

# 孤独症谱系障碍个体的视听言语整合

何建青<sup>1</sup> 胡耿丹<sup>1</sup> 胡金生<sup>2</sup>

(1. 同济大学人文学院, 上海, 200082 ; 2. 辽宁师范大学心理学院, 大连, 116021 )

**摘 要** 孤独症谱系障碍个体的视听言语整合存在异常, 但视听言语整合能力的发展是迟滞还是停滞的依然无法确定。注意、时间感知缺陷、唇动识别缺陷、噪音背景都可能是孤独症者产生视听言语整合缺陷的原因。神经生理研究表明, 孤独症者的颞叶区域的活动以及 N4、N1、P2 等脑电成分波幅存在异常, 大脑皮层之间的连通性较低。未来研究应该加强对语义整合的研究、开发以言语整合为基础的干预方法。

**关键词** 孤独症谱系障碍; 视听言语整合; 原因探究; 神经机制

**分类号** G760

## The Audiovisual Speech Integration in Autism Spectrum Disorders

HE Jianqing<sup>1</sup> HU Gengdan<sup>1</sup> HU Jinsheng<sup>2</sup>

(1.School of Humanities, Tongji University, Shanghai, 200082 ;2. College of Psychology,

Liaoning Normal University, Dalian, 116021)

**Abstract:** The alterations of audiovisual speech integration of Autism Spectrum Disorders (ASD) is a severe problem. There are two quite different views on their development of the audiovisual integration ability: One is that the development of this ability is delayed, while the other points out the development of this ability is arrested. The deficits on attention , time perception, the flaw of lip movements recognition, and noise background, have different effects on the audiovisual speech integration of individuals with ASD. The neurophysiological studies showed abnormal brain activity with ASD, such as temporal lobe activities , amplitudes of N4, N1, P2 ERP component, the weak connectivity between the cerebral cortexes. Future studies should further explore the audiovisual semantic integration, develop intervention techniques based on speech integration.

**Key words:** autism spectrum disorders; audiovisual speech integration; reason analysis; neural mechanism

何建青, 女, 博士在读, 同济大学人文学院。研究方向: 心理与认知科学。通讯地址: 上海市杨浦区四平路 1239 号同济大学。邮编: 200082。电话: 15241191767。E-mail: 1810592@tongji.edu.cn

通讯作者: 胡金生, 男, 教授, 博士生导师, 心理学院院长, 辽宁师范大学心理学院。研究方向: 儿童认知与发展。通讯地址: 辽宁省大连市沙河口区黄河路 850 号 辽宁师范大学南院。电话: 18640816916。E-mail: hu\_jinsheng@126.com

胡耿丹, 男, 教授, 博士生导师, 同济大学人文学院。研究方向: 心理与认知科学。通讯地址: 上海市杨浦区四平路 1239 号同济大学。邮编: 200082。E-mail: 13918689338@163.com

# 孤独症谱系障碍个体的视听言语整合

<sup>1</sup>何建青 <sup>1</sup> 胡耿丹 <sup>1</sup> 胡金生 <sup>2</sup>

(1. 同济大学人文学院, 上海, 200082 ; 2. 辽宁师范大学心理学院, 大连, 116021 )

**摘要** 孤独症谱系障碍个体的视听言语整合存在异常, 但视听言语整合能力的发展是迟滞还是停滞的依然无法确定。注意、时间感知缺陷、唇动识别缺陷、噪音背景都可能是孤独症者产生视听言语整合缺陷的原因。神经生理研究表明, 孤独症者的颞叶区域的活动以及 N4、N1、P2 等脑电成分波幅存在异常, 大脑皮层之间的连通性较低。未来研究应该加强对语义整合的研究、开发以言语整合为基础的干预方法。

**关键词** 孤独症谱系障碍; 视听言语整合; 原因探究; 神经机制

**分类号** G760

## 1 引言

孤独症谱系障碍 (Autism Spectrum Disorders, ASD) 是一组广泛性神经发育障碍, 主要表现为社会交流障碍、狭隘兴趣以及重复和刻板行为, 其病因尚不明确, 涵盖了从基因到特定医学领域的广阔范围<sup>[1]</sup>。ASD 个体在生命早期就已表现出对感觉通道信息加工的异常, 较容易被无关的情景、声音、触觉信息所吸引并伴随紧张、焦虑情绪的产生, 对包含自身名字的有意义声音较少作出反应<sup>[2-4]</sup>。在各通道信息同时存在的现实交流情景中, 交流者不只从一个通道收集信息, ASD 个体的感知特点会降低各通道协调运作的效率。个体通过各感觉通道表达自身的内心所想, 更能准确的传达给他人自己的意图。ASD 个体在多感觉通道信息整合过程中表现出的非典型性、特别是在视听言语整合中表现出的缺陷可能是产生社会交流障碍的原因之一<sup>[5-8]</sup>。

视听言语整合 (audiovisual speech integration) 是指通过知觉机制将视、听通道的言语线索有效地整合为连贯、统一的言语知觉过程<sup>[9-11]</sup>。该过程涉及高级语言和基础认知过程, 但整合过程中的加工机制目前仍不明确。目前, 存在一些相关理论能帮助个体多角度理解视听言语的加工过程。例如, 前标记整合模型 (Pre-Labeling Integration Model, PRE) 认为, 听觉和视觉言语信息被指定为特定的标签, 在反应发生之前, 连续的言语信息数据已经进行了跨通道的整合<sup>[12]</sup>。言语知觉运动理论 (the motor theory of speech perception) 则认为感知言语就是感知声道发音的运动, 言语感知依赖于对动作表征的理解<sup>[13]</sup>。除此之外, 感知模糊逻辑模型 (The Fuzzy Logic Model of Perception, FLMP) 是通过贝叶斯推理假设视觉和听觉信息融合的一种方法<sup>[14]</sup>。FLMP 假设视听言语加工中输入的视觉、听觉言语信息以及储存在长

# 孤独症谱系障碍个体的视听言语整合

**摘要** 孤独症谱系障碍个体的视听言语整合存在异常，但视听言语整合能力的发展是迟滞还是停滞的依然无法确定。注意、时间感知缺陷、唇动识别缺陷、噪音背景都可能是孤独症者产生视听言语整合缺陷的原因。神经生理研究表明，孤独症者的颞叶区域的活动以及 N4、N1、P2 等脑电成分波幅存在异常，大脑皮层之间的连通性较低。未来研究应该加强对语义整合的研究、开发以言语整合为基础的干预方法。

**关键词** 孤独症谱系障碍；视听言语整合；原因探究；神经机制

## 1 引言

孤独症谱系障碍（Autism Spectrum Disorders, ASD）是一组广泛性神经发育障碍，主要表现为社会交流障碍、狭隘兴趣以及重复和刻板行为，其病因尚不明确，涵盖了从基因到特定医学领域的广阔范围<sup>[1]</sup>。ASD 个体在生命早期就已表现出对感觉通道信息加工的异常，较容易被无关的情景、声音、触觉信息所吸引并伴随紧张、焦虑情绪的产生，对包含自身名字的有意义声音较少作出反应<sup>[2-4]</sup>。在各通道信息同时存在的现实交流情景中，交流者不只从一个通道收集信息，ASD 个体的感知特点会降低各通道协调运作的效率。个体通过各感觉通道表达自身的内心所想，更能准确的传达给他人自己的意图。ASD 个体在多感觉通道信息整合过程中表现出的非典型性、特别是在视听言语整合中表现出的缺陷可能是产生社会交流障碍的原因之一<sup>[5-8]</sup>。

视听言语整合（audiovisual speech integration）是指通过知觉机制将视、听通道的言语线索有效地整合为连贯、统一的言语知觉过程<sup>[9-11]</sup>。该过程涉及高级语言和基础认知过程，但整合过程中的加工机制目前仍不明确。目前，存在一些相关理论能帮助个体多角度理解视听言语的加工过程。例如，前标记整合模型（Pre-Labeling Integration Model, PRE）认为，听觉和视觉言语信息被指定为特定的标签，在反应发生之前，连续的言语信息数据已经进行了跨通道的整合<sup>[12]</sup>。言语知觉动觉理论（the motor theory of speech perception）则认为感知言语就是感知声道发音的运动，言语感知依赖于对动作表征的理解<sup>[13]</sup>。除此之外，感知模糊逻辑模型（The Fuzzy Logic Model of Perception, FLMP）是通过贝叶斯推理假设视觉和听觉信息融合的一种方法<sup>[14]</sup>。FLMP 假设视听言语加工中输入的视觉、听觉言语信息以及储存在长

时记忆中的代表性特征原型融合起来，提供评价等级，最后按照简单的贝叶斯规则将这些评价定义为最后的言语感知<sup>[15]</sup>。

Altieri、Pisoni 和 Townsend 以 Summerfield （1987）阐述的视听言语整合过程的解释为基础，进行了进一步的分析与总结，提出了一个新的理论框架。在此新的框架中，不仅对通道特异性理论（modality-specific theories）（图 1）和动作基础性理论（gestural-based theories）（图 2）作了进一步的补充解释，还通过提供行为和生理的研究证据支持了通道特异性理论。通道特异性理论强调了视听通道中具体的言语信息特征在视听整合过程中的作用，认为经过初步的感觉加工，认知系统将视觉和听觉信息转变成分散的语音特征，听觉特征转变成音素（phonemes），视觉特征转变成视素（visemes），整合过程是将分散的具体语音特征知觉为连贯、统一的过程。而动作基础性理论则重视声道的过滤功能和发音动作，认为言语发音动作是产生视听言语整合的关键，视听言语信息的神经认知表征组成了假定性声音结构（hypothetical vocal tract configurations）。早期感觉编码之后，相关的听觉和视觉信息转变成运动编码，随后映射到产生讲话方式的假定性声音结构，最后将听觉和视觉的动态信息同时进行整合<sup>[16]</sup>。

在日常交流中，个体不仅要分析交谈者的声音言语信息，还要随时观察其面孔以及肢体传递的言语信息，即重视视觉和听觉通道对言语信息的传递作用。例如，当观看视频资源时，将画面遮住或者将声音过滤都会影响对视频内容的把握和理解。视听言语信息的相互作用能够提高言语认知效率，影响语音识别、探测、可理解性和情绪韵律的解码，可以促成较高的信息整合效益，远高于视、听单通道收集的信息之和，这种整合优势在言语交流中表现的更为明显<sup>[17-19]</sup>。视听言语整合对个体沟通技能和言语技巧的发展产生影响。一方面，婴儿期的研究发现，ASD 高风险婴儿已经出现视听言语整合缺陷的雏形<sup>[20]</sup>。这种缺陷不仅影响婴儿期言语的产生，还会影响后期认知能力的发展。特别是在面对面的言语交流中，视听言语整合缺陷使 ASD 个体失去了全面提取包括讲话者的情绪状态、语调、意向性在内的言语内容的路径。另一方面，婴儿在出生后的第一年就能将注意集中到讲话者的嘴部，运用视觉线索帮助发展音素分类，加强对语言的理解。ASD 儿童在此时期缺少对视觉言语线索的敏感性，视觉言语信息感知不足，影响视听通道言语信息的结合，阻碍功能性语言技巧的发展<sup>[21]</sup>。此外，来自父母的报告称，ASD 个体在早期会出现感觉的特异性，在婴儿时期出现对面孔和言语社会性刺激的不敏感<sup>[4]</sup>。因此，在婴儿时期出现的视听言语整合模式的异常为 ASD 的早期诊断提供了重要参考。目前，ASD 个体视听言语整合能力的发展趋势依然没有明确的定论，这种能力的缺陷贯穿一生发展的始终还是在成年期逐渐接近普通人的水平还需进一步的

探讨。

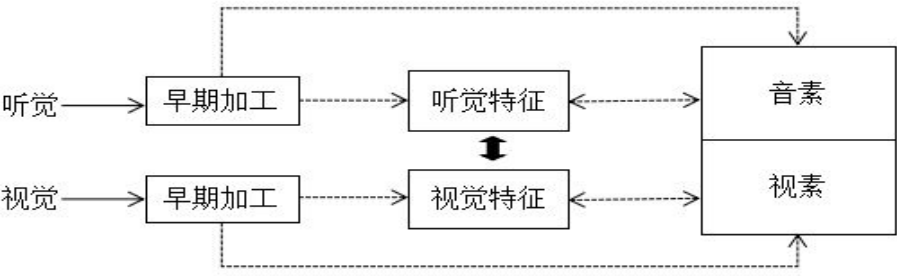


图 1 通道特异性理论 (Altieri, Pisoni, & Townsend, 2011)

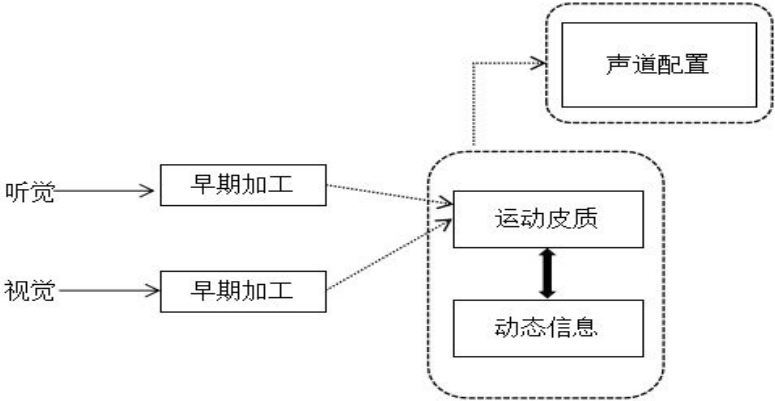


图 2 动作基础性理论 (Altieri, Pisoni, & Townsend, 2011)

2 ASD 个体视听言语整合缺陷

麦格克效应 (McGurk effect) 能够反映视觉和听觉通道言语信息的整合<sup>[22-23]</sup>。实验中给被试呈现听觉言语声音“ba”，视觉上呈现不一致的言语刺激“ga”，被试会做出“融合”反应，报告听到或者看到的是“da”或者“tha”，这种视、听通道呈现不一致言语刺激的范式即为麦格克实验范式，产生的错觉现象就是麦格克效应<sup>[24]</sup>。当视、听刺激来自不同性别的讲话者或者是个体没有意识到在观看一个面孔时，依然会产生这种错觉。

普通婴儿较早表现出对多感知言语线索的敏感性,2 个月时，婴儿能够将听到的元音字母与合适的发音面孔相匹配<sup>[25-26]</sup>。5 个月大的普通婴儿已经能够感知到麦格克效应<sup>[27]</sup>。随着年龄的发展，4 岁时对视觉言语线索的敏感性有了显著增强，在 5 岁时对麦格克效应的敏感性会降低，但在此之后又呈现增强趋势，特别是到 8 岁时不仅在麦格克任务上的反应更加可靠，而且可以较好的运用视觉线索从噪音中识别言语信号，青春期及其之后视听言语整合能力表现出逐步、稳点的发展趋势<sup>[28-29]</sup>。因此，普通个体 (Typical Development, TD) 视听言语整合能力呈现出接近“U”形的发展趋势。但对 ASD 个体视听言语整合能力的发展趋势目前没有清晰的定论。

ASD 个体在青春期之前的视听言语整合能力发展迟缓，但进入青春期后视听言语整合能力会以较快的速度发展，达到 TD 个体的水平，所以部分研究者认为 ASD 儿童视听言语整合能力的发展是迟滞的<sup>[30-31]</sup>。在研究中，年龄较小的 ASD 儿童视听言语整合能力显著弱于 TD 儿童，随着年龄的增长，ASD 个体的视听言语整合能力也在快速发展，并且速度要快于 TD 个体，所以在发展后期，ASD 个体的视听言语整合能力可以“追上”TD 个体。9 个月大的 ASD 高风险婴儿和低风险婴儿已经出现了麦格克效应敏感性的差异。低风险婴儿在视觉和听觉通道不匹配的情况下更长久的注视嘴部，表明低风险婴儿已经能够发现视觉和听觉言语线索之间的差异<sup>[20]</sup>。低风险婴儿虽然还不具备语言表达能力，但其反应体现了对麦格克效应的敏感，高风险婴儿不具有这样的敏感性。13 岁之后，ASD 个体和 TD 个体感知麦格克效应的敏感性趋于相近，ASD 个体感知麦格克效应的能力似乎得到了全面改善。ASD 个体在青春期视听言语整合的显著提高，展现出 ASD 儿童在进入青春期后这种整合能力快速、不连续的发展，表明 ASD 个体视听言语背后的神经机制可能依然具有可塑性，同时对从事 ASD 的教育者和临床工作者具有重要的参考价值。神经生理层面的研究表明，颞上沟区域包含与视听言语整合相关的镜像神经元，该区域镜像神经元的继续发展也为发展迟滞性观点提供了证据<sup>[32-33]</sup>。

部分研究者认为 ASD 个体感知麦格克效应能力的发展是停滞的<sup>[34]</sup>。ASD 个体在发展早期与 TD 个体感知麦格克的敏感性不存在差异，但是随着年龄增长所带来的认知以及神经发展的变化，在发展后期 ASD 和 TD 个体会出现行为上的较大差异。在 6-12 年龄段，感知麦格克的敏感性的水平没有较大差异；13-19 岁年龄段，ASD 个体感知麦格克效应的敏感性要低于 TD 个体。TD 个体在发展过程中，表现出感知麦格克效应正确率的提高，而 ASD 个体却没有表现出这种趋势。这表明 ASD 个体感知麦格克的能力可能存在永久的缺陷，贯穿毕生发展的始终。这种发展趋势凸显了早期干预的重要性。ASD 个体视听言语整合能力在早期处于正常水平，随着后期神经发展的不完善以及认知功能的不健全，可能导致了视听言语整合能力的发展停滞。因此，在 ASD 个体视听言语整合能力水平处在正常范围内的时期给予正确的指导与训练，可能会使这种能力一直保持在正常水平，从而使视听言语整合能力得到一定的发展。

上述研究中提到的 ASD 个体视听言语整合能力发展趋势的不一致可由诸多因素导致。ASD 个体表现出“弱中央统合”，具有特征和局部信息的加工偏好，在提取整体要点时存在问题，他们往往注重部分和细节，而忽略整体加工<sup>[35-36]</sup>。ASD 个体在视听言语整合的信息收集阶段，因缺乏整体的结构思维，出现偏重听觉通道言语信息现象，而不是将视听言语

信息当做一个整体进行感知、加工、整合，不能从整合的言语信息中获得更大信息量。研究表明，ASD 个体在听觉单通道的情况下表现越好，感知到麦格克效应的概率就越低<sup>[10]</sup>。由于 ASD 个体对声音信息有较高的敏感性和弱中央统合的认知特点，研究中呈现的声音类型、分贝不同，个体收集听觉信息量的差异都会对研究结果产生影响。

表 1 ASD 人群与健康人群麦格克效应敏感性的比较

资料来源	被试 及数量	年龄范围 (岁)	研究任务	实验刺激	结果
Mongillo et al., 2008	ASD: 15 TD: 21	8-19 11-19	麦格克实验范式	音节/ba/, /da/, /va/, /ða/	ASD 被试麦格克效应正确率（78.3%） 低于 TD 被试（94.9%）
Taylor, Isaac, & Milne, 2010	ASD: 24 TD: 30	7-16 7-16	麦格克实验范式	音节 /aba/, /ava/, /atha/, /ada/, /aga/	年轻 ASD 被试麦格克效应的正确率低 于 TD 被试, 年龄较大时没有发现两组 敏感性的显著差异
Saastai et al., 2011	ASD: 16 TD: 16	20-50 21-48	麦格克实验范式	音节/aka/, /apa/, /ata/	ASD 和 TD 组感知麦格克的模式存在 差异。反应“k”: ASD（49%），TD（95%）； 反应“t”: ASD（46%），TD（4%）
Wojnarowski et al., 2013	ASD: 18 TD: 18	8-17 8-17	麦格克实验范式	音节/ba/, /ga/, /da/, /tha/	ASD 被试麦格克效应的正确率 （37.94%）与 TD 被试（28.44%）没有 差异
Stevenson et al., 2014	ASD: 30 TD: 31	6-18 6-18	麦格克实验范式	音节/ga/, /ba/	低年龄组中 ASD 被试对麦格克效应的 正确率与 TD 被试没有差异；高年龄组 中 ASD 被试对麦格克效应的正确率低 于 TD 被试

3 ASD 个体视听言语整合缺陷的原因探究

3.1 注意的影响

假设视听言语整合是一个自动发生的过程，那么整个过程就不需要参与的参与。目前的研究倾向于视听言语整合需要注意调节的观点<sup>[37-38]</sup>。注意可以影响最初的信息收集以及后期的信息整合。在信息收集阶段，注意能够调节信息的收集内容与容量，也会影响最后的视听言语整合<sup>[39]</sup>。根据特雷斯曼的特征整合理论，在整合感知之前的早期信息加工阶段，个体感知到的视觉客体的特征会被自动的平行加工，在后期阶段，需要注意将这些特征绑定在一起<sup>[15]</sup>。因此，注意对视听言语整合的影响可能渗透在整个加工过程中，这种影响不仅存在于 TD 个体的研究中<sup>[40-41]</sup>，还存在于 ASD 个体中。在 ASD 个体中，注意对视听言语整合的影响主要集中在信息收集阶段，对注意的操作可参考 TD 群体的研究<sup>[42]</sup>。TD 个体的研究中，对

注意的操作可通过指导语的提示和变换实验刺激实现。通过指导语引导时,让被试注意单通道而忽视另一通道的信息,从而实现对视觉或者听觉的选择性注意<sup>[43]</sup>。视觉注意能够对 ASD 个体感知麦格克效应的正确率产生影响。一方面,ASD 个体较少注意社会性信息,视觉刺激通过面孔图片呈现会使视觉信息输入减少<sup>[44]</sup>。另一方面,ASD 个体中对听觉言语信息的高敏感性可能占用较多的注意资源<sup>[45]</sup>,在跨通道收集信息时注意力资源分配不均,信息收集不全面,影响后期整合的发生。Magne'e (2011) 等人的研究中,通过观看情绪性的面孔(快乐和害怕)和声音,发现 ASD 个体在分散注意环境中多感知整合能力是缺失的,证明了注意在 ASD 群体的视听整合中是一个重要因素<sup>[46]</sup>。在视听言语整合研究中并列的给被试两个面孔,使用箭头引导被试注意其中的一个面孔,ASD 个体和 TD 个体一样都会受到无关面孔的干扰而分散注意力,视听言语整合都受到了注意的影响<sup>[47]</sup>。目前的研究中,大多通过操纵单通道的视觉注意以探究 ASD 个体注意对视听言语整合的影响,此类研究只是从信息收集的角度侧面反映了注意对视听言语整合的影响,未来研究应该着重于对注意与视听言语整合影响的直接探究,利用功能性核磁共振等技术,清晰显示注意在整合发生阶段起到的作用,而不是只局限于信息收集阶段。

### 3.2 时间感知缺陷

时间感知是跨通道整合信息的重要能力,这种能力可以通过测量时间窗的宽窄来体现。时间窗(temporal window)是指两个单独刺激被感知为一个整合刺激的时间间隔<sup>[48]</sup>。TD 个体的研究表明,个体表现出越高的时间精确性(多感知整合中较窄的时间窗),视听言语整合的精确率也会增加<sup>[49-52]</sup>。相比 TD 个体,ASD 个体在整合视听言语刺激时,出现了较宽的时间窗。较宽的时间窗体现了 ASD 个体在视觉和听觉之间较弱的时间敏感性,时间作为一个线索来整合言语信息的可靠性是大打折扣的<sup>[7]</sup>。这种缺陷可以用时间绑定缺陷假说(the temporal binding deficit hypothesis)进行解释。时间绑定缺陷假说认为,ASD 个体存在时间绑定缺陷,不能灵活整合不同层面的加工信息。同一客体不同维度的特征会同时激活大脑中的神经元,这些神经元依靠时间关系将客体的特征绑定在一起,形成了对客体的整体认知,然而 ASD 个体神经元不能完全被同时激活,造成了特征绑定的不准确<sup>[53]</sup>。脑区的时间同步性在多感知信息整合中起到重要作用,时间绑定的缺陷很有可能导致 ASD 个体视听时间感知的缺损,使视觉和听觉的信息不能被整合为一个整体<sup>[54]</sup>。明确 ASD 个体的时间感知缺陷和言语整合的关系是探寻视听整合的关键。一方面,时间感知是多感知整合加工中的一个重要影响因素,ASD 个体时间感知缺陷会对视听言语整合的过程产生较大的负性影响。ASD 个体较宽的时间窗无法精确的整合视觉和听觉的言语信息,在整合过程中较容易绑定无关信



息，降低了视听言语信息的整合率。另一方面，对于言语信息和简单刺激反应的差异可能涉及到 ASD 个体视听整合的不同加工系统，其间的联系和差异值得进一步的探索。另外，通过训练减小时间窗宽度的也许会是 ASD 个体视听言语整合缺陷的有效干预措施。

### 3.3 唇动识别缺陷

ASD 个体的视听言语整合缺陷可能和唇动识别缺陷有关。在视听言语整合中，听觉系统接收言语的声音部分，而视觉系统主要通过对讲话者的唇部运动收集言语的视觉信息，特别是当听觉信息较弱或者是信息较难理解时，视觉信息在言语感知中会起到重要作用。这种通过观察讲话者的唇部运动获得视觉言语信息的能力称作视话（speech reading）能力<sup>[54]</sup>。在视话方面表现越好，感知到的麦格克效应就越多。言语知觉动觉理论（the motor theory of speech perception）也强调了言语感知依赖于对动作表征的理解。关于正常人的行为和事件相关电位研究表明，观察讲话者的唇部运动能够加快对语音的加工，并且能够提高加工效率<sup>[52]</sup>。多数研究认为 ASD 个体表现出视话能力的缺损，这极有可能是导致 ASD 个体视听言语整合困难的重要原因之一。ASD 儿童不能理解唇动所传达的视觉言语信息，他们较少注意面孔的嘴部区域，因而在观察面孔时更有可能注视脸部的非焦点区域，包括面颊、前额较少含有言语信息的区域<sup>[55]</sup>。这种视话能力缺陷在婴儿时期已初步显现。随着发展进程的推进，视话能力缺陷对 ASD 个体言语整合的影响可能会进一步加深<sup>[56]</sup>。

### 3.4 噪音背景的影响

日常生活中，噪音的影响无处不在。在言语整合实验中加入不同的噪音水平也是提高实验生态效度的重要手段。在以噪音为背景的社会交谈中，视听言语整合能力显得尤为重要。例如在繁华的街道或者是大风天，视听言语信息整合的优势在噪声环境中表现的更加明显。在噪音环境或者是容易使人分心的环境中，言语认知的能力是显著提高的，这种现象普遍存在于正常群体中<sup>[57]</sup>。可能原因是，在这种复杂的环境中，TD 个体会将焦点放在言语内容上，在此环境中会充分调动视听言语整合的功能。然而，这种噪声环境给 ASD 个体的视听言语整合造成很多困难。

在噪音背景下，ASD 个体表现出显著的视听言语感知缺陷，随着噪音水平的提高缺陷也会加重。但在单独听觉通道中，ASD 儿童和 TD 儿童具有相似的言语感知表现，并且不受噪音水平的影响。Foxe（2013）的研究中，探究了 5-17 岁被试噪音中视听言语整合的能力，发现 5-12 岁的 ASD 儿童的视听言语感知较容易受到噪音水平的影响，这和 Irwin（2011）等人的研究结果是一致的（被试年龄 5-15 岁）；而 13-15 岁的 ASD 儿童在噪音中的视听言语整合能力是完全改善的，这与已有研究的结果不符<sup>[58]</sup>。这种冲突的结果，一方面反映了随

着机体的成熟，ASD 个体视听言语整合能力可能也有所变化，年龄发展差异应该引起重视。另一方面，噪音背景和实验刺激材料的操作有很大差异，最简单的噪音背景是在言语感知中呈现粉红噪音，通过控制噪音响度来体现水平，而有些研究的噪音结构略显复杂，例如对一篇故事的录音进行处理，去除语言部分来作为噪音背景，这种噪音背景对于 ASD 个体可能是一个较大的挑战。

## 4 ASD 个体视听言语整合缺陷的神经生理机制

### 4.1 脑区或神经回路连接异常

ASD 个体视听言语感知缺陷可能是由于大脑皮层之间的连通性降低或者神经联结异常导致。潜在的神经机制水平上，ASD 个体的病理生理基础可能是来源于多重空间尺度上神经回路之间的异常联结<sup>[59]</sup>。来自神经和解剖的研究指出，ASD 可能与特定的远程神经联系完整性的降低相联系。联结损伤理论也认为，ASD 个体功能缺陷的皮层区域，功能性联结是降低的，从而导致结合视觉皮层和听觉皮层信息能力的降低<sup>[60]</sup>。一方面，ASD 群体左右语言联结偏侧化优势是不明显的，他们言语加工架构的皮层结构可能存在异常。另一方面，内侧前额叶皮质（medial prefrontal cortex, mPFC）是一个包括内侧壁（内侧额叶脑回和前扣带回）和大脑额叶腹侧面的区域。该区域与计划管理、决策和日常难题解决相关。mPFC 网络可以接收一些感觉的输入，该区域与丘脑、小脑皮层之间的异常联结，也可能是导致 ASD 个体视听言语整合困难的原因之一<sup>[61]</sup>。

颞上沟（superior temporal sulcus, STS）是掌控人类社会交流互动的重要脑区，在对该脑区的研究中发现，它与视听言语整合有重要联系。左侧颞上沟在视听言语整合中起着重要作用<sup>[62]</sup>。一项运用血氧水平依赖功能磁共振成像（blood-oxygen level dependent functional magnetic resonance imaging, BOLD fMRI）的研究表明，左侧 STS 区域的波幅反应是和感知麦格克效应的可能性相关的，视话活动进行时，STS 区域是被广泛激活的：较弱的 STS 反应意味着感知到麦格克效应的较低可能性，较强的 STS 反应意味着较高的可能感知到麦格克效应<sup>[63]</sup>。

颞上沟后部（posterior superior temporal sulcus, pSTS）也在视听言语加工中起到重要作用，包括对麦格克效应的感知。ASD 个体和 TD 个体的 pSTS 区域结构存在差别，体现在灰质体积和白质联结方面的不同<sup>[64-66]</sup>。大脑左半球的白质束与语言通路的韦尼克氏和布罗卡氏节点相联系，而 ASD 青少年在左半球表现出较低的白质联结<sup>[67]</sup>。pSTS 结构和功能的异常可能为 ASD 个体提供了言语整合缺陷的病因学基础。

### 4.2 特定脑波的异常

N4 的调节幅可以表现出言语意义的整合<sup>[52]</sup>。Megnin (2012) 等人在研究中测验了对口语单词的反应,来阐述从声音检测开始到意义整合的听觉言语加工阶段范围的视觉言语效应(唇部运动伴随着口语单词),发现视听交流效应的模式和波形是和颞上平面的活动相关的。此外,听觉的 N1 和 P2 等 ERP 成分,也是经常被提及的顶电位(vertex potential),对听觉言语有较大的反应并且受到视觉言语的调节<sup>[26]</sup>。研究表明,视听言语整合情景下,ASD 个体和 TD 个体在顶中部区域的 P2 波幅存在明显差异。目前研究中,对 ASD 个体和 TD 个体脑波活动差异进行比较的研究还较少,未来的研究中不仅要发现已经发现的异常脑波进行进一步的探究,还需扩大研究范围,找到 ASD 个体有别于 TD 个体的关键脑波活动指标。

## 5 展望

### 5.1 加强对语义整合的研究

目前视听言语整合研究中,多是以麦格克实验范式为基础,通过音素刺激的整合效率检验视听言语整合能力。面孔中嘴部区域的运动被认为是言语信息最多的区域,因此在此实验范式中,视觉刺激多是呈现一个讲话者的面孔。然而,用面孔呈现刺激存在一些局限性,面孔包含较多的社会性信息,是能够提供更多交流信息的区域,ASD 个体对社会性信息存在一定的排斥,所以这种对社会信息的排斥很可能影响 ASD 个体视觉言语信息的获取。除面孔之外,其他与言语相关的动作也对视听言语感知有重要影响,但是 ASD 个体对这种言语交流动作没有偏爱的倾向<sup>[68]</sup>。Hubbard 等人(2012)选取了 13 个高功能 ASD 和 TD 儿童,用 fMRI 测量这些儿童在整合声音和手势动作时的差异,发现 TD 儿童的脑区将敲击手势当做具有交流性意义的动作,能够成功的将它和同时发生的言语整合起来识别,ASD 儿童的脑区没有表现出对社会交流中多感觉通道模型信息的有效整合。在视听语义加工的研究中,运用视听启动范式,通过呈现现实生活中熟悉的物体以及声音,要求被试对目标刺激做出反应。该类范式排除了社会性的影响,并且从语义层面探究了 ASD 个体的视听言语整合能力<sup>[69-71]</sup>。

### 5.2 以言语整合为基础的干预方法

关于 ASD 的研究,最终目的是制定行之有效的干预办法,针对性最强的措施,往往会有事半功倍的效果。例如,针对 ASD 个体视听言语感知缺陷的面部言语(Listening to Faces, L2F)电脑程序。L2F 是一个交互式、自适应程序、适应 iPad 的一个应用程序。在这个程序中会在多样的听觉噪音水平下呈现讲话者产生多音节词的视频。该程序适应使用者的表现,随着表现的提高难度会加大。视频中的讲话者会在年龄和性别上变化,言语信号的变化也会越来越大。在运用追踪程序的研究中,4 个 8-10 岁的 ASD 儿童接受了每周 3 天、持续 12

周、每次大约持续 10 分钟的训练。数据表明，识别言语的任务中（伴随噪声），儿童没有经过前期的训练，但是儿童的表现是提高了。因此，L2F 能够帮助 ASD 儿童利用视听通道感知噪音中的言语信息<sup>[17]</sup>。虽然数据样本较小且没有控制组，但这种方式提供了干预的新方式。未来的研究要提出更多关于改进 ASD 个体视听言语缺陷的措施，使 ASD 个体的行为不仅在实验环境下有所提升，并能够迁移到实际生活中并保持。

## 参考文献

- 1 American Psychiatric Association ( APA) . Diagnostic and statistical manual of mental disorders ( DSM—V) ( 5th ed.) . Washington, DC: American Psychiatric Association, 2013
- 2 Baranek G T. Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of autism and developmental disorders*, 1999, 29(3): 213-224
- 3 Dawson G, Toth K, Abbott R, et al. Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, 2004, 40(2): 271
- 4 Talay-Ongan A, Wood K. Unusual sensory sensitivities in autism: A possible crossroads. *International Journal of Disability, Development and Education*, 2000, 47(2): 201-212
- 5 Katagiri M, Miya K, Matsui M. Difficulty of crossmodal processing in individuals with autism spectrum disorders: An audio-visual gap/overlap paradigm study. *Research in Autism Spectrum Disorders*, 2014, 8(4): 424-431
- 6 Ross L A, Del Bene V A, Molholm S, et al. Sex differences in multisensory speech processing in both typically developing children and those on the autism spectrum. *Frontiers in neuroscience*, 2015, 9: 185
- 7 Stevenson R A, Segers M, Ferber S, et al. Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Research*, 2016, 9(7): 720-738
- 8 Stevenson R A, Siemann J K, Woynaroski T G, et al. Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 2014, 44(6): 1470-1477
- 9 Altieri N, Townsend J T, Wenger M J. A measure for assessing the effects of audiovisual speech

- integration. *Behavior research methods*, 2014, 46(2): 406-415
- 10 Bebko J M, Schroeder J H, Weiss J A. The McGurk Effect in Children With Autism and Asperger Syndrome. *Autism Research*, 2014, 7(1): 50-59
  - 11 Stevenson R A, Zemtsov R K, Wallace M T. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 2012, 38(6): 1517
  - 12 Braida L D. Crossmodal integration in the identification of consonant segments. *The Quarterly Journal of Experimental Psychology Section A*, 1991, 43(3): 647-677
  - 13 Galantucci B, Fowler C A, Turvey M T. The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, 2006, 13(3): 361-377
  - 14 Massaro D W. Multiple book review of speech perception by ear and eye: A paradigm for psychological inquiry. *Behavioral and Brain Sciences*, 1989, 12(4): 741-755
  - 15 Navarra J, Alsius A, Soto-Faraco S, et al. Assessing the role of attention in the audiovisual integration of speech. *Information Fusion*, 2010, 11(1): 4-11
  - 16 Altieri N, Pisoni D B, Townsend J T. Some behavioral and neurobiological constraints on theories of audiovisual speech integration: a review and suggestions for new directions. *Seeing and perceiving*, 2011, 24(6): 513
  - 17 Irwin J, Preston J, Brancazio L, et al. Development of an audiovisual speech perception app for children with autism spectrum disorders. *Clinical linguistics & phonetics*, 2015, 29(1): 76-83
  - 18 Saalasti S, Kätsyri J, Tiippana K, et al. Audiovisual speech perception and eye gaze behavior of adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, 2012, 42(8): 1606-1615
  - 19 Teder-Sälejärvi W A, Russo F D, McDonald J J, et al. Effects of spatial congruity on audio-visual multimodal integration. *Journal of cognitive neuroscience*, 2005, 17(9): 1396-1409
  - 20 Guiraud J A, Tomalski P, Kushnerenko E, et al. Atypical audiovisual speech integration in infants at risk for autism. *PloS one*, 2012, 7(5): e36428
  - 21 Lewkowicz D J, Hansen-Tift A M. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, 2012, 109(5): 1431-1436
  - 22 McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature*, 1976, 264(5588): 746

- 23 Tiippana K. What is the McGurk effect?. *Frontiers in Psychology*, 2014, 5: 725
- 24 Woynaroski T G, Kwakye L D, Foss-Feig J H, et al. Multisensory speech perception in children with autism spectrum disorders. *Journal of autism and developmental disorders*, 2013, 43(12): 2891-2902
- 25 Bristow D, Dehaene-Lambertz G, Mattout J, et al. Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of cognitive neuroscience*, 2008, 21(5): 905-921
- 26 Knowland V C P, Mercure E, Karmiloff-Smith A, et al. Audio-visual speech perception: A developmental ERP investigation. *Developmental Science*, 2014, 17(1): 110-124
- 27 Rosenblum L D, Schmuckler M A, Johnson J A. The McGurk effect in infants. *Perception & Psychophysics*, 1997, 59(3): 347-357
- 28 Kawabe T, Shirai N, Wada Y, et al. The audiovisual tau effect in infancy. *PloS one*, 2010, 5(3): e9503
- 29 Ross L A, Molholm S, Blanco D, et al. The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, 2011, 33(12): 2329-2337
- 30 Taylor N, Isaac C, Milne E. A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *Journal of autism and developmental disorders*, 2010, 40(11): 1403-1411
- 31 Keane B P, Rosenthal O, Chun N H, et al. Audiovisual integration in high functioning adults with autism. *Research in Autism Spectrum Disorders*, 2010, 4(2): 276-289
- 32 Kilner J M, Blakemore S J. How does the mirror neuron system change during development?. *Developmental science*, 2007, 10(5): 524-526
- 33 Williams J H G, Massaro D W, Peel N J, et al. Visual–auditory integration during speech imitation in autism. *Research in developmental disabilities*, 2004, 25(6): 559-575
- 34 Mongillo E A, Irwin J R, Whalen D H, et al. Audiovisual processing in children with and without autism spectrum disorders. *Journal of autism and developmental disorders*, 2008, 38(7): 1349-1358
- 35 Happé F, Frith U. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 2006, 36(1): 5-25
- 36 Nuske H J, Bavin E L. Narrative comprehension in 4–7-year-old children with autism: testing

the Weak Central Coherence account. *International Journal of Language & Communication Disorders*, 2015: 1-12

37 Kaiser J, Hertrich I, Ackermann H, et al. Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. *Neuroimage*, 2006, 30(4): 1376-1382

38 Tiippana K, Andersen T S, Sams M. Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, 2004, 16(3): 457-472

39 Talsma D, Senkowski D, Soto-Faraco S, et al. The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, 2010, 14(9): 400-410

40 Talsma D, Doty T J, Woldorff M G. Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration?. *Cerebral cortex*, 2006, 17(3): 679-690

41 Fernández L M, Visser M, Ventura-Campos N, et al. Top-down attention regulates the neural expression of audiovisual integration. *NeuroImage*, 2015, 119: 272-285

42 Iarocci G, Rombough A, Yager J, et al. Visual influences on speech perception in children with autism. *Autism*, 2010, 14(4): 305-320

43 Buchan J N, Munhall K G. The influence of selective attention to auditory and visual speech on the integration of audiovisual speech information. *Perception*, 2011, 40(10): 1164-1182

44 Tanaka J W, Sung A. The “eye avoidance” hypothesis of autism face processing. *Journal of autism and developmental disorders*, 2016, 46(5): 1538-1552

45 Bhatara A, Quintin E M, Fombonne E, et al. Early sensitivity to sound and musical preferences and enjoyment in adolescents with autism spectrum disorders. *Psychomusicology: Music, Mind, and Brain*, 2013, 23(2): 100

46 Magnée M J C M, de Gelder B, van Engeland H, et al. Multisensory integration and attention in autism spectrum disorder: Evidence from event-related potentials. *PLoS One*, 2011, 6(8): e24196

47 Saalasti S, Tiippana K, Kätsyri J, et al. The effect of visual spatial attention on audiovisual speech perception in adults with Asperger syndrome. *Experimental brain research*, 2011, 213(2-3): 283-290

48 Richards M, Goltz H, Wong A. The audiovisual temporal binding window in unilateral amblyopia: monocular and binocular effects. *Journal of Vision*, 2016, 16(12): 143-143

49 Boer L D, Eussen M, Vroomen J. Diminished sensitivity of audiovisual temporal order in

- autism spectrum disorder. *Frontiers in integrative neuroscience*, 2013, 7: 8
- 50 Foss-Feig J H, Kwakye L D, Cascio C J, et al. An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*, 2010, 203(2): 381-389
- 51 Kwakye L D, Foss-Feig J H, Cascio C J, et al. Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in integrative neuroscience*, 2011, 4: 129
- 52 Megnin O, Flitton A, RG Jones C, et al. Audiovisual speech integration in autism spectrum disorders: ERP evidence for atypicalities in lexical-semantic processing. *Autism Research*, 2012, 5(1): 39-48
- 53 Brock J, Brown C C, Boucher J, et al. The temporal binding deficit hypothesis of autism. *Development and psychopathology*, 2002, 14(2): 209-224
- 54 Bhat J, Pitt M A, Shahin A J. Visual context due to speech-reading suppresses the auditory response to acoustic interruptions in speech. *Frontiers in Neuroscience*, 2014, 8: 173
- 55 Schelinski S, Riedel P, von Kriegstein K. Visual abilities are important for auditory-only speech recognition: evidence from autism spectrum disorder. *Neuropsychologia*, 2014, 65: 1-11
- 56 Smith E G, Bennetto L. Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, 2007, 48(8): 813-821
- 57 Ma W J, Zhou X, Ross L A, et al. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS One*, 2009, 4(3): e4638
- 58 Foxe J J, Molholm S, Del Bene V A, et al. Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex*, 2013, 25(2): 298-312
- 59 Vissers M E, Cohen M X, Geurts H M. Brain connectivity and high functioning autism: a promising path of research that needs refined models, methodological convergence, and stronger behavioral links. *Neuroscience & Biobehavioral Reviews*, 2012, 36(1): 604-625
- 60 Rippon G, Brock J, Brown C, et al. Disordered connectivity in the autistic brain: challenges for the 'new psychophysiology'. *International journal of psychophysiology*, 2007, 63(2): 164-172
- 61 Martínez-Sanchis S. Neurobiological foundations of multisensory integration in people with autism spectrum disorders: the role of the medial prefrontal cortex. *Frontiers in human neuroscience*, 2014, 8: 970
- 62 Calvert G A, Campbell R. Reading speech from still and moving faces: the neural substrates of



visible speech. *Journal of cognitive neuroscience*, 2003, 15(1): 57-70

63 Nath A R, Beauchamp M S. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage*, 2012, 59(1): 781-787

64 Shukla D K, Keehn B, Müller R A. Tract-specific analyses of diffusion tensor imaging show widespread white matter compromise in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 2011, 52(3): 286-295

65 Boddaert N, Chabane N, Gervais H, et al. Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *Neuroimage*, 2004, 23(1): 364-369

66 Lange N, DuBray M B, Lee J E, et al. Atypical diffusion tensor hemispheric asymmetry in autism. *Autism research*, 2010, 3(6): 350-358

67 Fletcher P T, Whitaker R T, Tao R, et al. Microstructural connectivity of the arcuate fasciculus in adolescents with high-functioning autism. *Neuroimage*, 2010, 51(3): 1117-1125

68 Hubbard A L, McNealy K, Scott-Van Zeeland A A, et al. Altered integration of speech and gesture in children with autism spectrum disorders. *Brain and behavior*, 2012, 2(5): 606-619

69 Cantiani C, Choudhury N A, Yan H Y, et al. From sensory perception to lexical-semantic processing: an erp study in non-verbal children with autism. *PloS one*, 2016, 11(8): e0161637

70 David N, R. Schneider T, Vogeley K, et al. Impairments in multisensory processing are not universal to the autism spectrum: no evidence for crossmodal priming deficits in Asperger syndrome. *Autism Research*, 2011, 4(5): 383-388

71 Silverman L B, Bennetto L, Campana E, et al. Speech-and-gesture integration in high functioning autism. *Cognition*, 2010, 115(3): 380-393

## **The Audiovisual Speech Integration in Autism Spectrum Disorders**

**Abstract:** The alterations of audiovisual speech integration of Autism Spectrum Disorders (ASD) is a severe problem. There are two quite different views on their development of the audiovisual integration ability: One is that the development of this ability is delayed, while the other points out the development of this ability is arrested. The deficits on attention, time perception, the flaw of lip movements recognition, and noise background, have different effects on the audiovisual speech integration of individuals with ASD. The neurophysiological studies showed abnormal brain activity with ASD, such as temporal lobe activities, amplitudes of N4, N1, P2 ERP component, the weak connectivity between the cerebral cortexes. Future studies should further explore the audiovisual semantic integration, develop intervention techniques based on speech integration.

**Key words:** autism spectrum disorders; audiovisual speech integration; reason analysis; neural mechanism