心理科学进展 2025, Vol. 33, No. 10, 1794–1804 https://doi.org/10.3724/SP.J.1042.2025.1794
Advances in Psychological Science https://cstr.cn/32111.14.2025.1794

视听协同发展在婴儿期语言习得中的作用：典型发育与高风险自闭症谱系障碍婴儿的比较分析*

靳孟可 [1] 严璘璘 [1] 刘少英 [1] 肖乃奇 [2]

([1] 浙江理工大学理学院心理学系, 杭州 310018)
([2] 麦克马斯特大学心理、神经科学和行为学系, 汉密尔顿 L8S4L8, 加拿大)

摘 要 多模态感知机制对婴儿期语言习得至关重要, 其中视听整合在典型发育婴儿的语言能力发展中起着关键作用。相比之下, 高风险自闭症谱系障碍(ASD)婴儿在此整合过程中常面临挑战。典型的言语感知发展轨迹强调面部特征加工的重要性——对眼部与嘴部区域的注意能促进语言学习。实证研究表明, 4.5 月龄的婴儿已具备视听整合能力, 这种能力可有效预测后期语言发展水平。而高风险 ASD 婴儿则表现出社会性注意减少和视听整合功能受损, 此类缺陷可能破坏常规语言习得路径。因此, 早期干预策略应优先采用基于生物学的感觉引导方法, 重点增强多感官整合能力, 而非仅针对注意行为进行训练。理解这些机制不仅能深化对典型语言发展的认知, 更能为制定早期干预方案提供实证基础, 从而支持高风险 ASD 群体的语言习得。

关键词 语音感知, 视听匹配, 多感官整合, 高风险自闭症谱系障碍婴儿, 语言发展, 早期干预
分类号 B844

1 引言

在生命的最初两年, 婴儿在接触周围语言时逐步掌握语音系统(李思瑾 等, 2023), 涵盖语音及其分布、组合、韵律信息(如音节、节奏、重音), 并逐步发展出将离散语音单位组合为词汇的感知能力(Curtin & Zamuner, 2014)。在与周围环境的互动过程中, 婴儿对语言的感知与习得并非单一感官的孤立参与, 而是多种感官协同作用的复杂过程(Rosenblum, 2008)。在面对面的交流中, 来自婴儿照顾者的视听言语信息, 特别是婴儿导向言语(Infant-directed speech, IDS), 通过提供丰富的多模态线索(如语言与面部表情的同步性), 显著促进了婴儿对语音单位的感知与词汇的习得, 从而对他们的语言发展产生了决定性的影响(Coffey & Snedeker, 2025; Weisleder & Fernald, 2013; Zhou et al., 2023)。

婴儿对语音和面孔的偏好与生俱来(Buiatti, et al., 2019; Wu et al., 2022), 在生命早期, 他们能够广泛辨别不同类别的语言和面孔。然而, 随着月龄的增长, 这种辨别能力逐渐趋于专门化, 具体表现为对熟悉语音(如母语)和面孔(如本族面孔)的感知能力增强, 而对不熟悉语音(如非母语)和面孔(如他族面孔)的辨别能力则逐渐减弱(Maurer & Werker, 2013)。值得注意的是, 语音知觉和面孔识别的专门化并非独立发展, 而是在婴儿成长过程中表现出动态的依赖关系: 6 个月时两者尚未相关, 9 个月时呈现负相关, 而到 12 个月时则转变为正相关(Xiao et al., 2018)。这一发现揭示了婴儿视听感知能力的跨感官协同发展及其随月龄增长的动态演变特征。鉴于婴儿认知资源有限, 其在视听匹配与视听整合过程中对眼部及嘴部的选择性注意随月龄增长而动态变化, 这种变化对其言语习得能力具有显著影响(Edgar et al., 2023; Pons et al., 2019)。自闭症谱系障碍(Autism Spectrum

收稿日期：2024-10-30
* 本文系浙江省高校重大人文社科攻关计划项目“中加婴儿视听整合能力的比较研究” (项目编号: 2024QN133)和浙江省教育科学规划课题“婴儿早期视听知觉整合的发展规律及干预研究” (项目编号: 2024SCG329)的研究成果之一。
通信作者：严璘璘, E-mail: yanlinlin@zstu.edu.cn

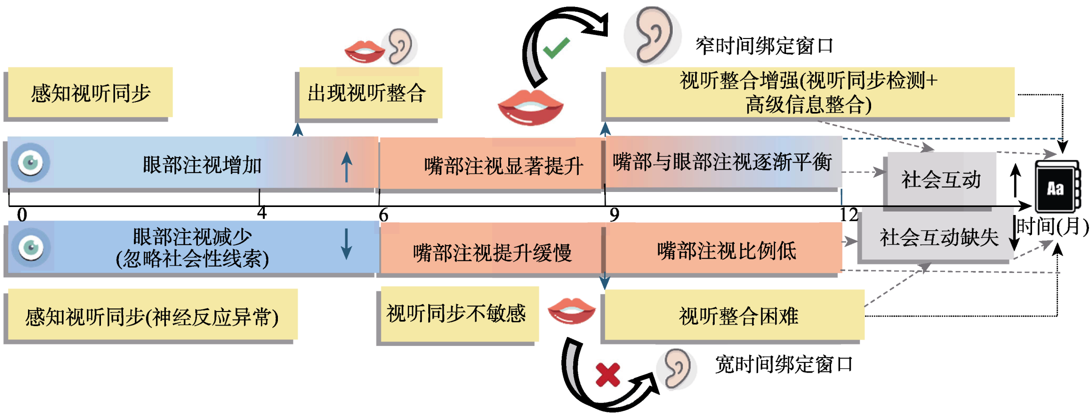

图 1　典型发育婴儿和高风险 ASD 婴儿视听整合能力发展轨迹的对比分析

Disorder, ASD)作为一种早发性神经发育障碍，通常伴随语言困难，但因自闭症婴儿极少在 2 岁前确诊(Leng et al., 2024; Sheldrick et al., 2017)，可能错失干预的关键期。近年来研究发现，高风险自闭症谱系障碍婴儿(以下简称高风险 ASD 婴儿)在早期发展中可能存在视听匹配及整合能力缺陷，并表现出眼部−嘴部区域选择性注意的异常模式(Lozano et al., 2025)。系统比较典型发育婴儿(以下简称婴儿)语言习得各阶段的关键行为标志与高风险 ASD 婴儿的特异性偏离特征，不仅能够深入揭示语言发展的内在机制，而且为高风险 ASD 婴儿的早期识别和精准干预提供重要依据。

本文通过系统梳理婴儿言语感知能力的发展轨迹(见图 1)，重点考察了视觉信息提取、匹配与整合的发育特征，创新性地揭示了眼部−嘴部选择性注意分配在言语习得过程中对视听整合的关键调控作用。本文从视听协同发展的视角，首次构建了婴儿视听整合能力的阶段性发展模型。并针对高风险 ASD 婴儿群体，指出他们在相应阶段发展过程中存在的特异性偏离，这些差异点为深入理解 ASD 早期发病机制及制定个体化早期干预方案提供了重要的理论支撑和实践指导。

2　婴儿的言语感知

新生儿自出生起便展现出对人类语音的显著偏好(Vouloumanos & Werker, 2007)，并具备初步的音素区分能力(Wu et al., 2022)。研究表明，新生儿的大脑能够辨别音素(如元音和辅音的区分)、音节及其序列(陈钰, 张丹丹, 2020; 李思瑾 等, 2023)，同时能够检测单词边界、学习词汇并分离语音流(Kujala et al., 2023)。这些能力使得婴儿能够从日常语言输入中快速且高效地习得语言。

婴儿语言能力的发展呈现出明显的阶段性特征，这一过程既受先天认知机制的影响，又与后天语言环境密切相关。研究表明，6~12 个月婴儿会经历“母语特异性”转变：对母语音素(如英语的/r/和/l/)的敏感度显著提升，而非母语音素的辨别能力下降(Kuhl et al., 1992, 2006; Werker & Tees, 1984)。值得注意的是，这种转变过程具有可塑性。研究发现 9 个月大的美国婴儿在与普通话使用者进行真实互动后，仍能保持对普通话音素的辨别能力(Kuhl et al., 2003)，但通过屏幕的互动则无法产生相同效果。这表明婴儿通过统计学习机制提取母语音系规律，并结合婴儿导向沟通中的多维信息(Kosie & Lew-Williams, 2024)，在尚未掌握大量词汇的情况下就能适应母语环境。在词汇加工层面，新生儿已表现出对音节序列的敏感性，研究显示其对 ABB 结构(如"mubaba")的神经激活强于 ABC 结构(Gervain et al., 2008)。随着月龄增长，6 个月大的婴儿能利用高频词(如自己名字)分割相邻新词(Bortfeld et al., 2005)，7.5 个月大时可突破韵律偏好来分割抑扬格单词(Sandoval & Gómez, 2016)。面对连续语音的复杂性，7~8 个月婴儿会选择关注中等可预测性的刺激，以优化认知资源的分配(Kidd et al., 2014)。这些发展不仅依赖于听觉系统的加工，还需要视觉信息的协同配合。

视觉信息在语言习得过程中发挥着重要的补充作用，与听觉系统协同增强婴儿的语言加工效率。说话者的面部运动(如口型变化)不仅能反映音

素发音特征，还能提供韵律结构和语言整体特征。研究表明，视觉线索对语音分割具有辅助作用：当听觉信号模糊时，婴儿通过观察说话者口型区分音素(Kuhl, 2010)。这种视听整合机制在早期即发挥作用，4~6个月婴儿已能匹配特定口型与对应元音(Lewkowicz & Hansen-Tift, 2012)。在词汇习得中，视觉情境与语言输入的时空共现为词汇−指称映射提供关键线索，如 Bergelson 和 Swingley (2012)发现 6 个月婴儿通过视觉焦点锁定与语音同步出现的物体。值得注意的是，视觉互动对语音感知的促进作用具有特异性：真实社交场景中的眼神接触、动态面部表情与语言输入形成多模态刺激(Kuhl, 2010)。这种多通道整合不仅解释了为何真人互动能有效改善非母语音素感知衰退，同时也表明婴儿语言习得本质上是一个多模态信息整合的学习过程。

3 婴儿的视听言语整合

3.1 典型发育婴儿的视听言语整合

3.1.1 典型发育婴儿对视觉信息的提取

婴儿天生具备面部感知能力，这种能力在出生后的第一年通过与照顾者的面对面交流得到特异性发展(Buiatti et al., 2019; Quinn et al., 2019)。研究表明，4个月大婴儿的视觉注意主要受物理显著性驱动，而 6~8 个月大的婴儿在视觉搜索任务中表现出稳定的面部注意偏好，其注意力受物理显著性的影响明显减弱(Kwon et al., 2016)。这一转变揭示了注意引导机制自下而上向基于社会信息的自上而下调节的发展轨迹。随着视觉系统的成熟，婴儿对面部的注意偏好进一步强化，尤其倾向于关注视听动态面孔(Bastianello et al., 2022)。动态面孔通过提供时空同步的视听线索，为婴儿社交与语言发展构建了关键学习界面。首先，眼睛注视引导的社会参照作用。婴儿通过追踪说话者视线建立词汇−客体映射：当婴儿通过凝视锁定于说话者注视的目标物体时，其新词的学习准确率显著提升(Çetinçelik et al., 2021)，且该能力与第二年词汇量增长呈正相关(Brooks & Meltzoff, 2015; Tenenbaum et al., 2014)。其次，嘴部运动的语音解码功能。随着语言加工复杂度增加，婴儿的注意资源逐步向说话者嘴部区域分配(Imafuku et al., 2019a, 2019b)。在听觉输入受损(如非母语刺激或者背景噪音)情境下，嘴部运动提供的视觉线索可有效补偿听觉模糊性(Lewkowicz & Hansen-Tift, 2012)。纵向研究显示，12 个月大婴儿对嘴部的注视时长及目光追随能力均可预测其在 18~24 个月时的词汇量(Tenenbaum et al., 2014)。

3.1.2 典型发育婴儿视听言语整合的发展变化

婴儿的视听言语整合能力发展是一个由感知偏好为基础，多模态经验塑造的复杂过程。其核心在于对视听时间同步性线索的敏感性提升与跨模态信息整合能力的动态发展。这一过程以月龄为节点，呈现从感官匹配到整合的连续过程。

(1) 0~3 个月：视觉偏好的形成。婴儿的视觉系统通过感知偏好与社会互动形成双重驱动机制。从出生起，婴儿就对直接眼神注视表现出特殊敏感性。行为实验显示，2 至 5 天大的新生儿能够区分直接注视和转移注视(Farroni et al., 2002)。这种注视模式通过注视追随建立共同注意基础(Ishikawa et al., 2022)，并借助眼睛运动解码社会互动意图(Belteki et al., 2022)。此阶段注意资源分配呈现单通道特征，婴儿更倾向于加工高对比度的眼部动态信息，而对嘴部运动仅维持基础感知。神经生理学证据表明，3 个月大婴儿的初级视听皮层已能够对同步的语音−面部刺激产生跨模态的反应，但整合精度限于较小时间窗口(Lewkowicz, 1996)。这种早期视听匹配能力优先加工社会性刺激(Suri et al., 2023)，为后续视听整合启动奠定基础。

(2) 3~6 个月：视听整合的启动。随着注意力定向功能的发展(Hurley & Oakes, 2015)，婴儿逐渐突破对眼部区域过多的依赖，形成动态注意力分配策略。这一阶段，两种认知驱动力显现：感知驱使婴儿关注具有显著视听冗余线索的嘴部运动进而促进感知(Tomalski, 2015)，而认知则推动其探索语言输入的规律性(Kidd et al., 2014)。多模态实验表明，4.5个月婴儿已具备跨模态整合能力，典型表现为 McGurk 效应(Burnham & Dodd, 2004)，即能将冲突的视听语音刺激(如视觉/ga/与听觉/ba/)整合为统一感知“da”。该效应具有经验依赖性与文化特异性(Ujiie et al., 2021; Yan et al., 2024)。此外，注意转移表现出情境依赖性——面对非母语刺激时，婴儿的嘴部注视时长增加 28% (Birulés et al., 2018)，表明视觉策略调整与语言输入复杂度密切相关。神经影像学研究进一步揭示，5月龄婴儿的颞上沟区域对不匹配的视听刺激产生

特异性响应(Riva et al., 2022), 为后续视听整合奠定了神经基础。

(3) 6~9 个月: 语言敏感期的开始。婴儿的视觉注意力模式从眼部向嘴部发生显著转移。多模态实验表明, 当呈现视听同步的说话面孔时, 6~8 月大的婴儿对嘴部的注视比例从与眼部相当逐渐显著超越(Lewkowicz & Hansen-Tift, 2012)。这一转变与三个关键认知机制的形成密切相关：首先, 视听冗余线索的利用。动态嘴部提供的视听冗余线索能有效补偿听觉输入的模糊性, 在背景噪音下, 视觉线索显著提升语音辨别的准确率(Lalonde & Werner, 2019)。其次, 语音−运动映射机制的成熟。语音−运动映射机制的成熟促进发音模仿, 婴儿通过观察说话者的构音动作, 更精确地控制发声器官运动(Imafuku et al., 2019a)。最后, 动态注视策略的调整。婴儿能根据任务需求灵活调整注视策略, 例如面对不符合语法规则的刺激时, 其嘴部注视时长显著增加(Birulés et al., 2022)。发展神经模型揭示, 此阶段跨模态信息的传递效率提升至 100 毫秒水平(Pons & Lewkowicz, 2014)。值得注意的是, 双语环境婴儿的嘴部注视优势持续时间较单语组延长 4~6 个月(Pons et al., 2015), 这一调整深刻体现了认知系统的可塑性。

(4) 9~12 个月: 互动意识的提升。婴儿的注视模式逐渐呈现眼部−嘴部动态平衡的特征。这种转变源于以下两个认知变化。首先, 随着母语语音范畴化基本完成, 嘴部运动的视觉信息价值相对降低, 促使注意资源向更具社会价值的眼部区域转移(Mercure et al., 2022)。其次, 心理理论能力的萌芽使婴儿能通过眼部线索推断说话者的交流意图, 从而显著提升互动质量。纵向研究表明, 12 月龄时建立的共同注意能力可显著预测 24 月龄的词汇量(Wu et al., 2014)。此时, 婴儿的跨模态整合机制已具备多层化特征：在基础层面, 时间同步性检测精度显著提高, 能处理复杂语境下的视听异步(Lewkowicz, 2010); 在高级层面, 能够利用音素特征(音调、音色)、情感表达、说话者的性别等非模态特征进行多源信息整合(Hillairet de Boisferon et al., 2016)。这种双重加工能力使婴儿能同时利用眼部传达的社会参照信息和嘴部提供的语音线索, 形成语言学习的双通道优化系统。

(5) 12 个月后: 视听协同的优化。随着语言系统的初步建立, 婴儿注意资源的分配根据任务需求呈现动态调整特征。在词汇习得情境中, 嘴部注视比例仍保持优势, 以强化语音−语义映射。在社会互动场景中, 眼部注视重新占据主导, 以捕捉非言语交流线索(Santapuram et al., 2022)。这种动态平衡机制的建立标志着跨模态整合能力的成熟。研究表明, 12 月龄的跨模态加工能力能够独立解释 24 月龄语言水平的变异(Edgar et al., 2022)。此外, 婴儿开始发展元认知监控能力, 能根据交流对象的注视方向主动调整自身注意策略(Wu et al., 2014)。这种双向互动能力的形成将语言学习推向更高的发展阶段。

综上所述, 婴儿视听整合能力的发展呈现出明显的阶段性特征。在初始阶段, 婴儿主要依赖对眼部和嘴部的生物性偏好建立视听匹配; 随后通过注视策略的调整, 逐步强化对视听线索的解码能力; 最终在社会认知的驱动下实现视听协同优化。这一过程既受时间同步性检测和统计学习的调控, 也在婴儿与照顾者的动态互动中持续塑造。贯穿整个发展进程的婴儿导向语言(IDS)作为一种特异性视听信号(Tan et al., 2023), 通过视听线索的夸张和强化, 持续优化婴儿的跨模态语言学习环境, 为其语言能力的发展提供了重要支持。

3.2　高风险自闭症谱系障碍婴儿的视听言语整合

3.2.1　高风险自闭症谱系障碍婴儿对视觉信息的提取

由于自闭症谱系障碍(ASD)极少在 2 岁前确诊, 因此本文关注的高风险 ASD 婴儿特指具有较高概率在后期被诊断为 ASD 的婴幼儿群体。这一群体通常包括以下特征：具有 ASD 家族史, 表现出与 ASD 相关的早期行为特征(如社交互动异常、语言发展延迟、重复刻板行为等), 具有 ASD 相关的生物标记物。

研究表明, 高风险 ASD 婴儿在生命第二年常表现出语言延迟, 这一现象与其对面部区域的非典型注意模式相关。事件相关电位(ERP)研究显示, 6 个月大高风险 ASD 婴儿(即后来被诊断为 ASD 儿童的兄弟姐妹)对面部表情的 ERP 反应显著减弱, 且对面部的视觉注意力持续时间较短(Kolesnik et al., 2019)。此外, 8 个月大高风险 ASD 婴儿对面部刺激的 N290 成分反应降低(van Noordt et al., 2022), 表明其无法有效利用视觉韵律和言语线索进行社会性学习。高风险 ASD 婴儿可能因早期感知偏差, 优先加工低水平物体特征而忽视社会性

线索(如缺乏对母语语音和熟悉面孔的优先加工), 从而形成感知狭窄特征(Campos, 2018; Johnson et al., 2015)。这种加工偏向导致两个关键发展障碍。一方面, 眼区注视减少削弱了社会互动中的情感联结和共同注意机会, 直接影响词汇习得的社会认知基础。另一方面, 由于眼神接触减少, 难以引发照顾者的正向反馈, 形成“社交回避-学习语言机会减少”的恶性循环。值得注意的是, 高风险 ASD 婴儿对嘴部的关注并未伴随语言能力的提升, 表明其利用视觉线索进行语言学习的能力存在缺陷(Chawarska et al., 2022)。这种“嘴部关注-语言能力”关联的断裂, 可能源于高风险 ASD 婴儿在整合视觉线索与语言信息上的障碍, 使其对嘴部运动的注意未能有效促进语言发展。

3.2.2 高风险自闭症谱系障碍婴儿视听言语整合发展的缺陷

高风险自闭症谱系障碍婴儿的视听加工能力发展呈现出独特的轨迹, 其社会性注意和跨模态整合能力在生命早期即表现出显著异常。以下从不同发展阶段阐述其视听加工特征及潜在机制。

(1) 0~3 个月: 早期社会注意的神经行为偏离。典型发育婴儿在出生后即表现出对人脸核心区域(如眼睛)的先天偏好(郭桐阳 等, 2024), 这种偏好在母婴互动过程中不断强化, 成为社会认知发展的基石。然而, 高风险 ASD 婴儿的视觉加工模式与典型发育婴儿存在细微但重要的差异。虽然新生儿普遍具有对直视的敏感性(Farroni et al., 2002), 但高风险 ASD 婴儿更倾向于优先加工局部客体特征(如明暗对比或运动轨迹), 而相对忽视眼神接触等社会性线索(Johnson et al., 2015)。纵向研究表明, 高风险 ASD 婴儿在出生后前两个月对眼睛的注视时间与典型发育婴儿相似, 但从 2 个月开始, 其注视时间会逐渐减少(Jones & Klin., 2013)。这种注视行为的衰减可能与早期视觉-社会网络的发育异常有关。神经影像学研究显示, 高风险 ASD 婴儿的下丘脑-枕叶白质通路存在特征性异常(Nair et al., 2021), 不过在这一阶段, 其行为表现尚未达到显著异常水平。特别需要关注的是, 母婴互动过程中的反馈循环可能会放大这种早期异常。当婴儿对眼神接触的反应较弱时, 父母可能会无意识地减少视觉反馈, 从而进一步限制婴儿视觉社会学习的机会(Hyvärinen et al., 2014)。这种双向互动过程可能会加剧高风险 ASD 婴儿在社会性注意发展方面的偏离轨迹。

(2) 3~6 个月: 社会性注意衰减与视听整合的早期异常。典型发育婴儿的注视模式呈现从眼部到嘴部的自然过渡, 4.5 月龄时已能通过 McGurk 效应展现初步的视听整合能力(Burnham & Dodd., 2004), 并通过母婴互动强化眼部注视形成共同注意基础。相比之下, 高风险 ASD 婴儿表现出显著的社会性注意衰减: 6 月龄时眼部注视时间持续减少(Elsabbagh et al., 2013; Jones & Klin., 2013), 在动态社会场景中对社会场景和面孔的关注度降低(Chawarska et al., 2013)。神经影像证据显示, 高风险 ASD 婴儿的颞叶皮层对社会性视听觉刺激的反应较弱(Zhang & Roeyers., 2019)。在视听整合方面, 高风险 ASD 婴儿虽然能检测视听同步, 但敏感性可能低于典型婴儿。此外, 高风险 ASD 婴儿在 5 个月时已出现左侧颞叶语言区对母语刺激的神经反应减弱, 缺乏对母语与非母语的神经区分能力(Wagner et al., 2025)。这些早期异常可能影响视觉-听觉跨模态整合的效率。

(3) 6~9 个月: 视听整合缺陷的显化阶段。典型发育婴儿的嘴部注视比例从 6 月龄的 35%提升至 12 月龄的 62%, 这有助于其整合视听线索, 促进语音解码(Lewkowicz & Hansen-Tift, 2012)。高风险 ASD 婴儿在面部注视模式上表现出非典型的发展轨迹。虽然部分研究观察到他们存在嘴部注视增长趋势, 但其增幅显著低于典型发育婴儿(Lozano et al., 2024), 直到 18 月龄时才出现明显的嘴部注视比例提升(Minagawa et al., 2019)。神经机制研究发现, 高风险 ASD 婴儿大脑区域之间的信息传递效率低下(Zhang & Roeyers., 2019)。行为上表现为:对视听同步刺激(如说话面孔)的敏感性降低(Suri et al., 2023)。9 月龄时无法像典型发育婴儿那样整合冲突的视听信息(McGurk 效应缺失)(Guiraud et al., 2012)。这种缺陷源于时间绑定窗口过宽和注意分配异常。值得注意的是, 女性高风险 ASD 婴儿可能通过更强的嘴部注视能力(Kleberg et al., 2019)部分弥补语言延迟, 而男性缺陷更为明显(Long et al., 2024)。

(4) 9~12 个月: 社会沟通缺陷。这一阶段典型发育婴儿的视听整合能力显著提升, McGurk 效应增强(Guiraud et al., 2012), 而高风险 ASD 婴儿仍表现出明显的视听整合困难。行为上, 高风险 ASD 婴儿对语音-嘴部运动同步性的检测阈值较典型

发育婴儿较宽(Suri et al., 2023), 导致实时语音解码效率较低。同时, 高风险 ASD 婴儿对嘴部的注视时间增长缓慢, 显著少于典型发育婴儿(Lozano et al., 2025), 导致嘴部注视与语言发展的关联断裂(Chawarska et al., 2022)。这些缺陷直接反映在语言发展中: 12 个月时高风险 ASD 婴儿表现出语言产出和理解的延迟(Belteki et al., 2022; Hudry et al., 2014)。在社会互动方面, 高风险 ASD 婴儿往往表现出眼神接触减少, 这会降低父母与之互动的频率(Jones & Klin., 2013)。与此同时, 这些婴儿更倾向于关注非社会性刺激(如几何图形)(Pierce et al., 2011), 从而形成“社交减少−互动机会减少−社交能力发展滞后”的恶性循环。值得注意的是, 这一阶段家长通常开始注意到孩子的异常表现(Ozonoff et al., 2009)。临床实践表明, 结合标准化评估工具可显著提高早期筛查的准确性(Talbott et al., 2025)。

(5) 12 个月后: 视觉−社会沟通缺陷进入稳定关联期。在视听整合方面, 与典型发育婴儿相比, 高风险 ASD 婴儿在语音模仿能力的发展上表现出明显延迟(Young et al., 2011), 且其在听觉信息加工存在显著的困难(Germani et al., 2014; Demopoulos & Lewine, 2016)。12 个月高风险 ASD 婴儿在接受性语言能力和联合注意行为上表现出明显的缺陷(Wagner et al., 2025)。这些缺陷导致高风险 ASD 婴儿难以有效利用视听整合线索来促进语言习得(Wang et al., 2025)。神经影像学研究发现, 高风险 ASD 婴儿在 12~24 个月之间脑容量过度生长(Hazlett et al., 2017), 且在 24 个月时表现出视觉网络与默认模式网络(DMN)连接减弱, 这种神经连接异常与社交缺陷显著相关(Lombardo et al., 2019)。这些缺陷最终导致：语言产出减少, 特别是表达性语言发展滞后(Tsang et al., 2018); 社会性事件的注视显著减少(Santapuram et al., 2022)。在 24 个月时, 这些儿童在自闭症诊断观察量表(ADOS)中表现出联合注意与社交互动能力低下(Gotham et al., 2009), 逐渐形成 ASD 的核心临床症状。

综上所述, 高风险 ASD 婴儿的语言发展障碍本质上源于社会认知与跨模态整合系统的异常。这一异常发展轨迹呈现出明显的阶段性特征: 在早期表现为社会性注意的衰减(如眼部注视减少、嘴巴注视增长缓慢), 随后逐步发展为视听整合能力的明显缺陷(如时间同步感知阈值升高、McGurk 效应缺失等)。这些阶段性特征为高风险 ASD 婴儿的早期识别和干预策略提供重要启示。

4　高风险自闭症谱系障碍婴儿的干预措施

基于高风险 ASD 婴儿在社会性注意衰减与跨模态整合障碍的核心缺陷, 以下干预策略旨在通过优化环境输入和互动模式促进语言和社会认知发展。

针对社会性注意衰减的干预。在 0~6 月龄阶段, 建议采用生物运动偏好引导的多模态干预：利用高对比度动态面部刺激(如夸张表情与同步语音组合), 结合眼动追踪反馈系统实现行为强化。当婴儿注视照顾者眼部区域时触发动态视觉奖励, 通过操作性条件反射建立注视联结。对于 6~12 月龄婴儿, 可引入联合注意闭环训练系统：通过可穿戴式眼动仪实时监测视线方向, 在检测到共同注意发起时自动播放目标物体的语音标签, 形成“注视−听觉反馈”强化环路。实施过程中需整合家庭场景, 指导照顾者采用婴儿导向言语(Infant-Directed Speech, IDS)并保持面对面交互姿势, 通过语速调控、面部表情增强及肢体动作同步性优化, 显著提升社会性刺激的知觉显著性。

针对视听时间同步性感知缺陷以及跨模态整合障碍的干预。6~12 月龄阶段实施视听冗余梯度训练：从简单音节−唇动同步配对(如/ma/-闭合唇形)的基础映射开始, 通过高频率重复强化语音−视觉关联；进一步引入噪音干扰情境, 训练婴儿通过视觉线索进行语音辨别, 从而增强视听补偿机制。为促进语言意图理解, 采用虚拟现实技术模拟眼神−语音协同指向(如虚拟形象转头注视物体并同步命名), 结合多模态韵律匹配游戏(如高兴语调与微笑表情动态联系), 系统提升情感−语义跨通道整合能力。值得注意的是, 母亲敏感性(maternal sensitivity)——即互动中表现出的偶发性、合作性与可及性特质——可显著影响婴儿的多感官统合效能(Rothbart et al., 2011)。最新纵向研究表明, 高敏感性母亲的婴儿在跨模态整合能力上具有显著优势(Bruce et al., 2022), 这为基于视频反馈的家庭干预提供了理论依据。通过记录分析亲子互动场景并给予实时反馈, 可使家长更精准识别婴儿的注意焦点与需求, 进而优化互动

策略。此外，婴儿导向语言(infant-directed speech)通过强化视听同步性及激发积极情绪(Alviar et al., 2023)，能进一步促进婴儿的语言理解与社会互动能力发展。

5 总结和展望

婴儿语言习得是一个依赖视听协同的动态发展过程，典型发育婴儿与高风险自闭症谱系障碍(ASD)婴儿在这一过程中存在明显差异。典型发育婴儿的发展呈现规律性阶段特征：0~6 个月主要通过注视眼部区域建立社会认知基础(Jones & Klin, 2013)，6~12 个月则转向嘴部注视主导的语音解码模式(Lewkowicz & Hansen-Tift, 2012)。这一转变使其能够通过精准的时间同步性检测和统计学习(如音节−嘴部共现规律)优化语言输入效率，最终形成灵活的注意策略(Edgar et al., 2023)。婴儿导向语言(IDS)的特征性夸张表达(如放慢语速、夸张的嘴部动作)为这一过程提供了重要支持(Kosie & Lew-Williams, 2024)。相比之下，高风险 ASD 婴儿表现出双重缺陷：早期社会性注意不足(2~6 个月时眼部注视时间显著减少)导致联合注意发展滞后(Jones & Klin, 2013)；视听整合异常表现为时间同步性感知阈值扩大(Suri et al., 2023)和 McGurk 效应缺失(Guiraud et al., 2012)。这些缺陷形成恶性循环：嘴部注视发展迟缓影响语音解码效率(Chawarska et al., 2022)，而眼神接触减少又进一步减少了与照顾者的语言互动机会。

未来研究可在研究方法和干预策略上继续完善。在方法学层面，应突破现有实验室范式的局限，通过可穿戴眼动技术结合自然情境观察(如喂食、游戏等)，深入探究母婴互动中注视行为与语言输入的动态关联机制，以更真实地揭示多模态语言习得的本质。在干预策略方面，建议基于 ASD 发展的阶段性特征构建精准干预模型：早期(0~6 个月)重点重建社会性注意，后期(6~12 个月)强化视听整合训练。同时，需要开发智能化的婴儿导向语言干预方案，在保持声学夸张特性的基础上，通过实时反馈技术优化照顾者的互动质量，包括嘴部运动幅度、表情−语调匹配度等关键参数。这种融合自然观察方法与智能化干预手段的研究范式，将有助于更全面地理解 ASD 的发病机制，并为早期干预提供更精准的指导。

参考文献

陈钰, 张丹丹. (2020). 新生儿语音感知的脑机制. *心理科学, 43*(4), 844−849. https://doi.org/10.16719/j.cnki.1671-6981.20200411

郭桐阳, 莫李澄, 张丹丹. (2024). 婴儿面孔和注视方向加工的认知神经机制. *心理科学进展, 32* (10), 1670−1679.

李思瑾, 王庭栋, 彭芝琳, 张丹丹. (2023). 新生儿对语音的感知、辨别和学习. *心理科学进展, 31*(12), 2295−2305. https://doi.org/10.3724/sp.J.1042.2023.02295

Alviar, C., Sahoo, M., Edwards, L. A., Jones, W., Klin, A., & Lense, M. (2023). Infant-directed song potentiates infants' selective attention to adults' mouths over the first year of life. *Developmental Science, 26*(5), e13359. https://doi.org/10.1111/desc.13359

Bastianello, T., Keren-Portnoy, T., Majorano, M., & Vihman, M. (2022). Infant looking preferences towards dynamic faces: A systematic review. *Infant Behavior and Development, 67*, 101709. https://doi.org/10.1016/j.infbeh.2022.101709

Belteki, Z., van den Boomen, C., & Junge, C. (2022). Face-to-face contact during infancy: How the development of gaze to faces feeds into infants' vocabulary outcomes. *Frontiers in Psychology, 13*, 997186. https://doi.org/10.3389/fpsyg.2022.997186

Bergelson, E., & Swingley, D. (2012). At 6−9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253−3258. https://doi.org/10.1073/pnas.1113380109

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2018). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science, 22*(3), e12755. https://doi.org/10.1111/desc.12755

Birulés, J., Martinez-Alvarez, A., Lewkowicz, D. J., de Diego-Balaguer, R., & Pons, F. (2022). Violation of non-adjacent rule dependencies elicits greater attention to a talker's mouth in 15-month-old infants. *Infancy, 27*(5), 963−971. https://doi.org/10.1111/infa.12489

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*(4), 298−304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology, 130*, 67−78. https://doi.org/10.1016/j.jecp.2014.09.010

Bruce, M., Panneton, R., & Taylor, C. (2022). Multisensory integration and maternal sensitivity are related to each other and predictive of expressive vocabulary in 24-month-olds. *Journal of Experimental Child Psychology, 214*, 105304. https://doi.org/10.1016/j.jecp.2021.105304

Buiatti, M., Di Giorgio, E., Piazza, M., Polloni, C., Menna, G., Taddei, F., … Vallortigara, G. (2019). Cortical route for facelike pattern processing in human newborns. *Proceedings of the National Academy of Sciences, 116*(10), 4625−4630. https://doi.org/10.1073/pnas.1812419116

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204−220. https://doi.org/10.1002/dev.20032

Campos, R. (2018). If you want to get ahead, get a good master. Annette Karmiloff-Smith: The developmental perspective. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, *41*(1), 90−137. https://doi.org/10.1080/02103702.2017.1401318

Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology*, *11*, 589096. https://doi.org/10.3389/fpsyg.2020.589096

Chawarska, K., Lewkowicz, D., Feiner, H., Macari, S., & Vernetti, A. (2022). Attention to audiovisual speech does not facilitate language acquisition in infants with familial history of autism. *Journal of Child Psychology and Psychiatry*, *63*(12), 1466−1476. http://10.1111/jcpp.13595

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, *74*(3), 195−203. https://doi.org/10.1016/j.biopsych.2012.11.022

Coffey, J. R., & Snedeker, J. (2025). How strong is the relationship between caregiver speech and language development? A meta-analysis. *Journal of Child Language*. Advance online publication. https://doi.org/10.1017/S0305000924000692

Curtin, S., & Zamuner, T. S. (2014). Understanding the developing sound system: Interactions between sounds and words. *WIREs Cognitive Science*, *5*(5), 589−602. https://doi.org/10.1002/wcs.1307

Demopoulos, C., & Lewine, J. D. (2016). Audiometric profiles in autism spectrum disorders: Does subclinical hearing loss impact communication? *Autism Research*, *9*(1), 107−120. https://doi.org/10.1002/aur.1495

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2022). Intersensory matching of faces and voices in infancy predicts language outcomes in young children. *Developmental Psychology*, *58*(8), 1413−1428. https://doi.org/10.1037/dev0001375

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2023). Intersensory processing of faces and voices at 6 months predicts language outcomes at 18, 24, and 36 months of age. *Infancy*, *28*(3), 569−596. https://doi.org/10.1111/infa.12533

Elsabbagh, M., Gliga, T., Pickles, A., Hudry, K., Charman, T., Johnson, M. H., & BASIS Team. (2013). The development of face orienting mechanisms in infants at-risk for autism. *Behavioural Brain Research*, *251*, 147−154. https://doi.org/10.1016/j.bbr.2012.07.030

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602−9605. https://doi.org/10.1073/pnas.152159999

Germani, T., Zwaigenbaum, L., Bryson, S., Brian, J., Smith, I., Roberts, W., … Vaillancourt, T. (2014). Brief report: Assessment of early sensory processing in infants at high-risk of autism spectrum disorder. *Journal of Autism and Developmental Disorders, 44*(12), 3264−3270. https://doi.org/10.1007/s10803-014-2175-x

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences, 105*(37), 14222−14227. https://doi.org/10.1073/pnas.0806530105

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693−705. https://doi.org/10.1007/s10803-008-0674-3

Guiraud, J. A., Tomalski, P., Kushnerenko, E., Ribeiro, H., Davies, K., Charman, T., & BASIS Team. (2012). Atypical audiovisual speech integration in infants at risk for autism. *PLOS One, 7*(5), Article e36428. https://doi.org/10.1371/journal.pone.0036428

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., … IBIS Network. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*(7641), 348−351. https://doi.org/10.1038/nature21369

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2016). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3), e12381. https://doi.org/10.1111/desc.12381

Hudry, K., Chandler, S., Bedford, R., Pasco, G., Gliga, T., Elsabbagh, M., … Charman, T. (2014). Early language profiles in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(1), 154−167. https://doi.org/10.1007/s10803-013-1861-4

Hurley, K. B., & Oakes, L. M. (2015). Experience and distribution of attention: Pet exposure and infants' scanning of animal images. *Journal of Cognition and Development*, *16*(1), 11−30. https://doi.org/10.1080/15248372.2013.833922

Hyvärinen, L., Walthes, R., Jacob, N., Chaplin, K. N., & Leonhardt, M. (2014). Current understanding of what infants see. *Current Ophthalmology Reports, 2*(4), 142−149. https://doi.org/10.1007/s40135-014-0056-2

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019a). Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Developmental Science, 22*(6), e12825. https://doi.org/10.1111/desc.12825

Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., & Myowa, M. (2019b). Audiovisual speech perception and language acquisition in preterm infants: A longitudinal study. *Early Human Development, 128,* 93−100. https://10.1016/j.earlhumdev.2018.11.001

Ishikawa, M., Senju, A., Kato, M., & Itakura, S. (2022). Physiological arousal explains infant gaze following in various social contexts. *Royal Society Open Science*, *9*(8), 220592. https://doi.org/10.1098/rsos.220592

Johnson, M. H., Jones, E. J., & Gliga, T. (2015). Brain adaptation and alternative developmental trajectories. *Developmental Psychopathology, 27*(2), 425−442. https://doi.org/10.1017/S0954579415000073

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature, 504*(7480), 427−431. https://doi.org/10.1038/nature12715

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development, 85*(5), 1795−1804. https://doi.org/10.1111/cdev.12263

Kleberg, J. L., Nyström, P., Bölte, S., & Falck-Ytter, T. (2019). Sex differences in social attention in infants at risk for autism. *Journal of Autism and Developmental Disorders*, *49*(4), 1342−1351. https://doi.org/10.1007/s10803-018-3799-z

Kolesnik, A., Begum Ali, J., Gliga, T., Guiraud, J., Charman, T., Johnson, M. H., … BASIS Team. (2019). Increased cortical reactivity to repeated tones at 8 months in infants with later ASD. *Translational Psychiatry, 9*(1), 46. https://doi.org/10.1038/s41398-019-0393-x

Kosie, J. E., & Lew-Williams, C. (2024). Infant-directed communication: Examining the many dimensions of everyday caregiver-infant interactions. *Developmental Science, 27*(5), e13515. https://doi.org/10.1111/desc.13515

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713−727. https://doi.org/10.1016/j.neuron.2010.08.038

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Inverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), 13−21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096−9101. https://doi.org/10.1073/pnas.1532872100

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. https://doi.org/10.1126/science.1736364

Kujala, T., Partanen, E., Virtala, P., & Winkler, I. (2023). Prerequisites of language acquisition in the newborn brain. *Trends in Neuroscience*, *46*(9), 726−737. https://doi.org/10.1016/j.tins.2023.05.011

Kwon, M.-K., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, *52*(4), 537−555. https://doi.org/10.1037/dev0000080

Lalonde, K., & Werner, L. A. (2019). Infants and adults use visual cues to improve detection and discrimination of speech in noise. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3860−3875. https://doi.org/10.1044/2019_jslhr-h-19-0106

Leng, L. L., Zhu, Y. W., & Zhou, L. G. (2024). Explaining differences in autism detection timing: Age of diagnosis and associated individual and socio-familial factors in Chinese children. *Autism, 28*(4), 896−907. https://doi.org/10.1177/13623613231187184

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1094−1106. https://doi.org/10.1037//0096-1523.22.5.1094

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology, 46*(1), 66−77. https://doi.org/10.1037/a0015579

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences, 109*(5), 1431−1436. https://doi.org/10.1073/pnas.1114783109

Lombardo, M. V., Eyler, L., Moore, A., Datko, M., Carter Barnes, C., Cha, D., … Pierce, K. (2019). Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. *eLife*, *8*, e47427. https://doi.org/10.7554/eLife.47427

Long, H. L., Ramsay, G., Bene, E. R., Su, P. L., Yoo, H., Klaiman, C., … Oller, D. K. (2024). Canonical babbling trajectories across the first year of life in autism and typical development. *Autism*, *28*(12), 3078−3091. https://doi.org/10.1177/13623613241253908

Lozano, I., Campos, R., & Belinchón, M. (2025). Sensitivity to temporal synchrony in audiovisual speech and language development in infants with an elevated likelihood of autism: A developmental review. *Infant Behavior & Development*, *78*, 102026. https://doi.org/10.1016/j.infbeh.2024.102026

Lozano, I., Viktorsson, C., Capelli, E., & Gliga, T. (2024, July). *Developmental pathways of audiovisual processing in infants at elevated likelihood of autism during the first year of life.* Paper presented at the meeting of the International Congress of Infant Studies, Glasgow, Scotland.

Maurer, D., & Werker, J. F. (2013). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154−178. https://doi.org/10.1002/dev.21177

Mercure, E., Bright, P., Quiroz, I., & Filippi, R. (2022). Effect of infant bilingualism on audiovisual integration in a McGurk task. *Journal of Experimental Child Psychology*, *217*, 105351. https://doi.org/10.1016/j.jecp.2021.105351

Minagawa, Y., Hata, M., Hakuno, Y., Yamamoto, E., & Abe, K. (2019, May). *Developmental trajectory of selective attention to a talking face in infants at-risk for ASD*. Poster session presented at the meeting of the International Society for Autism Research, Montreal, QC.

Nair, A., Jalal, R., Liu, J., Tsang, T., McDonald, N. M., Jackson, L., … Dapretto, M. (2021). Altered thalamocortical connectivity in 6-week-old infants at high familial risk for autism spectrum disorder. *Cerebral Cortex*, *31*(9), 4191−4205. https://doi.org/10.1093/cercor/bhab078

Ozonoff, S., Young, G. S., Steinfeld, M. B., Hill, M. M., Cook, I., Hutman, T., … Sigman, M. (2009). How early do parent concerns predict later autism diagnosis? *Journal of Developmental and Behavioral Pediatrics*, *30*(5), 367−375. https://doi.org/10.1097/dbp.0b013e3181ba0fcf

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, *68*(1), 101−109. https://doi.org/10.1001/archgenpsychiatry.2010.113

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, *26*(4), 490−498. https://doi.org/10.1177/0956797614568320

Pons, F., Bosch, L., & Lewkowicz, D. J. (2019). Twelve-month-old infants' attention to the eyes of a talking face is associated with communication and social skills. *Infant Behavior and Development*, *54*, 80−84. https://doi.org/10.1016/j.infbeh.2018.12.003

Pons, F., & Lewkowicz, D. J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica*, *149*, 142−147. https://doi.org/10.1016/j.actpsy.2013.12.013

Quinn, P. C., Lee, K., & Pascalis, O. (2019). Face processing in infancy and beyond: The case of social categories. *Annual Review of Psychology*, *70*, 165−189. https://doi.org/10.1146/annurev-psych-010418-102753

Riva, V., Riboldi, E. M., Dondena, C., Piazza, C., Molteni, M., & Cantiani, C. (2022). Atypical ERP responses to audiovisual speech integration and sensory responsiveness in infants at risk for autism spectrum disorder. *Infancy*, *27*(2), 369−388. https://doi.org/10.1111/infa.12456

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*(6), 405−409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Rothbart, M. K., Sheese, B. E., Rueda, M. R., & Posner, M. I. (2011). Developing mechanisms of self-regulation in early life. *Emotion Review*, *3*(2), 207−213. https://doi.org/10.1177/1754073910387943

Sandoval, M., & Gómez, R. L. (2016). Overriding the metrical bias with lexical information: English-learning 7.5-month-olds use *Mommy* to segment Iambic words. *Language Learning and Development*, *12*(4), 398−412. https://doi.org/10.1080/15475441.2016.1162722

Santapuram, P., Feldman, J. I., Bowman, S. M., Raj, S., Suzman, E., Crowley, S., … Woynaroski, T. G. (2022). Mechanisms by which early eye gaze to the mouth during multisensory speech influences expressive communication development in infant siblings of children with and without Autism. *Mind, Brain, and Education*, *16*(1), 62−74. https://doi.org/10.1111/mbe.12310

Sheldrick, R. C., Maye, M. P., & Carter, A. S. (2017). Age at first identification of autism spectrum disorder: An analysis of two US surveys. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(4), 313−320. https://doi.org/10.1016/j.jaac.2017.01.012

Suri, K. N., Whedon, M., & Lewis, M. (2023). Perception of audio-visual synchrony in infants at elevated likelihood of developing autism spectrum disorder. *European Journal of Pediatrics*, *182*(5), 2105−2117. https://doi.org/10.1007/s00431-023-04871-y

Talbott, M. R., Young, G. S., & Ozonoff, S. (2025). Can combining existing behavioral tools improve identification of infants at elevated likelihood of autism in the first year of life? *Autism*, *29*(2), 462−475. https://doi.org/10.1177/13623613241275455

Tan, S. H. J., Kalashnikova, M., Di Liberto, G. M., Crosse, M. J., & Burnham, D. (2023). Seeing a talking face matters: Gaze behavior and the auditory-visual speech benefit in adults' cortical tracking of infant-directed speech. *Journal of Cognitive Neuroscience, 35*(11), 1741−1759. https://doi.org/10.1162/jocn_a_02044

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F., & Morgan, J. L. (2014). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173−1190. https://doi.org/10.1017/s0305000914000725

Tomalski, P. (2015). Developmental trajectory of audiovisual speech integration in early infancy. A review of studies using the McGurk paradigm. *Psychology of Language and Communication*, *19*(2), 77−100. https://doi.org/10.1515/plc-2015-0006

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology, 169*, 93−109. https://doi.org/10.1016/j.jecp.2018.01.002

Ujiie, Y., Kanazawa, S., & Yamaguchi, M. K. (2021). The other-race effect on the McGurk effect in infancy. *Attention, Perception, & Psychophysics, 83*(7), 2924−2936. https://doi.org/10.3758/s13414-021-02342-w

van Noordt, S., Desjardins, J. A., BASIS Team, & Elsabbagh, M. (2022). Inter-trial theta phase consistency during face processing in infants is associated with later emerging autism. *Autism Research, 15*(5), 834−846. http://10.1002/aur.2701

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science, 10*(2), 159−164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Wagner, L., Banchik, M., Tsang, T., Okada, N. J., Altshuler, R., McDonald, N., Bookheimer, S. Y., Jeste, S. S., Green, S. A., & Dapretto, M. (2025). Atypical early neural responses to native and non-native language in infants at high likelihood for developing autism. *Molecular Autism, 16*, 6. https://doi.org/10.1186/s13229-025-00640-w

Wang, X., Bouton, S., Kojovic, N., Giraud, A. L., & Schaer,

M. (2025). Atypical audio-visual neural synchrony and speech processing in early autism. *Journal of Neurodevelopmental Disorders, 17*(1), 9. https://doi.org/10.1186/s11689-025-09593-w

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143−2152. https://doi.org/10.1177/0956797613488145

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. https://doi.org/10.1016/S0163-6383(84)80022-3

Wu, R., Tummeltshammer, K. S., Gliga, T., & Kirkham, N. Z. (2014). Ostensive signals support learning from novel attention cues during infancy. *Frontiers in Psychology*, *5*, 251. https://doi.org/10.3389/fpsyg.2014.00251

Wu, Y. J., Hou, X., Peng, C., Yu, W., Oppenheim, G. M., Thierry, G., & Zhang, D. (2022). Rapid learning of a phonemic discrimination in the first hours of life. *Nature Human Behavior, 6*(8), 1169−1179. https://doi.org/10.1038/s41562-022-01355-1

Xiao, N. G., Mukaida, M., Quinn, P. C., Pascalis, O., Lee, K., & Itakura, S. (2018). Narrowing in face and speech perception in infancy: Developmental change in the relations between domains. *Journal of Experimental Child Psychology, 176*, 113−127. https://doi.org/10.1016/j.jecp.2018.06.007

Yan, L., Liu, X., Hu, S., Liu, S., & Xiao, G. (2024, July). *Infants demonstrated the McGurk effect with own-race but not other-race faces.* Paper presented at the meeting of the International Congress of Infant Studies, Glasgow, Scotland.

Young, G. S., Rogers, S. J., Hutman, T., Rozga, A., Sigman, M., & Ozonoff, S. (2011). Imitation from 12 to 24 months in autism and typical development: A longitudinal Rasch analysis. *Developmental Psychology*, *47*(6), 1565−1578. https://doi.org/10.1037/a0025418

Zhang, F., & Roeyers, H. (2019). Exploring brain functions in autism spectrum disorder: A systematic review on functional near-infrared spectroscopy (fNIRS) studies. *International Journal of Psychophysiology*, *137*, 41−53.

Zhou, X., Wang, L., Hong, X., & Wong, P. C. M. (2023). Infant‐directed speech facilitates word learning through attentional mechanisms: An fNIRS study of toddlers. *Developmental Science, 27*(1), e13424. https://doi.org/10.1111/desc.13424

Audiovisual integration in infant language acquisition: Different patterns in typically developing infants and those at elevated risk for autism spectrum disorder

JIN Mengke[1], YAN Linlin[1], LIU Shaoying[1], XIAO Naiqi[2]

([1] *Department of Psychology, School of Science, Zhejiang Sci-Tech University, Hangzhou* 310018, *China*)

([2] *Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton* L8S4L8, *Canada*)

Abstract: Multimodal sensory mechanisms are critical for language acquisition in infancy. Audiovisual integration, in particular, plays a pivotal role in developing linguistic skills among typically developing infants. In contrast, infants at high risk for autism spectrum disorder (ASD) often encounter significant challenges in this integration process. Typical developmental trajectories of speech perception emphasize the importance of processing facial features. Specifically, attention to the eye and mouth regions facilitates language learning. Empirical evidence shows infants as young as 4.5 months demonstrate audiovisual integration abilities. These early abilities reliably predict later language development outcomes. In contrast, infants at high risk for ASD show reduced social attention and impaired audiovisual integration. Such deficits may disrupt typical language acquisition pathways. Therefore, early intervention strategies should prioritize biologically-informed, sensory-guided approaches. These approaches should focus on enhancing multisensory integration rather than exclusively targeting attentional behaviors. Understanding these mechanisms advances our knowledge of typical language development. Moreover, it provides empirical foundations for creating targeted early interventions to support language acquisition in high-risk ASD populations.

Keywords: speech perception, audiovisual matching, multisensory integration, high-risk autism spectrum disorder (ASD) infants, language development, early intervention