

《心理学报》审稿意见与作者回应

题目：汉语塞-元-塞音序列语境效应机制探讨

作者：刘文理 周详 张明亮

第一轮

审稿人 1 意见：

意见 1：本研究通过控制语境音与目标判断音之间的关键声学线索的异同，考察了语境音影响目标音识别的内在机制。研究结果支持了听觉理论对语境效应的解释。研究问题明确，研究结果清晰。但是本研究也有一些需要进一步澄清的问题。

第一，作者在引言部分详尽论述了语音识别的语境效应的两个影响因素：协同发音的知觉弥补；听知觉理论。但是在 3 个实验中，研究者并没有控制和考察“协同发音”这一因素。另外，“发音特征”和“频谱对比”都涉及到语境音的声学参数的变化，“发音特征”和“频谱对比”假设的异同是什么？

回应：在自然言语中，协同发音效应是存在的，如分别在音节/pa/、/pi/和/pu/后面产生的/ta/的发音部位和声学线索会有些细微的区别，这是由于前面三个音节的发音部位差异所产生的协同发音效应。同样/ka/的产生也会受到前面语境的影响。但在语境效应实验中，为了更好的控制声学线索的变化，研究者通常会合成一个声音连续体，如通过变化 F3 起始频率合成/ta-/ka/连续体，连续体上声音的其他声学参数都相同。然后以此连续体作为目标音，考察不同的语境对识别这个连续体上声音的影响（Mann, 1980; Lotto & Kluender, 1998）。因此实验中通常不再具体考察语境协同发音对目标音的影响，只关注不同语境对同一目标音识别的影响。

但在语境效应的理论解释中，一些研究者认为听者识别相同的目标音也受到前面语境的影响，正是由于考虑到了自然言语中语境对目标音所产生的协同发音效应，即“协同发音的知觉弥补”理论（Mann, 1980; Fowler, Brown, & Mann, 2000）。与之相对应的另一派理论认为关键因素是前面语境的声学特征，只要语境的声学特征中存在与目标音频谱上相对立的成分，就会有这种语境效应的产生，这就是听觉理论，听觉理论认为关键不是协同发音和发音特征，而是语境音的声学特征（Lotto & Kluender, 1998; Holt, Lotto, & Kluender, 2000）。

虽然“发音特征”和“频谱对比”理论都涉及到语境音的声学参数变化，但是两个理论解释语境效应的关注点不同。“发音特征”理论强调语境音的发音部位，如果语境音发音部位靠前，听者会知觉到更多发音部位靠后的目标音，如果语境音发音部位靠后，听者反而知觉到更多发音部位靠前的目标音，即知觉弥补了语境协同发音的影响。而“频谱对比”理论主要关注语境音的频谱特征，如果语境音有更高的频谱，听者会更多的以频谱较低的目标音进行反应，如果语境音频谱较低，听者会更多的以频谱较高的目标音进行反应，即是一种听觉上的频谱对比效应。在研究的最多的一个例子中，即/ar/和/al/对识别/da-/ga/连续体的影响，/l/有较前的发音部位，/r/有较后的发音部位，因此根据发音特征理论预期/ar/后会有更多的/da/反应，/al/后会有更多的/ga/反应，实验结果与此一致。而从声学特征来看，/r/有较低的 F3 频谱，/l/有较高的 F3 频谱，根据频谱对比效应假设，/ar/后会有更多的/da/反应（/da/有更高的 F3 起始频率），/al/后会有更多的/ga/反应（/ga/有更低的 F3 起始频率），实验结果也与这个假设一致。因此，就/r/和/l/对/da-/ga/的语境效应而言，两种理论都可以解释，难以区分。

当前实验采用了三个新的语境音，/pa/、/pi/和/pu/。这三个语境音从发音部位来看，塞

音/p/的唇音部位相对不变，元音/i/有最靠前的部位，其次是/a/，/u/的发音部位最靠后，根据发音特征理论的预期，/pu/语境会有更多的/ta/反应，其次是/pa/语境，最后是/pi/语境。从声学特征来看，三个音节主要区别是 F2 的轨迹，其中/pi/与目标音/ta/有对比性的频谱成分，其他两个音节都没有这类对比性成分。因此从频谱对比理论预期，/pi/语境会产生更多的/ta/反应，/pa/和/pu/两个语境效应不明确，也许与基线条件没有差别。这样通过新的语境音，两个理论的预期产生了矛盾，通过实验结果可以确定哪个理论有更好的解释力度。

参考文献：

- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407–412.
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60, 602–619.
- Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception & Performance*, 26, 877–888.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, 108, 710–722.

意见 2：第二，在实验 1 中，研究结果发现，在三种语境条件下都表现为“ga”的反应比率低于“pu”的反应比率，只是这一差异的程度在三种语境下不同。不知在无语境条件下，被试对“ta/ka”连续体的反应情况是怎么样的？是不是在无语境条件下，“ga”的反应比率就低于“pu”的反应比率？

当涉及到三个实验条件的两两比较，作者是如何控制多重比较效应的。

回应：基线条件下的实验结果就是无语境条件下的。从结果来看，/pi/音节作为语境时产生了最少的“ga”反应，在刺激 1-2 上的“ga”反应显著的少于/pu/语境条件，在刺激 2 上的“ga”反应与/pa/语境条件差异边缘显著。/pa/语境产生的“ga”反应数量居中，与基线条件没有显著差异。音节/pu/充当语境时被试有最多的“ga”反应，除在刺激 1-3 上多于/pi/语境外，在刺激 1 和 3 上也多于基线条件。因为实验目的是考察不同音节的语境效应与哪种理论的预期更吻合，潜在的机制是什么。因此没有对基线条件下的数据进行过多的解读。事实上，在一些语境效应实验中甚至不设置基线条件，只考察不同语境音的语境效应。

最初的多重比较主要采用 LSD 法来进行的，这种方法确实对 I 型错误的累积没有进行很好的校正，我们采用了 Sidak 方法重新进行了分析。实验 1 校正后的分析表明：对于刺激 1，配对比较表明基线和/pi/条件下被试的“ga”反应比率显著低于/pu/条件下的反应比率 ($p < .05$)。对于刺激 2，配对比较表明/pi/条件下被试的“ga”反应比率显著低于/pu/条件 ($p < .05$)，与/pa/条件下的差异边缘显著 ($p < .1$)。对于刺激 3，基线和/pi/条件下被试的“ga”反应比率与/pu/条件差异边缘显著 ($p < .1$)。因此差异的总体模式没有变化，只是显著性水平有所下降。另外我们对实验 2 和实验 3 也进行了类似的校正，讨论中的相关描述也进行了修改。所有的修改在正文中都用红色进行了标注。

审稿人 2 意见：《汉语塞-元-塞音序列语境效应机制探讨》一文通过三个实验试图考察语境音对汉语听者识别目标音的影响及潜在机制。研究问题具有一定科学依据和研究价值，但是在文章写作方面存在很大问题，研究问题的逻辑以及实验方法都交待得不够清楚，实验 3 的设计与结果也不对应，有必要重新收集实验数据。因此，建议进行大修之后重新评审。下面是详细的意见：

意见 1: 前言: 1. 第 1 段仅介绍了言语知觉中的语境效应, 探讨这个问题对言语知觉有什么意义? 是目前研究者广泛关注的问题吗?

回应: 我们在第 1 段中缩短了介绍语境效应的文字, 增加了对研究这个问题意义的论述。言语知觉语境效应自发现后一直是这个领域的热点问题之一, 从文献的发表可以看出, 从 80 年代一直到现在, 都有有关这个问题的实验和理论争论。具体修改在文中已用红色进行了标注。

意见 2: 2. 第 3 段, 内容太多, 可以考虑分几段从不同角度进行综述, 而且, 还需要考虑去掉一些罗列式的文献介绍, 如: “Stephens 和 Holt(2003)发现...。Holt(2006)发现...”, 需要一些必要的分析和综合, 进行更为概括性的综述。

回应: 我们将第 3 段分成了两段进行综述。去掉了许多罗列式的文献介绍, 如删除“Holt, Lotto 和 Kluender (2000)发现元音连续体([]-[])的识别受到双侧辅音语境([b]和[d])的影响, 且模拟辅音关键声学线索的非言语滑音或纯音也影响到中间元音连续体[]-[]的识别, 影响模式与辅音语境一致, 都是一种频谱对比效应”, 删除“关键因素是言语和非言语语境的频谱特征是否匹配。甚至非言语语境和言语目标之间有一个较长的时间间隔或存在许多干扰的中间频率纯音时, 非言语声音的语境效应也能延续 (Lotto, Sullivan, & Holt, 2003; Holt, 2005)”。精简后更突出对文献的概括和综合。

意见 3: 3. 各部分的内容逻辑性不强, 2-4 段都只是三种不同的解释, 第 5 段就是本研究的具体操作。那这些不同理论解释之间的分歧点究竟在哪里? 前人研究的局限是什么? 本研究的贡献为什么能解决或如何解决这个问题的? 即本研究的理论贡献在哪里?

回应: 我们对前言后半部分进行了彻底改写, 从第 6 段到第 10 段。第 6 段首先根据前面 4 段的介绍总结了三个理论观点的分歧, 然后重点讨论了前人研究的局限, 前人研究所采用的语境效应例子难以区分发音特征理论和听觉理论的解释力。

第 7-9 段重点介绍当前研究的创新和贡献, 通过采用同一塞音和三个元音语境结合而成的三个音节作为语境音, 发音特征理论和听觉理论对实验结果做出了不同的预期。这个实验材料可以区分两种理论的解释力度。具体修改见正文, 已用红色进行了标注。

意见 4: 4. 最后一段, 仅介绍了 3 个实验各自的问题出发点, 合在一起能解决什么样的理论问题? 对当前的研究有什么样的贡献?

回应: 对 3 个实验及相互之间的关系进行了更仔细的介绍。前言第 6-9 段重点介绍了实验 1 的目的和贡献, 最后 1 段介绍了实验 2 和实验 3 的目的。实验 1 考察发音特征理论和听觉理论哪个能够更好的解释三个音节的语境效应。实验 2 和实验 3 通过非言语声音模拟三个语境音节的声学线索, 考察实验 1 三个音节语境效应差异是否源于声学线索的差异, 是否能够由非言语声音进行模拟。如果能够进行模拟, 则为听觉理论的解释提供进一步的支持。实验 2 只模拟了三个音节的 F2 轨迹, 考察关键声学线索的作用。实验 3 进一步模拟了三个音节所有的共振峰轨迹, 考察除关键声学线索之外的其它线索对三个音节语境效应差异是否也起到一些作用。具体修改在文中已用红色进行了标注。

意见 5: 实验一: 1. 第 1 段有大量关于实验刺激的介绍, 让读者费解, 这部分需要在方法中进行详细介绍。建议作者概括实验的具体目的, 进行提炼之后明确表述。在实验三有类似的问题。

回应: 这一部分有很多的声学参数描述, 看起来是对实验刺激的介绍, 事实上这部分不是实

验刺激的介绍。因为发音特征理论要根据语境音的发音部位对实验结果做出预期，而听觉理论要根据语境音的声学特征对实验结果做出预期。这一部分只是仔细的分析了实验 1 所使用的语境音的发音部位和声学特征，然后据此提出发音特征理论和听觉理论对实验结果预期之间的矛盾。语境效应实验中常有类似的安排，如 Viswanathan, Fowler 和 Magnuson (2010) 研究中的实验 1。因为常规的实验报告这一部分通常比较简短，所以我们把这一部分放到了前言的第 7 和第 8 段进行介绍。我们对这一部分进行了简化，只简单介绍了实验 1 的材料和实验目的。具体修改已用红色进行了标注。

参考文献：

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1005–1015.

意见 6: 2. 在 2.1.3 中，实验程序不清楚，如“被试用鼠标点击相应的图标进行反应”，这里的图标是什么？刺激是如何呈现的？呈现时间多少？被试的反应是否设有超时？刺激间隔多长时间等实验细节信息不全。而且，语境音节与 /ta/-/ka/ 连续体感知任务是通过什么样的实验程序来实现的？这部分需要作者提供足够的信息以保证该实验能被重复出来。

回应：我们在实验 1 的程序中补充了相关信息，包括图标的内容，刺激呈现方式，被试如何反应，反应后间隔多久等。实验 2 和实验 3 的程序同实验 1。具体修改在文中已用红色进行了标注。实验程序和刺激呈现由 praat 软件完成，该软件是语音分析软件，有标准化的言语知觉任务模块，可以非常方便的完成声音刺激识别和分辨任务，自动记录被试的反应。

意见 7: 3. 在 2.2 部分，为什么只报告前 3 个刺激的简单效应和多重比较？从图 2 来看，刺激 4、8-10 的简单效应也可能显著，而且模式与前 3 个刺激不同。如果因为结果太多，建议使用表的形式把所有统计结果罗列出来，然后只报告有显著差异的结果。

回应：因为结果分析发现只有前两个刺激的简单效应是显著的，刺激 3 的简单效应边缘显著，其它的都不显著，因此只分析了前三个刺激的简单效应，并进行了多重比较。我们在正文中补充说明了其它刺激的简单效应都不显著。

意见 8: 4. 既然使用了 /ta/-/ka/ 的连续体感知任务，如果语境有效应的话，是否会影响到对分类知觉的感受性？能否使用感觉性的指标来考查不同语境的作用？设计到了相应的因素却并不考察是不太合适的。而且，这样的实验设计和数据分析都是不必要的，完全可以只考察被试在刺激 2/3 与 8/9 上的反应即可。

回应：当前研究采用的设计和数据分析方式与西方研究者 (Lotto & Kluender, 1998; Holt, Lotto, & Kluender, 2000; Holt, 2006; Viswanathan, Fowler, & Magnuson, 2010) 采用的模式完全相同，这也是语境效应实验标准的分析模式。因此这种设计与分析模式虽然看起来比较啰嗦，我们仍然沿用了西方研究者的方式。毕竟言语知觉研究中的众多范式都是国外研究者首创的。

至于感受性，语境效应实验包括范畴知觉实验一般不进行分析，我们也不太清楚原因。当前实验以“ga”反应百分比为因变量进行了分析，事实上也就分析了“da”反应百分比，因为二者是互补的，被试的反应不是“da”就是“ga”。因此这类实验中研究者通常以一个范畴的反应百分比为因变量进行分析。

参考文献：

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60, 602–619.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, 108, 710–722.

Holt, L. L. (2006). Speech categorization in context: Joint effects of non-speech and speech precursors. *Journal of the Acoustical Society of America*, 119, 4016–4026.

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1005–1015.

意见 9: 实验二: 1. 第 1 段仅仅说明实验操纵了什么, 没有说清为什么要这么做, 以及收集这个实验的意义在哪里。

回应: 我们在实验 2 的第 1 段增加了实验 2 的目的和意义。前言最后一段曾经详细的论述了实验 2 的目的, 实验 1 讨论部分最后一段也有关于实验 2 的衔接。因此这个地方只是进行了概括的说明。具体修改见正文中的红色标注。

意见 10: 2. 结果部分与实验 1 类似, 易读性比较差。

回应: 我们采用了这类实验常用的分析方式, 因为连续体有 10 个刺激, 当简单效应显著的刺激比较多时, 结果描述就会多一些。我们在讨论部分的第一段对结果分析进行了概括, 以增加易读性。分析后对结果进行概述, 这也是这类实验通常所使用模式。

意见 11: 实验三: 1. “此外,两个/pi/音节的发音部位几乎完全相等,如果它们展现出不同的语境效应,且这种效应能由关键的声学线索差异进行解释,这为语境效应的听觉理论提供了进一步支持,与发音特征理论的观点相矛盾。” 1)“此外”的表述不合适,这应该是本实验的重要目的,而不是“此外”的附带作用。2)这里是如何支持听觉理论,反对发音特征理论的?这也是本部分需要重点论述的,而使用了什么材料应该在方法部分进行详细说明。

2. 方法部分仍然不清楚,在图 5 中,是否/pi/1 来自实验 1,而/pi/1-F2 的数据来自实验 2?

3. 实验三的设计是考察新合成的音节/pi/1 及其 F2 模拟音,但结果部分却并没有考察这二者的差异,而是分别使用这两个条件与实验 1、2 的结果进行对比,而且是被试间的比较。建议实验 3 重新设计和收集新的数据。

回应: 鉴于实验 3 在设计和数据分析方面的不足,我们完全放弃了最初的设计,重新设计了实验 3。

新设计的实验 3 与实验 1 和实验 2 在逻辑上衔接的更好,且能更深入的探讨实验 1 三个言语音节语境效应的原因和潜在机制。实验 1 以/pa/、/pi/、/pu/三个音节为语境音,使听觉理论和发音特征理论的预期相矛盾,实验结果与听觉理论的预期更为一致。实验 2 以三个音节 F2 模拟音为语境音,考察三个音节所表现出的语境效应是否与 F2 模拟音所表现出的一致,如果一致,则表明三个音节的语境效应差异主要源于它们关键的声学线索差异。实验结果发现 F2 模拟音所表现出的语境效应总体上与三个音节较为一致,但在细节方面存在一些差异。这表明三个音节所表现出的语境效应不能完全由关键的声学线索差异来解释。那么至少还有两个可能的原因,一是其它声学线索可能也起到一些作用,另一个就是三个音节的语音范畴感知也能影响到它们的语境效应表现。实验 3 以三个音节的正弦波言语为语境音,正弦波言语与三个音节的所有共振峰轨迹都匹配,这样就进一步考察了其它声学线索在三个音节的语境效应中所起到的作用。结果分析发现实验 3 与实验 2 表现出类似的效应模式,这表明其它声学线索在其中所起到的作用非常有限。因此实验 1 与实验 2 (包括实验 3) 结果的差异不应该归于其它声学线索的作用,可能主要由于三个音节语音范畴感知的影响。

这样通过三个实验层层递进的考察了三个音节语境效应差异的起源,以及何种理论能够

更好的解释言语知觉中的语境效应。总的实验结果表明三个音节的语境效应差异主要源于它们之间关键声学线索的差异，这部分与听觉理论的观点一致（但是听觉理论的频谱对比假设没有预期到/pa/和/pu/及其模拟音的效应，这部分在总讨论有单独的论述）；但是声学线索差异（即使考虑到所有的声学线索，实验3）不能解释三个音节语境效应的全部模式，因为实验1和后面两个实验的结果还是有细微的差异，这表明三个音节的语音范畴感知也可能影响到它们的语境效应模式。因此，虽然相比发音特征理论，听觉理论给出了更好的预期和解释，但是也不能完全解释语境效应，还是要考虑语音范畴感知对语境效应的影响。

实验3的目的和意义在前言最后一段，实验2的讨论部分，整个实验3的部分，以及总讨论部分都有重复的论述。中英文摘要涉及到的部分也进行了修改。具体修改都用红色进行了标注。

意见 12：总讨论：1. 讨论不够深入，如“发音特征理论对三个/pV/音节语境效应的预期与实验结果相矛盾，没有得到当前实验结果的支持”，发音特征理论是什么样的预期，与本实验的哪个结果相矛盾？仅仅说它们是相矛盾是不够的，需要用具体的结果来分析并讨论才能让读者理解。

文章还有多处类似的语句，缺乏深入地分析和讨论。如：“这与频谱对比效应假设及Mitterer(2006)的建议不完全一致”，后面的分析也没说透彻。

这一段文字也让人读不懂：“Mitterer(2006)也认为只有语境音与目标音共享关键的声学特征，二者才有听觉交互作用的可能。这些理论能够较好的解释/pi/音节及其 F2 模拟音的语境效应,但是不能解释音节/pa/和/pu/及其 F2 模拟音的语境 效应。”

回应：我们对总讨论进行了大量的修改，对实验结果进行了深入的分析和讨论，修改了这些模糊的句子。总讨论前三段在对三个实验结果进行总结分析的基础上，重点讨论了当前研究中塞-元-塞音序列的语境效应的起源。第4段和第5段详细讨论了频谱对比效应解释/pa/和/pu/及其非言语模拟音的语境效应的不足，以及如何解释这种类型的语境效应。最后一段简单的论述了听觉神经科学关于语境效应的研究，这为当前的理论解释提供了一些支持。

具体修改都用红色进行了标注。

意见 13：2. 讨论不足，没有突现出本研究的意义，也没有结合本领域的前沿问题，进行扩展讨论，更缺乏对今后研究的展望。

回应：我们在最后一部分增加了总结和展望，概述了当前研究的结果，并指出了研究结果的意义。最后对今后研究的方向进行了展望。

意见 14：另外，还有一些语言表述方面的问题：

如，“根据 Mitterer(2006)的建议,...”，这里的“建议”挺奇怪。

“近乎完美”，这里的主观性表达：“完美”。

回应：已经修改了这些不规范的用法。

第二轮

审稿人 1 意见：

研究者详细回答了我的所有问题，没有进一步的审稿意见。

审稿人 2 意见：

作者对文章进行了较大修改，针对我上次的评审意见进行了很好地修改和回答，并且重新设计和收集了实验3的数据。因此，建议发表。