

身份识别的跨通道研究策略： 面孔—语音联合加工

周爱保，张荣华，马小凤，朱婧

(西北师范大学心理学院，甘肃兰州 730070)

摘要：对人身份识别是社会认知的重要方面，面孔—语音特征为对人身份识别提供了丰富的信息。已有的研究主要运用再认、识别范式探讨对面孔—语音的联合加工。对人面孔—语音加工的机制为独立的序列加工或者联合的平行加工，所涉及脑区主要包括颞上沟等区域，然而对二者的探讨并没有统一结论。通过改进研究范式、确定对比基线等方法可以对面孔—语音联合加工进行更深入的研究。

关键词：面孔—语音整合；身份识别；跨通道

中图分类号：G446

文献标识码：A

文章编号：1008-0627(2013)01-0038-04

一、引言

对人知觉是社会知觉的重要部分，指人通过感官对自我和他人的直接整体感知。通常而言，对人知觉包括对个人身份(区别性特征，如性别、年龄和体型等独特的信息)以及情绪状态的知觉，涉及对他人语音和面孔信息的整合。人们通过一个动态的系统对他人进行知觉，对面孔、语音和身体等线索的加工位于该系统的较低水平。这种加工通常涉及对多个感觉通道输入信息的整合，其中视觉—听觉是主要加工通道。

视觉—听觉联合加工对于人们如何编码他人信息具有重要意义，这种“自下而上”的加工为人们对人知觉提供了依据。面孔和语音提供了个体的独特身份信息，对于他人身份的识别基于对被识别者面孔、语音信息的分析，在对个人身份识别过程中，面孔—语音整合是所有知觉的基础。在社会交往中，动态的面孔和语音特征为信息接收者提供了系统而准确的时空信息，来自两个通道的信息可以通过互补和简化两种处理方式来提高人们对人知觉时的速度和准确性。

二、面孔—语音加工的主要研究范式

(一) 再认范式

对人的再认(recognition)，是当前输入的信息与信息接收者曾经接收过的信息之间的连

结被激活的过程。这种激活可以通过不同感觉通道的信息(例如一个人的面孔、轮廓或者语音)同时或者分别呈现而实现，因此可以通过考察同时呈现的刺激与单独呈现的刺激之差来探索人们面孔—语音的整合方式。再认范式主要包括学习和测试两个阶段：在学习阶段，通常需要被试学习并记忆人物的面孔语音信息，并通过对被试进行正确率测试或者问卷来确保被试完全掌握了匹配；在测验阶段，要求被试对于单独或者同时呈现的不同通道刺激进行再认。目前的面孔—语音跨通道对人身份识别研究大多采用该范式进行。然而在实际生活中，对已知他人再认只是一个方面，更多的时候，我们需要对陌生人进行识别，因此提出了识别范式。

(二) 识别范式

当人们面对他人时，会自动触发关于他人的社会分类(例如性别、年龄或者种族等)，而面孔和语音是这种分类自下而上的线索。与再认范式类似，基于同时呈现的视听刺激与单通道刺激和之差 $[S_{A+V} - (S_A + S_V)]$ 的逻辑，设计S1-S2识别范式。单通道条件下，S1和S2均为相继呈现的听觉刺激；跨通道条件下，S1为面孔刺激，S2为之后呈现的听觉刺激，S1和S2组合形成人物一致组和相异组。实验要求被试对语音所表现

收稿日期：2012-10-30

第一作者简介：周爱保(1962-)，男，甘肃武山人，教授/博导，主要研究方向：实验心理学，认知心理学。E-mail: zhouab@nwnu.edu.cn

的人物社会属性进行分类（例如年轻或年老）。

通过两种范式，可以肯定面孔、语音信息在对人识别时相互影响，但仍然不能确定这种影响的方式，也无法量化影响的大小，因此需要改进实验范式，例如模糊信息或切分变量，使视觉、听觉信息的影响可以量化。另外，对于面孔语音联合知觉的主要研究范式都涉及单通道与视听双通道的对比，以期通过对不同通道间识别效果差异来说明跨通道识别的本质。然而在视听跨通道研究中，两种通道接收信息的类型和特质不同，所识别信息的基线亦有所差异，正是这种固有的物理属性差异导致视听通道信息识别的不同。采用对比方法进行研究固然可以在一定程度上阐释不同通道的信息加工，然而，是否可以直接对视听加工本质进行探索，有待新范式的出现。

三、面孔—语音加工的神经机制

（一）单感觉通道的加工

对说话者识别的单通道研究表明，对人识别可以通过对面孔或语音的单通道加工进行，例如仅通过看照片或者打电话就可以知道个体身份，对面孔和语音进行加工的脑区分别为梭状回面孔区（FFA）和颞上沟（STS）区域。

截至目前，研究者在面孔知觉的研究领域颇有建树，他们发现，面孔加工是基于分布广泛的、序列式的神经网络，这些网络在大脑两个半球均有分布。具体而言，不同的面孔特征激活不同的脑区：恒定的面孔特征加工脑区位于右侧梭状回面孔区；而对于动态面孔特征（如眼动和唇动）的知觉则与颞上沟关联。^[1]

语音同样提供个人身份信息，研究表明，语音刺激可以激活STS区域。相比其他声音，人的语音在STS脑区有相对更大的激活，这种激活受说话者的身份信息影响（反应说话者的语音特征），而与所说的内容无关。具体而言，接近颞叶语音区域（TVAs）的右侧颞上回后部，在对说话者进行识别的过程中有重要作用。^[2]

面孔—语音联合加工至少在一定程度上依赖于单通道视、听觉区域的激活，例如FFA区域。然而对于视、听觉通道在身份识别时所占权重比例，尚未有定论。语音和面孔信息以不同的方式为对人识别提供线索：语音可以远距离传递信息，而面孔却能对信息有更精细的表达。纵观已

有研究，似乎表明对于动态的、以时间维度呈现的刺激，如动态面孔或者语音，STS区域更加敏感；而对于静态的、以空间维度呈现的刺激，如面孔信息，则更多地依赖于FFA区域。尽管对面孔和语音的加工不同，但是二者并不是完全独立的，当两者同时呈现时，他们之间有交互作用。

（三）面孔—语音联合加工机制

当同时呈现相关的面孔时，语音更容易被识别；而当出现不一致的面孔时，语音的识别受到阻碍。这表明，听者不能忽视与语音同时呈现的面孔，面孔和语音的结合在行为方面表现出易化和抑制两种效果。易化表现为当说话者面孔可见（“唇读”）时，语言会得到更好的理解。而抑制则出现在由实验控制的非一致刺激条件下，通常伴随正确识别率的减少以及相应反应时的增加：如在“麦克格效应”（McGurk Effect）中，语音和口型的不一致所制造出的错误知觉。

1. 交互作用发生的时间机制

面孔—语音联合是否发生于单通道信息充分得到加工之后，不同研究有不同结论。Burton等人指出，对于姓名和面孔的联合加工单元，会合成一个对他人身份识别的超通道结点，人们通过这种节点来对他人身份进行加工。^[3]基于Burton的观点，有研究认为，只有当视听通道的信息分别得到加工后，才会出现视听联合加工，因此提出对语音和面孔特征的分离序列加工过程。^[4]在该模型中，面孔和语音信息首先分别得到独立编码，继而形成面孔—语音知觉单元，这些单元存储并提供了个人诸如专业或者姓名之类的自传体信息——人们可以通过他人的面孔或/和语音获得。而有研究提出质疑，认为面孔—语音加工并非各自独立，即跨通道联合加工并不需要单通道加工作为基础。呈现面孔会对语音的加工产生影响，并且这种跨通道交互作用在加工早期和晚期均有表现，表明知觉加工和语义加工阶段都有跨通道交互作用。因此，面孔和语音加工并非各自独立的加工过程，而是在知觉建构编码阶段就已经出现整合。^[5]

2. 交互作用发生的脑区定位

与联合加工的时间机制争议相同，跨通道加工是否存在一个将不同通道进行联合加工的区域，是跨通道加工脑区定位的研究重点。部分研

究认为,多通道感觉的加工过程基于输入刺激的单通道充分加工,即存在一个联合区域,对得到充分加工的单通道刺激进行整合。^[6]

然而,有研究认为,主要的视觉、听觉脑区域可能直接涉及多感觉交互,并且这些区域同样参与了对人知觉。例如听觉皮层会受到视觉输入信息的影响——听觉皮层可以由看见说话者的唇动而引发激活;同样,熟悉的人的语音可以激活视觉区域。^[7]

颞侧区域,尤其是STS区域被普遍认为是视听多感觉整合区域。对人类而言,颞上沟联接来自视觉和听觉的信息,并且该区域的激活受到感觉通道可靠性的影响:越可靠的通道会获得更大的权重,当听觉通道更加可信时,STS和听觉皮层的功能性相关有所增强,而当视觉皮层更加可信时,STS和视觉皮层的功能性相关增强。左侧角回(angular)和后侧海马也被认为参与了面孔和语音的整合。^[8]表明对于面孔和语音联合加工涉及多种脑功能,例如感觉、注意以及记忆。

早期面孔-语音联合知觉的研究集中于讨论是否存在一个超通道(supramodal),将视听通道的刺激联合进行加工。随着计算机技术的发展,神经图像技术在心理学研究中的应用,大量研究采用ERP、fMRI及PET等技术,试图揭示跨通道加工的机制。当前一个一致的发现是,存在高水平的联合皮层,作为异质性通道的融合区域,可以结合感觉皮层的信息——尤其是颞上沟后侧区域(PSTS),这些区域在接受跨通道和单通道刺激或者一致和不一致的面孔-语音配对时,表现出不同的激活。^[8]

(三) 面孔-语音加工模型

来自面孔和语音的身份信息如何在脑内进行加工?当前的争议主要集中在面孔-语音信息的加工是否在面孔、语音信息分别得到加工后进行。据此,主要存在三种解释模型。

平行加工模型认为,多感觉整合只有在单感觉得到充分加工后才会出现。相比于单通道感觉加工,多感觉整合是更高序列的加工。例如,面孔失认症患者虽然失去了对面孔的认知能力,但仍然保留对语音的识别能力,证明了面孔和语音通过独立的序列进行加工。而以相反的损伤模式为特征的语音失认症案例表明同样支持了面孔

语音的独立加工过程。然而,大量行为研究证明,面孔-语音的联合加工是自动发生的,视、听通道的信息彼此影响,并不能完全独立;而对脑神经的研究发现,跨通道对人识别同时激活视觉、听觉加工区域和第三区域,说明跨通道自动加工的生理基础;并且就时间进程而言,跨通道发生在知觉阶段的早期。平行加工模型并不能解释这种自动产生的联合加工,而新的模型认为,跨通道识别是广泛存在于认知过程中,具有广泛生理基础的加工。

第二种模型指出,尽管不同的感觉系统可以各自独立进行加工,多感觉整合是快速的、广泛的并且遍布于各种物种。即存在通道特异性的分布回路亚系统,它们联系在一起,共同组成了多通道知觉系统。例如对猴子的感觉加工系统研究表明:听觉系统加工来自同伴的呼唤,视觉系统加工同伴的面孔和身体,它们的感情系统对同伴的表情和动作进行情绪加工;联合区域捕捉这些动作,将它们存储以用作之后的表征,当其中的某些动作被激活,那么将同时激活不同通道的信息。^[9]

第三种为结构解剖模型。这种观点认为,数目众多的神经元都参与多感觉加工,能够对视觉或者听觉刺激做出反应,并且能够在神经元水平对两种一致的信息进行整合。可以认为,大部分大脑新皮层是多感觉的。因此,知觉发展并没有在一个感觉通道发生,而是在一开始就进行了整合。与第一种模型完全不同,这种模型将跨通道整合解释为存在于整个加工过程。

与第一种加工模型相比,后两种模型都强调了跨通道加工是一种复杂的加工过程,并不是简单的单通道加工的累加,并且广泛分布于大脑皮层,在认知早期就参与识别过程。这两种模型的区别在于是否承认单通道加工的独立存在:第二种模型认为,同时存在对于单通道信息的识别和双通道信息的联合加工;而第三种模型则更加强调双通道的加工作用。虽然实验研究表明,广泛的神经元都可以进行跨通道知觉,然而,第三种模型无法解释面孔失认症的例子。因此,我们可以推测,单通道和双通道在对人身份识别中都参与信息加工,这种加工的权重会因为信息本身的可靠性而有所不同。

四、未来研究展望

视觉和听觉是人类认识世界最主要的感觉通道,通过两种通道获得的信息似乎比单通道信息更加可靠。视觉可以进行空间加工,而听觉可以进行时间加工,这种多维的加工方式可以让对人的知觉加工更加立体。然而,也有研究认为,两种通道有可能发生冲突而影响对人识别。时至今日,对于人们使用视听通道来知觉他人的最优解仍然没有定论,有一部分原因来自于比较基线的差异。

当前研究中,主要有视听单通道相加作为基线,单通道基线以及视听联合加工三种方式。部分研究以单通道之和或者单通道与双通道加工之间的差异来表征双通道加工,这种用线性模型来表征的双通道加工从研究伊始就将双通道和单通道的关系进行了简单线性定位,而降低了研究结果的效度;视听联合加工基线更适用于脑电及脑磁图等脑机制探索,而没有办法从行为上分离双通道加工,选择某种更为有效的基线对于研究结果的可靠性尤为重要。未来研究中,对于基线的确定应该更加灵活,可以将单通道基线与联合基线相结合,补充说明视听通道的加工方式。视听联合加工研究在多学科多领域中都有所涉及:计算机科学,医学,物理学,心理学等多个领域都受益于视听联合加工的研究成果;同时视听联合加工也需要从不同学科加以阐释。对视听联合加工的研究应该是多学科研究成果的交叉,未来研究应该从多学科领域进行整合及应用。

参考文献

- [1] ERNST M O, BULTHOFF H H. Merging the senses into a robust percept [J]. *Trends In Cognitive Science*, 2004, 8 (4): 162-169.
- [2] KAYSER C, LOGOTHETIS N K. Do early sensory cortices integrate cross-modal information? [J]. *Brain structure and function*, 2007, 212 (2): 121-132.
- [3] BURTON A M, BRUCE V, JOHNSTON R A. Understanding face recognition with an interactive activation model [J]. *British Journal of Psychology*, 1990, 81 (3): 361-380.
- [4] ELLIS H D, JONES D M, MOSDELL N. Intra- and inter-modal repetition priming of familiar faces and voices [J]. *British Journal of Psychology*, 1997, 88 (1): 143-156.
- [5] FÖCKER J, HÖLIG C, BEST A, RÖDER B. Cross modal interaction of facial and vocal person identity information: An event-related potential study [J]. *Brain Research*, 2011, 1385 (18): 229-245.
- [6] CAMPANELLA S, BRUYER R, FROIDBISE S. Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations [J]. *Clinical Neurophysiology*, 2010, 121 (11): 1 855-1 862.
- [7] CAMPANELLA S, BELIN P. Integrating face and voice in person perception [J]. *Trends in Cognitive Sciences*, 2007, 11 (12): 535-543.
- [8] JOASSIN F, MAURAGE P, Campanella S. The neural network sustaining the crossmodal processing of human gender from faces and voices: An fMRI study [J]. *Neuroimage*, 2010, 54 (2): 1 654-1 661.
- [9] GHAZANFAR A A, CHANDRASEKARAN C, LOGOTHETIS N K. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys [J]. *The Journal of Neuroscience*, 2008, 28 (17): 4 457-4 469.

Interaction of Face and Voice in Status Identification

ZHOU Ai-bao, ZHANG Rong-hua, MA Xiao-feng, ZHU Jing

(School of Psychology, Northwest Normal University, Lanzhou 730070, China)

Abstract: Status identification, an important issue of social cognition, can obtain abundant information from personal face and voice. The existant literature has mostly used the identification and recognition paradigm to investigate the integral process of voice and face. The vertical and horizontal mechanisms have been applied to look at the STS area of the brain. However, no consensus has been reached. The paper, thus, suggest that further indepth research be needed by innovating the paradigms and contrastive baselines.

Key words: face-voice interaction, status identity, cross-modal

(责任编辑 夏登武)