

# 社会心理语言学视域下

## 言者个体与群体身份的编码和解码

**摘要：**言语交流中，听者如何快速有效地感知言者的身份和个性是社会心理语言学的重要问题。关注言者间身份变异解码的传统研究发现听者区分言者间身份的正确率受听者音系知识及言者基频和声道长度的影响。新近研究发现，言者会因交际意图变化而调整发声策略（语言结构、语言风格和发声生理基础），听者能通过适应言者内部的变异进而识别言者身份。本文回顾了音系规则对身份编码的特殊制约，梳理了底层声学参数如何表征言者间及内部身份变异、进而影响言者身份感知；引入了内/外群体概念，探讨了言者在群体身份渗透意图下会采用不同发声策略这一现象如何支持交际调节理论；基于以上提出了言语互动场景下的言者身份编码及解码模型，并展望了三个研究方向。

**关键词：**言者身份；嗓音表情；交际意图；社会分组；社会心理语言学

## A Social Psycholinguistics Perspective:

## Encoding and Decoding Mechanisms for Speakers' Individual and Group Identities

Wenjun Chen,<sup>1</sup> Yanbin Hu<sup>1</sup>, Xiaoming Jiang,<sup>1\*</sup>

<sup>1</sup> *Institute of Linguistics, Shanghai International Studies University, Shanghai 201620, China*

**Abstract:** How listeners quickly and effectively perceive speakers' identity and personality in verbal communication remains a widely researched topic for social psycholinguistics. Traditional research focusing on the perception of between-speaker identity variation reported that the correct rate for between-speaker differentiation is subject to listeners' phonological knowledge and speakers' Fundamental Frequency (F0) and Vocal Tract Length (VTL). Recent research has found that speakers modulate their vocalisation strategies (language structure, language style and physiological basis of vocalisation) according to their changing communicative intentions, whereas listeners could adapt to within-speaker variations and recognise speakers' identities. This article reviews the unique constraints on speaker identity encoding imposed by phonological rules and unpacks how underlying acoustic parameters characterise within- and between- speaker identity variations that influence speaker identity perception. It further introduces the concept of in-/out-group and explores how the phenomenon where speakers would adopt varied vocalisation strategies when motivated by group identity permutation intentions support the Communication Accommodation Theory (CAT). Based on such, it proposes Speaker Identity Encoding and Decoding Model for Verbal Interaction Scenarios and calls for future research's attention in three directions.

**Keywords:** speaker identification; vocal expression; communication intention; social grouping; social psycholinguistics

\* The corresponding author is to be contacted at [xiaoming.jiang@shisu.edu.cn](mailto:xiaoming.jiang@shisu.edu.cn).

## 1. 引言

马看四蹄，人看四相。《红楼梦》中林黛玉能快速通过王熙凤的声音感知到王氏飞扬跋扈的性格和在贾府的显赫地位正是凭借了“音相”。言语交流中，人声不仅传递了语言信息，也包含了言者身份和情绪信息 (Belin et al. 2004)。听者不仅可以从声音中听出对方“是谁”，也能对其“是什么样的人”形成大致印象。人声和人脸一样承载了身份信息，也被称为“听觉人脸” (Schirmer 2018)。言者身份包括性别、年龄、体型等信息 (Campanella & Belin 2007)，其由以基频和声道长度为主的语音信号组合编码 (Frühholz & Schweinberger 2021; Lavan et al. 2019c)，而听者对其解码则主要依靠右侧颞上沟 (right anterior superior temporal sulcus, right STS) (Formisano et al. 2008)。言者身份信息和言语中的语言信息乃至表示语用目的的重音强调等共享基频这样的语音信号 (Frühholz & Schweinberger 2021; Tang et al. 2017)，即言者身份会随着发声的言语任务不断改变。然而大量的言者身份研究没有考虑言语交互场景下的言者身份编码和解码的动态性，即对社会互动维度的关注远少于对认知心理维度的关注 (束定芳 & 张立飞 2021)，故本文特别探讨了言语交际互动维度下言者身份和语言信息编码和解码之间的交互关系。

社会心理语言学语言观下，言语交际是一种有意识的言语活动，对其具体的语用模式和社会心理的言语机制研究需要整合社会语言学、心理语言学、和工程语言学等学科 (王德春 & 孙汝建 1992a, 1992b)。束定芳 (1992) 列举了都柏林的教师可以通过学生言语中的语言线索 (如，不标准的发音) 推断贫民学生家庭的社会地位，进而降低对学生的评价这一现象 (即，对言者群体身份的判断影响社会互动)；介绍了归因 (causal attribution) 以及集团特征 (group distinctiveness) 这两个概念，可用于解释 Giles 等 (1991) 的适应理论 (accommodation theory) 下的语言趋同现象 (convergence)；进而指出言语交际中编码过程的社会心理作用以及解码过程中的社会心理因素是社会心理语言学的主要研究对象。束定芳 (1992) 在展望中呼吁语言学界探究语言变化、语言结构、语言风格、以及集团语和社会心理之间的关系。因此有关言者身份的研究是一个社会心理语言学视角下的语言跨学科问题，需要综合来自心理语言学、社会语言学、言语交际科学、实验心理学、实验语用学、和认知神经科学交叉学科的证据进行探讨。对言者身份编码和解码研究将促进对语言和社会心理之间的关系、人工智能克隆语音的应用、以及语言学习和跨语种加工等问题的理解。因此，本文在语言跨学科的视角下探讨言者个体和群体身份在动态的言语互动中的编码和解码机制。

本文关注口语交际中语言规则 (音系结构和句法结构) 和语言风格与听者对言者身份的解码和后续社会互动方案决策的关系的一个关键动力是，经典的语言学理论没有考虑多模态互动场景下，例如口语交流时声学信息的可变性将如何影响听者对言者个体和群体身份的认知。例如，乔姆斯基认为语言使用者有一种先天能力，这使得他们可以产出和理解无限多的句子；这种能力使得他们作为听者并听到同一个句子时，即使这些听者又不同的背景和经历，他们也能够用相同的方式理解该句子 (Chomsky 2014)。言语行为理论任务，言者会采用话语以实现特定的言语行为 (如提出要求和下达命令)，听者会根据交际场景中的语境来理解言者的言语行为，此时，言者是谁并不影响听者对言者言语行为的理解 (Austin 1975)。格

莱斯认为言者字面意思以外还有暗示的更多内容，而听者会利用会话中的隐含意义来理解句子；这一理论表明，语境和非语言信息在理解话语方面起着关键作用，而没有强调言者个体内部身份 (Grice 1975)。

然而以上语言学理论是以语言本体结构（并未考虑多模态的互动）为中心的观点，即听者对话语的理解不会因为言者是谁而有所改变，大脑似乎不会因为言者身份不同而区别化地加工句子。然而，语音为载体沟通互动下的心理语言学和神经语言学实验却表明，听者对言者的话语的加工会受到言者个体间和个体内身份差异的影响。例如，社会范畴下的地位高低不同的双方在进行社会互动时，大脑神经活动指标会敏感于对敬语“您”和“你”的使用出现违反的条件 (Jiang et al. 2013)。句法结构范畴下，例如德语中 SOV 和 OSV 句法结构的对立研究也有类似的发现，即听者预期言者会讲简单的 SOV 句子，但听到言者讲难的 OSV 句子时实验者观察到 P600 (Kroczek & Gunter 2021)。语用范畴下，阅读实验中，个体是否常用讽刺的方式表达自己，也有类似的 P600 效应被观察到 (Regel et al. 2010)。由此，言者角度看，由音系结构和语言风格（侧重于语音层次的沟通）组成的语言规则会影响言者的言语产出，产出的差异具体体现在精密的声学分析上（例如，基频、声道长度参数、音强、时长、jitter、和 shimmer 等参数的组间差异上，此类的组可以是：自信、中性、怀疑的“知道感”驱使下的言语韵律差异 (Jiang & Pell 2017))。而听者会对言者产出的语音中的声学变化敏感。其中，基频和声道长度关键表征了言者个体内部身份和言者个体与其他言者个体之间的身份差异；由此，个体会敏感于声音中的言者身份变化。以上推论的一个证据来自于探索听者解码不同口音英语使用者产出的语音中的自信水平的脑电研究，结果发现在 ERP 早期阶段就能加工包括音系结构的语言结构和言者身份信息 (Jiang et al. 2020)。因此，本文重点关注了语音传递信息的背景下个体对言者产出的言语的理解，侧重于听者如何解读言者的身份进而影响听者的发声策略参与社会互动。

言者身份编码与解码的研究多关注个体在控制条件下（即用中性嗓音表情录制实验刺激）的发声所涉及的认知加工模式。基于人声线索的身份解码实验主要通过两个范式探究听者对不熟悉声音和熟悉声音的识别机制：说话人辨别 (AX discrimination)，即听者基于所听的两个句子判断陌生说话者是否为同一个人；再认 (speaker identification)，此范式源于司法实践中的嫌犯指认，即当事人听一个阵列的语音后调用记忆以指出嫌犯身份 (Frühholz & Belin 2018; Levi 2019)。国内有多篇相关综述：伍可等 (2020) 从感知的神经机制角度介绍了人声言语、情绪及身份加工所涉及的双通路模型、多阶段模型和整合模型；周爱保等 (2021) 区分了获得性和发展性人声失认症患者受损脑区的差异性；明莉莉和胡学平 (2021) 对比了正常视力者和盲人加工人类嗓音身份不同的脑机制。此外，陈忠敏 (2021) 按照听觉器官以上神经系统的解剖生理构造讨论了语音感知的特点以及所涉及的解剖生理机制，提示了言者身份编码的生理基础。

这些综述多为探讨基于同一群体成员个体间的人声身份编码与解码，而较少探讨基于社会群体之间的人声身份差异与加工机制的问题。然而，个体还会在中性嗓音表情以外更复杂地发声。个体发声会随着讲话风格、环境和社会场景（例如，模仿他人）、所处认知发展阶段、情绪、生理和心理状态等产生内部变异 (Lavan et al. 2019b)，即听者从声音中感知言者“是谁”和“什么样的人”的结果不是恒定的，会受到这些因素影响。

在语言结构和语言风格与发声生理基础共同影响发声策略的基础上，本文回顾了言者的个体和群体身份编码与解码相关文献，提出了一个整合的言语互动场景下的言者身份编码及解码模型，并基于此进行了研究展望。

## 2. 言者个体身份的生理基础和语言声学编码

和 Hauser、Chomsky、和 Fitch 乔姆斯基等（2022）提出的广义语言官能（FLB）类似，声音身份编码和解码是物种的普遍能力；也和他们提出的狭义语言官能（FLN）中的递归性类似，人类的言者身份解码和编码因为语言的进化更加复杂。言者在言语计划阶段确定语言结构和语言风格后，将通过调用发声生理基础执行特定的“发声策略”，于此产出携带声音身份的语音。其中，语言结构包括音系结构（可由口音或文体变异（stylistic variant）带来）和句法结构（例如实验设计下言者偏好使用 SOV 或 OSV 结构（Kroczek & Gunter 2021））；语言风格包括性别二元的说话风格、和特定语用选择（例如倾向于说讽刺的话(Regel et al. 2010)）等。

### 2.1 声音身份的编码与识别是普遍能力

生命体的体型大小是声音身份编码的关键要素，比如野狼在捕猎或产仔时会通过嚎叫向同伴昭示领地，对同类体型感知是许多物种社会组织中的普遍现象（Harrington and Mech 1979）。身份编码的后果与具有进化意义的繁殖后果密切相关。Reby 和 McComb（2003）分析了 24 头雄性红鹿的吼叫声、体重和繁殖成功率数据后发现基于共振峰间距计算的声道长度与红鹿体重正相关，且普通状态下吼叫对应的最大声道长度与繁殖成功率正相关。人类社会中也有类似的发现。Šebesta（2019）等邀请了 84 名来自巴西和 68 名来自捷克的异性恋参与者朗读短句和唱歌、并报告社会化性生活情况，发现女性短演讲中声道长度更短及歌唱中声道长度更长可以预测女性的性行为。由此可知，物种对同类身份的判断是一种普遍能力。

人类从声音中解码对方身份的能力先于语言交流出现：Polka 等（2022）合成了共振峰间距较大的婴儿（F2-F1=3761）和间距较小的成人女性（F2-F1=2315）元音/i/音频，发现平均年龄 220 天的婴儿对模拟婴儿发声状态的元音有偏好。而该能力与语言习得有密切关系：Fecher 和 Johnson（2019）等给平均年龄 136 天的大部分时间仅曾暴露在英语环境下的婴儿呈现了由 4 名双语者（2 名讲英语和波兰语；2 名讲英语和西班牙语）女性录制的英语、波兰语和西班牙语句子；以注视时间为因变量、言者（异/同）和语言（母语/非母语）为主要拟合参数的混合线性模型，结果显示言者和语言均无主效应但其间有交互作用，这表明刺激是否为婴儿母语会调节婴儿听同一言者或不同言者音频时的注视时间。

基于其他物种与早期语言习得的研究表明人类在进化过程中保留了从声音中识别对方的能力，且该能力早于语言交流出现，然而语言这一人类特有的现象使该能力较于其他物种更加复杂化。

### 2.2 言者个体身份编码与解码的特异性：语言音系规则的制约

听者能够在互动中出于交际功能的目的整合言者的身份信息（即，社会目的，social goal）以及话语中的内容信息，从而形成语言目的（linguistic goal）（Kuhl 2011），而正是听者对语言目的的解码使得人类言者识别不同于普通动物识别同类。Perrachione 等（2011）发现损伤了（由 Comprehensive Test of

Phonological Processing, 或 CTOPP 的 Phonological Memory 与 Phonological Awareness 两个子测试得分表征) 英语音系规则的、由临床诊断判断的阅读障碍群体在识别用母语英语和完全陌生的中文编码的言者身份时准确率相当, 且均远低于健康人群对照组。其中, 从音系学的角度看, 如果个体完全不知道一门语言, 那么个体讲较难准确地听辨出、并产出该语言的声音 (sounds) 和声音模式 (sound patterns), 由此, 个体缺乏对特定语言的音系规则的了解 (Goldsmith et al. 2014)。英语单语阅读障碍患者缺失了对中文音系规则的了解, 阅读障碍又使他们损伤了英语的音系规则, 这导致其对母语的音系规则知识仅处于陌生语言水平, 此损伤使他们在母语条件下言者身份识别的准确率和陌生语言水平的条件下相似。人类言者身份解码高度依赖听者的音系规则知识, 正是对母语有更多音系知识导致了“语言熟悉效应 (language familiarity effect)”: 就算是听倒放而无语义通达的句子 (Fleming et al. 2014; Goggin et al. 1991), 单语听者会在母语条件下更准确识别言者身份 (Perrachione and Wong 2007)。值得注意的是, Orena 等 (2015) 发现加拿大蒙特利尔的英语单语成年人比美国康涅狄格州英语单语者能更快更准确地学习和识别法语使用者的言者身份, 这表明被暴露于特定音系规则的使用场景也可助于语言熟悉效应的出现。

以上证据提示, 人类听者对言者话语中身份和语音信息的整合加工与普通动物识别同类可能具有加工机制上的差异。

### 2.3 言者个体身份编码与解码的特异性: 句法结构和语言风格的绑定

Kroczeck 和 Gunter (2021) 训练被试暴露在特定说话人讲 70% OSV 和 30% SOV 句子或相反的设计下, 被试由此建立了对特定言者是 OSV 言者或 OSV 言者的映像; 测试时被试听到 SOV 言者讲 OSV 句子时的 ERP 成分中发现了中后部的 P600--表征了句法重分析或修复。类似的实验操作下也发现听者能对言者的高/低句法依附模式 (syntactic attachment style) 有所预期 (Kamide 2012)。

更多研究也表明听者会将言者身份和特定的语言风格绑定。例如, Regel 等 (2010) 采用与以上类似设计让读者形成对两个人名使用讽刺的倾向的绑定和预期, 当读者读到非讽刺风格者的讽刺语用时, 也有一个类似的 P600 被发现。

值得注意的是, Walker 和 Perry (2022) 操纵了男女特异的韵律 (一名女性用习惯的韵律 vs 模仿男性韵律) 和语言风格 (男/女性化的用词场景), 被试听到该女性用男性韵律讲话时的 ERP 诱发了表征语义不一致或非预期的刺激的更大的 N400, ERP 结果也报告了韵律和语言风格的交互作用, 即与女性言者身份及女性韵律和女性言语风格一致条件时会比其他不一致条件时会产生更大的负波。该研究表明社会范畴 (男女二分的韵律) 和语言本体范畴 (性别特异的词汇) 会共同影响言语产出的结果, 而听者对两个范畴的加工会与听者对言者身份的解码联系起来。

由上可推知, 个体能在严格控制的实验室中学习他人句法结构和语言风格的使用倾向, 而这一能力的由来与个体长期在自然互动场景下对他人语言使用的特征和他人的身份进行组合的实践密切相关。

### 2.4 基于言者间变异参数的个体身份编码与解码

语言是人类物种高级进化的产物, 其使得言者身份编码相较于其他物种更复杂。但人声身份的表征仍然依赖其进化过程中朴素的更底层的发声基础。Winters 等 (2008) 招募了英语单语被试, 在熟悉阶段给他

们呈现了由英语-德语双语者录制的德语元-辅-元单词，即听者需要将德语单词背后的身份和对应的名字联系起来；经由 8 轮次的熟悉-再熟悉-再认，被试能够将德语音频后的 10 个身份和相应的名字关联；最后在测试与泛化阶段给他们呈现由双语者录制的另一套英语单词，并要求他们指出该音频背后的言者身份；结果发现听德语单词熟悉言者身份后，听者能远高于机会水平分辨出双语言者英语发声背后的身份。这说明言语中有独立于音系结构之外的声学线索稳定地编码了人声身份，即基频（Matsumoto et al. 1973; Xu et al. 2013）以及表征声道长度（vocal tract length）的共振峰间距（Ghazanfar and Rendall 2008; Johnson 2020），基频和声道长度以交互地方式影响音色（Timbre）并编码言者身份（von Kriegstein et al. 2006）。

关于发声生理基础，发声时气流自呼吸系统（肺部）经过气管传至发声系统（声带），而后经喉部到达构音系统（由口腔、咽腔和鼻腔组成的调音区，即声道），最终产出语音（Nakagawa et al. 1995）。首先，声带震动频率被表征为听觉可感的音高（pitch），即基频。基频是个体间的声门脉冲率（glottal-pulse rate），由于其声带构造差异造成的不同声学表征，尽管其与个体身材大小关系不紧密，但具有明显的性别差异：成人男性声带比女性的长约 60%且更宽更粗，导致男性声门脉冲普遍比女性低，因此男性基频一般比女性低一个八度（Titze 1989; Künzel 1989）。其次，共振峰之间的距离与声道长度有统计关系，共振峰间距越小，声道长度越长，且声道长度与个体体型有直接关系（Lee et al. 1999; Johnson 2020），个体出生时声道长度约为 8 厘米，成年人的声道长度从 13 到 20 厘米不等（Lammert and Narayanan 2015）。同时，基频和声道长度交互影响言者身份编码。声门脉冲率较低时，从同等长度的声道发出的声音会由此听起来体型更大；如果声门脉冲率较低，则听起来体型更小（Smith and Patterson 2005）。

除了基于共振峰计算的基频和声道长度这样的频谱信息，还有更多参数也表征了言者身份差异。Chen 等（2022）对男女各半的 300 名来自白人、黑人、亚洲人和拉美裔群体的言者发出的 10 秒长的“Aaaaah”就以下参数进行了分析：反映时间信息的均方根能量（RMS energy），反映频谱信息的频谱质心（spectral centroid）与声谱衰减（spectral roll-off），反映频谱包络的形状梅尔频率倒谱系数（MFCCs），以及反映信号中信息量的熵相关值-谱熵（spectral entropy）、概率密度函数（PDF entropy）、排列熵（permutation entropy）、以及奇异值分解（SVD entropy）。结果发现种族间在以上参数上的差异不显著，然而男女（不区分种族）间在多个指标上有显著差异。

总的来说，生理结构导致的个体间基频和声道长度参数差异主要表征了言者间身份，且更广泛的时间、频谱、倒谱、及熵相关参数也可以反映性别群体身份。

## 2.5 基于言者内变异参数的个体身份编码与解码

人声身份识别领域多采用言者中性嗓音下的发声作为刺激以探索听者对声音身份的识别机制（Perrachione et al. 2011; Fleming et al. 2014）。然而所使用的语种、场景下的副语言信息表达的要求（比如，言者的自信和怀疑时的情态）使得言者个体内部有较大变异。例如，Voigt 等（2016）分析了 25 名德法、20 名德意双语者在轻松话题采访时的录音，发现德法双语女性说法语时使用的基频平均值更高，德意双语女性讲意大利语基频更低。在针对展现言者不同水平“知道感”的陈述语的分析中发现，当观察句首和句中成分时，自信的基频相较于不自信高，然而观察句尾成分时不自信的基频却更高（Jiang and Pell

2017), 这对言者身份仅通过表征言者间变异的声学参数进行编码的观点形成了挑战, 即听者解码言者身份时依赖的线索更复杂。

研究发现, 听者具有适应言者身份内部变异的能力。不同言者有基于不同的原型的声音身份, 这些身份分布在多维声音空间中 (Latinus and Belin 2011)。基于此, Lavan 等 (2019c) 通过调整半音的方式操纵了以基频和声道长度为二维空间上的身份, 通过在以声道长度为横轴、基频为纵轴的空间中, 向上/下移动 1.6 个半音的基频和向左/右移动 2.36 个半音的声道长度建立了 4 个独立于源声的声音身份 (左下角的声音因不自然而被排除)。围绕 3 个新声音原型的中点, 在声道长度左右 2.25 半音、和基频 3.6 半音范围内各自新建了距离原型较近的 16 个内周点、以及较远的 18 个外周点; 内/外周点反映了 3 个声音的内部变异。听者在训练阶段仅听过外周点上的声音, 然而他们在测试阶段判断声音“新/旧 (old/new)”时却报告曾听到过外周点的声音。这表明听者对言者声音身份有原型的认识并可以适应言者身份的内部变异。

然而, 听者适应言者身份内部变异的能力是有限的。Lavan 等 (2019a) 将来自美剧《绝命毒师》的两位男主角的 1.2 秒到 4 秒长的情绪发声刺峰值标准化至 0.400 帕并在 10kHz 处进行低通滤波后, 要求听者对经合成操作后音频背后的言者身份进行拖拽分类; 结果发现, 即使是熟悉该剧集 of 的听者仍然很难将音频准确地归类为两位言者, 而是归类为更多位言者。Xu 和 Armony (2021) 准备了 12 名言者分别产出的 4 个句子 (2 种情绪韵律: 害怕/中性 \* 2 种语义内容), 在听者熟悉阶段给他们呈现了其中 6 名言者的中性/害怕韵律下的各 1 个句子, 而后在测试阶段播放所有 12 名录音者的 48 个句子并要求听者判断该声音身份是否出现过, 结果发现听相同内容的句子时, 如果韵律相同, 被试准确率普遍高于 80%, 然而若韵律不匹配时其准确率仅在机会水平附近。因此, 听者对言者内部变异的适应仅限于特定阈限内。

以上表明言者身份会随使用的语言或具体情态而产生内部变异, 而听者能够在一定阈限内适应这样的变异, 将声音身份有所改变的言者身份归一化地识别为同一个人。

### 3. 言者群体身份渗透意图调节发声策略

社会心理学从原型理论的角度解释了社会群体的划分, 即态度、行为、习俗等个体相关属性的模糊集合会在交流个体的心智表征中形成一个原型化的人类群体观念, 该原型所代表的属性可最大限度地提高群体的实体性进而导致刻板印象的产生; 其中, 语言和言语风格是个体所处群体的身份象征之一 (Hogg 2016), 人们据此划分社会互动对象为内/外群体成员 (Jiang et al. 2020); 且社会群体划分具有可渗透性, 即成员可以改变其身份表征 (Hogg 2016)。下文回顾言者群体身份如何被编码及个体如何通过调节发声进行群体渗透, 而后提出一个基于群体互动视角的言者身份表征解码的整合性框架。

#### 3.1 言者群体身份解码

言者所用语言是听者划分内/外群体的标准之一。肯尼亚人认为使用斯瓦希里语和吉里阿马语不同地定义了语言使用者的自我、权利、权益、及宗教 (Kinzler 2021)。有关婴儿的实证研究为母语者的内群体偏好提供了印证。例如, 12 个月大的婴儿更倾向于取用母语内群体者递过的食物 (Shutts et al. 2009); 10 个

月大的英语母语的婴儿会优先选择英语内群体讲解员展示过的玩具，且 2.5 岁左右的英/法单语儿童也更多把物品交给内群体母语者以进行游戏互动 (Kinzler et al. 2012)。Begus 等 (2016) 给 11 个月左右的婴儿呈现了其母语内群体 (英语) 和外群体 (西班牙语) 女性言者指着婴儿不熟悉物品进行名词教学的视频，观察了婴儿的 3-5Hz  $\theta$  频段活动 ( $\theta$  频段的神经振荡通常用以表征信息处理和学习，成年人的  $\theta$  频段为 4-8Hz)，发现婴儿在内群体言者条件下有更强烈活动的  $\theta$  振荡。这都表明听者对特定语种的语言结构敏感并会由此判断言者的群体身份，进而差异化地进行社会互动。

口音规则是语言的一部分，言者的口音被听者用于群体划分。Rubin (1992) 给北美大学生播放了标准南方美音的授课音频并同时匹配了亚洲或白人面孔图片，发现当标准南方美音和亚洲面孔建立联系时，学生认为该言者有更重的非标准口音、更差的教学资质、讲课内容更难懂。Jiang 等 (2020) 招募了 44 名有相当法语能力且对澳大利亚英语口语了解的来自加拿大魁北克省的英语母语听者，被试在听了由加拿大英语口语、澳大利亚英语口语、和带魁北克法语口音者录制的，带有自信或怀疑嗓音表情的音频后，进行了可信度评分，发现在自信条件下，加拿大英语口语和澳大利亚英语口语相比法语口音的音频听上去更可信；然而当句子带有怀疑嗓音时，加拿大英语口语相比澳大利亚英语或法语口音者听起来更不可信。Tamagawa 等 (2011) 提供了真人言者以外的证据：使用新西兰口音的被试听了基于英国、美国、和新西兰口音训练的，介绍同一款血压计的机器人的合成语音后，认为美国口音相比新西兰口音的合成声更像机器声，且比带有新西兰合成口音的机器人性能更差。此实验中机器人的口属于语言结构中对不同的音系结构的表现。以上三项实验表明，言者对音系结构的选择关键地编码了他们被听者感知到的群体身份；而听者将基于不同言者身份的感知会不同程度地改变互动决策。

基于性别二分态的言语风格也标志群体身份。男性言语风格的典型特征是使用俚语 (粗俗语)、讲话更加直率、声音低沉、咄咄逼人、和显得更有权威，而女性的特征是速率和基频变化幅度大、讲话更加温和、开放、自我揭露、和显得更加情绪化 (Giles et al. 1983; Hogg 1985)。Martin 等 (2021) 提出的“大二” (big two) 模型认为个体对他人特质判断和评价遵循了两个与性别角色有很大重叠的维度：a. 代理/男性气质，即自信、竞争、支配、独立、自我利益、目标追求；b. 社区/女性气质，即养育、温暖、表达、关注他人、社会导向。其中男性气质与他人感知到的“能力”密切相关，例如，男性化的面部特征使个体看起来更有能力 (Oh et al. 2019)，声音更低的、即更加男性化的女性被认为更具主导性，而女性化的声音则被与幼稚及性不成熟联系起来 (Borkowska and Pawlowski 2011)。因此，听者基于男女言语风格差异区分其群体身份并将此划分结果与特定刻板映像联系起来。

言语交际中所使用的具体语言种类、所展现的口音、以及男性化或女性化的程度都编码了言者的群体身份，听者基于此将互动对方识别为内群体或外群体成员并差异化地调整互动方案。值得注意的是，大量现有文献已经关注了听者对言者的口音感知如何影响社会互动，但大多是基于真人之间的言语沟通，少有研究关注到近年新出现的人工智能克隆人声与语音合成技术。

### 3.2 言者身份编码的内/外群体渗透机制

言者的群体身份可以通过其基于意图调整发声策略而得到改变。一方面，交际调节理论



(communication accommodation theory, CAT) 提出言者会在交际过程中为了缩短交际双方的社会距离、促进相互理解并提高交际效率而在口音、语速、音量、停顿与语言内容使用上与对方趋同 (Coupland et al. 1988; Bernhold and Giles 2020), 这样的韵律层次的趋同现象在汉语自然会话场景下也被发现 (夏志华 & 马秋武 2019)。典型的社会范畴下的趋同的一个例子是高低地位等级的双方互相采用对方管用的言语模式 (束定芳 1992)。另一个案例是成人会使用儿向语和婴儿交流, 具体特征有基频变化更丰富 (Stern et al. 1982)、基频更高和末音节延长 (Albin and Echols 1996)、重复更多 (Hills 2013)、语篇更短 (Soderstrom et al. 2008)。Sorokowski 等 (2019) 录制了 27 名男性和 24 名女性在大学工作的科学家谈论日常话题 (问路) 和权威性话题 “如何成为科学家且这样值得吗” 的音频, 发现男女言者在提供专业建议时基频都降低了, 且女性 (33Hz) 比男性 (14Hz) 做出了更多降低的努力。Harrington 等 (2000) 调查英女王伊丽莎白二世在上世纪 50 到 80 年代讲话音频中的元音后, 发现其发声方式有向年轻群体和平民靠近的趋势。以上案例说明言者为了显得亲切或更专业会策略性地改变对语言风格的选择, 而这样的改变不仅体现在对发声基础的调控努力程度上, 也体现在对文体变异 (stylistic variant) 方式的计算和执行。

另一方面, Pisanski 等 (2021) 指出人类有声语言的元音复杂性 (vocalic complexity) 可能起源于在动物界中这样一个普遍现象: 物种会降低声道共鸣 (即降低共振峰) 以达到声音体型夸张, 这一现象存在于具有复杂语言系统的人类中, 也常见于不具有类似人类语言系统的其他动物群体中。人类言语场景中, 言者咄咄逼人时声道长度相较于中性发声更长 (Pisanski et al. 2022), 言者开心时发声的声道长度比愤怒、悲伤或中性时更短 (Kim et al. 2020), 且言者产出自信相比不自信录音时的基频更低 (Jiang and Pell 2017), 这样的声学参数暗示了言者编码自己身份时拉长/缩短声道的、调控生理基础的努力; 而这些发声策略的改变将直接影响听者感知到的言者身份。也就是说, 人类在进化出语言系统后仍然保留了通过拉长/缩短声道、提高/降低基频改变声音体型的能力, 并在具体语用场景中潜意识进行了声音体型改变 (如, 咄咄逼人时言者使自己听起来体型更大)。由上可知, 更短时的副语言信息传递可能和人类长时的语言结构、语言风格的行成有关, 类似于广义语言官能 (FLB) 和狭义语言官能 (FLN) 之间的关系。

本节表明言者会基于特定交际目的, 遵循特定的语言规则改变自己的言语发声方式进而使自己听起来属于特定群体, 这一机制可能起源于动物界中普遍存在的体型夸张化现象因而具有进化意义。然而, 对此类发声调控的精细观察以及语言规则是否具有跨文化间一致性 (考虑到语种多样性) 仍然有待回答。同时, 对语言规则习得或使用异常群体如何编码和解码言者身份的研究将有助于理解声音中身份、语言、副语言信息间的关系。

### 3.3 言语社会互动场景下的言者身份编码和解码的理论框架

以上综述表明, 言者会在交流意图下调整发声策略以影响听者得到的 “言者是谁以及是什么样的人” 的印象。互动场景下, 听者接下话轮时将基于所整合的关于言者的身份和语言的信息调整言语产出。因此, 本文整合了 (1) Belin 等 (2004) 提出的人声加工模型指出人脑中的三条神经通路会在识别声音为人声后被分别激活, 用以分别精细化地加工言语、情绪和身份信息; (2) Cummings (2015) 从听者言者话轮循环角度提出的人类言语沟通循环; (3) Jiang 等 (2020) 从声音身份和情绪加工时间进程角度提出的噪

音表情的认知处理模型，其中听者在 ERP 的早期阶段会加工言者语音中的嗓音信息结构（vocal structure），包括嗓音身份信息 and 嗓音言语信息（包括了语言结构）；最终提出了社会互动场景下言者身份编码和解码框架（图 1）。

框架中两个基础元素是：（1）情绪信息有惊讶、高兴、生气、悲伤、害怕、厌恶、中性这样的基础情绪，也有表现“知道感”的自信嗓音信号（蒋晓鸣 2020）；（2）身份信息由性别、年龄、受教育程度、吸引力、能力、群族等变量表征（Frühholz and Belin 2018）。

话轮伊始，言者在渗透到对方群体身份的交际意图的驱动下选择听者特有的音系和句法结构完成语言编码（如，是否采用典型男性化的言语风格）；在言语运动编码阶段，言者基于语言结构信息（如，改变对[r]音的卷舌程度（Labov 2006））和副语言信息的表达规则（如，是否延长声道长度、降低基频以显得自己更自信（Jiang and Pell 2017））完成如何调用舌头、嘴唇和声带等发声基础的计划；在言语运动执行阶段，言者脑中发出神经信号控制发声的解剖基础完成言语发声，传出声波。

听者的听觉系统将机械波震动转为神经信号传递至听觉中枢，完成听觉信息接受。在言语感知阶段，听者在 100 毫秒左右进行了嗓音结构分析，同步加工了嗓音身份信息、情绪信息以及话语内容信息（理解表示句子功能的句法信息结构）；在 200 毫秒左右，听者进行了嗓音重要性侦测（比较言者语气、与听者自身的口音相似度等）以决定注意力分配的多少；在 250 毫秒往后，听者进入了语言理解阶段，对模糊的语义再确认/消歧、基于身份信息进行语用推理、并将身份信息整合到语境中。而后言者接过话轮并基于意图策略性地发声，话轮周而复始。

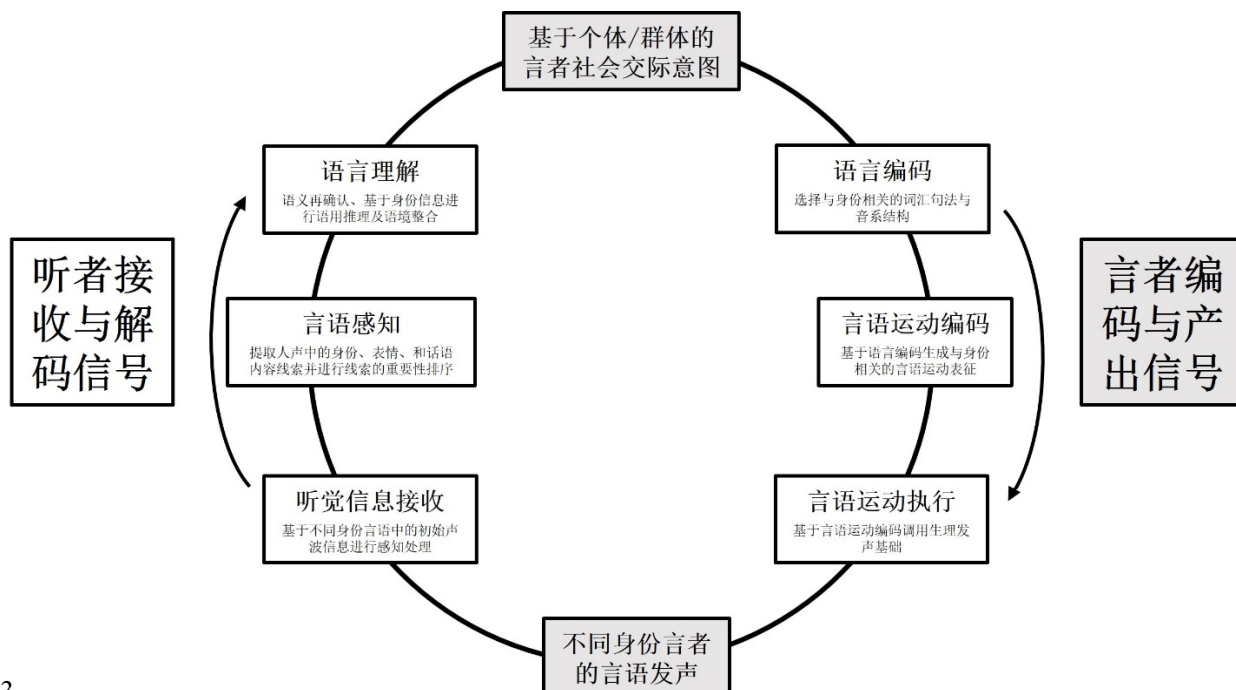


图 1. 言语互动场景下的言者身份编码及解码模型

## 4. 研究展望

基于学界有关言者个体和群体身份编码和解码的成果，未来研究可以探讨（1）什么指标可以表征言者调整发声运动以达到群体渗透目的时的努力；（2）听者对人工智能克隆人声加工的脑机制及情态韵律线索对跨群体发声解码的调节作用；（3）语言规则缺损者编码和解码言者身份的内部变异。

第一，个体如何基于语言规则调控发声策略尚需探讨。社会规范与行为习惯是个体通过言语的和非言语的社会交互习得的，其具有跨文化的共性，因此互动参与者可以跨文化范畴地理解言者的文化符号（蒋晓鸣 2020）。那么，特定语用功能指导下的言语产出，比如自信、支配、顺从等社会态度的发声，是否可以通过对言者音频的声学参数和成像技术对发声模态进行跨文化的比较？即，是否可以观察到自信发声拉长声道而怀疑缩减声道长度，而这样的模式是否有文化间一致性？可能的研究手段有声学参数分析、声门生理运动测量技术、和磁共振成像技术。

第二，人类对克隆人声的认知加工机制亟待探索。首先，技术已经可以基于几秒长的音频克隆真人的声音模型，并以此伪造个体的言者身份（Jia et al. 2018），那么听者在感知真人声源和其对应的克隆声时，将如何定义克隆声的内/外群体身份并进行社会互动决策呢？例如，Pernet 等（2015）发现颞叶声音区有三个斑块选择性地对人声敏感，且 Zhang 等（2021）通过皮层脑电 ECoG 发现癫痫病人左侧前颞叶的特定电极点仅对母语人声有反应，更值得注意的是，Di Cesare 等（2022）给听者呈现了传递社会意图的词汇“你好”后，真人声较于中性噪音的合成声特异地激活了听者的背侧-中央脑岛区域。那么 是否有特定神经相关物表征个体不同地加工真人和其对应的克隆声呢？其次，语音合成技术可以产出富含哭、笑、打哈欠等表达性发声的音频（Kharitonov et al. 2022），甚至让两个合成声模型进行自发而实时的、带有自然的重叠和停顿的闲聊（Kreuk et al. 2021）；如果以上技术和克隆人声相结合，即克隆声变得更加“拟人化”时，听者是否仍会对克隆声的群体归类会受到改变吗？将来的研究可结合电生理和磁共振成像从时间进程和空间维度上探索听者对不同条件下言者身份的差异化感知。同时，经典的语音适应范式（adaptation paradigm）的理论基础是听者特定神经元反应会随暴露在同一类刺激下的时间增加而减弱，若刺激特征改变，神经元反应将更加强烈，因此该范式可被用于进一步探索语音韵律如何调节人类听众对真人和克隆声的认知（Belin and Zatorre 2003; Grill-Spector et al. 2006）。再次，声纹识别技术将如何应对克隆人声带来的言者身份危机呢？正如升级声道长度归一化算法可提升人声身份识别产品的准确率（Tan 2021），但尚不清晰前文提及的更广泛的频谱、倒谱等参数在改变识别准确率的作用。

第三，语言规则缺损者的言者产出和听者感知机制也需要研究。首先，跨性别者为了使发声达到渗透至与生理性别对立的性别群体，需要通过（辅以视觉的）口腔共鸣语音治疗习得对应规则或接受环甲切除术（Neumann and Welzel 2004; Hardy et al. 2016; Dahl and Mahler 2020），这些干预措施会影响语音规则的表征与应用，这将如何启示对性别认同产生危机群体的矫正方案实施？其次，在社会交流/互动方面存在持续性障碍的自闭症谱系人群的“社交脑”存在异常，这些病人通常表现出语用规则运用的障碍，这与他们的额下回（inferior frontal gyrus, IFG）、颞上回（superior temporal gyrus, STG）和杏仁核（amygdala）过度激

348 活密切相关 (Peng et al. 2020)。然而颞上回 (STG) 参与了陌生嗓音身份加工, 额下回 (IFG) 通过与前  
349 颞上沟 (anterior superior temporal sulcus, anterior STS) 的功能连接参与了熟悉嗓音身份加工 (伍可 et al.  
350 2020), 且双侧中、后部颞上沟 (posterior STS/ superior STS) 表征了个体对声音中情感韵律信息的解码  
351 (Leipold et al. 2022); 由此, 自闭症谱系群体在整合真人和克隆言者身份与典型被试相比存在差异? 引入  
352 韵律信息这一言者内发声变量后将如何调节行为结果和相对应的神经相关物表征? 规则缺损导致的行为后  
353 果差异将进一步检验语言规则在言者身份编解码中的因果机制。  
354

355

## 356 参考文献

- [1] Albin, D.D., Echols, C.H. Stressed and word-final syllables in infant-directed speech [J]. *Infant Behavior and Development*, 1996(4): 401-418.
- [2] Austin, J.L. *How to do things with words*: Oxford university press, 1975.
- [3] Begus, K., Gliga, T., Southgate, V. Infants' preferences for native speakers are associated with an expectation of information [J]. *Proceedings of the National Academy of Sciences*, 2016(44): 12397-12402.
- [4] Belin, P., Fecteau, S., Bedard, C. Thinking the voice: Neural correlates of voice perception [J]. *Trends in Cognitive Sciences*, 2004(3): 129-135.
- [5] Belin, P., Zatorre, R.J. Adaptation to speaker's voice in right anterior temporal lobe [J]. *Neuroreport*, 2003(16): 2105-2109.
- [6] Bernhold, Q.S., Giles, H. Vocal accommodation and mimicry [J]. *Journal of Nonverbal Behavior*, 2020(1): 41-62.
- [7] Borkowska, B., Pawlowski, B. Female voice frequency in the context of dominance and attractiveness perception [J]. *Animal Behaviour*, 2011(1): 55-59.
- [8] Braber, N., Cummings, L., Morrish, L. *Exploring language and linguistics* [M]. Cambridge: Cambridge University Press, 2015.
- [9] Campanella, S., Belin, P. Integrating face and voice in person perception [J]. *Trends in cognitive sciences*, 2007(12): 535-543.
- [10] Chen, X., Li, Z., Setlur, S., et al. Exploring racial and gender disparities in voice biometrics [J]. *Scientific Reports*, 2022(1): 1-12.
- [11] Chomsky, N. *Aspects of the theory of syntax*: MIT press, 2014.
- [12] Coupland, N., Coupland, J., Giles, H., et al. Accommodating the elderly: Invoking and extending a theory1 [J]. *Language in Society*, 1988(1): 1-41.
- [13] Dahl, K.L., Mahler, L.A. Acoustic features of transfeminine voices and perceptions of voice femininity [J]. *Journal of Voice*, 2020(6): 961-e919.
- [14] Di Cesare, G., Cuccio, V., Marchi, M., et al. Communicative and affective components in processing auditory vitality forms: An fmri study [J]. *Cerebral Cortex*, 2022(5): 909-918.
- [15] Fecher, N., Johnson, E.K. By 4.5 months, linguistic experience already affects infants' talker processing abilities [J]. *Child Development*, 2019(5): 1535-1543.
- [16] Fleming, D., Giordano, B.L., Caldara, R., et al. A language-familiarity effect for speaker discrimination without comprehension [J]. *Proceedings of the National Academy of Sciences*, 2014(38): 13795-13798.
- [17] Formisano, E., De Martino, F., Bonte, M., et al. "Who" is saying "what"? Brain-based decoding of human voice and speech [J]. *Science*, 2008(5903): 970-973.
- [18] Frühholz, S., Belin, P. *The oxford handbook of voice perception* [M]. Oxford: Oxford University Press, 2018.
- [19] Frühholz, S., Schweinberger, S.R. Nonverbal auditory communication—evidence for integrated neural systems for voice signal production and perception [J]. *Progress in Neurobiology*, 2021: 101948.
- [20] Ghazanfar, A.A., Rendall, D. Evolution of human vocal production [J]. *Current Biology*, 2008(11): R457-R460.
- [21] Giles, H., Coupland, N., Coupland, I. 1. Accommodation theory: Communication, context, and [J]. *Contexts of accommodation: Developments in applied sociolinguistics*, 1991.
- [22] Giles, H., Scholes, J., Young, L. Stereotypes of male and female speech: A british study [J]. *Central States Speech Journal* 1983(4).
- [23] Goggin, J.P., Thompson, C.P., Strube, G., et al. The role of language familiarity in voice identification [J]. *Memory Cognition*, 1991(5): 448-458.
- [24] Goldsmith, J.A., Riggle, J., Alan, C.L. *The handbook of phonological theory*: John Wiley & Sons, 2014.
- [25] Grice, H.P. Logic and conversation *Speech acts*. Brill: 41-58, 1975.

- [26] Grill-Spector, K., Henson, R., Martin, A. Repetition and the brain: Neural models of stimulus-specific effects [J]. *Trends in Cognitive Sciences*, 2006(1): 14-23.
- [27] Hardy, T.L.D., Boliek, C.A., Wells, K., et al. Pretreatment acoustic predictors of gender, femininity, and naturalness ratings in individuals with male-to-female gender identity [J]. *American Journal of Speech-Language Pathology*, 2016(2): 125-137.
- [28] Harrington, F.H., Mech, L.D. Wolf howling and its role in territory maintenance [J]. *Behaviour*, 1979(3-4): 207-249.
- [29] Harrington, J., Palethorpe, S., Watson, C.I. Does the queen speak the queen's english? [J]. *Nature*, 2000(6815): 927-928.
- [30] Hauser, M.D., Chomsky, N., Fitch, W.T. The faculty of language: What is it, who has it, and how did it evolve? [J]. *science*, 2002(5598): 1569-1579.
- [31] Hills, T. The company that words keep: Comparing the statistical structure of child-versus adult-directed language [J]. *Journal of Child Language*, 2013(3): 586-604.
- [32] Hogg, M.A. Masculine and feminine speech in dyads and groups: A study of speech style and gender salience [J]. *Journal of Language and Social Psychology*, 1985(2): 99-112.
- [33] Hogg, M.A. Social identity theory [M]. Shelley McKeown, R. H., Neil Ferguson. *Understanding peace and conflict through social identity theory: Contemporary global perspectives*. Switzerland: Springer: 3-17, 2016.
- [34] Jia, Y., Zhang, Y., Weiss, R., et al. Transfer learning from speaker verification to multispeaker text-to-speech synthesis [J]. *Advances in Neural Information Processing Systems*, 2018.
- [35] Jiang, X., Gossack-Keenan, K., Pell, M.D. To believe or not to believe? How voice and accent information in speech alter listener impressions of trust [J]. *Quarterly Journal of Experimental Psychology*, 2020(1): 55-79.
- [36] Jiang, X., Li, Y., Zhou, X. Is it over-respectful or disrespectful? Differential patterns of brain activity in perceiving pragmatic violation of social status information during utterance comprehension [J]. *Neuropsychologia*, 2013(11): 2210-2223.
- [37] Jiang, X., Pell, M.D. The sound of confidence and doubt [J]. *Speech Communication*, 2017: 106-126.
- [38] Johnson, K. The  $\delta f$  method of vocal tract length normalization for vowels [J]. *Laboratory Phonology*, 2020(1).
- [39] Kamide, Y. Learning individual talkers' structural preferences [J]. *Cognition*, 2012(1): 66-71.
- [40] Kharitonov, E., Copet, J., Lakhota, K., et al. Textless-lib: A library for textless spoken language processing [J]. *arXiv preprint arXiv:2202.07359*, 2022.
- [41] Kim, J., Toutios, A., Lee, S., et al. Vocal tract shaping of emotional speech [J]. *Computer Speech Language*, 2020: 101100.
- [42] Kinzler, K.D. Language as a social cue [J]. *Annual Review of Psychology*, 2021: 241-264.
- [43] Kinzler, K.D., Dupoux, E., Spelke, E.S. 'Native' objects and collaborators: Infants' object choices and acts of giving reflect favor for native over foreign speakers [J]. *Journal of Cognition Development*, 2012(1): 67-81.
- [44] Kreuk, F., Polyak, A., Copet, J., et al. Textless speech emotion conversion using decomposed and discrete representations [J]. *arXiv preprint arXiv:2111.07402*, 2021.
- [45] Kroczeck, L.O.H., Gunter, T.C. The time course of speaker-specific language processing [J]. *Cortex*, 2021: 311-321.
- [46] Kuhl, P.K. Who's talking? [J]. *Science*, 2011(6042): 529-530.
- [47] Künzel, H.J. How well does average fundamental frequency correlate with speaker height and weight? [J]. *Phonetica*, 1989(1-3): 117-125.
- [48] Labov, W. *The social stratification of english in new york city* [M]. Cambridge: Cambridge University Press, 2006.
- [49] Lammert, A.C., Narayanan, S.S. On short-time estimation of vocal tract length from formant frequencies [J]. *PloS one*, 2015(7): e0132193.
- [50] Latinus, M., Belin, P. Anti-voice adaptation suggests prototype-based coding of voice identity [J]. *Frontiers in Psychology*, 2011: 175.
- [51] Lavan, N., Burston, L.F., Ladwa, P., et al. Breaking voice identity perception: Expressive

- voices are more confusable for listeners [J]. *Quarterly Journal of Experimental Psychology*, 2019a(9): 2240-2248.
- [52] Lavan, N., Burton, A.M., Scott, S.K., et al. Flexible voices: Identity perception from variable vocal signals [J]. *Psychonomic Bulletin Review*, 2019b(1): 90-102.
- [53] Lavan, N., Knight, S., McGettigan, C. Listeners form average-based representations of individual voice identities [J]. *Nature communications*, 2019c(1): 1-9.
- [54] Lee, S., Potamianos, A., Narayanan, S. Acoustics of children's speech: Developmental changes of temporal and spectral parameters [J]. *The Journal of the Acoustical Society of America*, 1999(3): 1455-1468.
- [55] Leipold, S., Abrams, D.A., Karraker, S., et al. Neural decoding of emotional prosody in voice-sensitive auditory cortex predicts social communication abilities in children [J]. *Cerebral Cortex*, 2022.
- [56] Levi, S. Methodological considerations for interpreting the language familiarity effect in talker processing [J]. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2019(2): e1483.
- [57] Martin, A.E., Slepian, M.L. The primacy of gender: Gendered cognition underlies the big two dimensions of social cognition [J]. *Perspectives on Psychological Science*, 2021(6): 1143-1158.
- [58] Matsumoto, H., Hiki, S., Sone, T., et al. Multidimensional representation of personal quality of vowels and its acoustical correlates [J]. *IEEE Transactions on Audio Electroacoustics*, 1973(5): 428-436.
- [59] Nakagawa, S., Shikano, K., Tohkura, Y.i. *Speech, hearing and neural network models* [M]. Amsterdam: IOS Press, 1995.
- [60] Neumann, K., Welzel, C. The importance of the voice in male-to-female transsexualism [J]. *Journal of Voice*, 2004(1): 153-167.
- [61] Oh, D., Buck, E.A., Todorov, A. Revealing hidden gender biases in competence impressions of faces [J]. *Psychological Science*, 2019(1): 65-79.
- [62] Orena, A.J., Theodore, R.M., Polka, L. Language exposure facilitates talker learning prior to language comprehension, even in adults [J]. *Cognition*, 2015: 36-40.
- [63] Peng, Z., Chen, J., Jin, L., et al. Social brain dysfunctionality in individuals with autism spectrum disorder and their first-degree relatives: An activation likelihood estimation meta-analysis [J]. *Psychiatry Research: Neuroimaging*, 2020: 111063.
- [64] Pernet, C.R., McAleer, P., Latinus, M., et al. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices [J]. *Neuroimage*, 2015: 164-174.
- [65] Perrachione, T.K., Del Tufo, S.N., Gabrieli, J.D. Human voice recognition depends on language ability [J]. *Science*, 2011(6042): 595-595.
- [66] Perrachione, T.K., Wong, P.C. Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex [J]. *Neuropsychologia*, 2007(8): 1899-1910.
- [67] Pisanski, K., Anikin, A., Reby, D. Static and dynamic formant scaling conveys body size and aggression [J]. *Royal Society Open Science*, 2021(1): 211496.
- [68] Pisanski, K., Anikin, A., Reby, D. Vocal size exaggeration may have contributed to the origins of vocalic complexity [J]. *Philosophical Transactions of the Royal Society B*, 2022(1841): 20200401.
- [69] Polka, L., Masapollo, M., Ménard, L. Setting the stage for speech production: Infants prefer listening to speech sounds with infant vocal resonances [J]. *Journal of Speech, Language*, 2022(1): 109-120.
- [70] Reby, D., McComb, K. Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags [J]. *Animal Behaviour*, 2003(3): 519-530.
- [71] Regel, S., Coulson, S., Gunter, T.C. The communicative style of a speaker can affect language comprehension? Erp evidence from the comprehension of irony [J]. *Brain research*, 2010: 121-135.
- [72] Rubin, D.L. Nonlanguage factors affecting undergraduates' judgments of nonnative english-speaking teaching assistants [J]. *Research in Higher Education*, 1992(4): 511-531.

- [73] Schirmer, A. Is the voice an auditory face? An ale meta-analysis comparing vocal and facial emotion processing [J]. *Social Cognitive Affective Neuroscience*, 2018(1): 1-13.
- [74] Šebesta, P., Mendes, F.D.C., Pereira, K.J. Vocal parameters of speech and singing covary and are related to vocal attractiveness, body measures, and sociosexuality: A cross-cultural study [J]. *Frontiers in Psychology*, 2019: 2029.
- [75] Shutts, K., Kinzler, K.D., McKee, C.B., et al. Social information guides infants' selection of foods [J]. *Journal of Cognition Development*, 2009(1-2): 1-17.
- [76] Smith, D.R.R., Patterson, R.D. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age [J]. *The Journal of the Acoustical Society of America*, 2005(5): 3177-3186.
- [77] Soderstrom, M., Blossom, M., Foygel, R., et al. Acoustical cues and grammatical units in speech to two preverbal infants [J]. *Journal of Child Language*, 2008(4): 869-902.
- [78] Sorokowski, P., Puts, D., Johnson, J., et al. Voice of authority: Professionals lower their vocal frequencies when giving expert advice [J]. *Journal of Nonverbal Behavior*, 2019(2): 257-269.
- [79] Stern, D.N., Spieker, S., MacKain, K. Intonation contours as signals in maternal speech to prelinguistic infants [J]. *Developmental Psychology*, 1982(5): 727.
- [80] Tamagawa, R., Watson, C.I., Kuo, I.H., et al. The effects of synthesized voice accents on user perceptions of robots [J]. *International Journal of Social Robotics*, 2011(3): 253-262.
- [81] Tan, Z.-H. Vocal tract length perturbation for text-dependent speaker verification with autoregressive prediction coding [J]. *IEEE Signal Processing Letters*, 2021: 364-368.
- [82] Tang, C., Hamilton, L.S., Chang, E.F. Intonational speech prosody encoding in the human auditory cortex [J]. *Science*, 2017(6353): 797-801.
- [83] Titze, I.R. Physiologic and acoustic differences between male and female voices [J]. *The Journal of the Acoustical Society of America*, 1989(4): 1699-1707.
- [84] Voigt, R., Jurafsky, D., Sumner, M. *Between-and within-speaker effects of bilingualism on f0 variation* [C]. Interspeech. San Francisco, The United States, 2016:1122-1126.
- [85] von Kriegstein, K., Warren, J.D., Ives, D.T., et al. Processing the acoustic effect of size in speech sounds [J]. *Neuroimage*, 2006(1): 368-375.
- [86] Walker, M., Perry, C. It's the words you use and how you say them: Electrophysiological correlates of the perception of imitated masculine speech [J]. *Language, Cognition and Neuroscience*, 2022(1): 1-21.
- [87] Winters, S.J., Levi, S.V., Pisoni, D.B. Identification and discrimination of bilingual talkers across languages [J]. *The Journal of the Acoustical Society of America*, 2008(6): 4524-4538.
- [88] Xu, H., Armony, J.L. Influence of emotional prosody, content, and repetition on memory recognition of speaker identity [J]. *Quarterly Journal of Experimental Psychology*, 2021(7): 1185-1201.
- [89] Xu, M., Homae, F., Hashimoto, R.-i., et al. Acoustic cues for the recognition of self-voice and other-voice [J]. *Frontiers in psychology*, 2013: 735.
- [90] Zhang, Y., Ding, Y., Huang, J., et al. Hierarchical cortical networks of “voice patches” for processing voices in human brain [J]. *Proceedings of the National Academy of Sciences*, 2021(52): e2113887118.
- [91] 陈忠敏. 语音感知的特点及其解剖生理机制 [J]. *中国语音学报*, 2021(1): 8-24.
- [92] 蒋晓鸣. 文化互鉴视角下非言语表情的嗓音编码和解码 [J]. *《同济大学学报》(社会科学版)*, 2020(1): 116-124.
- [93] 明莉莉, 胡学平. 人类嗓音加工的神经机制——来自正常视力者和盲人的脑神经证据 [J]. *心理科学进展*, 2021(12): 2147.
- [94] 束定芳. 《语言与社会心理学》评介——兼论社会心理语言学的研究对象、目标及方法 [J]. *外国语(上海外国语学院学报)*, 1992(03): 10-14.
- [95] 束定芳, 张立飞. 后“经典”认知语言学: 社会转向和实证转向 [J]. *现代外语*, 2021(03): 420-429.
- [96] 王德春, 孙汝建. 社会心理语言学的理论和方法论基础 [J]. *外国语(上海外国语学院学报)*, 1992a(04): 3-7+82.



- [97] 王德春, 孙汝建. 社会心理语言学的学科性质和研究对象 [J]. *外国语(上海外国语学院学报)*, 1992b(03): 3-9+82.
- [98] 伍可, 陈杰, 李雯婕, et al. 人声加工的神经机制 [J]. *心理科学进展*, 2020(5): 752-765.
- [99] 夏志华, 马秋武. *同济博士论丛: 汉语对话中韵律趋同的实验研究* [M]. 上海: 同济大学出版社, 2019.
- [100] 周爱保, 胡砚冰, 周滢鑫, et al. 听而不“闻”? 人声失认症的神经机制 [J]. *心理科学进展*, 2021(3): 414.