

《心理学报》审稿意见与作者回应

题目：语境性别影响元音和辅音范畴感知的计算建模研究

作者：刘顺，王穗苹，黄健

第一轮

审稿人 1 意见：

本文通过两个听辨实验，分别考察环境音性别对元音共振峰和辅音 VOT 范畴感知的影响，采用 LBA 模型验证混合效应模型，对实验作答的选项和反应时进行计算建模分析。论文整体语言流畅，格式规范，数据详实。以下几点意见和建议供作者参考修订：

意见 1：传统的语音学实验证明，元音是不存在范畴感知的辨别和区分边界的，跟本文的结果相反，建议对照 Strange (1995) Figure 2，解释本文中图 3 的结果。

回应：非常感谢审稿人提出的这个核心理论问题。您指出的关于元音范畴感知的经典观点，以及具体提到的 Strange (1995)文献，促使我们对本研究结果的理论定位进行更清晰的阐述和必要的文献澄清。

首先，关于您提到的 Strange (1995) 的 Figure 2，我们经过仔细核查，在 Language and Speech 期刊上找到题为“Learnability Constraints on Theories of Speech Perception: From Spoken Words to Speech Sounds”的这篇短文（1995, 38(2), 217-222）中，并无图表。而另一本著作《Speech perception and linguistic experience: issues in cross-language research》（York Press, Baltimore, 1995）我们目前无法获取全文。因此，我们无法针对该特定图表进行直接回应。

我们理解，您的核心关切在于引述支持“元音感知更具连续性、范畴性较弱”这一经典论断的权威观点，确实，有相关的研究表明元音的范畴感知相对辅音的范畴感知边界没那么分明(如, Fry et al., 1992)。不同元音对的连续体的范畴感知程度也不同，确实有一些元音对的感知函数更接近线性(如, Steffman, 2023)。我们完全认同这确实是语音感知领域一个长期存在的理论争论。然而，相比于塞音 VOT 的范畴化程度较高(陈飞 等, 2019)的现象，元音感知模式可能高度依赖于任务要求(McMurray, 2022)。元音感知时表现出多大程度的“范畴性”，可能更容易会受到实验任务、声学连续体的设计、所对比的元音范畴在特定语言中的关系等多种因素的高度调节。例如早期采用基于辨别任务的经典研究常发现元音的

感知连续性较强，范畴边界不如辅音陡峭(Fry et al., 1962)。而在更为复杂的范畴化分类任务（即本研究采用的任务）中，某些特定元音对的连续体的识别函数完全可能表现出显著的范畴化特征(Steffman, 2023)。而我们研究中的图 3 显示出的元音的范畴化的结果可能也是由于采用了类似的范畴化分类任务的结果，这与许多采用类似任务的元音研究结果是一致的(例如，Fry et al. 1962; Pisoni, 1973)。范畴化分类任务要求被试强制将刺激归入给定的离散语音标签（如[a]或[e]），这可能是导致了元音识别中更容易出现陡峭的、范畴化的边界的元音。这个结果表明，元音识别过程中可能更容易收到外在因素的影响。因此元音和辅音识别对于外在环境的影响的不同差异是否同样会反映在说话人归一化效应上，则并不清楚，这也构成了本研究的一个重要的研究出发点。

为了更全面、公允地呈现这一理论背景，并直接回应您的宝贵意见，我们在论文的前言部分增加了一段专门的论述，内容如下（见正文第 3 页，第 1 段）：

“此外，说话人归一化是否会对不同的音素产生不同的影响，同样是一个尚未解决的问题。语音系统中存在两大核心音素：元音和辅音。元音主要依赖于前三个共振峰的频率信息 (Oganian et al., 2023)，本质上是对频率线索的感知。相比之下，辅音的区分线索更为多样。例如，区分塞音清浊的浊音起始时间（VOT）则代表了语音信号中另一种关键的声学线索—时间信息(Toscano & McMurray, 2015)。之前的研究发现，塞音 VOT 的范畴化程度较高，可能不容易受环境影响(陈飞 等, 2019)，但关于元音感知的范畴化则更会受到外在因素的影响产生不同的表现。具体来说，基于辨别任务的经典研究常发现元音的感知连续性较强，范畴边界不如辅音陡峭(Fry et al., 1962)，而后续研究认为，元音感知模式高度依赖于任务要求(McMurray, 2022)，在识别或分类任务中，某些特定元音对的连续体的识别函数完全可能表现出显著的范畴化特征(Steffman, 2023)。因此元音和辅音之间对于外在因素影响的这种差异，是否同样会反映在说话人归一化的效应上，依然是不清楚的研究问题。”

意见 2：从汉语“g“到”k“的 VOT 递增连续统，只做了五阶，样本量不足以表明范畴感知的边界，普通话”k“的 VOT 长到 100ms 都没问题，建议再增加四阶，这样也与元音共振峰九阶平衡。

回应：感谢审稿人的宝贵建议。我们完全认同 5 阶连续体在描绘范畴感知边界时的局限性。为此，我们严格按照审稿人的建议，制作了 VOT 跨度更广的 9 阶 g-k 连续体（VOT 步长与元音共振峰等级保持平衡），并重新招募被试收集了数据。

结果显示，9 阶连续体产生的感知曲线更完整地覆盖了范畴转换区域，更重要的是，

其效应模式与原 5 阶实验高度一致，这进一步验证了本研究结论的稳健性。我们在修订稿的“方法”和“结果”部分已全部替换为 9 阶实验的数据。并同时将 5 阶连续体的数据都放在开放科学框架网站。

意见 3：本文研究环境音性别对范畴感知的影响，但是元音的目的音只用了男性发音人，辅音目的音只用了女性发音人，完整的做法应该是两个实验的目的音均用上两个性别发音人，并且元音和辅音的目的音均由这两位发音人录制。

回应：感谢审稿人的建设性意见。我们已根据要求补充了缺失的实验条件（女声元音与男声辅音）。

由于未能再次招募到原始录音人，此外，考虑到元音和辅音实验，男性和女性目的音实验均采用的是被试间设计，因此，我们招募了新的录音人制作了缺失的实验材料，并招募了足够数量的被试进行实验。我们认为招募不同的录音人可能也会对该研究产生另外的优势。具体来说，不同的录音人存在不同的声学特点，如果我们可以发现类似的结果，或许可以更好地说明本研究的结论可能更具有推广性。

在这个基础上，我们通过将被试在四种条件下的反应进行对比分析，我们发现前置环境音的性别效应对不同性别的目的音均产生了显著影响，且未发现目的音性别与环境音性别的显著交互作用。这一发现增强了本研究结论的泛化性。修订稿中已呈现了这四个实验的完整数据集及对比图表。

意见 4：实验一的听者女性 16 人，男性 8 人，实验二的听者女性 18 人，男性 6 人，被试性别不平衡。社会语音学发现男女听者分别听辨男女发音是有差异的，更何况本文讨论的是环境音性别对目的音范畴感知的影响，更需要在被试上平衡性别。

意见 5：基于上述第 4 点，建议在混合效应模型里把听者性别作为一个固定因素考虑。LBA 的建模分析，也要考虑这个因素。

回应：诚如审稿人所言，以往研究确实发现被试性别对语音感知存在一定的影响。我们已针对这一问题补充了实验数据，目前各实验的男女被试比例已平衡（具体人数见修订稿各实验的“被试”部分，正文第 6、13、15、18 页）。

补充数据后的结果表明，我们的核心实验效应（环境音性别效应）在平衡了性别比例后，与之前的结果相比，并没有明显变化。此外，我们接受审稿人的建议，通过在统计模型中将被试性别作为控制变量进行处理，有效地剥离了听者自身性别带来的方差，从而更

纯粹地验证了前置环境音的跨性别普适效应。

意见 6：LBA 模型并没有得出与 mixed-effects 模型相反的结果，那么第 20 页第一行的结论“对辅音 VOT 的感知则未发现环境音性别的显著影响”是如何得出的，表 4 中性别主效应高度显著，而且不存在 VOT 和性别的交互效应，结论和统计结果不相符。

回应：非常抱歉给您造成的困扰。根据最新的实验结果，我们已经修改了相关的讨论内容。这里提到的“对辅音 VOT 的感知则未发现环境音性别的显著影响”已经不再适用，因此修订稿中已经没有相关的表述。

意见 7：元音目的音时长 120ms，先行语后 140ms 播放目的音，这些数据最好由参考文献支撑，或者说明是否通过预实验获得。

回应：非常感谢您提出的这一关于实验参数细节的专业建议。您所关注的刺激时长和呈现间隔确实是确保实验有效性和可比性的关键参数。

目标刺激的时长（120 毫秒）和先行语与目标音之间的间隔（140 毫秒）的确定基于之前的相关研究（唐俊杰，2015，Liu & Liu, 2025）。基于这些设置，前人研究展示了对语音连续体的典型范畴化反应和显著的环境性别效应。因此我们制作了时长为 120ms 的语音连续体，并将前置音与目标音的设置为 140 毫秒。

以上内容，我们在 2.1.3 实验材料和程序部分进行了补充说明（见正文第 7-8 页）：
“最终形成 9 个水平的元音连续体，并将声音标准化为长度 120 毫秒(参考 S. Liu & Liu, 2025) “
“目标刺激长度和先行语和目标音之间的时间间隔参考了前人相关研究的设置(S. Liu & Liu, 2025)，相同的设置观察到了显著的前置性别效应。”

意见 8：文中辅音目的音的[ge]和[ke]到底是汉语拼音还是国际音标？如果是汉语拼音，请增加国际音标标注；如果是国际音标，那么[g]必须是手写体的，然而汉语普通话没有浊音的[g]，汉语拼音 g 的国际音标是不送气清音[k]，而汉语拼音 k 的国际音标是送气清音[kʰ]。

回应：非常感谢您的建议。我们已经将文中的语音标注统一为国际音标，即[k]和[kʰ]（例如正文第 15 页，倒数第 2 段）。

.....

审稿人 2 意见：

这篇论文在探讨语境性别信息对语音范畴感知的影响方面提供了重要的理论贡献，采用计算建模方法揭示了多重认知机制，突破了传统的“说话人归一化”理论框架。研究结果不仅丰富了语音感知领域的理论基础，还为理解语音感知的实时认知计算过程提供了创新视角。作者对实验设计和数据分析的严谨态度也值得肯定。但是这项研究在被试量、实验刺激的设计、及语言表达等方面需要进行大修。希望作者根据以下意见进行修改，以提高论文的质量。

意见 1：在审阅过程中，我注意到作者在确定被试量时是依赖于他们团队以往一篇脑电研究的样本量，而并未进行功效分析。虽然该样本量是基于 Liu & Liu (2025)的研究，该研究招募了 24 名参与者并发现了环境性别对语音感知的显著影响，但考虑到说话人归一化的感知结果存在较大个体性差异，24 名被试是否足以支持行为学研究的结论仍需进一步探讨。建议作者考虑增加被试量，以提高研究结果的可靠性和普适性，并进行相应的功效分析，以确保样本量能够有效捕捉到环境性别对语音感知的影响。这样不仅能增强研究的科学性，也能为后续研究提供更坚实的基础。

回应：感谢审稿人的建设性意见。我们完全认同行为学研究在处理说话人归一化等具有较大个体差异的课题时，需要更充足的样本量以确保结果的可靠性。

遵照建议，我们扩大了样本规模，目前每个实验的有效被试量均达到了 30 人以上（总样本量显著增加）。同时，我们使用 R 语言的 `simr` 包进行了功效分析。结果表明，在当前的样本量下，前置性别主效应的统计功效超过了 0.80。此外，增加样本量后的结果与原实验结果保持高度一致，这进一步增强了本研究结论的可靠性。相关统计描述与功效分析结果已补充至修订稿的“被试”部分（见正文第 6、13、15、18 页）。

意见 2：作者在两项研究中所使用的目标刺激的发音人性别存在不一致性。具体而言，实验 1 中使用了男性的目标元音语音，而实验 2 则使用了女性说话者录制的目标 VOT 刺激。这种不一致性可能会影响实验结果的可比性和解释性。（作者解释这是为了增强结果的泛化性，这个解释是行不通的）

对于元音的目标刺激而言，男性和女性的共振峰频率确实存在显著差异。因此，为了确保实验设计的全面性和结果的可靠性，建议实验 1 同时制作男性和女性版本的目标元音刺激。这将有助于更全面地考察环境性别对语音感知的影响，并能够更好地控制发音人性别所带

来的潜在干扰。

回应：感谢审稿人的深刻评述。我们诚恳地接受您的建议，并认识到在原设计中，目标刺激性别的跨实验不一致性确实可能削弱结果的可比性，尤其是男女声在共振峰和 VOT 基础特征上的生理差异可能引入额外的干扰变量。

为此，我们重新设计并补充了完整的平衡实验。在修订稿中，我们将研究扩展为四个子实验：

元音感知实验（实验 1a 和 1b）：男声目标音和补充录制的女声目标音。

辅音感知实验（实验 2a 和 2b）：女声目标音和补充录制的男声目标音。

这种（前置环境性别：男/女） $\times$ （目标音性别：男/女）的全交叉设计，使我们能够独立析取出环境性别的效应，并排除了目标音性别作为混淆变量的可能性。

结果显示，无论目标音是男声还是女声，前置环境性别均诱发了显著且方向一致的范畴感知偏移。这不仅解决了实验设计的一致性问题的，也通过多维度的验证，真正实现了研究结论的泛化性。相关数据和对比分析已在修订稿的“实验方法”和“结果”部分详细呈现。

意见 3：在讨论部分，结果发现对辅音 VOT 的感知则未发现环境音性别的显著影响。这其实有可能是因为目标刺激连续统本身的范畴化程度高低是有差异的。（参考：陈飞, 张昊, 王士元, 彭刚. (2019). "内部因素与元音范畴化感知". 《语言科学》. 18(4): 410-425.）。一般而言，塞音 VOT 的范畴化感知程度较高，被试对于其判断自然较少受到语境信息的影响；而单元音的范畴化感知程度较低（甚至有很多学者认为是元音的感知是非范畴化的），因此受到语境信息的影响可能较大。这一点作者可以在结果讨论部分做补充。

回应：感谢审稿人提供这个深刻理论视角。我们非常赞同关于“辅音与元音范畴化程度差异”的论述。参考陈飞等（2019）提到的，塞音 VOT 的强范畴化特征可能使其感知边界比元音更具鲁棒性，从而对语境信息的敏感度相对较低。我们认为审稿人的这个观点，也为本研究提供了另外一个重要研究出发点：研究说话人归一化这一外在语境因素是否在元音和辅音范畴化是否存在差别，我们也将这一重要的观点放在前言部分，内容如下（见正文第 3 页，第 1 段）：

“此外，说话人归一化是否会对不同的音素产生不同的影响，同样是一个尚未解决的问题。语音系统中存在两大核心音素：元音和辅音。元音主要依赖于前三个共振峰的频率信息 (Oganian et al., 2023)，本质上是对频率线索的感知。相比之下，辅音的区分线索更为多样。例如，区分塞音清浊的浊音起始时间 (VOT) 则代表了语音信号中另一种关键的声

学线索—时间信息(Toscano & McMurray, 2015)。之前的研究发现, 塞音 VOT 的范畴化程度较高, 可能不容易受环境影响(陈飞 等, 2019), 但关于元音感知的范畴化则更会受到外在因素的影响产生不同的表现。具体来说, 基于辨别任务的经典研究常发现元音的感知连续性较强, 范畴边界不如辅音陡峭(Fry et al., 1962), 而后续研究认为, 元音感知模式高度依赖于任务要求(McMurray, 2022), 在识别或分类任务中, 某些特定元音对的连续体的识别函数完全可能表现出显著的范畴化特征(Steffman, 2023)。因此元音和辅音之间对于外在因素影响的那种差异, 是否同样会反映在说话人归一化的效应上, 依然是不清楚的研究问题。”

在我们的第一版稿件中, 由于 VOT 连续体阶数较少 (5 阶) 且样本量有限 (24 人), 统计效能可能不足以稳定捕捉到环境音性别对于辅音 VOT 的影响, 导致混合效应模型结果仅呈边缘显著。

为此, 我们在修订版本中进行了两项关键改进: 一是优化刺激物: 将 VOT 连续体由 5 阶扩展至 9 阶, 更精细地刻画了范畴转换区域; 二是提升统计效能: 将被试量增加至 30 人。

新的实验结果显示, 在提升了实验灵敏度后, 混合效应模型与 LBA 建模分析均一致发现了显著的环境性别效应。这表明环境音性别对辅音感知同样存在影响, 只是其效应量级可能受辅音强范畴化特性的制约。与此同时, 我们发现环境音性别对元音和辅音的影响有相似的机制, 同时也会存在语音特异性的不同机制。结合我们的发现, 我们也重新组织了总讨论 (见正文第 21-23 页)。

意见 4: 该论文的语言表达存在一些问题, 部分内容显得偏向于英文的直接翻译, 导致语句不够流畅。此外, 某些表达方式显得口语化, 缺乏学术论文应有的正式性和严谨性。

回应: 非常感谢审稿人的建议。我们对全文进行了重新地修改, 以使表达更加流畅、正式、严谨。

意见 5: 该论文的篇幅过长, 可以考虑做一些精简和删除。另外, “确定他们听到的刺激是 [ge] 还是 [ke]”, 按照语音学规范, [] 中的标注应该是国际音标 IPA, 作者却使用的是拼音。

回应: 非常感谢审稿人的建议。我们删除了一些不必要的图表, 并且将一部分数据结果放到了补充材料里面 (存放于数据公开网站), 以便读者需要获取相关内容。此外, 我们精简了方法和结果的陈述, 在保证透明性的前提下尽量简洁。最后, 我们对前言, 尤其是讨论进行了一定的精简, 在确保不损失论文完整性的前提下尽量缩短篇幅。

另外，我们已经将文中的语音标注统一为国际音标，即[k]和[kʰ]（例如正文第 15 页，倒数第 2 段）。

第二轮

审稿人 1 意见：

该论文的修改稿在样本量、数据量、文献综述和分析讨论各层面都较上一稿有了显著的提高，审稿人的几点修改意见主要集中在讨论部分。

意见 1：该文采用的是范畴感知的方法测试传统混合效应模型和 LBA 模型的语音环境性别感知效应。范畴感知用在辅音上主要就是塞音的 VOT 或者 POA（place of articulation）。既然作者选择了调整元音/a/到/e/的 F1 做范畴感知，是有意还是无意选择作者所说的对识别说话人性别信息作用不大的 VOT，为什么不做 POA 的 F2 调整，比如/pa/到/ta/的调整？这样不就是同一个共振峰层面上的测试了吗？这个问题，需要在实验设计出现前就说清楚，同时在讨论部分解释。

回应：感谢审稿专家提出这一关键问题。本研究选择元音 F1 连续体和辅音 VOT 连续体是有意设计，而不是无意忽略 POA/F2。我们的核心目的并不是比较元音和辅音在同一声学维度上的语境性别效应，而是考察语境性别线索是否会影响依赖不同基础声学线索的语音范畴感知过程。元音 /a/-/e/ 的 F1 连续体主要反映共振峰频率线索，而塞音 VOT 连续体主要反映时间线索。二者分别代表语音感知中较典型的频率维度和时间维度，因此有助于检验语境性别效应是否具有语音特征特异性。

同时，审稿专家指出的 POA/F2 连续体确实是一个非常有价值的设计。若采用 /ba/-/da/ 等通过 F2 或共振峰过渡线索构建的 POA 连续体，就可以检验语境性别对辅音共振峰相关线索的影响，从而区分“元音-辅音差异”与“频率线索-时间线索差异”。但这将对应另一个更细化的问题，即在同属频率/共振峰线索的条件下，语境性别效应是否仍因音类不同而变化。我们将在修订稿中将这一点作为本研究的边界和未来研究方向加以说明。

根据该意见，我们在引言进入实验设计前补充说明：

“我们关注了语音系统中两个核心且具有代表性的元素：元音共振峰（实验 1a, 1b）和辅音 VOT（实验 2a, 2b）。本研究选择元音共振峰和辅音 VOT 作为考察对象，是基于两方面考虑。首先，元音和塞音分别是语音系统中最常被用于范畴感知研究的两类材料；其次，二者所依赖的主要声学线索不同：元音/a/-/e/连续体主要通过 F1 变化体现共振峰频率

线索，而塞音清浊范畴主要依赖 VOT 这一时间线索。因此，本研究的目的是并不是在元音和辅音之间匹配完全相同的声学维度，而是考察语境性别线索是否会影响依赖不同基础声学线索的语音范畴感知。”

其次，我们在讨论（6.1）中再次说明了“为什么 VOT 对性别识别作用不大，仍值得做”：

“需要说明的是，VOT 并不是说话人性别识别中稳定的、主要的线索。正因如此，VOT 连续体在本研究中具有特殊意义：它可以检验语境性别效应是否只局限于与性别高度相关的频率线索，还是也会扩展到与性别关系相对较弱、但在塞音范畴感知中具有核心作用的时间线索。如果语境性别只通过直接匹配性别相关声学维度发挥作用，那么 VOT 连续体上的效应应当较弱或不稳定；而本研究发现 VOT 范畴判断同样受到语境性别影响，说明语境性别对语音感知的作用可能并不限于低水平声学匹配(Johnson & Sjerps, 2021)。”

最后，我们在研究展望中补充了 POA/F2 的研究方向及潜在收益：

“此外，本研究尚不能区分辅音内部不同声学线索的语境性别效应。未来研究可以进一步采用/ba/-/da/等 POA 连续体，通过操控 F2 过渡或相关共振峰线索，检验当辅音范畴同样依赖频率/共振峰维度时，语境性别效应是否仍不同于元音范畴。这将有助于进一步区分“元音-辅音差异”和“频率线索-时间线索差异”在说话人归一化中的作用。”

意见 2：建议题目修改成“语境性别影响元音和辅音范畴感知的计算建模研究”。原题的“环境”指的是语音环境，需要表述清楚；范畴感知本身就需要设计“连续体”作为实验刺激，无需指出。

回应：感谢审稿专家的建议。根据该建议，我们已将题目由“环境性别影响元音和辅音连续体范畴感知的计算建模研究”修改为“语境性别影响元音和辅音范畴感知的计算建模研究”，以更准确地概括本文的研究对象和方法。

意见 3：讨论部分很唐突地冒出新文献，并且没有展开解释。比如，6.2 最后一段“它既支持了“早期作用观”(Watkins & Makin, 1994)，也支持了“晚期作用观”(Bladon et al., 1984)”，前后都没有就这两个文献对实验结果的解释作用展开讨论。

回应：感谢审稿专家指出这一问题。根据该意见，我们修改了在引言中对这两个文献的介绍，以更明确地说明两类理论观点的核心差异：Watkins 和 Makin(1994)代表的早期作用观点强调前置语音的声学特征可能在目标语音形成明确类别判断之前影响其听觉表征；

Bladon 等人(1984)代表的晚期作用观点则强调听者会根据说话人特征调整语音表征或范畴判断标准。

“关于说话人归一化的作用阶段，已有研究存在不同解释。一类观点强调较早期的听觉-感觉编码机制，认为前置语音的声学特征会改变随后目标语音的听觉表征，从而在语音类别判断之前即产生作用(e.g., Watkins & Makin, 1994)。另一类观点则强调较晚期的表征调整或范畴决策过程，认为听者会根据说话人特征调整语音范畴边界或判断标准，从而影响最终的类别反应(e.g., Bladon et al., 1984)。因此，仅凭类别反应比例很难判断语境性别究竟影响的是早期感觉编码、后期类别决策，还是二者之间的信息转化过程。”

同时，我们也扩展了总讨论中 6.2 节的相关内容，将 LBA 模型参数与上述理论解释更清楚地对应起来。具体而言，非决策时间包含感觉编码和运动反应等成分。由于本研究中不同语境性别条件下的反应方式和按键要求完全相同，我们参照 Tillman 等人(2017)的做法，将语境性别对 t_0 的影响主要解释为目标语音早期编码过程的变化。基于这一解释，当前 t_0 结果与 Watkins 和 Makin(1994)所强调的早期听觉加工观点相一致。另一方面，语境性别对决策阈值 (b) 的影响表明，听者在形成类别反应时所需的证据量也会随语境变化而调整，这与 Bladon 等人(1984)所强调的说话人信息参与语音表征调整和类别判断过程的观点相一致。此外，语境性别对漂移率 (v) 的影响还说明，语境信息会影响听觉信息转化为类别决策证据的效率。通过上述修改，我们希望更清楚地说明本研究结果与既有理论观点之间的关系，并避免原稿中简单罗列文献而讨论不足的问题。

“其中，非决策时间包含感觉编码和运动反应等成分。由于本研究中不同语境性别条件下的反应方式和按键要求完全相同，我们参照 Tillman 等人(2017)的解释，将语境性别对 t_0 的影响主要解释为目标语音早期编码过程的变化。基于这一解释，当前 t_0 结果与 Watkins 和 Makin(1994)所强调的早期听觉加工解释相一致，即前置语音的声学特征可以改变随后目标语音的听觉表征。与此同时，语境性别对决策阈值的影响表明，听者在形成类别反应时所需的证据量也会随语境变化而调整，这与 Bladon 等人(1984)强调的说话人信息参与语音表征调整和类别判断过程的观点相一致。此外，漂移率的变化说明语境性别还会影响听觉信息转化为类别决策证据的效率。由此看来，语境性别对语音范畴感知的影响并不局限于单一加工阶段，而可能同时体现在早期感觉编码、证据累积和后期类别决策等多个环节。”

意见 4：讨论的句子表达不够严谨，6.3 第三段“元音识别中往往依赖共振峰频率信息，而

女声语音相比于男性语音，具有更大的声学空间和更高的基频动态(Hillenbrand & Clark, 2009)”，既然前面提到的元音识别依赖共振峰频率，后面又为何提到基频？男女声差异确实集中在元音声学空间和基频高度，但是元音识别与基频高度并没有关系。

回应：感谢审稿专家指出这一问题。原文在讨论女声目标元音可能提供更充分声学信息时，同时提到“元音声学空间”和“基频动态”，容易造成基频直接参与元音类别识别的误解。事实上，本文此处想强调的是：元音范畴识别主要依赖 F1、F2 等共振峰频率及其相对分布；女性说话人的元音通常具有更高的共振峰频率和更扩展的元音声学空间，因此女声目标元音在共振峰维度上可能提供了更充分的类别区分线索。基频可以作为说话人性别识别的重要线索，但并不是本文解释元音范畴识别差异的核心依据。

根据该意见，我们已删除该处关于“基频动态”的表述，将讨论重点限定在元音声学空间和共振峰线索上。相应句子修改为：

“元音识别主要依赖 F1、F2 等共振峰频率及其相对分布。相比男性语音，女性语音的元音通常具有更高的共振峰频率和更扩展的元音声学空间(Hillenbrand & Clark, 2009)，因此当被试识别女声目标元音时，目标语音本身可能已经提供了较为充分的共振峰类别线索，前置语境性别线索的额外调节作用因而可能减弱。”

意见 5：文中多处表达不清或者选字错误。不知道作者是否借助 AI 写作或者是机器翻译的结果，比如 6.3 的第一段“因此这些跨语言，跨音素的一致结果似乎提示了一种较为一般的语音感知机制：环境语音对于语音感知的影响，是通过直接影响语音识别中将听觉信息转为为类别决策证据的效率而实现的”，这一句里“跨语言”后面应该是顿号，“跨音素”应该是“跨音类”，“似乎提示”不是规范的学术写作语言，还有“一种较为一般的”这样的表达也很不规范，最后一个分句的修饰语太长，很难断句，像是英文的句法。6.1 第二段“这一结果一定程度上支持了”的第一个“支”和 6.3 第二段“只依赖少量的证据既可以做出准确的判断”的“既”字都是错别字。

回应：感谢审稿专家指出文稿中的语言问题。根据该意见，我们对全文进行了校对，并重点修改了讨论部分中表达不够严谨和不符合中文句法的句子。

具体而言，我们已将“跨语言，跨音素”修改为“跨语言、跨音类”；

将“似乎提示”、“一种较为一般的”等不规范用语删除；

将原本的长句“环境语音对于语音感知的影响，是通过直接影响语音识别中将听觉信息转为为类别决策证据的效率而实现的。”重写为“语境性别对语音感知的影响可能部分体

现为对证据转化效率的调节 “;

将“这一结果在一定程度上支持了”修改为“这一结果在一定程度上支持了”;

将“只依赖少量的证据既可以做出准确的判断”的 “既 “修改为“即”。

同时，我们也拆分并重写了 6.3 节相关长句，使讨论中的逻辑关系更加清楚。例如，

“尽管 LBA 模型在四个实验中一致支持 vbt0 模型，表明环境性别对元音和辅音感知均具有多阶段调制作用，但进一步分析参数变化的具体模式后可以发现，元音与辅音之间也存在一定共性。首先，漂移率在元音共振峰和辅音 VOT 任务中均受到语境性别调节，说明环境性别可以影响听觉信息转化为类别决策证据的效率。Tillman 等人(2017)在英文元音识别研究中也发现，声学线索主要通过调节漂移率影响元音类别判断。结合本研究结果来看，这种跨语言、跨音类的一致性提示，语境性别对语音感知的影响可能部分体现为对证据转化效率的调节。”

“基于 LBA 模型，非决策时间的延长表明被试在早期的感觉编码阶段时间更长，而更低的决策阈值则表明，被试在后期决策过程中只需积累较少的证据即可做出判断，表现为决策标准降低，反应效率更高。”

.....

审稿人 2 意见:

作者对我的意见做出了较大的修改，包括增加了被试量以提高统计效力，补充了新的实验材料和实验数目（保证元音和辅音实验都有男性和女性发音人的刺激），我对于作者的认真修改非常满意，建议接受发表。

回应：非常感谢审稿人对我们工作的认可。

第三轮

编委意见:

作者针对两位评审专家的意见进行了两轮认真而全面的修改，很好地回应了评审专家的问题，并对文章做了相应的修改。修改后的文章具有较好的理论价值，建议接受发表。

主编意见:

同意发表。