

《心理学报》审稿意见与作者回应

题目：统计学习和感觉-运动联结：婴儿语音感知的双重路径

作者：肖容，梁丹丹

第一轮

审稿人 1 意见：

本文考察了婴儿语音感知中的统计学习和感觉-运动联结机制。作者通过在不同实验中分别操纵输入语音的分布规律(单峰/双峰)和舌尖运动反馈(抑制/允许)，发现 8 月龄汉语环境婴儿在学习印地语辅音对比/d/-/d̪/时，语音分布规律和舌尖运动状态都会影响其语音区分表现，而在非语音条件下未观察到类似效应。研究沿用了该领域的经典范式，并尝试在汉语婴儿样本中检验相关效应，同时加入非语音对照条件，具有一定的新意和研究价值。

不过，本文在理论框架和结果解释上仍有一些需要进一步斟酌之处。具体意见如下。

意见 1：作者将统计学习和感觉-运动联结的对比置于“物种特异性”框架下，认为前者更体现跨物种的一般学习机制，而后者更体现人类特异性。对此表述建议更加谨慎。当前比较认知与 vocal learning 的研究表明，感觉-运动耦合、感觉反馈以及与发声学习相关的机制，在其他物种(尤其是 vocal learners，如鸣禽)中也有较充分证据。因此，不宜将这两种学习机制的对比简单解释为“物种特异性”与“一般性”之间的区分。

回应：感谢审稿人的重要提醒。原文中将统计学习与感觉-运动联结的对比过于直接地置于物种特异性与一般性对比的框架下，确有简化之嫌，也容易引发误解。根据您的建议，现已对引言和讨论中的相关表述进行了修订。

我们希望澄清的是，本研究所讨论的“感觉-运动联结”并非指广义的感觉-运动耦合机制。广义的感觉-运动耦合显然并非人类所独有，在多类动物的知觉-动作系统中均可观察到。我们真正关注的，是与发声学习和语音知觉相关的听觉-运动耦合，即听觉输入与发音动作表征之间的功能联结，以及这种联结在语音知觉中的调节作用。就此而言，我们同意审稿人的判断：即使是这一更具体层面的机制，也不宜被简单表述为严格意义上的“人类特异机制”。虽然语音感知领域的感觉-运动联结观的思想根源，即言语感知的运动理论(Liberman &

Mattingly, 1985)在提出之初有很强的物种特定和先天主义倾向,但近年来的实证研究发现了其他物种的发声学习者(尤其是鸣禽)中,同样存在与发声学习、听觉反馈和感觉-运动整合相关的机制证据(Arnon et al., 2025; Hisey et al., 2018; Kearney et al., 2019; Kim et al., 2021; Ritters et al., 2019)。

根据这一意见,我们对引言及讨论中涉及“物种特异性”的相关表述进行了修改,主要体现在以下两个方面:第一,删除或弱化了将两种机制简单对应为“跨物种一般性”与“人类特异性”的二元对立表述;第二,将论述重点由“物种范围上的差异”调整为“作用方式上的差异”,以更准确地刻画两种机制在婴儿语音习得中的互补而非对立的立场。具体而言,统计学习更侧重于对外源性输入规律的提取——婴儿通过感知环境中语音输入的分布特征,逐步调整感知空间并习得语音范畴边界,其作用依赖持续的外部分布信息;感觉-运动联结则更侧重于内源性加工系统的主动调节——即时的发音相关运动状态可直接参与语音感知过程,在缺乏特定语音输入经验的条件下仍可发挥作用。

具体而言,原文引言第2段关于跨物种一般性与人类特异性对立的论述现已修改如下(见正文P1,第24~29行的红色标记内容):

在语言发展的最早期阶段,一个核心理论问题是:婴儿语音感知敏感性的形成与发展,究竟更多依赖于对外源性输入特征的提取,还是更多依赖于与语音加工相耦合的内源性系统(Kuhl, 2021)? 围绕这一问题,学界形成了两条主要解释路径:统计学习理论与感觉-运动联结观。前者强调输入分布信息对感知空间的塑造作用,将语音习得视为广义规律提取机制作用于语音领域的结果,后者则更强调发音相关运动状态对语音感知的主动调节作用。

同时,在引言的理论阐述(见正文P1第30行的红色标记内容)和讨论部分(见正文P21第10~14行的红色标记内容)也相应调整了相关表述,指出统计学习与感觉-运动联结代表的是两类性质互补的语音习得机制,其核心区别不在于是否属于某一物种所独有,而在于它们作用于语音发展的路径不同。其中,统计学习强调对输入统计分布的追踪与提取;感觉-运动联结则强调运动系统、感觉反馈与知觉加工之间的双向耦合。

Arnon, I., Carmel, L., Claidière, N., Fitch, W. T., Goldin-Meadow, S., Kirby, S., ... & Fisher, S. E. (2025). What enables human language? A biocultural framework. *Science*, 390(6775), eadq8303.

Hisey, E., Kearney, M. G., & Mooney, R. (2018). A common neural circuit mechanism for internally guided and externally reinforced forms of motor learning. *Nature Neuroscience*, 21, 589–597.

Kearney, M. G., Warren, T. L., Hisey, E., Qi, J., & Mooney, R. (2019). Discrete evaluative and premotor circuits enable vocal learning in songbirds. *Neuron*, 104, 559–575.

Kim, Y., Kwon, S., Rajan, R., Mori, C., & Kojima, S. (2021). Intrinsic motivation for singing in songbirds is enhanced by temporary singing suppression and regulated by dopamine. *Scientific Reports*, 11, 20350.

- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36.
- Riters, L. V., Spool, J. A., Merullo, D. P., & Hahn, A. H. (2019). Song practice as a rewarding form of play in songbirds. *Behavioural Processes*, 163, 91–98.

意见 2: 与此相关, 本文目前对两种机制的区分, 在某些地方也容易被写成“先天/后天”的对立, 这一点建议避免。无论是统计学习还是感觉-运动联结, 都不宜被简单等同为先天或后天机制; 二者都可能同时受到发育基础与环境经验的共同作用。例如, 统计学习依赖外界语音输入, 但其对统计规律敏感的能力本身具有较强的生物学基础; 同样, 感觉-运动联结虽然强调运动机制的参与, 其作用发挥也仍然依赖语音经验的输入。

回应: 感谢审稿人指出这一重要的概念澄清问题。将统计学习与感觉运动联结对立为“后天/先天”两极, 确实是一种过度简化。在本研究的实验设计下, 也无法对这种强二分框架提供明确证据。

更准确地说, 两种机制都体现为生物学基础与环境经验的共同作用。统计学习依赖外界语音输入才能发挥作用, 但其对统计规律敏感的能力在生命早期即可观察到, 提示其具有一定的生物学准备基础(Saffran et al., 1996; Maye et al., 2002); 感觉-运动联结强调运动系统在语音知觉中的参与, 但其作用并非完全独立于经验输入, 而是受到发育成熟与早期语音经验的塑造。婴儿期发声-运动协调及规范咿呀语的逐步建立(Lee et al., 2017; Lang, 2019)提示, 与语音相关的感觉-运动耦合会随发展过程不断加强。因此, 将两者简单对立为先天与后天, 不仅不够准确, 也容易掩盖两种机制在发展过程中可能存在的动态交互关系。

根据这一意见, 除意见 1 回应中所述修改外, 我们进一步对引言和讨论中可能引发“先天/后天”二元对立联想的表述进行了删除或调整 (见正文 P2 第 32~34 行、P3 第 1 行、P3 第 24~25 行和 27~28 行、P23 第 13~21 行的红色标记内容)。

- Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical babbling: A marker for earlier identification of late detected developmental disorders? *Current Developmental Disorders Reports*, 6(3), 111–118.
- Lee, C.-C., Jhang, Y., Relyea, G., Chen, L.-M., & Oller, D. K. (2018). Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development*, 50, 140–153.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928

意见 3: 本文设置非语音条件是一个很有价值、也是最有理论潜力的设计创新，建议作者在引言中进一步突出这一设计的重要性，并明确提出该条件下的理论预期。例如，在统计学习框架下，若该机制具有较强的领域一般性，是否应预期其对非语音刺激也产生一定影响；而在感觉-运动联结框架下，若该机制更偏向语音加工，则是否应预期运动操纵不会影响非语音区分。若能在前言中将这一点更清楚地提出，将更有助于凸显非语音条件在区分两种机制中的理论贡献。

回应: 感谢审稿人的建议。现已对引言相关部分进行了修改，主要体现在以下四个方面。

第一，进一步突出了非语音条件设计的创新性和理论重要性。在引言中补充指出：两种理论涉及一个更广泛的问题，即语音加工在早期发展中是否表现出相对于一般听觉加工的特殊性。本研究将这一问题操作化为：在相同声学变化维度和相同任务范式下，统计分布信息与感觉-运动状态对婴儿听觉区分的影响，是否会因刺激具有语音属性而不同。统计学习通常被视为可作用于多种信息类型的广泛性规律提取机制，而感觉-运动联结更强调发音器官状态与语音动作表征之间的对应关系。由此，非语音条件可作为检验两种机制作用边界的重要设计，有助于判断观察到的效应更接近一般听觉变化检测，还是更接近语音相关加工调节 (修改见正文 P3 第 16~23 行、P3 第 32~34 行、P4 第 1~4 行的红色标记内容)。

第二，在引言中补充并明确提出了两种机制在非语音条件下的理论预期。在统计学习框架下，若该机制在生命早期体现出较强的一般性，则应预期其对结构上可比的非语音刺激也产生一定影响；若效应仅出现在语音条件，则更可能说明统计学习在该发展阶段受到输入类型的系统性调节。在感觉运动联结框架下，若该机制对语音信号具有选择性支持，则运动操纵应对非语音区分无效；若非语音条件同样受到运动操纵的影响，则对其语音选择性假设构成挑战。因此，“语音条件受操纵调节而非语音条件不受调节”的分离模式，将为感觉运动联结的语音选择性提供支持；而两种条件均受调节，则更支持统计学习的领域一般性立场。非语音条件由此成为区分两种理论机制的关键判准 (修改见正文 P4 第 31~34、P5 第 1~2 行的红色标记内容)。

第三，对原先较强的“领域特殊性”表述作了收敛处理。在引言和问题提出中，我们不再将两种机制直接概括为“领域特殊性”，而改用更为审慎的“语音选择性”等表述，以避免理论推断过强，并与讨论部分的相关论述保持一致。同时在修改后的讨论中进一步强调了本研究的发展性意义。我们指出，本研究聚焦于 8 月龄婴儿，提供了来自生命早期的新证据，说明统计学习和感觉-运动联结在语音与非语音条件下的差异，可能在婴儿期就已初步显现 (见 4.2 节，即正文 P22 第 8~22 行、P23 第 1~2 行和 6~10 行的红色标记内容)。

意见 4: 关于“领域特异性”的结论, 建议作者更加谨慎。目前的结果至多说明, 在本研究条件下显著效应仅出现在语音材料中, 而非语音条件下未观察到显著效应。但这并不足以直接证明两种机制具有严格的语音领域特异性。相关结论应建立在对语音与非语音材料中学习效率差异的直接统计检验基础之上, 而不能仅根据“语音显著、非语音不显著”来推断。

回应: 感谢审稿人的重要提醒。审稿人指出, 仅依据“语音条件显著、非语音条件不显著”的结果模式, 尚不足以直接推出两种机制具有严格的语音领域特异性, 相关结论应建立在语音与非语音条件之间效应差异的直接统计检验基础之上, 而不能依赖显著性与否的表面比较。根据该意见, 现已对数据分析和结果部分作了相应修订。在修改后的分析框架中, 已将刺激类型作为固定效应之一, 并在同一模型中检验其与试次类型、组别的全部交互项, 从而实现了语音与非语音条件之间的直接统计比较。

在实验 1(统计学习)中, 修改后的模型以刺激类型(辅音/非语音)、试次类型(交替/非交替)、组别(单峰/双峰)及其全部交互项为固定效应, 以试次序列为协变量, 被试设为随机截距并包含试次序列的随机斜率。结果显示, 试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互显著, $F(1, 592.98) = 4.13, p = 0.043$, 固定效应系数方向为 $b = -0.28, SE = 0.14, 95\% CI [-0.56, -0.01]$ 。这表明, 试次类型与组别之间的关系会随刺激类型不同而发生显著变化。进一步的简单效应分析显示, 在双峰组中, 试次类型 $\times$ 刺激类型的简单交互显著, $b = -0.28, SE = 0.10, t = -2.86, p = 0.004$; 而在单峰组中, 该简单交互不显著, $p = 0.991$ 。进一步分解发现, 仅在双峰组-辅音条件下, 交替试次的注视时间显著长于非交替试次, $b = -0.30, SE = 0.07, t = -4.24, p < 0.001$; 而在双峰组-非语音条件下, 该差异不显著, $p = 1.000$ 。此外, 在双峰组-交替试次中, 辅音条件的注视时间显著长于非语音条件, $b = 0.31, SE = 0.07, t = 4.36, p < 0.001$ 。上述结果表明, 统计学习效应在辅音与非语音条件之间存在直接的统计学差异, 且该差异主要体现在双峰分布输入对辅音刺激的促进上。

在实验 2(感觉-运动联结)中, 采用与实验 1 相同的分析框架, 以刺激类型(辅音/非语音)、试次类型(交替/非交替)、组别(抑制运动/允许运动)及其全部交互项作为固定效应。结果显示, 试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互达到边缘显著, $F(1, 554.00) = 3.71, p = 0.055$, 固定效应系数为 $b = -0.27, SE = 0.14, 95\% CI [-0.55, 0.01]$ 。后续简单效应分析显示, 在允许运动组中, 试次类型 $\times$ 刺激类型的简单交互显著, $b = -0.23, SE = 0.10, t = -2.35, p = 0.019$; 而在抑制运动组中不显著, $p = 0.709$ 。进一步分解表明, 仅在允许运动组-辅音条件下, 交替试次的注视时间显著长于非交替试次, $b = -0.24, SE = 0.07, t = -3.36, p < 0.001$; 在允许运动组-非语音条件下, 该差异不显著, $p = 1.000$ 。此外, 在允许运动组-交替试次中, 辅音条件

的注视时间显著长于非语音条件, $b = 0.27, SE = 0.07, t = 3.82, p < 0.001$ 。实验 2 也呈现出与语音选择性一致的模式。

需要特别说明的是, 重新分析并未推翻一审稿件中的主要结果方向, 而是使原先基于分别分析得出的判断获得了更严格的统计检验。具体而言, 实验 1 中, 原稿已观察到双峰分布输入后婴儿主要在辅音条件下表现出区分效应, 而非语音条件未出现稳定区分。修订后的统一模型进一步显示, 试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互显著, 说明这种语音与非语音之间的差异本身达到统计显著水平。因此, 实验 1 的主要结论未发生方向性改变。实验 2 中, 原稿已观察到允许运动组主要在辅音条件下出现交替优势, 而非语音条件和运动抑制条件下未出现稳定效应。修订后的统一模型显示, 关键三阶交互达到边缘显著, 简单效应模式与原稿方向一致。

尽管如此, 我们认同审稿人的意见: 上述交互效应所提供的证据仍属间接, 不足以支持“严格的语音领域特异性”这一强结论。若要进一步支持严格的领域特殊性, 还需在刺激匹配、任务难度控制、注意资源控制、多种非语音材料重复验证以及多指标收敛证据等方面提供更强支持, 以排除熟悉度、生态效度和统计功效等替代解释。据此, 现已对摘要、引言及讨论中的相关表述进行修改, 将“领域特异性”统一收敛为更为审慎的“语音选择性”, 并在讨论中明确说明。

除此之外, 相较于一审稿件, 本次修改稿在统计分析方法上还进行了以下改进, 并对 2.2 节和 3.2 节的结果部分做了全面替换:

其一, 试次序列的编码方式由原始编码改为中心化协变量。一审稿件将试次按 1-8 编码作为固定效应, 修改稿将其重新编码为条件内呈现顺序(1-4), 并进行中心化处理(取值-1.5、-0.5、0.5、1.5), 以连续协变量的形式纳入模型以控制习惯化效应, 同时在随机效应结构中加入试次序列的随机斜率, 使模型对个体间习惯化速率差异有更好的估计。

其二, 因变量处理方式更为规范。一审稿件对 $\log$ 变换的说明较为简略; 修改稿明确说明采用 $\log(x+1)$ 变换以满足残差正态性假设, 并注明描述统计与图表均以原始注视时间(秒)呈现, 以保持可解释性。

其三, 局部变化的补充分析中, 同样不再将语音与非语音条件分别分析, 而是将刺激类型纳入同一模型框架。具体而言, 修改稿将试次序列进一步划分为前半段与后半段, 并以试次阶段(前半段/后半段)、刺激类型(辅音/非语音)、试次类型(交替/非交替)、组别及其交互项作为固定效应, 被试设为随机截距, 以检验不同刺激类型下随试次推进的注视变化模式是否存在差异。该分析的目的是作为对习惯化效应和局部注意变化趋势的补充检验。

上述修改使分析策略与研究假设更加匹配, 2.2 节和 3.2 节的结果报告已据此全面替换, 相关统计量(F 值、 b 值、 SE 、95% CI)均基于修改后的模型重新计算和报告。

意见 5: 基于以上几点, 建议作者相应调整摘要、引言及讨论中关于本文创新性和理论意义的表述, 使之与当前证据强度更加匹配。

回应: 感谢审稿人的提醒。意见 1~4 分别指出了原稿中有关“物种特异性”“先天/后天对立”“非语音条件理论预期”以及“领域特异性结论”的表述需要更加谨慎。这些意见已经分别在相应部分作出回应, 我们进一步认识到, 这些问题共同指向一个更总体的写作问题: 原稿中摘要、引言、讨论中关于创新性和理论意义的表述, 整体上存在一定程度的概括偏强, 容易超出当前实验结果所能直接支持的证据范围。因此, 在逐条回应意见 1~4 的基础上, 又对全文相关表述进行了整体性降调和一致性修改。

在摘要中, 将原先较强的结论性表述进行了收敛。原稿中曾表述为“统计学习与感觉-运动联结均能影响婴儿语音习得, 两种机制均表现出语音加工特异性”, 修订后改为“统计学习与感觉-运动联结均可能调节婴儿语音区分表现, 并表现出对语音输入的相对优势”。这一修改避免将当前实验结果直接上升为“领域特异性”的强结论, 而是将结论限定在“当前任务中语音条件更容易表现出调节效应”这一证据范围内(修改见正文 P1 第 8~9 行红色标记内容)。

在引言中, 我们不再将统计学习与感觉-运动联结简单置于“跨物种一般机制”与“人类特异机制”或“后天输入”与“先天内源机制”的二元对立之中, 而是将两者概括为两条并非互斥的解释路径: 统计学习更强调外部输入分布信息对感知空间的塑造作用, 感觉-运动联结更强调发音相关运动状态对语音感知的主动调节作用。同时, 我们将原先较大的理论问题收敛为可由本研究检验的操作性问题, 即在相同声学变化维度和相同任务范式下, 统计分布信息与感觉-运动状态对婴儿听觉区分的影响是否会因刺激是否具有语音属性而不同, 以此回应非语音条件的理论意义, 同时避免将本研究直接解释为对“人类语音特殊性”或“严格领域特异性”的证明(相关修改见引言红色标记内容)。

在讨论中, 对理论解释进行了降调处理。原稿中将结果解释为“两种机制均具有语音加工特异性”或“感觉-运动机制具有低环境依赖性”, 修订后改为: 在 8 月龄婴儿、当前刺激设计和交替/相同注视范式下, 语音输入可能更容易获得统计学习和感觉-运动状态的支持; 感觉-运动操纵结果提示, 即时口腔运动状态可能在未提供目标语音短时训练输入的条件下调节婴儿语音区分, 但这一结论仍需结合实验 2 中三阶交互边缘显著等情况谨慎解释。

我们也在讨论中进一步强调，未来仍需通过更大样本、多种刺激材料、神经生理指标以及同时操纵分布信息和运动反馈的设计来检验两种机制的协同关系和作用边界(相关修改见讨论部分红色标记内容)。

综上，现已对审稿人前四条意见中直接涉及的具体表述进行了修改，也进一步对全文中关于创新性和理论意义的表述进行了主动降调和统一处理。修订稿现在将本文贡献更谨慎地界定为：本研究为 8 月龄婴儿语音区分可能受到短时输入分布和即时口腔运动状态共同调节提供了初步发展证据，并显示这种调节在当前任务条件下主要体现在语音刺激中；但本文并不主张由此证明严格的语音领域特异性、物种特异性或先天机制。

.....

审稿人 2 意见：

本文围绕婴儿语音感知发展的核心理论争议，考察统计学习机制与感觉-运动联结机制在语音加工中的作用，并进一步检验两种机制是否具有领域特异性。研究选题具有一定理论意义，实验设计整体较为系统。然而，尽管作者在理论上明确提出了关于机制领域特异性的假设，当前的统计分析尚不足以对这一核心问题作出直接检验。具体而言，语音与非语音条件被分别作为独立实验进行单独分析，而未在同一统计框架下进行直接比较。因此，现有结果无法有效检验机制效应在不同刺激之间是否存在显著差异，也难以支持关于两种机制均表现出语音加工特异性的结论。

鉴于领域特异性问题是本文理论贡献的关键所在，数据分析上的局限不足以支撑当前研究的核心结论。建议作者在修订中采用更为合理的统计模型，对语音与非语音条件进行直接比较。若该问题能够得到充分解决，本文有望为理解婴儿语音感知机制提供更具说服力的证据。以下为具体问题和建议：

意见 1： 统计分析方法不足以支撑“两种机制均表现出语音加工特异性”的核心结论。作者旨在检验两种机制的领域特异性，即统计学习是否同样影响非语言听觉加工，以及感觉-运动机制是否专属语音处理。语音与非语音条件下的结果对比是回答该问题的关键。然而，在当前的统计分析中，作者并未对语音与非语音条件进行直接比较，而是将二者分别作为实验 1a(2a)和实验 1b(2b)进行独立分析检验。这种分析策略无法直接检验交互效应，也因此难以对两种机制语音加工的特异性假设作出明确的统计推断。仅基于语音与非语音条件下分别显著或不显著的结果，尚不足以支持机制具有语音加工特异性的结论。鉴于分析方法的这一局限，本文相关结果及结论仍有待进一步验证。建议作者采用更为合理的统计模型，对语音与

非语音条件进行直接比较，以更有效地检验研究假设。

回应：感谢审稿人指出这一重要的统计分析问题。根据这一意见，现已对统计分析策略进行了实质性调整。在修改稿中，不再分别对语音与非语音条件进行割裂分析，而是将两类刺激统一纳入同一线性混合效应模型中，以刺激类型(辅音/非语音)作为固定效应之一，并同时检验其与试次类型(交替/非交替)和组别的全部交互项，从而在同一模型框架内对语音与非语音条件之间的效应差异进行直接统计检验。重新采用统一模型后，本研究的核心结果模式并未推翻一审稿件中的主要发现，而是使原先基于“语音显著、非语音不显著”的间接判断得到了更直接的统计检验。

实验 1 中，原稿已观察到双峰分布输入后婴儿主要在辅音条件下表现出区分效应，而非语音条件未出现稳定区分。修订后的统一模型进一步显示，试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互显著，说明这种语音与非语音之间的差异本身达到了统计显著水平。具体而言，试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互显著， $F(1, 592.98) = 4.13, p = 0.043$ ，固定效应系数方向为 $b = -0.29, SE = 0.14, 95\% CI [-0.56, -0.01]$ 。这表明，试次类型与组别之间的关系会随刺激类型不同而发生显著变化。进一步的简单效应分析显示，在双峰组中，试次类型 $\times$ 刺激类型的简单交互显著， $b = -0.28, SE = 0.10, t = -2.86, p = 0.004$ ；而在单峰组中，该简单交互不显著， $p = 0.991$ 。进一步分解发现，仅在双峰组-辅音条件下，交替试次的注视时间显著长于非交替试次， $b = -0.30, SE = 0.07, t = -4.24, p < 0.001$ ；而在双峰组-非语音条件下，该差异不显著， $p = 1.000$ 。此外，在双峰组-交替试次中，辅音条件的注视时间显著长于非语音条件， $b = 0.31, SE = 0.07, t = 4.36, p < 0.001$ 。上述结果表明，统计学习效应在辅音与非语音条件之间存在直接的统计学差异，且该差异主要体现在双峰分布输入对辅音刺激的促进上。

在实验 2(感觉-运动联结)中，原稿同样已观察到允许运动组主要在辅音条件下出现交替优势，而非语音条件和运动抑制条件下未出现稳定效应。修订后的统一模型显示，试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互在 log 尺度主模型中达到边缘显著，简单效应模式与原稿结果方向一致。具体而言，结果显示，试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互达到边缘显著， $F(1, 554.00) = 3.71, p = 0.055$ ，固定效应系数为 $b = -0.27, SE = 0.14, 95\% CI [-0.55, 0.01]$ 。后续简单效应分析显示，在允许运动组中，试次类型 $\times$ 刺激类型的简单交互显著， $b = -0.23, SE = 0.10, t = -2.35, p = 0.019$ ；而在抑制运动组中不显著， $p = 0.709$ 。进一步分解表明，仅在允许运动组-辅音条件下，交替试次的注视时间显著长于非交替试次， $b = -0.24, SE = 0.07, t = -3.36, p < 0.001$ ；在允许运动组-非语音条件下，该差异不显著， $p = 1.000$ 。此外，在允

许运动组-交替试次中,辅音条件的注视时间显著长于非语音条件, $b = 0.27$, $SE = 0.07$, $t = 3.82$, $p < 0.001$ 。由此可见,实验 2 也呈现出与语音选择性一致的模式。

在本研究所采用的任务条件、刺激设计和发展阶段下,语音输入相比非语音输入更容易激活统计学习机制,并可能更容易激活感觉-运动联结机制。二者在当前条件下均表现出对语音输入的相对优势,其中感觉-运动机制的语音选择性仍有待在更大样本和更高统计功效的研究中进一步验证。

上述修改已在方法、结果与讨论部分统一落实。修订后的分析策略使语音与非语音条件能够在同一模型框架内进行直接统计比较,从而为核心研究假设提供了更充分、也更严格的统计推断依据。

除此之外,相较于一审稿件,本次修改稿在统计分析方法上还进行了以下改进,并对 2.2 节和 3.2 节的结果部分做了全面替换:

其一,试次序列的编码方式由原始编码改为中心化协变量。一审稿件将试次按 1~8 编码作为固定效应,修改稿将其重新编码为条件内呈现顺序(1~4),并进行中心化处理(取值-1.5、-0.5、0.5、1.5),以连续协变量的形式纳入模型以控制习惯化效应,同时在随机效应结构中加入试次序列的随机斜率,使模型对个体间习惯化速率差异有更好的估计。

其二,因变量处理方式更为规范。一审稿件对 $\log$ 变换的说明较为简略;修改稿明确说明采用 $\log(x+1)$ 变换以满足残差正态性假设,并注明描述统计与图表均以原始注视时间(秒)呈现,以保持可解释性。

其三,局部变化的补充分析中,同样不再将语音与非语音条件分别分析,而是将刺激类型纳入同一模型框架。具体而言,修改稿将试次序列进一步划分为前半段与后半段,并以试次阶段(前半段/后半段)、刺激类型(辅音/非语音)、试次类型(交替/非交替)、组别及其交互项作为固定效应,被试设为随机截距,以检验不同刺激类型下随试次推进的注视变化模式是否存在差异。该分析的目的是作为对习惯化效应和局部注意变化趋势的补充检验。

上述修改使分析策略与研究假设更加匹配,2.2 节和 3.2 节的结果报告已据此全面替换,相关统计量(F 值、 b 值、 SE 、95% CI)均基于修改后的模型重新计算和报告。

意见 2: 被试月龄选择的依据不充分。文章中提到“关键变化集中于婴儿 6 至 12 月龄阶段”,但实验仅选择约 8 个月龄婴儿作为被试。作者给出的依据是:8 月龄处于/d/-/d/感知调谐的关键窗口(6~12 月龄),且此阶段运动神经初步成熟而统计学习效应尚未减弱(Werker & Tees, 1984; Yoshida et al., 2010)。然而,上述参考文献并未直接提供支持“8 月龄”这一具体时间

点的证据，也未明确说明为何该月龄是检验统计学习与感觉-运动机制交互或并存的理想阶段。建议作者更具体地阐明选择 8 月龄婴儿的理论与经验依据。

回应：感谢审稿人的提醒。根据审稿人的建议，我们已在修改稿引言中进一步补充选择 8 月龄的理论与经验依据。同时，我们也认识到，单一月龄设计只能提供一个关键发展时间点上的证据，若要更完整地揭示统计学习与感觉-运动联结在语音感知重组过程中的发展变化，未来采用跨月龄比较或纵向追踪设计将是更理想的方案。因此，修改稿在说明 8 月龄选择合理性的同时，也避免将其表述为唯一或最优时间点，而是更谨慎地界定为“较为理想的观测时点”。

具体而言，本研究选择 8 月龄主要基于以下三方面考虑。

1. 语音感知重组的关键过渡阶段

在出生后第一年内，婴儿的语音感知经历显著重组：注意焦点从普遍区分母语与非母语语音，逐步转向聚焦母语内部的对比。这一转变主要表现为三种模式：保持对母语语音对比的敏感性、逐渐丧失对非母语对比的区分能力，以及提升对母语内部更细微差异的辨别力(Kuhl et al., 2006)。这种向母语特征靠拢的感知调整称为“感知调谐”，被视为母语习得的关键阶段(Curtin et al., 2011; Reh et al., 2021; Werker, 2024)，其关键变化集中于 6 至 12 月龄阶段，并在辅音、元音及声调等不同语音成分中均已得到验证(Kuhl et al., 1992; Liu & Kager, 2014; Yeung et al., 2013)。

Werker & Tees(1984)的经典研究表明，婴儿对非母语语音对比的感知能力在 6 至 12 月龄间逐渐下降，这一感知重组过程在 10 至 12 月龄时基本完成。这意味着 6 至 12 月龄构成一个动态的过渡窗口：在窗口早期(6 月龄左右)，婴儿仍保有较强的跨语言感知灵活性；在窗口晚期(10 至 12 月龄)，感知系统已高度母语化，对非母语对比的区分能力大幅下降。

8 月龄恰好处于这一过渡期的中间阶段，具有独特的检验优势。本研究的核心目标是考察统计学习与感觉运动联结机制在语音习得中的作用，这要求被试的语音感知系统仍具备足够的可塑性，若习得已基本完成、感知系统趋于稳定，实验操纵便难以产生可检测的效应，习得机制本身也无从考察。若选择更早的月龄(如 4 至 5 月龄)，感知重组尚未启动，婴儿对几乎所有语音对比均保有普遍敏感性，难以区分实验操纵的作用与早期固有的感知灵活性；若选择更晚的月龄(如 10 至 12 月龄)，感知系统已高度母语化，非母语对比的感知区间大幅压缩，统计学习与感觉运动机制的塑造效果可能因“地板效应”而难以检测，习得机制的作用窗口实际上已经关闭。8 月龄婴儿的感知系统尚处于主动调谐阶段，对环境输入与运动反

馈均保有充分的响应能力,是在感知灵活性仍存与可塑性窗口尚开之间取得平衡的理想观测时间节点。

2. 统计学习效应的有效检测窗口

已有研究表明,婴儿在 6~8 月龄阶段能够稳定地利用分布频率信息调整辅音感知。Maye 等(2002)首先发现,输入频率分布会影响 6~8 月龄英语背景婴儿的语音感知,单峰分布暴露会削弱其对非母语/d/-/t/对比的敏感性。随后,Maye 等(2008)进一步表明,双峰分布输入能够增强 8 个月婴儿对更困难辅音对比(/g/-/k/)的区分能力。相比之下,Yoshida 等(2010)发现,10 月龄婴儿虽然仍可从双峰分布输入中获益,但所需输入量显著增加。具体而言,6~8 月龄婴儿只需约 2.3 分钟的单峰分布暴露即可改变语音区分敏感性,而到 10 月龄时,则需要 4 分钟以上的输入才能重建已处于下降过程中的非母语辅音对比区分能力。上述结果说明,辅音统计学习效应在 10 月龄左右已开始减弱,因此 8 月龄处于统计学习仍可靠可测、但尚未明显衰减的发展窗口。

3. 感觉运动系统的初步成熟及其调节作用

已有研究表明,感觉运动操纵在婴儿语音感知中具有因果性作用。限制 6 月龄婴儿的舌尖运动会削弱其对印地语/ɖa/-/ɗa/对比的区分敏感性(Bruderer et al., 2015);舌尖运动抑制也会特异性削弱 3 月龄婴儿对该对比的神经反应,而不影响不涉及舌尖运动差异的/ba/-/da/区分(Choi et al., 2021)。这些结果表明,运动系统对语音知觉的调节作用并不完全依赖特定语音范畴的既有经验。然而,现有直接证据主要集中于 3-6 月龄阶段,对于 6 月龄之后、尤其是在统计学习能力已可稳定激活的发展阶段,感觉-运动机制是否仍然发挥作用,尚缺乏系统检验。与此同时,口腔运动能力在 6 至 9 月龄间快速发展(Ejiri & Masataka, 2001),8 月龄婴儿已具备较为稳定的口腔运动模式,使感觉-运动操纵在这一阶段具有较好的操作可行性和发展适切性。

因此,8 月龄同时满足两个关键条件:一方面,婴儿在这一阶段仍保有足够的感知灵活性,统计学习效应可以被可靠检测;另一方面,其感觉-运动系统又已发展到足以接受有效操纵。本研究希望在语音感知尚具可塑性的阶段,考察统计学习与感觉-运动机制的作用。由此,8 月龄构成了检验两种机制及其关系的理想平衡点。未来研究若能采用跨月龄比较设计,例如比较 6 月龄、8 月龄、10 月龄和 12 月龄婴儿,或采用纵向追踪设计,将更有助于揭示统计学习和感觉-运动联结在语音感知重组过程中的作用强度、时间窗口及其相互关系。这样的设计能够更直接回答两种机制是同步变化、先后发挥作用,还是在不同发展阶段具有不同权重。

根据审稿意见，我们已在引言中补充如下论述(正文P4 第18~25 行的红色标记内容)：

选择8 月龄作为目标月龄，主要基于两方面考虑。其一，该月龄处于/d/-/d/感知调谐的关键窗口内，6 个月婴儿仍能区分/d/-/d/，而10 个月后其敏感性显著下降(Werker & Tees, 1984)。其二，8 月龄同时满足统计学习效应可检测、且感觉-运动操纵具有发展適切性的条件。一方面，此阶段统计学习效应已可被可靠检测(Maye et al., 2002, 2008)，而该效应在10 个月后开始减弱(Yoshida et al., 2010)；另一方面，婴儿口腔运动系统已具备一定成熟度，能够对感觉-运动操纵作出有效响应(Bruderer et al., 2015; Ejiri & Masataka, 2001)。因此，8 月龄构成了同时检验统计学习与感觉-运动机制的较为理想的观测时点。

- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, 112(44), 13531–13536.
- Choi, D., Dehaene-Lambertz, G., Peña, M., & Werker, J. F. (2021). Neural indicators of articulator-specific sensorimotor influences on infant speech perception. *Proceedings of the National Academy of Sciences*, 118(20).
- Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, 39, 492–504.
- Ejiri, K., & Masataka, N. (2001). Co-occurrences of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, 4(1), 40–48.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9, F13–F21.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608.
- Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, 133(2), 385–394.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1), 122–134.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111.
- Reh, R. K., Hensch, T. K., & Werker, J. F. (2021). Distributional learning of speech sound categories is gated by sensitive periods. *Cognition*, 213, 104653.
- Werker, J. F. (2024). Phonetic perceptual reorganization across the first year of life: Looking back. *Infant Behavior and Development*, 75, 101935.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63.
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139.
- Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, 15(4), 420–433.

意见 3：引言中提到“统计学习是领域一般性机制，这一理论立场倾向于将语音习得视为普遍认知学习能力的体现”，与实验结果统计学习对非语音无效是否存在潜在矛盾。在讨论中应进行充分的讨论。

回应：感谢审稿人指出这一重要问题。我们认同，本研究中统计学习效应主要出现在语音条件，而非语音条件未观察到稳定效应，需要在理论上加以澄清。根据该意见，现已在引言(正文P2 第6~10 行的红色标记内容)和讨论部分(正文P22 第7~22 行的红色标记内容)均进行了补充说明。

首先，澄清本研究结果的理论含义：实验中仅在语音条件下观察到统计学习效应，并不意味着统计学习本身是语音专属机制，而是表明在 8 月龄婴儿、当前刺激设计和任务条件下，统计学习对语音输入表现出相对优势。统计学习虽然常被视为领域一般性机制，但其具体表现并不均一，可能受到刺激类型、结构属性、任务要求及发展阶段的系统性调节。因此，本研究结果更支持“领域一般性机制在特定发展条件下呈现语音优势”，而非“统计学习仅服务于语音加工”。同时，本研究的价值还在于提供了来自生命早期的新证据。现有关于统计学习语音优势或模态约束的讨论多数来自成人或较大儿童，而本研究聚焦于 8 月龄婴儿，说明在生命早期阶段，统计学习对语音输入的敏感性就可能已经高于非语音输入，为理解统计学习与语音加工之间关系的发生基础提供了新的发展性证据。

具体修订涵盖以下四个方面，按逻辑层级依次展开。

第一，在引言第三段中增加了限定性表述，明确指出“广义学习机制”并不意味着统计学习在所有输入类型上均等有效，其实际效应仍可能受到刺激类型、任务情境和发展阶段的调节。

第二，讨论 4.2 的第二节中引入统计学习存在领域、模态与刺激类型约束的证据。补充引用了 Conway 与 Christiansen(2005)关于统计学习模态约束的系统比较研究，以及 Siegelman 与 Frost(2015)关于不同统计学习任务之间相关性较低的证据。后者尤为关键：它表明统计学习不能被简单理解为单一、统一的规律检测能力，而是同时表现出刺激特异性。基于这一框架，本研究中语音条件显著、非语音条件不显著，并不构成对“领域一般性”观点的直接否定，而更可能反映了统计学习在不同输入条件下受到的约束和调节。

第三，现有关于统计学习是否具有语音选择性的证据尚不一致。例如，Sweet 等(2024)在严格控制刺激声学特征的条件后发现，统计模式的跨音节提取并不会仅仅因为材料被知觉为“语音”而显著增强，这提示统计学习本身可能更接近一种广义的、自动的刺激驱动机制。基于这一研究现状，本研究在理论表述上采取了更为审慎的措辞，将原先的“领域特异性”调整为“语音选择性”。

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 24–39.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105–120.

Sweet, S. J., Van Hedger, S. C., & Batterink, L. J. (2024). Of words and whistles: Statistical learning operates similarly for identical sounds perceived as speech and non-speech. *Cognition*, 242, Article 105649.

Toro, J. M., Sinnott, S., & Soto-Faraco, S. (2005). *Specificity of motor contributions to auditory statistical learning*. *Journal of Cognition*, 7(1), Article 25.

意见 4: 实验 2 通过牙胶限制舌尖运动来操纵发音运动反馈，但这一操作同时可能引入注意分配变化、口腔不适或情绪状态改变等非特异性因素。尽管已有研究支持该范式的有效性，建议作者在讨论中更明确地承认这一操纵的潜在局限性。

回应: 感谢审稿人提出的这一建设性意见。我们认同牙胶操纵范式可能引入非特异性因素，具体而言，本文补充指出实验 2 中牙胶操纵范式存在潜在的非特异性效应。本研究通过使用软质条状牙胶与硬质扁平牙胶的对比设计，而非无牙胶与有牙胶的对比，在一定程度上控制了口腔异物感本身带来的混淆。然而，两种牙胶在质地与形状上的差异，仍可能引入额外变量：硬质牙胶对舌尖运动的限制可能伴随一定程度的口腔不适，进而影响婴儿的情绪状态；牙胶的存在与看护者协助扶持的行为，亦可能引发注意分配的变化，使婴儿的认知资源在不同条件间产生系统性差异。尽管 Bruderer 等(2015)的超声成像研究为该范式的选择性抑制效应提供了解剖层面的支持，且本研究在注视时间分析中未发现两组婴儿在非语音条件下存在整体行为差异，但上述混淆因素仍无法被完全排除。未来研究可引入生理指标(如心率、面部表情编码)以监测不适程度，或设计不依赖口腔操纵的替代范式，以更纯粹地分离感觉-运动反馈的特异性贡献。

现已在修改稿的第 4.4 节(正文 P24 第 10~15 行的红色标记内容)中补充了相关讨论。

意见 5: 自检报告答“计划样本量为每个被试组 20 人”，但正文未说明此样本量是如何确定的。建议补充认定样本量的依据。

回应: 感谢审稿人的建议。本研究每组样本量设定为 20 名，依据来自以下两个方面。第一，参照同领域同类研究的惯例。在婴儿统计学习、感觉运动语音感知及视听言语感知研究中，每组样本量通常在 8 至 36 名之间(Maye et al., 2002: 每组 12 名; Liu & Kager, 2014: 每组 16 名; Yeung & Werker, 2013: 每组 20 名; Bruderer et al., 2015: 每组 24 名; Choi et al., 2021: 每组 22 名等，如下表所示)，每组 20 名处于该范围的中间水平，在保证统计效能的同时兼顾了婴儿研究的现实可行性。第二，基于先验效能分析。在 G*Power 中将目标效应近似为一个 2(组别)× 2(试次类型)× 2(刺激类型)的混合设计三阶交互，并假设中等偏小效应($f = 0.15$)至中等效应($f = 0.20$)进行先验样本量估计。在 $\alpha = 0.05$ 、统计效能($1 - \beta$)= 0.80 的标准下，每组至少需要 12 至 31 名被试。综合以上两方面依据，本研究最终将每组样本量设定为 20 名。

研究领域	文献	月龄	实验/条件设置	各组样本量	总样本量
统计学习	Maye et al., (2002)	6 月龄; 8 月龄	单峰-双峰; 单峰-双峰	12, 12, 12, 12	48
统计学习	Maye et al., (2008)	7-9 月龄	单峰; 双峰; 对照	28, 36, 33	97
统计学习	Liu & Kager (2014)	11-12 月龄	单峰; 双峰	16, 16	32
统计学习	Wanrooij et al., (2014)	2-3 月龄	单峰-双峰; 单峰-双峰	8, 8, 8, 8	32
统计学习	Yeung & Werker (2009)	9 月龄	仅听觉-无区分; 物体映射; 物体映射-无区分	20, 20, 20	60
统计学习	Fennell & McDonald (2024)	6-7 月龄	单峰-双峰,单语-双语; 单峰- 三峰, 单语-双语	15, 15, 11, 12; 12, 16, 11, 20	112
功能性学习	Yeung & Nazzi (2014)	10 月龄	训练一致匹配; 训练非一致 匹配	24, 24	48
功能性学习	Yeung, Chen, & Werker(2014)	9 月龄	物体映射-无区分;增加前训 练阶段	20; 40, 38	98
功能性学习	Teinonen et al. (2008)	6 月龄	视听匹配; 视听不匹配	24, 24	48
视听言语感 知	Danielson et al. (2017)	6 月龄; 9 月龄; 11 月龄	6 月龄-视听一致; 6 月龄-视 听不一致; 9 月龄-视听一致; 9 月龄-视听不一致; 11 月龄 -视听一致; 11 月龄-视听不 一致	16, 16, 16, 16, 16, 16	96
视听言语感 知	Schure et al. (2016)	8 月龄	视听; 仅视觉; 仅听觉	36, 29, 28	93
视听言语感 知	Panneton et al. (2024)	11 月龄	男性面孔; 女性面孔	22, 24	46
感觉运动	Yeung & Werker (2013)	9 月龄; 4-5 月龄	允许唇部运动; 无关元音	20, 32	52
感觉运动	Bruderer et al. (2015)	6-7 月龄	仅听觉; 抑制舌尖运动;允许 舌尖运动	24, 24,24	72
感觉运动	Choi et al. (2021)	3 月龄	仅听觉; 抑制舌尖运动	22, 22	44
非母语音辨 别/感知重 组	Kotera et al. (2024)	5-6 月龄; 8-9 月 龄; 12-13 月龄	元音对比	38; 30; 28	96

现已在两个实验的被试部分(正文P5 第15~24 行, 正文P15 第16~17 行的红色标记内
容)补充了样本量确定的依据, 修改内容如下:

每组样本量设定为 20 名, 主要基于以下两方面依据。第一, 参照同领域同类研究的惯
例。在婴儿语音感知与习得研究中, 单组样本量通常在约 8 至 36 名之间(Bruderer et al., 2015;
Choi et al., 2021; Maye et al., 2002; Liu & Kager, 2014; Yeung & Werker, 2013), 每组 20 名被

试处于该范围内, 属于较为常见且具有现实可行性的样本量水平。第二, 基于先验效能分析。在 *G*Power* 中将目标效应近似为一个 $2(\text{组别}) \times 2(\text{试次类型}) \times 2(\text{刺激类型})$ 的混合设计三阶交互, 并假设中等偏小效应($f = 0.15$)至中等效应($f = 0.20$)进行估计。在 $\alpha = 0.05$ 、统计效能($1 - \beta$) = 0.80 的标准下, 所需样本量约为每组 12 至 31 名。综合以上两方面依据, 本研究最终将每组样本量设定为 20 名。

- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). *Sensorimotor influences on speech perception in infancy. Proceedings of the National Academy of Sciences of the United States of America*, 112(44), 13531–13536.
- Choi, D., Bruderer, A. G., & Werker, J. F. (2021). *Neural indicators of articulator-specific sensorimotor influences on auditory speech discrimination in 3-month-old infants. Proceedings of the National Academy of Sciences of the United States of America*, 118(22), Article e2025043118.
- Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*, 42, 37–48.
- Fennell, C. T., & McDonald, M. (2024). Distributional learning of bimodal and trimodal phoneme categories in monolingual and bilingual infants. *Infant Behavior and Development*, 77, Article 101983.
- Kotera, H., Khattab, G., Boll-Avetisyan, N., & Hohl, B. (2024). German infants' discrimination of the English /æ/-/ɛ/ contrast: Evidence from a cross-sectional and a longitudinal study. *Infant Behavior and Development*, 77, Article 101984.
- Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, 133(2), 385–394.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1), 122–134.
- Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111.
- Panneton, R., Ostroff, W. L., Bhullar, N., & Netto, M. (2025). Plasticity in older infants' perception of phonetic contrasts: The role of selective attention in context. *Infancy*, 30(1), e12620.
- Schure, S., Ter Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, 7, Article 525.
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108(3), 850–855.
- Wanrooij, K., Boersma, P., & Van Zuijlen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, 5, Article 77.
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2014). Referential labeling can facilitate phonetic learning in infancy. *Child Development*, 85(3), 1036–1049.
- Yeung, H. H., & Nazzi, T. (2014). Object labeling influences infant phonetic learning and generalization. *Cognition*, 132(2), 151–163.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113(2), 234–243.
- Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, 24(5), 603–612.

意见 6: 请注意写作与排版规范, 检查文章细节问题, 如避免中英文括号混用, P3 (如 James et al., 2020)。

回应: 感谢审稿人的细致审阅。我们已对全文进行系统检查, 统一了中英文括号的使用规范: 正文括号一律改为英文括号(), 并修正了审稿人指出的第 3 页及其他存在混用情况的位置。此外, 我们同时对全文标点符号、字体及格式等细节问题进行了全面核查与统一。

附: 基于原始尺度的敏感性分析

为检验主分析结果是否依赖因变量尺度，在 log 转换注视时间主模型之外，进一步以原始注视时间为因变量建立同结构线性混合模型，作为敏感性分析。总体来看，原始尺度模型与 log 尺度主模型在核心结果方向和简单效应模式上基本一致，说明主要结论并非由 log 转换本身造成。相关模型输出见附表 1~表 4。

1. 统计学习实验

表 1 统计学习测试阶段注视时间的固定效应系数表

固定效应	<i>b</i>	SE	<i>t</i>	95% CI
截距	2.76	0.26	10.70***	[2.25, 3.26]
试次序列	-0.54	0.06	-9.48***	[-0.65, -0.43]
试次类型: 交替 vs 相同	0.27	0.25	1.07	[-0.23, 0.76]
组别: 双峰 vs 单峰	-0.34	0.36	-0.93	[-1.05, 0.38]
刺激类型: 非语音 vs 辅音	0.03	0.25	0.12	[-0.46, 0.53]
试次类型 × 组别	1.02	0.36	2.85**	[0.32, 1.72]
试次类型 × 刺激类型	-0.06	0.36	-0.16	[-0.76, 0.64]
组别 × 刺激类型	0.07	0.36	0.20	[-0.63, 0.77]
试次类型 × 组别 × 刺激类型	-1.16	0.50	-2.29*	[-2.15, -0.17]

注: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。参照水平为: 相同试次、单峰组、辅音刺激。因此，截距表示抑制运动组在辅音相同试次中的平均注视时间；正值表示相对于参照水平注视时间更长，负值表示注视时间更短。

表 2 统计学习测试阶段注视时间的 ANOVA 检验表

效应	DenDF	<i>F</i>	<i>p</i>
试次序列	397.38	89.93***	< 0.001
试次类型	593.00	13.30***	< 0.001
组别	38.02	0.08	0.782
刺激类型	593.00	3.97*	0.047
试次类型 × 组别	593.00	3.03	0.082
试次类型 × 刺激类型	593.00	6.37*	0.012
组别 × 刺激类型	593.00	4.03*	0.045
试次类型 × 组别 × 刺激类型	593.00	5.26*	0.022

注: 自由度采用 Satterthwaite 近似, 仅报告分母自由度。* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

为检验结果是否依赖因变量尺度，本研究进一步以原始注视时间作为因变量进行敏感性分析。结果显示，原始尺度模型与 log 尺度主模型的核心结果模式一致：试次序列主效应显著，试次类型主效应显著，刺激类型主效应显著，且试次类型 × 组别 × 刺激类型三阶交

互同样显著, $F(1, 593.00) = 5.26, p = 0.022$; 固定效应系数为 $b = -1.16, SE = 0.50, t = -2.29, p = 0.022$ 。简单效应结果同样显示, 关键效应主要来源于双峰组辅音条件下的交替优势, 而非语音条件和单峰组均未出现稳定交替优势。

需要说明的是, 包含被试随机截距和试次序列随机斜率的模型出现 singular fit, 提示随机斜率部分估计不稳定。原始尺度模型中, 随机截距标准差为 0.83, 试次序列随机斜率标准差为 0.04, 二者相关达到-1.00; ICC = 0.214, 说明注视时间存在一定被试间差异。由于 log 尺度模型与原始尺度模型的因变量不同, 二者 AIC 不宜直接比较。

2. 运动控制实验

由于 log 转换后的主模型中, 试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互呈边缘显著, $F(1, 554) = 3.71, p = 0.055$, 固定效应结果也显示该交互为边缘显著, $b = -0.27, SE = 0.14, t = -1.93, p = 0.055$, 因此进一步进行原始尺度敏感性分析。

原始尺度模型结果显示, 试次序列主效应显著, $F(1, 39) = 38.07, p < 0.001$, 试次类型主效应显著, $F(1, 554) = 9.47, p = 0.002$, 刺激类型主效应显著, $F(1, 554) = 20.30, p < 0.001$ 。更重要的是, 原始尺度模型中试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互达到显著水平, $F(1, 554) = 5.63, p = 0.018$; 固定效应结果同样显示该三阶交互显著, $b = -1.15, SE = 0.48, t = -2.37, p = 0.018$ 。该结果与 log 尺度主模型的方向一致, 说明关键交互并非由 log 转换本身造成。

简单效应结果也支持这一点。在原始尺度模型中, 允许运动组的辅音条件下, 交替试次注视时间显著长于相同试次, $b = -0.24, SE = 0.07, t(554) = -3.36, p = 0.005$; 而允许运动组的非语音条件下, 交替与相同试次差异不显著, $b = -0.002, SE = 0.07, t(554) = -0.03, p = 1.000$ 。抑制运动组中, 辅音和非语音条件下交替与相同试次差异均不显著。与此同时, 在允许运动组的交替试次中, 辅音注视时间显著长于非语音, $b = 0.27, SE = 0.07, t(554) = 3.82, p = 0.001$ 。由此可见, 原始尺度模型与 log 尺度主模型在简单效应模式上保持一致: 关键效应主要表现为允许运动组在辅音条件下出现交替优势, 而非语音条件下未出现稳定区分为。

从模型可接受性来看, 运动控制实验的原始尺度模型未出现收敛失败或 singular fit 警告, 随机效应结构基本可接受。其随机截距标准差为 0.64, 试次序列随机斜率标准差为 0.28, 二者相关为-0.50, 残差标准差为 1.53, ICC = 0.15, 说明注视时间存在一定被试间差异, 但随机效应估计未表现出明显异常。由于 log 尺度模型与原始尺度模型的因变量不同, AIC 不作直接比较。

表 3 运动控制测试阶段注视时间的固定效应系数表

固定效应	<i>b</i>	SE	<i>t</i>	95% CI
截距	2.54	0.22	11.49***	[2.11, 2.97]
试次序列	-0.43	0.07	-6.17***	[-0.57, -0.29]
试次类型: 交替 vs 相同	0.15	0.24	0.62	[-0.32, 0.62]
组别: 允许运动 vs 抑制运动	-0.15	0.31	-0.49	[-0.76, 0.45]
刺激类型: 非语音 vs 辅音	-0.60	0.24	-2.48*	[-1.07, -0.13]
试次类型 × 组别	0.86	0.34	2.51*	[0.19, 1.53]
试次类型 × 刺激类型	0.16	0.34	0.47	[-0.51, 0.83]
组别 × 刺激类型	0.52	0.34	1.53	[-0.15, 1.19]
试次类型 × 组别 × 刺激类型	-1.15	0.48	-2.37*	[-2.10, -0.20]

注: * $p < 0.05$, *** $p < 0.001$ 。参照水平为: 相同试次、抑制运动组、辅音刺激。因此, 截距表示抑制运动组在辅音相同试次中的平均注视时间; 正值表示相对于参照水平注视时间更长, 负值表示注视时间更短。

表 4 运动控制测试阶段注视时间的 ANOVA 检验表

效应	DenDF	<i>F</i>	<i>p</i>
试次序列	39.00	38.07***	< 0.001
试次类型	554.00	9.47**	0.002
组别	38.00	1.22	0.277
刺激类型	554.00	20.30***	< 0.001
试次类型 × 组别	554.00	1.38	0.241
试次类型 × 刺激类型	554.00	2.92	0.088
组别 × 刺激类型	554.00	0.04	0.835
试次类型 × 组别 × 刺激类型	554.00	5.63*	0.018

注: 自由度采用 Satterthwaite 近似, 仅报告分母自由度。[†] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

3. 总体结论

综上，基于原始注视时间的敏感性分析总体支持 log 尺度主模型的结果方向。统计学习实验中，原始尺度模型与主模型均显示关键三阶交互显著，且简单效应均指向双峰组辅音条件下的交替优势。运动控制实验中，虽然 log 尺度主模型的关键三阶交互仅达到边缘显著，但原始尺度模型中该交互达到显著水平，且两种尺度下的简单效应模式一致，均显示允许运动组主要在辅音条件下出现交替优势。由此可见，两个实验的核心结果模式并不依赖于因变量尺度选择。

当前修订稿正文仍以 log 尺度模型作为主分析模型进行报告，未将原始尺度敏感性分析纳入正文结果部分。该分析仅作为回复信中的补充材料呈现，用于说明结果方向的稳健性以及实际注视时间差异的可解释性。正文中的结果解释和理论推论均基于 log 尺度主模型。

第二轮

审稿人 1 意见：

作者在修订中已较好回应了我此前的主要意见。我仅有一个后续问题需要作者进一步澄清。实验一的前/后半段试次分析中，作者称前期辅音与非语音条件无差异，后期才出现差异。但从图示结果看，尤其是在双峰组的交替试次中，trial 1 和 trial 2 的辅音条件在描述性上似乎已经高于非语音条件。因此，目前这一结论是否主要基于跨组别、跨试次类型合并后的刺激类型比较，仍需进一步说明。考虑到理论上最关键的是双峰组-交替试次中的辅音与非语音差异，建议作者补充该条件下前/后半段的直接统计比较，并根据相应结果适当调整文中关于“前期无差异、后期出现差异”的表述。

回应：感谢审稿人的细致意见。我们认同审稿人的判断：原稿中关于“前半段辅音与非语音无差异、后半段才出现差异”的表述不够准确。该结论主要来自前/后半段补充模型中跨组别、跨试次类型合并后的刺激类型比较，即在总体层面比较辅音与非语音条件的注视时间差异。正如审稿人所指出的，本研究理论上最关键的是双峰组-交替试次中的辅音与非语音差异，该条件直接关系到双峰分布输入后婴儿是否主要在语音刺激中表现出对交替变化的区分。

根据审稿人的建议，我们补充进行了双峰组-交替试次内前/后半段辅音与非语音条件的直接统计比较。需要说明的是，该前/后半段模型中与此相关的高阶交互均未达到显著水平，因此该比较并非由模型交互结构所支持的确证性分解，而是基于理论预期对特定条件的针对

性考察，其结果应作探索性解读。具体而言，在前/后半段模型的基础上，使用 emmeans 进一步估计双峰组-交替试次中前半段和后半段的辅音与非语音边际均值，并对两个阶段内的刺激类型差异分别进行检验；针对这两个比较采用 Bonferroni 法进行多重比较校正。

补充分析结果(见正文 P14, 第 14~25 行的红色标记内容)显示，在双峰组-交替试次中，前半段辅音条件的注视时间显著高于非语音条件， $\text{estimate} = 0.31$ ， $SE = 0.12$ ， $t(266) = 2.56$ ，校正后 $p = 0.022$ ；后半段辅音条件的注视时间同样显著高于非语音条件， $\text{estimate} = 0.30$ ， $SE = 0.12$ ， $t(266) = 2.44$ ，校正后 $p = 0.031$ 。为进一步澄清该优势是否随测试推进而变化，我们还检验了该条件内的阶段 $\times$ 刺激类型交互，结果显示辅音相对非语音的优势量在前、后半段之间无显著差异，交互对比 $\text{estimate} = 0.02$ ， $SE = 0.17$ ， $p = 0.928$ ，95% CI [-0.33, 0.36]。综合上述结果，在理论关键条件下，辅音相对于非语音的优势并非仅在后半段才出现，而是在前半段和后半段均可观察到，且其优势量在两个阶段之间未见显著差异。这与审稿人根据图示提出的观察是一致的。

据此，我们已相应调整结果部分的表述(见正文 P15, 第 18~27 行的红色标记内容)。修订稿中明确区分了两类结果：第一，在跨组别、跨试次类型合并后的总体比较中，前半段辅音与非语音差异不显著，后半段辅音注视时间显著高于非语音；第二，在理论关键的双峰组-交替试次中，辅音条件在前半段和后半段均显著高于非语音条件，且优势量在两阶段间无显著差异。因此，我们删除或弱化了“前期无差异、后期才出现差异”的绝对表述，改为说明：跨条件合并分析所呈现的“后半段才出现总体刺激类型差异”仅反映跨条件平均后的趋势，而在关键条件下，双峰组-交替试次中的辅音优势自测试早期即已出现并保持稳定。

这一补充分析并未改变实验 1 的核心结论，而是进一步细化了主分析所揭示的结果模式：双峰分布输入后的交替优势主要体现在辅音刺激中，且辅音相对于非语音的优势在测试前半段已经出现，并延续至后半段。

.....

审稿人 2 意见：

总体而言，作者已经基本回应了上一轮提出的主要问题，并在本轮修订中对原稿进行了较为实质性的修改。尤其是通过统一模型直接比较语音与非语音条件，较好地解决了原稿中最主要的统计分析问题。

意见 1: 仍需指出的是, 实验 2 的证据强度相对有限。在感觉-运动操纵实验中, 试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互效应仅达到边缘显著水平。因此, 关于“感觉-运动联结机制表现出语音选择性”或“两种机制的效应均主要出现在语音条件中”的结论仍应保持谨慎。

回应: 感谢审稿人的提醒。我们认同, 实验 2 中感觉-运动操纵的证据强度相对弱于实验 1。主模型中试次类型 $\times$ 组别 $\times$ 刺激类型三阶交互仅达到边缘显著水平, $F(1, 554.00) = 3.71$, $p = 0.055$, 因此不足以支持“感觉-运动联结机制表现出稳定语音选择性”的强结论。根据该意见, 我们已将实验 2 的相关结果定位为趋势性或探索性证据, 并系统降低了摘要、实验 2 小结、总讨论和结论中的表述强度。

与此同时, 我们希望补充以下几点说明, 以表明这一边缘显著的结果仍具有一定解释价值。以下补充结果仅用于说明实验 2 结果模式的方向一致性与个体层面的稳定性, 并不改变主模型三阶交互仅达到边缘显著、相关结论仍需谨慎解释的判断。

log 尺度主模型中的简单效应分析呈现出较为清晰且与理论预期一致的模式。尽管三阶交互在 log 尺度主模型中仅达边缘显著, 但简单效应分析所呈现的模式较为清晰: 运动反馈的促进效应仅见于允许运动组的辅音条件($p = 0.005$), 允许运动组的非语音条件($p = 1.000$)及抑制运动组的全部条件($ps = 1.000$)均无显著差异。四个条件中仅一个显著、且恰好是理论预期最关键的条件, 这种高度选择性的模式在方向上与感觉-运动联结机制语音特异性假设吻合。

第一, 我们对该效应进行了留一被试诊断, 并已将结果写入修订稿 3.2 节(见正文 P18, 第 27~32 行的红色标记内容)。结果显示, 在 40 次留一被试分析中, 有 39 次三阶交互保持在 $p < 0.10$ 的范围内, 表明该交互趋势并非由某一名被试单独驱动。此外, 允许运动组辅音条件下的关键简单效应在全部 40 次分析中均保持显著($ps < 0.005$), 表现出较好的稳定性。

第三, 第一轮修改中报告的原始尺度敏感性分析也呈现出方向一致的结果。原始注视时间模型中的三阶交互达到显著水平, $F(1, 554.00) = 5.63$, $p = 0.018$, 其简单效应模式与 log 尺度主模型一致。需要强调的是, 修订稿仍以残差分布更符合模型假设的 log 尺度模型作为主分析, 并如实将三阶交互报告为边缘显著; 原始尺度分析和留一被试诊断仅作为稳健性补充, 不用于替代主模型或提高结论强度。

根据审稿人的意见, 我们具体进行了以下表述修改。

1. 摘要中的修改

原稿将两个实验统一概括为“统计学习与感觉-运动联结均可能参与婴儿语音感知发展, 并表现出对语音输入的相对优势”, 同时使用了“感觉-运动联结机制具有低环境依赖

性”等相对较强的表述。根据审稿意见，修订稿对两个实验的证据强度作了区分，并降低了有关语音选择性和环境依赖性的结论强度。现修改为：

结果显示，双峰分布输入后婴儿表现出对目标辅音对比的区分敏感性，而单峰条件下未观察到稳定区分；允许运动组在辅音条件下呈现交替优势，而运动抑制组未呈现相应模式。总体上，非语音刺激在各实验条件下均未诱发稳定区分行为。研究提示，短时输入分布能够调节婴儿的语音区分表现；感觉-运动状态也可能影响婴儿对辅音差异的感知。总体而言，两项实验的结果在一定程度上均呈现出向语音输入倾斜的趋势。两条路径或存在功能差异：感觉-运动状态可能在较少依赖当前实验输入的情况下发挥即时调节作用；统计学习过程则可能通过输入分布信息影响语音感知表征的短时调整，并可能在更长时间尺度的语音感知调谐中发挥作用。（正文P1，第6~14行的红色标记内容）

修订稿通过分别使用“能够调节”和“也可能影响”区分两个实验的证据强度，将原有的“语音相对优势”调整为向语音输入倾斜的趋势，并将感觉-运动作用限定为对当前实验输入的相对较少依赖，同时区分统计学习的短时调整作用与可能的长期发展意义，从而使摘要结论更加审慎，并与整体统计证据保持一致。

2. 实验2小结中的修改

原稿第一点将“言语运动反馈可能调节婴儿的辅音感知”作为较明确的机制性结论。修订稿将“调节”改为较为审慎的“可能与……有关”，并删除重复的因果性表述。现修改为：

第一，在未提供目标语音短时熟悉或训练输入的条件下，言语运动反馈可能与婴儿的辅音感知有关。本实验未设置熟悉或训练阶段，婴儿在正式测试前未接受目标辅音刺激的额外实验暴露。在印地语辅音/d/-/ɖ/的感知任务中，使用软质条状牙胶的婴儿在交替条件下表现出对辅音差异的更高敏感性，而硬质扁平牙胶组则未出现显著区分。这提示，即使在未提供目标语音额外短时训练的条件下，言语运动反馈仍可能影响8月龄婴儿对新异辅音差异的区分表现。（正文P21，第11~16行的红色标记内容）

第二点中，原稿容易被理解为“非语音加工不依赖感觉-运动机制”。修订稿增加“当前任务条件下”“可能较少受到调节”和“仍需进一步验证”等限定：

第二，口腔运动状态对婴儿非语音区分未产生显著影响。在未设置非语音刺激短时熟悉或训练阶段的情况下，对于非语音，无论是否抑制言语运动反馈，婴儿在声音交替出现条件下都未表现出区分行为。这一结果提示，在当前任务条件下，非语音加工可能较少受到言语运动状态的调节，但该结论仍需在更高统计功效下进一步验证。（正文P21，第17~21行的红色标记内容）

3. 总讨论 4.1 节中的修改

修订稿将其调整为关联性和趋势性表述：

在允许运动条件下，婴儿在辅音刺激中表现出相对更明显的交替试次优势；而在运动抑制条件下未观察到这一模式。这一结果与Bruderer 等(2015)和Choi 等(2021)的发现方向一致，提示发音相关感觉-运动状态可能与早期语音感知存在一定关联。(正文P22，第15、17~18行的红色标记内容)

并在该段中明确补充证据限制：

然而需要指出的是，本研究关键三阶交互仅达到边缘显著水平，因此该结果更适合被视为趋势性证据，而非感觉-运动联结语音选择性的确定性证明。(正文P22，第26~27行的红色标记内容)

4. 总讨论 4.2 节中的修改

原稿将两个实验统一概括为“两种机制的效应均主要出现在语音条件中”。修订稿不再对两个实验作同等强度的概括，而是区分实验1和实验2的证据水平。现改为：

实验1中，统计学习效应主要出现在语音条件下；实验2中，感觉-运动操纵也呈现类似趋势。总体而言，在当前刺激设计与任务条件下，语音输入相较非语音输入更容易呈现出可检测的行为差异。(正文P23，第22~23行的红色标记内容)

同时增加跨实验收敛证据，但避免将其解释为确定性的语音选择性：

在当前刺激设计与任务条件下，语音输入相较非语音输入更容易呈现出可检测的行为差异。值得注意的是，两个实验均在各自的理论关键条件下观察到辅音交替试次相对于相同试次的注视优势，即实验1的双峰组和实验2的允许运动组呈现了方向一致的结果；其中，实验2的关键简单效应在留一被试分析中表现出较好的稳定性。(正文P23，第2~6行的红色标记内容)

这里将原稿中的“均获得了高度稳健的效应”改为“呈现方向一致的结果”和“初步收敛证据”，以避免用实验1的显著三阶交互强度替代实验2较弱的总体证据。

5. 总讨论 4.3 节中的修改

修改稿对统计学习与感觉-运动联结的具体作用路径及其整合作了较为充分的理论推演。修订稿在保留“两类因素可能作用于不同时间尺度”这一基本解释的同时，进一步降低了结论强度：将“共同调节语音习得”改为“可能分别从不同时间尺度影响婴儿语音区分表现”，将研究结论限定在本研究直接考察的语音区分层面；同时删除了对“语音范畴线索提

取”“听觉-发音映射”等具体机制路径的较强推断，并增加“可能”“更可能反映”等限定语，以避免将两个独立实验直接解释为已经得到验证的机制协同或整合：

综上，统计学习过程和感觉-运动状态可能分别从不同时间尺度影响婴儿语音区分表现：前者更强调输入分布对感知系统的持续塑造，后者更可能反映当前感知-运动状态的即时调节作用。多类因素在不同发展时间尺度上的作用，可能共同支持婴儿在生命早期逐步发展出精细化的语音感知能力。（正文P29，第29~32行的红色标记内容）

6. 结论中的修改

原稿在结论第一点中将两个实验统一概括为“短时输入分布和口腔运动状态均可能调节婴儿的语音区分表现”，未充分体现两项实验的证据强度差异；在第二点中，“两种机制的效应均主要出现在语音条件中”则是全文对实验2语音选择性最直接、概括强度最高的表述。根据审稿意见，修订稿分别调整了两点的表述：第一点将统计学习表述为“能够调节”，将感觉-运动状态表述为“也可能产生一定影响”，以区分两个实验的证据等级；第二点则将原有的统一结论拆分为分别对应两个实验的两项判断：

第一，短时输入分布能够调节8月龄婴儿的语音区分表现，口腔运动状态也可能对其辅音区分产生一定影响，提示婴儿语音敏感性的发展可能受到多类因素的影响。第二，在当前刺激设计和任务范式下，统计学习效应主要出现在语音条件中，未稳定推广至声学匹配的非语音刺激，提示其作用可能具有一定语音特异性；感觉-运动操纵的结果模式也在一定程度上呈现出向语音输入倾斜的趋势，但其是否具有语音特异性仍需进一步验证。（正文P25，第18~23行的红色标记内容）

意见2：此外，修订稿虽然在中文正文中已将“领域特异性”调整为“对语音输入的相对优势”。英文摘要中仍出现“Both mechanisms exhibited domain specificity for speech stimuli”等表述。建议作者在完成中文正文修改后，同步更新英文摘要，避免中英文摘要在理论结论强度上不一致。

回应：根据审稿人意见，将英文摘要中涉及“domain specificity”的表述强度与中文正文保持一致，改为条件性、初步性的表达。具体修改包括：Results部分末尾新增一句，明确注明实验2三阶交互仅达边缘显著、相关结果应作探索性解读，与中文正文的处理方式一致。Conclusion部分将“Both mechanisms exhibited domain specificity for speech stimuli”改为“the distributional-learning effect was primarily observed for speech stimuli, whereas the sensorimotor manipulation showed a similar but less conclusive tendency”，以“相对优势”替代“领域特异

性”，与中文修改方向吻合，并在后续补充说明该解释属初步性结论(正文 P29，第 15~21 行的红色标记内容)，有待未来研究验证，避免过度概括。

意见 3：最后，修订稿在写作和格式方面仍有少量细节需要进一步检查。例如，正文 2.1.1 被试章节中“Choi et al., 2021Maye et al., 2002”两个引文之间缺少分号。

回应：感谢审稿人的细致提醒。我们已根据该意见对正文 2.1.1 节中的引文格式错误进行了修正，将“Choi et al., 2021Maye et al., 2002”修改为“Choi et al., 2021; Maye et al., 2002”。同时，我们也对全文的写作和格式进行了再次检查，重点核对了文内引文、标点符号、统计符号、表格注释以及参考文献格式等细节。

第三轮

审稿人 2 意见：感谢作者的认真修改，修订稿已较好回应了此前的主要意见，我没有其他意见了。

编委意见：已达到发表水平，同意录用。

主编意见：同意录用。