中说研究采用的是“混合实验设计”，为何是混合实验设计？谁与谁混合呢？

回应：感谢审稿专家的意见。在摘要中，我们重新表述了实验设计部分内容“本研究采用 2（转换概率：高转换概率，低转换概率）*2（词长期待：两音节，三音节）的混合实验设计，转换概率是被试间变量，词长期待是被试内变量，以此考察转换概率和词长期待对语音统计学习的影响”，让实验设计看起来更加清晰。

意见 2：问题提出部分，关于文章的关键问题及创新性的描述不清晰。比如，从行文看，该文首要关注的问题似乎是转换概率变化的影响，但并没有重分引用有关研究文献（在讨论 4.1 第五段有提及），也没有清晰阐释在这些已有研究基础上提出的新问题或新假设是什么。作者又提到“本研究关注的第二个问题是语言经验对语音统计学习机制的影响。”但实验设计如何体现对“语言经验”的影响也不清晰。

回应：感谢审稿专家的意见。文章中关注的焦点是转换概率的变化，但确实在前言部分存在表述不清楚、不充分的问题。根据审稿专家的意见，我们重新组织了关于转换概率变化的相关内容。具体修改可见 P2，我们首先说明转换概率越高，音节间的连结关系越紧密，那么个体对它们的记忆会更加牢固，人工语言的学习效果会更好；但目前还没有研究对目标词转换概率发生变化时，被试概率词切分成绩是否会发生变化进行研究；对这一问题进行考察可以与现有相关研究形成互补，进一步证明转换概率在个体大脑中具有心理现实性；然后我们结合视觉统计学习中转换概率变化导致被试迫选成绩的显著差异，提出在语音模态下研究转换概率变化对人们统计学习效果的影响的必要性。

在语言经验的变量上，本研究关注的是词长期待，它是众多语言经验中的一种。我们对这一部分的论述进行了重新组织，具体解释了“词长期待”并介绍了相关研究，具体修改可见 P3-4。在韵律经验方面，Lew-Williams 和 Saffran（2012）发现如果先帮助婴儿产生韵律期待，那么这些经验会影响他们完成统计学习任务的效果。参考语言学的相关定性和定量研究，汉

语普通话中两音节词占主导地位，因此，我们提出汉语母语者具有的两音节词长期期待很可能会引导概率词切分的结果。

意见 3：关于“事先对比”，假设矩阵和对比矩阵分别是什么含义？二者是什么关系？至少从矩阵内容来看二者分别检验的是不同的研究假设。“事先对比”结果的有效性也存疑。另外，“事先对比”通常针对研究者关心的先验假设进行检验，在优先性上似应该置于综合分析之前。

回应：感谢审稿专家的意见。假设矩阵和对比矩阵是进行事先对比的核心概念，假设矩阵指的是研究者所关心的假设，在本研究中，我们最关心的是在 LTP 条件下被试完成三音节迫选试次的正确率是不是显著低于两音节迫选试次，以及对于完成三音节迫选试次的两组被试来说，在 HTP 条件下正确率是否要高于 LTP 条件。由于事先对比是依靠线性模型完成的，因此将假设矩阵进行变换则可以形成对比矩阵，换句话说，对比矩阵是假设矩阵的参数设置。在修改过程中，我们删除了对比矩阵和假设矩阵的表格，在前言部分和结果部分重新表述了我们进行事先对比的逻辑。

关于事先对比的位置问题，如您所述，其应该置于综合分析之前，但是在统计学习研究中，研究者习惯先对被试在不同情况下的正确率进行检验，即在相应条件下被试的迫选正确率是否显著高于随机水平，随后再做一些其他的检验分析，因此，我们才将事先对比放在综合分析之后。

意见 4：“词长期期待”的确切含义不清晰。正文中第一次出现“词长期期待”时，应对此概念进行适当解释。另外，“在有关词长期期待的研究中，有一篇研究支持我们的猜想”，不清楚“我们的猜想”是什么，应先介绍猜想再介绍支持猜想的研究。

回应：感谢审稿专家的意见。这里的表述不到位、不准确的地方我们已在原文相应位置进行了修改。文中的“词长期期待”指的是汉语母语者对普通话两音节词语的偏好。“在有关词长期期待的研究中，有一篇研究支持我们的猜想”这里的表述不准确，我们已经删除并进行了修改，让它变得清晰一些，便于读者阅读，您可以参考意见 2 的回应。

意见 5：“在迫选任务中，我们以非词（转换概率为 0）作为干扰刺激，而且保证目标词和非词在两种人工语言完全一致。”这句话难以理解。

回应：感谢审稿专家的意见。我们的迫选任务中，在 HTP 和 LTP 两组实验中选项都是一样的，用的是同样的目标词和非词进行迫选，如下表 1。由于研究中的实验音节都是无意义音节，如果在 HTP 和 LTP 两种条件下使用不同的迫选试次，那么迫选正确率的差异很可能是由于迫选刺激不同，因此我们控制了这一因素，从而保证在 HTP 和 LTP 两个水平下，实验效应是来自于不同的转换概率。

表 1 实验中使用的目标词和非词

目标词	非词
nveruote	nvegeilai
liageirua	liafote
diafolai	diaruorua
remei	rerou
rouse	meise

意见 6：“前因效应或后因效应”的说法似乎不常见，是指“首因效应或近因效应”？

回应：感谢审稿专家的意见。这里的表达不够到位，已经在文中相应位置更换为“首因效应或近因效应”。

意见 7：图 3 的必要性可再斟酌。

回应：感谢审稿专家的意见。图 3 的作用是帮助我们更好地进行结果分析，由于统计学习的认知过程比较复杂，包括学习阶段和测试阶段，因此我们觉得采用图例的方式比较清晰。

意见 8：有些字词错误或表达不精确。稍举几例：page8 第一段，“其他规则与 LTP 一致”，疑为“HTP”。同一段，数字序号 2 重复。“人工语言播放过程中音量恒定无法调整”中，“无法调整”有些多余可删去。

回应：感谢审稿专家的意见。已经修改这些错误，并重新认真阅读全文，改正相关文字表达不到位的地方。

.....

审稿人 2 意见：

文章探讨转换概率和词长期待在词汇分割中的作用。选题有意义。文章写作简洁清楚，逻辑性强。文章存在以下问题：

意见 1：文章探讨的两个因素转换概率和词长效应，作者把词长效应当作语言经验的作用。那么，转换概率的高与低虽然是一种客观存在，但词汇学习的个体对连续语流中词汇界限的感知，或者对转换概率的学习，也是语言经验的结果吧？作者把这两个因素放在一起探讨的理论设想是什么？发现了转换概率和词长期待的交互作用的理论价值和意义是什么？作者没有明确论述。

回应：感谢审稿专家的意见。如您所属，转换概率高低很可能也是一种语言经验，可能在某一种语言体系下当转换概率特别低时，如 0.01 时，头脑才认为存在词边界，而在另外一种语言体系下当转换概率仍旧较高时，如 0.2 时，头脑就认为存在词边界。不过在语料库的研究中，研究者使用多种语言的真实语料，通过设计算法来实现统计学习机制，在不同语言下

都发现当使用绝对边界（转换概率需要达到一个预先设定的阈限值时才被认为是词边界）而非相对边界（转换概率在某一个段语音序列附近较低时被认为是词边界）时统计学习机制对词语的切分效率最高（Saksida et al., 2017）。这一结果在一定程度上说明转换概率高低对统计学习效果的影响很可能独立于语言经验，而是一种自下而上的影响过程。

与此相对，词长期待对统计学习机制的影响是一种自上而下的影响，这一观点在 Lew-Williams 和 Saffran（2012）实验中就被提及，他们发现先给婴儿听一些三（两）音节词后，婴儿只能切分由三（两）音节词合成的人工语言（详见修改后的 P2）。目前来看，从自下而上视角（Palmer & Mattys, 2016; Potter et al., 2017）和自上而下视角（Bonatti et al., 2005; Palmer et al., 2019）都有不少研究关注影响统计学习机制的因素，但鲜有研究将两类因素合并共同探讨。本研究综合考察自下而上的转换概率线索和自上而下的词长期待如何影响语音统计学习任务的效果。从实验结果来看，我们发现两个因素的主效应和交互作用均不显著，但被试在 LTP 条件下的三音节迫选试次中的正确率与随机水平差异不显著，表明在此条件下被试无法有效完成统计学习任务，这暗示词长期待和转换概率很有可能共同对语音统计学习的效果产生影响。在我们的事先对比中，我们也发现了类似的结果，即（1）在三音节迫选试次中，HTP 条件下的被试迫选正确率更高；（2）在 LTP 条件下，被试的完成两音节迫选试次的正确率更高；（3）在 LTP&三音节迫选试次中，被试的迫选正确率也显著低于 HTP&两音节迫选试次中，简而言之，本研究的核心贡献是发现转换概率和词长期待会协同影响语音统计学习效果。

Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive science*, 41(S4), 913–927.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology*, 69(12), 2390–2401.

Palmer, S. D., Hutson, J., White, L., & Mattys, S. L. (2019). Lexical knowledge boosts statistically-driven speech segmentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 139–146.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science*, 16(6), 451–459.

Saksida, A., Langus, A., & Nespor, M. (2017). Co-Occurrence Statistics as a Language-Dependent Cue for Speech Segmentation. *Developmental Science*, 20(3). doi: 10.1111/desc.12390.

意见 2：研究材料虽然都是选取了汉语中不存在一声发音的音节。但是从材料看起来，从音段信息来讲，有些是有更高频的同音段信息的，比如“gei”，“te”，有些是没有的，“rua”，这些是否在三音节、两音节材料中进行了平衡？如果没有平衡是否会对实验结果有影响？

回应：感谢审稿专家的意见。确实如您所说，研究材料虽然都是选择第一声不存在音节，但在音段信息方面有些音节在两音节或者三音节中的出现频率更高，这种情况在目前的实验设计里难以解决。首先，在第一声条件下不存在的音节就不多，应该在 30 个左右，如果匹配

音段信息出现在不同长度词中的频率，那么所剩的音节数量将会更少，很难支撑实验研究。其次，本文参照了为数不多的声调语言相关实验设计，他们也并没有考量无意义音节在真实语言环境中的频率情况，如 Gómez 等人（2017）在粤语的基础上对比了声调变化和声调固定两种条件下统计学习的切分效果，声调变化下每个元音匹配不同的声调，声调固定条件下每个元音匹配第一声，这与本研究一样；同时作者还以普通话音节为材料重复同样的实验，实验中也并未报告实验音节在真实语言环境中的使用情况。最后，我们在实验中设计在 HTP 和 LTP 条件下迫选试次中是同样的目标词和非词，如果某一个音节在三音节词中出现的频率较高或较低，那么在两个转换概率条件下对实验结果的影响应该一致，从一定程度上避免了您所提到的问题。

Gómez DM, Mok, P., Ordin, M., Mehler, J., & Nespors, M. (2017). Statistical speech segmentation in tone languages: the role of lexical tones. *Language & Speech*, 61(1), 84-96.

意见 3：三音节成绩比两音节的成绩差，是否能完全说明词长期待问题？有没有记忆负担的解释？除了需要控制记忆的可能解释之外，如果具有汉语两音节词长语言经验的被试去感知一种外语词汇依然更倾向于在两音节词学得更好，可能更能说明这种语言经验的影响。

回应：感谢审稿专家的意见。我们认为如果这里存在记忆负担因素的影响，那么在 HTP 条件下三音节成绩和两音节成绩应该也存在显著差异，但从实验结果可知 HTP 条件下三音节成绩和两音节成绩没有差异，说明这里记忆负担不是导致出现在 LTP 条件下三音节成绩比两音节的成绩差的主要原因，相关的解释我们在文中补充。同时非常感谢您关于此实验不完善的地方提出的意见，我们也做了一些设想，比如对比普通话母语者完成声音材料的统计学习任务 and 语音材料的统计学习任务，随后厘清记忆负担和词长期待在其中发挥的作用。

意见 4：不知道为什么要把高低转换概率设置成被试间的因素。

回应：感谢审稿专家的意见。在设计实验之初，我们也尝试将转换概率高低设置成被试内因素，在节省的被试的同时还可以有效的减少实验误差。可是，我们发现在普通话中符合实验要求的音节并不多（匹配第一声为无意义音节，匹配其他声调时为有意义音节）。如果还采用被试内设计，那么被试先后完成两种转换概率实验条件间可能会产生干扰，而如果要要求被试间隔一段时间再来完成实验，则会导致被试大量流失，因此我们最终将转换概率高低设置成被试间因素。

意见 5：3.2 用 R 建一个线性模型的作用是什么？比进行方差分析有什么特殊的贡献吗？

回应：感谢审稿专家的意见。在本研究中，我们使用 R 语言建立了两个线性模型，一个是传统的事后检验方差分析，一个是事先对比。前者中，由于方差分析也可以看成是特殊的线性模型，因此我们使用的线性模型结果与方差分析结果没有任何差异，之所有使用线性模型来完成方差分析的工作，主要是考虑到和后面的事先对比一致。

在后一个线性模型中，我们使用了三个对比假设来完成事先对比，分别为（1）在 LTP 条件下，被试完成三音节迫选正确率显著低于两音节迫选，（2）HTP 条件下被试完成三音节迫选的正确率显著高于 LTP 条件下的被试，（3）由于高转换概率和两音节试次都是有助于统计学习效果的因素，因此我们还假设在 HTP 条件下的两音节试次中，被试的正确率显著高于在 LTP 条件下的三音节迫选试次。这三个假设均是来自于我们在前言中阐述的相关内容。

事先对比与常见的事后检验最大的好处在于，可以充分体现研究者的假设构想，可以只针对感兴趣的假设进行检验（Abelson & Prentice, 1997; Wilkinson, 1999），而避免了检验所有假设需要进行一类错误矫正而降低检验力的情况。在本研究中，我们可以看到两个因素的主效应和交互作用均不显著，但事实上从图 3 中可以看到，被试在 LTP 条件下的三音节试次中，迫选正确率与随机水平差异不显著，而且明显要低于其他三种情况，如果仅使用事后检验，在交互作用不显著情况下完成简单效应分析时需要进行多次 α 错误矫正，会降低统计检验力（Schad et al., 2020），因此本研究主要使用了事先对比的方式。

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110. doi: 10.1016/j.jml.2019.104038

Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypothesis. *Psychological Methods*, 2(4), 315-328.

Wilkinson, L. (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, 54(8), 594-604.

第二轮

审稿人 1 意见：修改稿的质量有明显提升，有以下意见或建议。

意见 1：把正确率与随机水平（0.5）进行比较的单样本 T 检验称作“事后检验”是否恰当？事后检验一般指在方差分析结果显著时，进行进一步组间比较或简单效应分析。文章中的“事后检验”似乎不是这个意思，建议重新斟酌这一表达，以避免引起误解。

回应：感谢审稿专家的意见。本文中的“事后检验”确实与方差分析中的事后检验一词冲突，造成误解。本文的事后检验并非方差分析中的事后检验，而是与事先对比相对的概念，包括方差分析、t 检验、卡方检验等一系列统计检验方法。前言部分最后一段第一次出现事后检验的名词，我们在此处增加一处脚注，强调其是与事先对比相对的概念，避免读者混淆。

意见 2：实验中的“非词”是如何生成或设置的，“非词”和“干扰词”是指同一种条件还

是不同条件不清楚。而且，把“目标词”和“非词”放在一起表达，容易让人把“目标词”理解为是真词。

回应：感谢审稿专家的意见。在相关英文文献中，目标词的英文为“target words”，跨界词的英文为“partwords”，非词的英文是“nonwords”，这三种词都是由人工语言中的无意义音节组成的。在统计学习实验的迫选任务范式中经常会使用的两种干扰刺激是跨界词和非词。两者都是基于音节间转换概率而来，跨界词音节之间转换概率不为0，非词内音节间的转换概率为0。在正文第2页脚注部分，我们对非词和跨界词的合成方式进行了简要的介绍。同时我们在材料部分也进行了补充，具体表述为“迫选任务中，两个试次分别为目标词和非词，其中非词的每个音节均来源于不同的目标词，非词音节间的转换概率为0”，例如非词“rerou”中的两个音节分别来自目标词“remei”和“rouse”的第一个音节，在人工语言中这两个音节没有先后连结的可能性，因此转换概率为0。

之前文章“2.2 材料”部分中出现的“干扰词”并不是迫选任务中提到的干扰刺激，这里的干扰词作用是为了降低 LTP 中目标词的转换概率。为了避免读者混淆，我们将本研究中的“干扰词”表述修改为“填充词”。

意见 3：个别文字问题：“‘nue4’为有意以音节”，疑为“有意义音节”。

回应：感谢审稿专家的意见。这里的表述有错误，“nue4”确为“有意义音节”，文字有错误的地方我们已在原文相应位置进行了修改。

审稿人 2 意见：

该研究具体探讨了语音统计学习中转换概率和词长期待的作用。研究有一定理论意义，研究设计合理，结果可靠。修改稿对上次提出的问题进行了阐述和解释，没有进一步问题，同意发表。

回应：衷心感谢审稿专家提出的修改意见以及对修改稿的肯定。

第三轮

编委意见：本文经过两轮审稿，审稿人均认为达到学报发表的水平，有以下修改意见：

意见 1：前言部分对研究的理论意义与实践价值揭示的不够。为什么要研究这一问题，两个自变量在汉语加工中有何作用，研究这一问题对于汉语学习和汉语教学有何价值？

回应：感谢编委专家的意见。

本文考察两个因素对语音统计学习的影响，从“统计学习”的名字来看，这一过程是一

个学习机制，但事实上，这一名称有一定迷惑性，统计学习并非完全是一种学习机制，正文前言中提到 Saffran (1996) 首次在实验研究中证实个体可以利用输入语言音节间的概率关系对语流进行切分，从而实现切分连续口语的过程，并将其定义为统计学习机制 (statistical learning)，从这个角度来看，统计学习机制是口语词切分的过程，这一过程以音节间的概率信息为线索。虽然，大量研究认为统计学习机制是个体 (婴儿、二语学习者) 习得词语含义的必备过程 (Erickson, L. C., Thiessen, E. D., & Estes, 2014)，但是本研究重点关注的是个体利用概率信息对人工语言的切分过程。因此，本研究的价值并非是为汉语学习和汉语教学提供理论支持。本文的意义和创新在于对两个自变量的考察，具体如下：

(1) **转换概率**：转换概率是统计学习机制的核心概念，其公式为，举例子来说，对于词语“葡萄”来说，第一个音节“pu2”后面出现“tao2”的可能性高，这是因为以“pu1”为词语第一个音节的词较少，因此“葡萄”两个音节的转换概率较高；而对于“打水”，这类的词语，“da3”后面还可能出现“人”、“球”等音节，因此“打水”的转换概率较低。

$$TP_{(X|Y)} = frequencyXY / frequency X$$

在统计学习的研究框架内，学者普遍认为个体依据转换概率的高低来切分语流，在所设置的人工语言中，目标词内音节的转换概率较高，目标词与目标词连接处的转换概率较低 (参考图 1)，因此个体会切分出人工语言中的目标词。不过，这类研究均是在一个人工语言内部通过设置目标词和跨界词不同的转换概率来验证统计学习机制的存在，**相关研究还没有考察当转换概率发生变化时，个体是不是对高转换概率的目标词切分效果更好，从统计学习理论来看，转换概率既然是统计学习发生的根本 (个体识别音节间的概率分布关系)，那么我们有理由推断这一因素会影响统计学习或词切分的效果。**同时，现有一些视觉统计学习研究对这一问题进行考察，但大量研究都证实视觉统计学习和听觉统计学习的相关性较低，**两种模态下的统计学习机制很可能存在领域特异性，本研究的目的是在语音统计学习任务中，考察转换概率高低统计学习机制的影响，从而加深对统计学习机制的理解。**这些内容在前言第三段中，都有相应的介绍，这次修改中，我们进一步进行了阐释，期待达到编委专家的要求。

(2) **词长期待**：词长期待是本文考察的第二个因素，是纳入在语言经验这个大的主题下的。在经典的统计学习研究中，研究者的目的是关注个体能否利用人工语言中编制的概率信息实现词切分过程，因此往往假设被试没有任何语言经验，或者语言经验对结果并没有明显的影响 (于文勃等，印刷中)。但是今年以来，越来越多的学者关注语言经验对统计学习机制的影响，这类研究更多是考察语言经验如何影响个体利用概率信息切分语流的过程。以往的研究中也有一些研究者关注到词长期待对统计学习机制影响 (Lew-Williams & Saffran, 2012)，但迄今为止还没有研究者关注普通话母语者的词长期待是否会影响统计学习机制。对这一问题的考察一方面可以看作是对语言经验和统计学习相关研究的一个补充，另一方面是对普通话“双音化”倾向的一种佐证。

最后，从加工方式来看，研究者普遍认为个体需要追踪音节间的转换概率来切分语流，那么转换概率是一种自下而上的信息；与此相对，语言经验是较为高级的线索，也就是自上而下的信息，本研究结合这两个因素可以看作是对统计学习机制研究中自上而下信息和自下而上信息的综合考量。

相关内容，我们在正文中进行重新表述并标注，希望达到编委专家的要求。

Erickson, L. C., Thiessen, E. D., & Estes, K. G. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language*, 72, 49–58.

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: expectations about word length guide infant statistical learning. *Cognition*, 122(2), 241-246.

意见 2：事先对比与事后检验的概念还是显得混乱，读后不知所云，建议作者清楚表述。

回应：感谢编委专家的意见。

总体来看，数据分析方法可以分为事先对比和事后检验两种。事先对比是对研究提出的实验假设进行检验的方式，并不需要对所有假设进行检验，而事后检验是对所有可能出现的对比都进行检验，但其实研究者可能只对其中某一对或者几对对比感兴趣，事后检验带来的额外检验可能会使统计检验力受到影响。

在本研究中使用了事先对比和事后检验两种方式。基于对以往相关研究的理论（Lew-Williams & Saffran, 2012）和普通话母语者韵律期待的思考，我们认为：三音节词违反了普通话母语者的词长期待，因此在概率词切分的过程中更难被切分出来；在低转换概率的条件下音节间的连结较弱，目标词切分出来的难度更大，由此在这两个前提下形成了三个事先假设，并使用事先对比对本研究的这三个假设进行验证。在事后检验中，由于统计学习研究需要将被试迫选正确率与随机水平（0.5）比较，以判断被试是否能够切分出人工语言中的目标词，因此我们需要使用事后检验来完成相关数据分析。

从结果来看，在事先对比中，LTP 条件下三音节词的正确率要显著低于 HTP 条件下三音节词的正确率，表明低转换概率确实会增加概率词切分的困难；LTP 条件下三音节词正确率边缘显著低于 LTP 条件下两音节词正确率，同时显著低于 HTP 条件下两音节词的正确率，说明特别是在 LTP 条件下三音节词更难被切分出来。在事后检验中，被试只有 LTP 条件下的三音节试次迫选正确率和随机水平没有显著差异，说明被试对于低转换概率条件下的、不符合词长期待的目标词学习效果较差。由此可以看到事先对比和事后检验得到的结果是一致的。

意见 3：文中的中文文献作者之间不要用&。

回应：感谢编委专家的意见。已将文中引用的所有中文文献中的作者之间&改为逗号。

意见 4：图 4 没有必要。

回应：感谢编委专家的意见。我们已经在文中删除图 4，并在相应部分（P12）重新表述了这张图片所反映内容。

意见 5：文章的标点符号运用很不规范，许多应该有逗号的地方没有，建议作者找一中文好的人审读。

回应：感谢编委专家的意见。我们再次对全文进行审读，修改了其中标点符号不规范的地方，并请一位文学院语言学专业背景学生进行了审读和修改。

意见 6：对自下而上加工和自上而下加工的表述不清晰，词长预期和转换频率各属于什么加工，应该指出。

回应：感谢编委专家的意见。转换概率是在语音统计学习领域可以体现为一个音节刺激（X）后接一个特定音节刺激（Y）的概率，反映的是两个相邻音节共同出现的可能性。一个单词内部的转换概率要比两个单词之间的音节转换概率高，在研究中被试可以通过追踪词内和词间的转换概率来完成对词的切分，我们认为转换概率是一种外部刺激推动了概率词切分的加工过程，这可以被认为是一种自下而上的加工。

与转换概率相对，词长期待属于人的语言经验，比如普通话母语者的“双音化倾向”，这是人们在特定语言环境中习得的、普遍存在于这一群体中的知识或者说经验。词长期待作为人们已有的知识和经验，人们可以利用这种自身已有的信息来帮助他们进行概率词切分，由此词长期待可以被认为是一种自上而下的加工。

我们在前言中对此进行了补充说明，希望能达到编委专家的要求。

第四轮

编委意见：作者修改了论文。建议刊用。

回应：非常感谢编委专家对修改稿提出的宝贵意见，以及对修改稿的肯定。

第五轮

主编意见：这篇文章经过修改后已经比较成熟了，没有额外意见，建议发表。

回应：非常感谢主编对修改稿提出的宝贵意见，以及对修改稿的肯定。