

## 早期视听整合加工——来自 MMN 的证据\*

辛 昕<sup>1</sup> 任桂琴<sup>1</sup> 李金彩<sup>2</sup> 唐晓雨<sup>1</sup>

(<sup>1</sup> 辽宁师范大学心理学院, 大连 116029) (<sup>2</sup> 上海交通大学外国语学院, 上海 200240)

**摘 要** 视听整合是指当呈现的视觉和听觉信号在时间、空间上大致接近时, 视觉和听觉系统倾向于整合的加工过程。失匹配负波(Mismatch Negativity, MMN)作为反映大脑早期加工的成分, 表征偏差的信息输入与感觉记忆痕迹之间的神经失匹配。以 MMN 作为探测指标的视听整合加工研究, 主要包括 MMN 在阅读理解中字母和语音、韵律信息、麦格克效应等方面的视听整合加工, 以及分析跨通道视听整合相互竞争、相互补充的关系。未来研究需聚焦于其他通道的跨通道整合加工, 同时应拓展 MMN 的诱发范式。

**关键词** 听觉; 视觉; 视听整合; MMN; ERP

**分类号** B842

失匹配负波(Mismatch Negativity, MMN)最早由 Näätänen, Gaillard 和 Mäntysalo (1978)率先发现并报道。典型的实验范式是双耳分听 oddball 范式, 即大概率的标准刺激和小概率偏差刺激声音随机呈现给被试的左或右耳, 要求被试注意一只耳而不注意另一只耳的声音。结果发现, 无论是注意耳还是非注意耳, 偏差刺激都比标准刺激引起更高的负波。前注意条件和注意条件下都能诱发 MMN (Alain, & Woods, 1997; Näätänen, Paavilainen, Titinen, Jiang, & Alho, 1993; Sams, Paavilainen, Alho, & Näätänen, 1985)。该成分可用于探测音位、语调、语言等加工(Näätänen et al., 1997; 任桂琴, 韩玉昌, 周永垒, 任延涛, 2011), 也可用于探测注意和知觉加工, 目前常见于研究临床上儿童和成人的对比研究。关注从儿童到成人注意分配能力的发展, 即从儿童对复杂分心任务表现出强烈的不随意注意, 到成人对复杂分心任务控制能力增强的变化过程(Choudhury, Parascando, & Benasich, 2015; Ruhnau et al., 2013; Ruhnau, Wetzl, Widmann, & Schröger, 2010)。而

且, 由于神经失调(Neurological Disorders)没有明确的标记, MMN 还可作为理解认知因素的标志, 应用于精神分裂症、失语症等问题的研究(Csukly, Stefanics, Komlósi, Czigler, & Czobor, 2013; Näätänen, Sussman, Salisbury, & Shafer, 2014; Särkämö et al., 2010)。

自 MMN 被发现以来, 研究者们已在视觉通道、听觉通道, 以及躯体感觉等多个通道中观察到了该成分。起初主要是在听觉通道中通过探测刺激变化诱发 MMN, 后续众多学者对听觉 MMN (auditory mismatch negativity, aMMN)进行了大量研究(Choudhury et al., 2015; Erlbeck, Kübler, Kotchoubey, & Veser, 2014; Weise, Grimm, Trujillo-Barreto, & Schröger, 2014)。视觉 MMN (visual mismatch negativity, vMMN)是源于 Näätänen (1988)在一系列 ERP 研究中发现, 非注意条件下视觉环境中的刺激变化, 能够诱发出在 N2b 晚期范围内的负成分, 而且这种 N2 类似波中含有自动成分。经过长时间的探索, 研究者们发现视觉刺激的形状、颜色、空间频率、面孔表情等偏差变化都能诱发 vMMN (Cleary, Donkers, Evans, & Belger, 2013; Hedge et al., 2015; Kreegipuu et al., 2013; Stefanics, Kimura, & Czigler, 2011)。后续逐渐在躯体感觉等通道观察到 MMN (Kekoni et al., 1997; Strömmer, Tarkka, & Astikainen, 2014)。Kekoni 等人 (1997)在 oddball 条件下呈现高频率的触觉偏差刺激,

收稿日期: 2016-05-26

\* 国家自然科学基金(31471075; 31100732; 31600882), 国家自然科学基金委员会与英国皇家学会合作交流项目(31611130183), 国家留学基金委公派留学项目(201508210098)。

通讯作者: 任桂琴, E-mail: renguiqin@126.com

发现有 N250 偏转, 和失匹配负波类似。这是首次提供证明躯体感觉能够诱发 MMN 的证据。

在这些单通道 MMN 研究基础上, 近年来研究者们进一步提出跨通道感觉模式, 如视听整合加工 MMN, 在视觉听觉整合环境下使用 oddball 范式诱发 MMN, 发现视听整合加工 MMN 有颞上皮质等区域的激活, 其中颞上回和颞上沟作为言语中枢, 对视听刺激的反应更为显著。颞上回具有调整自己语言和理解别人语言的作用, 受损将导致视听整合困难, 引起感觉性失语症 (Arnal, Wyart, & Giraud, 2011; Holloway, van Atteveldt, Blomert, & Ansari, 2015; Ikeda et al., 1996; Lu, Paraskevopoulos, Herholz, Kuchenbuch, & Pantev, 2014; van Atteveldt, Blau, Blomert, & Goebel, 2010; 文小辉, 李国强, 刘强, 2011)。同时发现, 在一个通道下呈现刺激, 另一个通道能够提供补充利用 (Bottari et al., 2014; Matusz, Retsa, & Murray, 2016; McDonald, Störmer, Martinez, Feng, & Hillyard, 2013); 在一个通道下执行的主任务, 另一个通道下出现分心任务, 视听跨通道之间又存在相互竞争的关系 (Andrés, Guerrini, Phillips, & Perfect, 2008; Leiva, Parmentier, & Andrés, 2015; Regenbogen et al., 2012; Wiens, Szychowska, & Nilsson, 2016)。近年来视听整合的特质主要体现在阅读理解中字母与语音信息的整合、视觉与听觉韵律信息整合、以及麦格克效应等领域的相关研究。

## 1 视觉、听觉通道中的 MMN

Näätänen 等人 (1978) 最先在听觉通道诱发出 MMN。aMMN 通常采用 oddball 范式, 传统 oddball 范式是在标准刺激序列中呈现单一的偏差刺激, 如果想要获得多种不同的 MMN, 就要执行多个程序, 从而导致较长的持续时间。为解决这一问题, Näätänen, Pakarinen, Rinne 和 Takegata (2004) 提出了一种可以同时获得多种 aMMN 的新范式, 即在一个刺激系列中同时呈现多个不同的偏差刺激。结果发现, 该 MMN 新范式可以在较短时间内同时获得多种 aMMN, 且这些 aMMN 与传统 oddball 范式获得的单一 aMMN 波幅相似。大量研究表明, aMMN 通常表现为额中央的头皮分布特点 (Näätänen, 1990; Weise et al., 2014), 其中也有颞叶皮层的参与。当探测听觉变化的刺激序列中, 偏差刺激相对的以大、中、小的概率呈现时, 额

叶和颞叶皮层将以不同的顺序激活, 呈现颞-额, 额-颞-额, 颞-额的激活顺序模式 (Tse, Rinne, Ng, & Penney, 2013)。

vMMN (Näätänen, 1988) 出现在视觉刺激呈现后的 150~400 ms, 可以在视觉偏差特征的反应中获得 (Cleary et al., 2013; Kreegipuu et al., 2013; Shtyrov, Goryainova, Tugin, Ossadtchi, & Shestakova, 2013; Stefanics et al., 2011)。最初该成分的存在性引起了巨大争议。主要是由于视觉信息呈现方式的限制, 尽管可以采用掩蔽技术, 但仍然很难保证被试不注意所看到的视觉刺激。为解决这一问题, Wei, Chan 和 Luo (2002) 提出了“跨通道延迟反应”oddball 范式。实验中要求被试分别注意视觉通道或者听觉通道中的信息, 而忽视另一通道中的信息。听觉信息中标准刺激是 800Hz 的短音, 偏差刺激是 1000Hz 的短音, 同时加入微弱的点击声作为听觉命令信号。视觉信息中标准刺激是对比度较弱的彩色图片, 偏差刺激是对比度较强的彩色图片, 视觉命令信号是红色十字图案。声音和图片在一个序列中随机呈现。但是只要求被试关注其中一种类型的刺激而忽视另一种类型的刺激。在只注意视觉刺激时, 要求被试先对出现的视觉信息做标准和偏差刺激判断的按键准备, 一半的被试对标准刺激用左拇指, 偏差刺激用右拇指反应, 另一半被试则相反。在看到命令信号后, 尽可能又快又准的开始按键判断。在要求被试只注意听觉刺激时程序相同。由于在命令信号出现前, 被试需要集中注意力去捕捉信号, 因此不易分心关注另一通道下的刺激, 这就有效的保证了被试对视觉刺激的非注意状态, 使得 vMMN 的获得更加可信。关于 vMMN 的神经发生位置, 有研究使用 LORETA (Low-resolution electromagnetic tomography) 定位技术, 报告主要激活区域为枕叶视觉纹状体区 (Pascual-Marqui, Michel, & Lehmann, 1994)。也有研究使用功能性磁共振成像 (fMRI, functional magnetic resonance imaging) 分析, 发现 vMMN 的激活区域主要在左侧额下回 (IFG) 和额中回 (MFG), 说明与听觉变化有关的额叶区, 也在视觉变化探测中有着重要的作用 (Hedge et al., 2015)。最近 Yang 等人 (2016) 对比 aMMN 和 vMMN, 发现性别之间在探测听觉刺激变化上没有显著差异, 但是在探测视觉变化时有不同的加工, 男性对这种变化的探测能力明显优于女性。

## 2 视听整合加工的 MMN

视听整合是呈现视觉和听觉信号的时间、空间大致接近时,视觉和听觉系统倾向于整合信息的加工过程,整合的前提是时间接近而不是空间接近(Koelewijn, Bronkhorst, & Theeuwes, 2010; van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008)。视听整合 MMN 是在视觉听觉整合环境下诱发 MMN。目前以 MMN 为探测指标的视听研究中,主要采用经典的 oddball 范式。但是最近一项研究中, Nair, Sasidharan, John, Mehrotra 和 Kutty (2016)提出一种新的实验范式,是通过游戏化实验逻辑评估神经认知(Assessing Neurocognition via Gamified Experimental Logic, ANGEL),是一种视觉 oddball 范式的游戏化改编(gamified adaptation of the visual oddball paradigm)。共有学习、识别、理解三种复杂度递增的游戏任务。在屏幕中央呈现加号,用 oddball 范式在加号左右两边分别呈现一张黑白相间的图片和一张关键图片,为 Mooney 面孔(Mooney face)或者 Kanizsa 三角形(Kanizsa triangle),或者这两种图片的碎片化形式。在呈现图片的同时用 oddball 范式呈现声音干扰,标准刺激 1000 Hz,偏差刺激 1500 Hz 的纯音。出现在视觉刺激开始前的 240 ms、40 ms 或者视觉刺激后的 160 ms。一共执行 8 个 block,在学习阶段,奇数 block 只要求被试在心里识别关键图片在左还是右侧,不需要按键反应,剩下的 block 要求进行按键反应。在识别阶段,所有 block 都需要按键反应,而且只要按键就会在随机延迟后出现短音。要求被试忽视声音,专注执行视觉任务。在理解阶段,要求被试对关键图片进行意义判断,有意义按右键,模糊意义按左键,另一半被试做相反操作。结果发现,这一范式能够同时诱发包括 MMN 在内的多种 ERPs,有视觉早期成分 C1,面孔识别成分 N170,以及识别靶刺激抑制无关刺激诱发的 N2pc 等。因此可以评估多种神经加工,例如听觉和视知觉、注意转换、记忆等多种水平的决策。

### 2.1 字母和语音视听整合 MMN

视听整合最突出的作用就是有助于阅读的理解,良好的阅读理解需要字母和语音之间的正确联系,如果无法取得关联,将导致发展性阅读障碍(Froyen, Willems, & Blomert, 2011; Shaywitz,

Morris, & Shaywitz, 2008; Žarić et al, 2014)。Žarić 等人(2014)采用被动视听 oddball 范式,对比不同程度的阅读障碍儿童和典型阅读能力儿童(typical readers)在字母和语音整合加工上的差异。以 3 种方式呈现实验材料,包括单独听觉呈现,视听同步呈现(AV-0)以及字母在听觉刺激前 200 ms (AV-200)呈现。单独听觉条件下,标准刺激是元音声音 /a/, 偏差刺激是元音声音 /o/。两种视听条件下的听觉呈现方式与单独听觉条件下的呈现序列一致,视觉呈现方式是在屏幕上呈现元音字母 a,这就形成了视听标准刺激为一致的声音和字母,视听偏差刺激为不一致的声音和字母。结果发现,在单独听觉条件下,严重阅读障碍儿童、一般阅读障碍儿童和典型阅读能力儿童都有明显的 aMMN 和晚期负波(late Negativity, LN)反应。在 AV-0 和 AV-200 两种条件下,典型阅读能力儿童都能诱发出波幅较为强烈的 MMN,而两种不同程度的阅读障碍儿童所诱发 MMN 的波幅都表现出一定程度的减弱。在视听同步条件下,发现儿童的阅读流畅性和 MMN 的潜伏期相关,阅读能力更好的儿童 MMN 潜伏期更长。结果表明,虽然在单独听觉条件下,不同程度的阅读障碍儿童和典型阅读能力儿童都有明显的 aMMN 反应,但是在视听整合条件下,典型阅读能力儿童表现为强烈的字母和语音的神经整合,诱发出波幅更强的 MMN,阅读障碍的儿童则表现出较弱的相对应的神经整合,诱发出的 MMN 波幅较弱。也就是说,视听条件下,阅读能力的差异和字母与语音神经整合的强弱有关。

Žarić 等人(2015)进一步提出问题,阅读训练能否改善阅读障碍儿童较弱的字母和语音神经整合。因此使用同样的范式,对阅读障碍儿童进行为期 6 个月的阅读教学和训练。对比训练的前后测结果发现,训练后字母和语音的神经整合能力有一定的改善,尤其是在整合 LN 的晚期窗口出现更早的跨通道效应。后测中也发现 LN 出现更早的潜伏期与联系字母-语音的正确率提高显著相关。也就是说, LN 的潜伏期能较好的预测儿童阅读能力的获得。也有研究者使用 Audilex 训练在阅读障碍儿童中进行考察,证实训练能够达到改善阅读障碍的目的(Huotilainen et al., 2011)。该实验选取的被试是出生时体重过低的儿童,出生时体重过低不利于神经发展,易导致发展性阅读

障碍(Roberts, Anderson, & Doyle, 2009; Stephens, & Vohr, 2009)。Audilex 训练是一种非文字视听任务,通过屏幕上呈现矩形的高度、厚度和长度,表示音高、音强和声音的持续时间,分别在阅读障碍组和控制组中执行实验程序。结果发现,训练后的实验组反应频率和持续时间的 MMN 波幅增大,但在控制组没有这种表现。MMN 波幅的增大通常解释为能够更容易和更可靠的探测变化,这反映了大脑皮层识别频率和持续时间变化的能力提高。有研究发现(Leppänen et al., 2002),阅读障碍可能和识别持续时间的缺陷有关。因此可以说明,在干预后这种阅读障碍缺陷有一定程度的改善。

以上研究是以阅读障碍儿童为被试,综合以上结论,通过不同阅读能力儿童所诱发 MMN 的波幅差异,可以证明阅读障碍儿童字母和语音神经整合能力与典型阅读能力儿童相比较弱。进一步发现实验组的阅读障碍儿童在训练后反应频率和持续时间的 MMN 波幅增大,也就是说这种较弱的神经整合通过系统化的训练能够随着时间发生变化和改善。还可以通过 LN 潜伏期的改变预测儿童阅读能力的获得,这为以后如何改善儿童阅读障碍提供了启示。

完整的单词是由字母所组成, Jost, Eberhard-Moscicka, Frisch, Dellwo 和 Maurer (2014)却发现,完整单词与语音的视听整合和阅读流畅性之间没有显著相关。由于这其中还涉及一系列复杂的加工编码过程,如将字母整合成完整单词的加工过程,因此导致这一结论的具体机制还有待深入研究。鉴于以往研究的结论,最近的一项研究, Jones, Kuipers 和 Thierry (2016)就以成人阅读障碍者为被试,考察在发展性阅读障碍中视听整合的核心缺陷是来自视觉-正字法还是视觉-语音的加工受损。采用主动 oddball 范式,要求被试对字母和语音的一致配对进行按键反应。其中偏差刺激的呈现方式是有策略的失匹配,以视觉上的相近字形作为正字法因素的考察,以语音上的发音相似作为语音因素的考察。结果发现,在信息输入基础上,高度功能性阅读障碍者能够发展正字法表征,而且可以获得充分的激活。但是他们在选择单独的表达形式输出时,不能自动的进行视觉和语音的联系。以上结果说明阅读的加工过程不仅依赖于高度特异性的正字法表征,更要依赖于强烈的视觉语音联系,视觉语音联系受损是导致阅读障

碍的主要因素。

由以上研究可以看出,在探究阅读障碍的核心原因时,虽然得出结论认为阅读障碍者能够发展正字法表征,而不能自动联系视听信息,但是该结论的获得是以成人为被试,随着年龄的增长正字法表征的发展与成熟和经验的交互作用无法排除,因此还无法将此结论推广到阅读障碍儿童中去,阅读障碍儿童的缺陷原因目前还无法得出结论。

## 2.2 韵律信息视听整合 MMN

良好的视听整合能力不仅有助于阅读理解,同样能够促进社会交往中双方的语言理解。语言理解是脑神经对所呈现的视觉信息和语言中的语法和语义信息的整合(Willems, Özyürek, & Hagoort, 2008),然而语言中不仅包含语法、语义等信息,还包含韵律信息。韵律信息也被称作超音段信息(suprasegmental information),包括声调轮廓、语调音高、重音等线索(任桂琴,刘颖,于泽,2012)。Li, Yang 和 Ren (2009)针对语言中的韵律信息能否在视觉情境(visual scene context)中获得即刻整合进行了探究。该研究采用重音作为韵律的表达形式。重音是一个韵律结构中的某一音节、词或短语相对的突出,是由一些声学特征,如音长、音高和音强的变化来实现(于泽,韩玉昌,任桂琴,2010)。实验分为两个部分均采用被动 oddball 范式,其一是视听结合条件,其二是单独的听觉条件。视觉刺激是在屏幕的左侧出现一个红色或者蓝色的三角形,再在屏幕右侧出现红色的矩形。听觉刺激是韵律上有不同重音的中文口语句子。纯听觉条件下,标准重音是红色,偏差重音是矩形。视听结合条件下的声音刺激和纯听觉条件下的声音一致,此时标准刺激是匹配的图片 and 声音,偏差刺激是不匹配的图片 and 声音。结果发现,在纯听觉条件下,语音的韵律变化诱发显著的 aMMN 反应;视听结合条件下,语音韵律和不匹配的图片也诱发显著的 MMN 反应。进一步对比,发现视听条件下诱发的 MMN 波幅显著大于 aMMN,但是视听结合条件下对偏差刺激的探测和纯听觉条件下的探测一样快。说明在非注意条件下,语音中的韵律信息和视觉信息同样能够即刻整合。

在言语交流中,同样需要利用韵律信息和视觉信息的即刻整合。语音中的韵律信息和视觉中

的非言语线索的整合,是帮助听者理解讲话者口语中的态度和意图的重要加工过程(González-Fuente, Escandell-Vidal, & Prieto, 2015; Prieto, Pugliesi, Borrás-Comes, Arroyo, & Blat, 2015)。非言语视觉线索是指不依赖于音调高低和语音停顿等语音信息,通过视觉直接观察得到的外部行为表现,如眼神交流、面部表情、姿态动作等(梁静等, 2014)。有研究表明,手势能够帮助听者识别讲话者口语中的讽刺语气(González-Fuente et al., 2015), Esteve-Gibert, Prieto 和 Pons (2015)发现即使是 9 个月大的婴儿,在产生韵律和手势之前,都已经具备良好的感知同步的韵律和手势的能力。也有研究提供证据,非言语线索中的面部信息,如嘴唇的移动也能对听觉信息提供知觉线索,促进视听知觉的整合(Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009)。而且听觉语音加工中的视觉线索不仅仅是嘴唇的移动,也包括头和下巴姿势和眉毛的移动(Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004)。最近的一项研究,Strelnikov, Foxton, Marx 和 Barone (2015)使用 oddball 范式,探究面孔韵律移动对视听整合的影响。选取母语为法语的被试,采用的听觉材料是 3 个法语单词(译成汉语为一、二、三),分别进行重音(with emphasis)和非重音(with no emphasis)朗读,且重音在二上。视觉材料是面孔的移动视频,一个是以面孔移动抬眉毛为重音,另一个是面孔不移动没有重音,在视听整合和单独视觉条件下进行。结果发现,面孔移动的视听条件下,在视觉重音开始移动之后,听觉刺激呈现之前,有显著的 MMN 的类似激活。在面孔移动单独视觉条件下,尽管有相同的重音移动但是没有这种负波的出现。这些结果表明视听条件下,脑神经反应的是听觉预期的特异性反应。在多重感觉条件下,面孔的移动传递语音字词中的重音韵律,视觉形式能够促进听觉语音的预测编码。该结论为面孔移动和语音的整合作用提供了新的证据。

总结以往听觉韵律信息与视觉信息整合的研究发现,不同类别的非视觉线索都能较好的促进与听觉知觉的整合,而且同样能够诱发 MMN,但是不同种类的视觉线索在帮助听者理解口语过程中,哪种线索的促进效果更为显著的区分性研究较为匮乏。而且,语言线索与视觉知觉相整合,是

否与非语言线索的视听整合存在时间进程上的差异也需要进一步探究。

### 2.3 麦格克(McGurk)视听整合 MMN

虽然大部分情况下,研究者们倾向关注良好视听整合对言语交流的促进作用,但是同样不能忽视视听整合中独特的错觉干扰现象,即麦格克效应(McGurk Effect) (McGurk & MacDonald, 1976),它主要表现在语音感知过程中听觉和视觉之间的相互作用。当看到的一种声音与听到的另一种声音相重叠时,会让人们觉察到第三种声音。例如,一个音节“ga”在配音时发出了“ba”的声音,被试就会称听到的音节是“da”。这一效应表明,人的脑神经无法对较为接近的失匹配的视觉和听觉线索进行探测,而是自动的整合它们为最接近的另一种合法语音。该效应不仅在成人中存在,同样在 5 个月大的婴儿中得到了验证(Kushnerenko, Teinonen, Volein & Csibra, 2008)。

Stekelenburg 和 Vroomen (2012)也在自己的研究中提出了 McGurkMMN。采用 oddball 范式,同时呈现视觉和听觉刺激,将视觉上的唇语发音/onso/和听觉上的复制正弦波(sine-wave replicas, SWS)声音/onso/一致作为标准刺激,将视觉上的唇语发音/omso/和听觉上 sws 声音/onso/不一致作为偏差刺激,结果发现被试能够识别出偏差刺激与标准刺激之间的失匹配,而且这种失匹配只存在于语音模型中,而不存在于非语音模型中。语音模型是指人等自然生物发出的声音,非语音模型是指电脑等人工机器发出的声音。Stekelenburg 等人(2012)将这种语音模型中,唇语信息与声音不一致的偏差刺激诱发出的 MMN 叫做 McGurkMMN。研究者认为,这种现象是由于语音模型能引发行为和神经水平的激活,引起对听觉刺激本质的期待,强烈的影响了视听语音的整合。只有把复制的正弦波(sine-wave replicas, SWS)解释为语音,唇语信息影响听觉知觉,才表现为麦格克效应和 McGurkMMN。该研究的这一发现证明,即使是视听不匹配的情况下,同样支持知觉的语言特异性模型(Remez, Rubin, Pisoni, & Carrell, 1981)。鉴于前注意和 MMN 的知觉本质,可以认为这个视听整合的知觉模型完成在早期加工阶段。

最近, Eskelund, MacDonald 和 Andersen (2015)又将撒切尔效应和麦格克效应相结合来进行视听

整合研究。撒切尔效应(Thatcherization) (Thompson, 1980)是指倒着的脸上一些局部特征变化很难被人发现,而这样的变化如果发生在正立的脸上则非常明显。运用这两个效应发现了一系列有趣的结果。对正常直立面孔有强烈的麦格克错觉和相对应的MMN;对于一般的倒转面孔,有麦格克错觉和相对应的MMN出现,但是没有出现撒切尔效应。而当直立面孔是撒切尔效应时,有较弱的麦格克错觉,但是没有McGurkMMN,而且发现麦格克错觉的强烈程度和McGurkMMN的波幅之间没有相关。研究者解释认为,这可能是一种阈限效应,需要达到一定强度的麦格克错觉才能诱发McGurkMMN。以上结论说明,面孔结构的确能够影响视听知觉的整合。

结合之前研究结论中所提到的语音特异性模型,也适用于解释该研究结论。可以解释为当面孔是直立的情况下,被试能够较为容易的将发出的声音结合面孔感知为语音模式,从而产生了麦格克错觉和McGurkMMN。而当面孔是撒切尔错觉时,破坏了正常的面部结构,被试很难在短时间内将声音感知为语音模式,而倾向于较快的感知为非语音模式,从而只有较弱的麦格克错觉。因此可以认为,再一次的验证了语音特异性模型适用于视听不匹配的麦格克效应。

## 2.4 视听整合MMN的加工特异性

那么,视听整合MMN的加工又具有什么样的特点? Andres, Cardy 和 Joannis (2011)要求被试忽视采用oddball范式呈现的听觉刺激,专注于视觉任务,对目标刺激进行按键反应。被动呈现的听觉刺激与视觉刺激有一致和不一致两种情况。研究发现,这两种情况下偏差刺激都比标准刺激诱发更大的负波,都存在显著的aMMN反应,但是在视听不一致情况下诱发的aMMN波幅呈减弱的趋势。视听整合大致发生在字母和声音刺激出现后的180 ms左右,是前注意加工阶段,但是视听整合的效应能持续影响到加工的晚期阶段。Mittag, Alho, Takegata, Makkonen 和 Kujala (2013)使用同样的范式,认为视听整合发生在注意阶段。以辅音-元音音节为听觉刺激,和书面音节或者破碎的图片同时呈现。要求被试注意目标,对持续时间更长的目标做出按键反应,分别在听觉注意、视觉注意、视听注意条件以及忽视视听的情况下进行,通过设计倒着默数数字的任务创造

忽视条件。结果发现,听觉上的辅音变化和书面音节同时呈现时,引起更大的负反应,认为字母和语音的整合在视听的早期阶段和视觉的晚期阶段,但在单独听觉刺激以及忽视视听刺激情况下没有视听整合的出现。因此得出结论,认为视听注意才能促进视听整合。

Nakagawa, Hoshiyama, Uemura 和 Jomori (2012)在听觉、视听同步和视听异步的3种条件下,考察视觉刺激的呈现方式是否影响听觉失匹配反应。结果发现,3种条件下,只有视听同步条件诱发MMN的潜伏期显著小于其他条件下MMN的潜伏期。同时呈现视觉刺激能够增强听觉MMNs (Mismatch magnetic fields)。说明视听失匹配反应与刺激的呈现方式有关,同步呈现视听刺激整合的失匹配反应最强烈。Mingjin, Hasko, Schulte-Körne 和 Bruder (2012)在探究中文的字和声调整合加工的时间进程时得出了有趣的结论。研究者提出假设,中文和英文等字母语言不同,是一种声调语言,因此可能具有不同的整合加工。研究采用oddball范式,在4种条件下呈现刺激,条件分为只有听觉,视听同步,视听SOA 100 ms,视听SOA 200 ms。听觉刺激是/yi3/和/yi1/,视觉刺激是汉字“以”,是中文中的介词,能够避免语义加工。结果显示,中国被试在字音异步100 ms后才能自动整合字和声调,而不是同步整合。研究者解释认为主要的原因和中文整体复杂性的提高,以及处理中文和英文的神经机制不同有关。

综上所述,视听整合既可以发生在前注意加工阶段,也可以发生在注意加工阶段。字母和语音的视听整合在视听同步条件下整合最强烈,而中文的字和音的整合在视听异步100 ms条件下才能获得。

## 3 跨通道视听整合的关系

### 3.1 跨通道视听整合的相互补充

无论是阅读还是社会交往,自然情境中普遍涉及多感觉通路的跨通道交互作用,以视觉和听觉通路为例,就有研究证实,二者之间存在相互补充利用的关系(Bottari et al., 2014; Matusz et al., 2016; McDonald et al., 2013)。

McDonald 等人(2013)设计实验,在呈现视觉刺激之前,呈现突然的声音,且声音的位置保证随机出现不可预测。随后为了排除视觉任务类型

对结果的干扰,又将视觉刺激任务替换为听觉刺激任务进行考察。结果发现,在视觉刺激之前呈现突然的干扰声音,能够激活对侧枕叶(auditory-evoked contralateral occipital positivity, ACOP)的纹外视觉皮层;即使替换视觉目标刺激为听觉目标刺激,也发现了类似的视觉皮层激活。说明突出但是与任务无关的声音激活视觉皮层,且与观察目标和意图之间相互独立,即认为是突出的声音自动激活了视觉皮层。证明视觉和听觉跨通道的相互补充利用。但是随后关于这一补充过程是否自动发生,Matusz 等人(2016)提出了质疑。他们认为仅靠之前的研究还不足以判定这个过程的自动性,之前得出的自动性结论都是基于声音位置的不可预测性,如果改变声音位置的可预测性能否得到同样的结论还有待验证。从而提出假设,认为突出的声音激活视觉皮层的过程并不是自动的。因此设计被动 oddball 范式,让被试观看无声电影,并告知没有任何声音出现。但是用两种方式呈现声音,一种是空间位置规律性情境,在左侧和右侧出现声音的概率相等;一种是空间位置无规律性情境,用 oddball 范式呈现声音。结果发现,只在声音位置不可预测的情境下发现 ACOP 的激活,而在声音位置可预测的情境下没有类似激活。得到了与之前的结论相反,有力地证实了猜想,也就是说,这种跨通道激活显然不是完全自动的。这为跨通道信息加工的进一步研究开拓了视野。

更典型的视觉和听觉跨通补充利用的例子,存在于早期听力障碍者之中。Bottari 等人(2014)设计高阶视觉加工(high-order visual processes)任务,是一种动态的视觉变化探测任务,图形的形状改变时移动方向也发生变化。发现早期听力障碍者发生了听觉皮层的补充。该视觉事件所诱发的“听觉”反应出现的时间与听力正常者执行相同任务出现 vMMN 的时间一致,在视觉刺激出现后的 150~300 ms 左右。而且,对视觉变化探测的听觉皮层的补充反应和视觉系统内的反应减少是成对出现的。结果说明,早期听力障碍者中,有从视觉皮层到听觉皮层的部分加工转换。因此表明,对即将发生的事件,听力障碍者的听觉皮层同样参与提取和存储视觉信息,视觉和听觉皮层之间能够跨通道补充利用。

由此可见,视觉和听觉皮层在自然情境中相

互补充利用,共同协作,但是这种协作过程的自动性相关问题还需要更多的研究进行探讨。虽然听觉皮层的核心能力是建构有关听觉情境的记忆痕迹,继而自动探测声音情境内的差异(Näätänen, Astikainen, Ruusuvirta, & Huotilainen, 2010),但是我们了解到,在早期的听力障碍者中,即使听觉功能受损,然而通过视觉皮层的补充,听觉皮层也能得到补偿激活,获得参与收集动态视觉任务的变化和形式信息的能力。这种跨通道的补充利用具有广泛的生物学的意义。

### 3.2 跨通道视听整合的相互竞争

视觉和听觉通道之间除了能够相互补充利用,也有研究提供证据表明,它们之间还存在着相互竞争认知资源分配的关系(Leiva et al., 2015; Regenbogen et al., 2012)。

Regenbogen 等人(2012)就通过研究发现了这一竞争关系。实验是以年轻人为被试,探究增加视觉任务的工作记忆负载对听觉信息加工的影响。要求被试执行 4 种难度递增 n-back 视觉任务,同时以被动 oddball 范式呈现高频和低频纯音。结果发现,随着视觉任务的难度上升,被试的反应时更长,正确率更低,证明视觉任务难度的提高有效操控了记忆负载的增加。基本听觉加工的诱发电位波幅也持续增强,能够系统的表明与跨通道的认知负载有关。同时发现,增加视觉工作记忆负载,主要听觉皮层逐渐失活,也就是说,当中枢执行系统增加认知负载时,会持续的为脑神经加工主要任务的区域分配资源,而且还出现了与视觉工作记忆负载的脑区相连接的听觉皮层。说明,视觉和听觉的跨通道既有竞争又有补充利用的关系。

除了某一通道下任务难度的增加,最典型影响认知资源分配的因素就是年龄(Andrés et al., 2008), Leiva 等人(2015)为解决以往研究对于是在单一模式还是跨通道模式下年龄影响认知资源分配这一争论(Guerreiro, Murphy, & van Gerven, 2010; Guerreiro, Murphy, & van Gerven, 2013; Guerreiro, & van Gerven, 2011),设计实验进行了两种通道模式的对比研究。实验被试根据年龄分为年轻和年老组,执行的任务是要求被试对听到或看到的数字进行奇偶判断。任务分为两种类型,一种是听听,另一种是视听。听听类型下,在和数字相对应的声音出现前有一个短促的无关要求忽视的声音;

视听类型下,在屏幕上随机出现1~6个数字,耳机里呈现无关要求忽视的声音。两种条件的无关声音都是以被动oddball范式呈现。结果发现,当目标是视觉任务时,年龄的增长使听觉的干扰效果增加,但是当目标是听觉任务时,听觉干扰并没有类似的影响。也就是说,在跨通道条件下,年龄的增长能够影响分心任务的执行,而在单一通道下不表现为这种趋势。证明视觉和听觉跨通道之间存在相互竞争认知资源分配的关系。

既然如此,增加视觉的知觉负载能否引起aMMN波幅、潜伏期等属性变化?为探究这一问题,Wiens等人(2016)设计了两种负载不同的字母探测任务。在屏幕中央呈现注视点,字母环在周围,要求被试集中注意注视点,余光注意字母环,并对字母X进行探测。低负载条件下,每次呈现相同的6个字母,高负载条件下,由7个HKMN VWZ中的5个随机字母和字母X形成字母环。在呈现字母环的同时使用被动oddball范式呈现两种纯音。原本假设随着知觉负载的增加,会影响任务无关纯音所诱发的aMMN。结果发现,即使高知觉负载影响行为上执行任务的反应时和正确率,也仍然可以观察到强烈的aMMN。在高、低负载知觉下,aMMN并不表现出差异。

这些研究表明,视觉和听觉之间除了相互补充利用还存在相互竞争的关系,在增加视觉主任务知觉负载的同时,呈现听觉干扰将导致认知资源的分配,影响主任务的执行。若主任务是听觉任务时,视觉干扰能否影响听觉主任务的执行,还有待进一步探讨。

#### 4 小结

视听整合的MMN研究,是在升华aMMN和弥补vMMN局限性等方面的一种创新,这种整合加工具有更高的生态效度,更贴近实验室外的多通道加工情境。这一问题的研究不仅能为多感觉通路的作用和机制提供解答,也能为今后更多的跨通道研究提供思路与依据。

首先,在视觉和听觉的跨通道整合中,存在相互补充又相互竞争的关系。早期耳聋者虽然听觉功能受损,但是视觉皮层能够激活补充听觉皮层,使听觉皮层获得参与收集和存储信息的能力。跨通道整合补充的加工效果优于单通道的加工效果,显示出跨通道整合的优势。这种视听的

跨通道整合帮助我们更有效的获取信息(McDonald et al., 2013; Bottari et al., 2014)。但是当执行某一通道下需要集中注意力的主任务时,跨通道整合又显示出分心、不利于任务执行的一面。以往研究发现,当逐渐增加视觉主任务的负载时,听觉分心任务的干扰力随着年龄的增加而增强,而在单一的听觉通道下,随着年龄的增长,干扰任务的影响并不显著(Leiva et al., 2015)。然而,如果将主任务类型变换为听觉或触觉等其他任务,能否获得这一结论仍不得而知。

其次,听觉信息和视觉信息对视听整合的效果有不同的影响。在言语交流中,有研究表明(Li et al., 2009),当韵律信息负载在听觉言语信息上时,单独听觉条件下能够诱发MMN成分。当听觉信息叠加到视觉刺激上时,视听整合条件下诱发MMN和单独听觉条件下诱发MMN一样快。也有研究发现(Strelnikov et al., 2015),当韵律信息负载在视觉线索上(如眉毛的移动),单独视觉条件下,并没有类似MMN的负波出现。然而当听觉信息和视觉信息同时呈现时,视觉韵律线索能够帮助被试做出关于听觉韵律的预测,达到视听整合的效果。

最后,需要进一步加强特殊群体的视听整合加工研究。Žarić等人(2014)考察阅读障碍儿童字母和语音神经整合加工特点发现,和典型发展儿童相比,阅读障碍儿童字母和语音神经整合能力较弱。同时Žarić等人(2015)发现,儿童这种较弱的视听神经整合,通过训练和教学能够有所改善。然而,对于成人阅读障碍者而言,存在哪些视听整合加工特点?在字母和语音的神经整合中,成人阅读障碍者是否也可以通过训练获得提高?研究(Jones et al., 2016)发现,视觉语音联系受损是导致成人阅读障碍者阅读障碍的主要因素之一。这一因素是否也影响儿童阅读障碍者的阅读行为,阅读过程中儿童阅读障碍者如何整合视听信息?对这些问题的研究将有助于深化人们对阅读障碍群体视听整合加工过程的理解,同时也有助于为改善该群体的整合加工训练提供依据。

#### 参考文献

- 梁静,李开云,曲方炳,陈宥辛,颜文靖,傅小兰. (2014). 说谎的非言语视觉线索. *心理科学进展*, 22, 995-1005.
- 任桂琴,韩玉昌,周永奎,任延涛. (2011). 汉语语调早期



- 加工的脑机制. *心理学报*, 43, 241–248.
- 任桂琴, 刘颖, 于泽. (2012). 汉语口语韵律的作用及其神经机制. *心理科学进展*, 20, 338–343.
- 文小辉, 李国强, 刘强. (2011). 视听整合加工及其神经机制. *心理科学进展*, 19, 976–982.
- 于泽, 韩玉昌, 任桂琴. (2010). 韵律在语言加工中的作用及其神经机制. *心理科学进展*, 18, 420–425.
- Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, 34, 534–546.
- Andres, A. J. D., Cardy, J. E. O., & Joanisse, M. F. (2011). Congruency of auditory sounds and visual letters modulates mismatch negativity and p300 event-related potentials. *International Journal of Psychophysiology*, 79, 137–146.
- Andrés, P., Guerrini, C., Phillips, L. H., & Perfect, T. J. (2008). Differential effects of aging on executive and automatic inhibition. *Developmental Neuropsychology*, 33, 101–123.
- Arnal, L. H., Wyart, V., & Giraud, A. -L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, 14, 797–801.
- Bottari, D., Heimler, B., Caclin, A., Dalmolin, A., Giard, M. -H., & Pavani, F. (2014). Visual change detection recruits auditory cortices in early deafness. *NeuroImage*, 94, 172–184.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5, e1000436.
- Choudhury, N. A., Parascando, J. A., & Benasich, A. A. (2015). Effects of presentation rate and attention on auditory discrimination: A comparison of long-latency auditory evoked potentials in school-aged children and adults. *PLoS One*, 10, e0138160.
- Cleary, K. M., Donkers, F. C. L., Evans, A. M., & Belger, A. (2013). Investigating developmental changes in sensory processing: Visual mismatch response in healthy children. *Frontiers in Human Neuroscience*, 7, 922.
- Csukly, G., Stefanics, G., Komlósi, S., Czigler, I., & Czobor, P. (2013). Emotion-related visual mismatch responses in schizophrenia: Impairments and correlations with emotion recognition. *PLoS One*, 8, e75444.
- Erlbeck, H., Kübler, A., Kotchoubey, B., & Veser, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400. *Frontiers in Human Neuroscience*, 8, 654.
- Eskelund, K., MacDonald, E. N., & Andersen, T. S. (2015). Face configuration affects speech perception: Evidence from a McGurk mismatch negativity study. *Neuropsychologia*, 66, 48–54.
- Esteve-Gibert, N., Prieto, P., & Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behavior and Development*, 38, 126–129.
- Froyen, D., Willems, G., & Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: An electrophysiological study of letter–speech sound processing. *Developmental Science*, 14, 635–648.
- González-Fuente, S., Escandell-Vidal, V., & Prieto, P. (2015). Gestural codas pave the way to the understanding of verbal irony. *Journal of Pragmatics*, 90, 26–47.
- Guerreiro, M. J. S., Murphy, D. R., & van Gerven, P. W. M. (2010). The role of sensory modality in age-related distraction: A critical review and a renewed view. *Psychological Bulletin*, 136, 975–1022.
- Guerreiro, M. J. S., Murphy, D. R., & van Gerven, P. W. M. (2013). Making sense of age-related distractibility: The critical role of sensory modality. *Acta Psychologica*, 142, 184–194.
- Guerreiro, M. J. S., & van Gerven, P. W. M. (2011). Now you see it, now you don't: Evidence for age-dependent and age-independent cross-modal distraction. *Psychology and Aging*, 26, 415–426.
- Hedge, C., Stothart, G., Jones, J. T., Frias, P. R., Magee, K. L., & Brooks, J. C. W. (2015). A frontal attention mechanism in the visual mismatch negativity. *Behavioural Brain Research*, 293, 173–181.
- Holloway, I. D., van Atteveldt, N., Blomert, L., & Ansari, D. (2015). Orthographic dependency in the neural correlates of reading: Evidence from audiovisual integration in English readers. *Cerebral Cortex*, 25, 1544–1553.
- Huotilainen, M., Lovio, R., Kujala, T., Tommiska, V., Karma, K., & Fellman, V. (2011). Could audiovisual training be used to improve cognition in extremely low birth weight children? *Acta Paediatrica*, 100, 1489–1494.
- Ikeda, K., Akiyama, H., Iritani, S., Koichi, K., Arai, T., Niizato, K., .... Kosaka, K. (1996). Corticobasal degeneration with primary progressive aphasia and accentuated cortical lesion in superior temporal gyrus: Case report and review. *Acta Neuropathologica*, 92, 534–539.
- Jones, M. W., Kuipers, J. R., & Thierry, G. (2016). ERPs reveal the time-course of aberrant visual-phonological binding in developmental dyslexia. *Frontiers in Human Neuroscience*, 10, 71.
- Jost, L. B., Eberhard-Moscicka, A. K., Frisch, C., Dellwo, V., & Maurer, U. (2014). Integration of spoken and written words in beginning readers: A topographic ERP study. *Brain Topography*, 27, 786–800.
- Kekoni, J., Hämäläinen, H., Saarinen, M., Gröhn, J.,

- Reinikainen, K., Lehtokoski, A., & Näätänen, R. (1997). Rate effect and mismatch responses in the somatosensory system: ERP-recordings in humans. *Biological Psychology*, 46, 125–142.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134, 372–384.
- Kreegipuu, K., Kuldkepp, N., Sibolt, O., Toom, M., Allik, J., & Näätänen, R. (2013). vMMN for schematic faces: Automatic detection of change in emotional expression. *Frontiers in Human Neuroscience*, 7, 714.
- Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 11442–11445.
- Leiva, A., Parmentier, F. B., & Andrés, P. (2015). Aging increases distraction by auditory oddballs in visual, but not auditory tasks. *Psychological Research*, 79, 401–410.
- Leppänen, P. H. T., Richardson, U., Pihko, E., Eklund, K. M., Guttorm, T. K., Aro, M., & Lyytinen, H. (2002). Brain responses to changes in speech sound durations differ between infants with and without familial risk for dyslexia. *Developmental Neuropsychology*, 22, 407–422.
- Li, X., Yang, Y., & Ren, G. (2009). Immediate integration of prosodic information from speech and visual information from pictures in the absence of focused attention: A mismatch negativity study. *Neuroscience*, 161, 59–66.
- Lu, Y., Paraskevopoulos, E., Herholz, S. C., Kuchenbuch, A., & Pantev, C. (2014). Temporal processing of audiovisual stimuli is enhanced in musicians: Evidence from magnetoencephalography (MEG). *PLoS One*, 9, e90686.
- Matusz, P. J., Retsa, C., & Murray, M. M. (2016). The context-contingent nature of cross-modal activations of the visual cortex. *NeuroImage*, 125, 996–1004.
- McDonald, J. J., Störmer, V. S., Martínez, A., Feng, W. F., & Hillyard, S. A. (2013). Salient sounds activate human visual cortex automatically. *The Journal of Neuroscience*, 33, 9194–9201.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- Mingjin, H., Hasko, S., Schulte-Körne, G., & Bruder, J. (2012). Automatic integration of auditory and visual information is not simultaneous in Chinese. *Neuroscience Letters*, 527, 22–27.
- Mittag, M., Alho, K., Takegata, R., Makkonen, T., & Kujala, T. (2013). Audiovisual attention boosts letter-speech sound integration. *Psychophysiology*, 50, 1034–1044.
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, 15, 133–137.
- Näätänen, R. (1988). Implications of ERP data for psychological theories of attention. *Biological Psychology*, 26, 117–163.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13, 201–233.
- Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory-cognitive core of cognitive processes. *Brain Research Reviews*, 64, 123–136.
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313–329.
- Näätäneiv, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., ... Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–434.
- Näätänen, R., Paavilainen, P., Titinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, 30, 436–450.
- Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, 115, 140–144.
- Näätänen, R., Sussman, E. S., Salisbury, D., & Shafer, V. L. (2014). Mismatch negativity (MMN) as an index of cognitive dysfunction. *Brain Topography*, 27, 451–466.
- Nair, A. K., Sasidharan, A., John, J. P., Mehrotra, S., & Kutty, B. M. (2016). Assessing neurocognition via gamified experimental logic: A novel approach to simultaneous acquisition of multiple ERPs. *Frontiers in Neuroscience*, 10, 1.
- Nakagawa, Y., Hoshiyama, M., Uemura, J.-I., & Jomori, I. (2012). Auditory MEG mismatch responses modified by visual stimulation accompanying auditory stimulation. *Neurophysiology*, 44, 247–254.
- Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, 18, 49–65.
- Prieto, P., Pugliesi, C., Borrás-Comes, J., Arroyo, E., & Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *Journal of Phonetics*, 49, 41–54.
- Regenbogen, C., De Vos, M., Debener, S., Turetsky, B. I.,

- Möbhang, C., Finkelmeyer, A., .... Kellermann, T. (2012). Auditory processing under cross-modal visual load investigated with simultaneous EEG-fMRI. *PLoS One*, 7, e52267.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947–949.
- Roberts, G., Anderson, P. J., & Doyle, L. W. (2009). Neurosensory disabilities at school age in geographic cohorts of extremely low birth weight children born between the 1970s and the 1990s. *The Journal of Pediatrics*, 154, 829–834.e1.
- Ruhnau, P., Herrmann, B., Maess, B., Brauer, J., Friederici, A. D., & Schröger, E. (2013). Processing of complex distracting sounds in school-aged children and adults: Evidence from EEG and MEG data. *Frontiers in Psychology*, 4, 717.
- Ruhnau, P., Wetzel, N., Widmann, A., & Schröger, E. (2010). The modulation of auditory novelty processing by working memory load in school age children and adults: A combined behavioral and event-related potential study. *BMC Neuroscience*, 11, 26.
- Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62, 437–448.
- Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H. M., Laine, M., .... Pihko, E. (2010). Auditory and cognitive deficits associated with acquired amusia after stroke: A magnetoencephalography and neuropsychological follow-up study. *PLoS One*, 5, e15157.
- Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology*, 59, 451–475.
- Shtyrov, Y., Goryainova, G., Tugin, S., Ossadtchi, A., & Shestakova, A. (2013). Automatic processing of unattended lexical information in visual oddball presentation: Neurophysiological evidence. *Frontiers in Human Neuroscience*, 7, 421.
- Stefanics, G., Kimura, M., & Czigler, I. (2011). Visual mismatch negativity reveals automatic detection of sequential regularity violation. *Frontiers in Human Neuroscience*, 5, 46.
- Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological evidence for a multisensory speech-specific mode of perception. *Neuropsychologia*, 50, 1425–1431.
- Stephens, B. E., & Vohr, B. R. (2009). Neurodevelopmental outcome of the premature infant. *Pediatric Clinics of North America*, 56, 631–646.
- Strelnikov, K., Foxton, J., Marx, M., & Barone, P. (2015). Brain prediction of auditory emphasis by facial expressions during audiovisual continuous speech. *Brain Topography*, 28, 494–505.
- Strömmer, J., Tarkka, I., & Astikainen, P. (2014). Somatosensory mismatch response in young and elderly adults. *Frontiers in Aging Neuroscience*, 6, 293.
- Thompson, P. (1980). Margaret thatcher: A new illusion. *Perception*, 9, 483–484.
- Tse, C. Y., Rinne, T., Ng, K. K., & Penney, T. B. (2013). The functional role of the frontal cortex in pre-attentive auditory change detection. *NeuroImage*, 83, 870–879.
- van Atteveldt, N. M., Blau, V. C., Blomert, L., & Goebel, R. (2010). fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. *BMC Neuroscience*, 11, 11.
- van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1053–1065.
- Wei, J. H., Chan, T. C., & Luo, Y. J. (2002). A modified oddball paradigm “cross-modal delayed response” and the research on mismatch negativity. *Brain Research Bulletin*, 57, 221–230.
- Weise, A., Grimm, S., Trujillo-Barreto, N. J., & Schröger, E. (2014). Timing matters: The processing of pitch relations. *Frontiers in Human Neuroscience*, 8, 387.
- Wiens, S., Szychowska, M., & Nilsson, M. E. (2016). Visual task demands and the auditory mismatch negativity: An empirical study and a meta-analysis. *PLoS One*, 11, e0146567.
- Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, 20, 1235–1249.
- Yang, X. X., Yu, Y. M., Chen, L., Sun, H. L., Qiao, Z. X., Qiu, X. H., .... Yang, Y. J. (2016). Gender differences in pre-attentive change detection for visual but not auditory stimuli. *Clinical Neurophysiology*, 127, 431–441.
- Žarić, G., González, G. F., Tijms, J., van der Molen, M. W., Blomert, L., & Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PLoS One*, 9, e110337.
- Žarić, G., González, G. F., Tijms, J., van der Molen, M. W., Blomert, L., & Bonte, M. (2015). Crossmodal deficit in dyslexic children: Practice affects the neural timing of letter-speech sound integration. *Frontiers in Human Neuroscience*, 9, 369.

## The characteristics and mechanisms of audiovisual integration: Evidence from mismatch negativity

XIN Xin<sup>1</sup>; REN Gui-Qin<sup>1</sup>; LI Jin-Cai<sup>2</sup>; TANG Xiao-Yu<sup>1</sup>

(<sup>1</sup> College of psychology, Liaoning Normal University, Dalian 116029, China)

(<sup>2</sup> School of Foreign Languages, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract:** Audiovisual integration refers to the cognitive process in which auditory and visual events that are presented at the same time and the same location tend to integrate with each other. Mismatch Negativity (MMN), as an index of early cortical processing, is usually used to explore whether there is mismatch between deviant information and traces of sensory memory. Previous studies on audiovisual integration mainly focused on the integration of letters and phonology, integration of non-language visual information and prosodic auditory information, McGurk Effect, as well as the mechanisms underlying the integration processing. This paper reviewed those recent studies and analyzed the possible factors that would influence the integration of auditory and visual processing. Future research should focus on the integration of information from multiple modalities.

**Key words:** auditory; visual; audiovisual integration; MMN; ERP