

《心理科学进展》审稿意见与作者回应

题目：听到“牛黄”能想到“黄牛”吗？——口语识别中的语音位置编码机制

作者：韩海宾，李兴珊

第一轮

审稿人 1 意见：

意见 1：作者在引言部分列举了语音位置编码的相关研究，但还是无法较好说明语音的时序位置编码与口语识别过程的关系，探究口语识别过程中的语音位置编码的重要性。

回应：谢谢审稿人的意见。为了更加突出语音位置编码的重要性，在修改稿中，我们按照专家的建议在引言部分的第一段中加入了有关语音的时序位置编码与口语识别过程的关系的论述。同时，我们使用了“语音转文字”的例证将阅读理解与口语识别区分开来，更好地说明它们之间的关系。

意见 2：语音位置编码从提出到发展过程中，不同理论之间是否存在异同点？更全面的阐述可以帮助读者更好地客观理解这个领域的现状。

回应：谢谢专家的建议！在修改稿中，我们根据专家的意见主要做出了以下几点修改：

- （1）修改了有关已有 Cohort 模型、TRACE 模型的描述，以突显模型之间的差异；
 - （2）增加了对另外两个交互激活模型 Shortlist 模型（Norris, 1994）以及 NAM 模型（Luce et al., 2000）的介绍；
 - （3）在第二部分倒数第三段强调了模型之间的共同特性；
 - （4）增加了一个对照表格（表 1. 早期词汇识别的代表性序列模型对比），旨在帮助读者更好地区分几种模型的差异并理解该领域的现状。
- 相关内容已添加到修改稿中，并使用红色字体标示。

意见 3：英文和汉字词语音位置编码是否存在差异？毕竟英文的字母单元与汉语中的字单元存在差异。

回应：谢谢建议！在原稿中，我们在第 3.4 部分关于影响因素以及第 4 部分展望中提到了汉语口语识别的一些特殊性，但并不够系统。在修改稿中，根据专家的建议，我们调整了文章结构，新增了第四部分“语音位置编码汉语口语识别中的特殊性”，并补充了有关汉语口语识别特殊性的相关内容。在该部分，我们主要强调了汉语口语识别中的三个特殊之处：（1）汉字词形音分离特性；（2）汉字词语音加工单元；（3）汉字词语音加工单元与语义的关系。这有助于凸显在口语识别中，汉语作为表意文字系统与拼音文字系统的不同之处。此外，我们结合 TISK 模型，对汉语口语识别可能的加工机制也进行了预测。

意见 4：文章既然以机制为主，是否可以添加些关于位置编码的神经机制的研究。比如文章中所说的听到“sub”之后，“bus”也会得到激活，文字的激活之外，大脑中激活的是语言区还是视觉区？或许能更好的解决语言的加工机制问题。

回应：据笔者所知，关于考察语音时序位置编码的神经科学研究尚不多见，而结论的一致性也存在一定的不足。我们查阅了相关文献，发现颞上回（STG）可能参与了语音特征在空间和时间上在大脑中的整合过程（Yee et al., 2008; Yi et al., 2019; Scott, 2009）。例如，Yee 等人

(2008)发现额下回(IFG)的病变使得词首起始音位的竞争效应变得非常小,而颞上回(STG)的病变则导致该竞争效应比对照组更大。

在修改稿中,为了更突出加工机制与神经机制,我们在**3.2 部分**(以往研究不管在音位还是音节水平其实都是在音位水平的不同研究问题进行探讨,所以我们只在3.2部分统一阐述了音位、音节的加工机制与神经机制)、**3.3 部分**以及**展望部分**分别增加了有关神经机制的描述。有趣的是,神经机制的部分结果也支持了语音内容特征和位置信息单独编码的方式。相关内容已添加到修改稿中。

意见 5: 可能需阐述一下语音位置编码的局限性以及文章整体上的不足之处。

回应: 谢谢建议,我们在修改稿第5部分的最后一段统一阐述了语音位置编码的局限性以及文章整体上的不足之处。

.....

审稿人 2 意见:

意见 1: 不同层面的语音位置转置效应是否存在不同的机制。作者系统地综述了音位、音节和句子层面的相关语音位置转置效应,然而这些不同层面的效应,似乎在机制上可能存在明显的差异。作者可以考虑增加部分的论述。

回应: 谢谢审稿专家的意见和建议。

目前来看,早期序列位置编码机制和相关理论模型已相对成熟,并得到大量研究支持。然而,近年来提出的语音灵活位置编码方式仍然缺乏充分的定量支持,因此在机制的挖掘上仍有不足。梳理现有研究,我们发现有关语音位置编码的讨论主要集中在音位水平,而对音节水平的讨论实际上也在考察音位的加工机制,结果也都支持了一种独立于输入时序位置的编码机制。正如 TISK 模型的假设一样,我们会对输入的语音序列进行编码,形成一系列与位置无关的音位或者二音位矩阵,然后在词汇水平层面或词汇决策阶段再识别出语音重叠程度更多的词汇。

讨论认知过程的内在机制离不开神经机制的生理证据支持,然而,目前关于语音时序位置编码的神经科学研究仍相对较少。在音节序列编码的研究中,发现颞上回(STG)可能参与了语音特征在空间和时间上在大脑中的整合过程(Yee et al., 2008; Yi et al., 2019; Scott, 2009)。

为了更突出加工机制,在修改稿中,我们在3.2部分(以往的研究不管在音位还是音节水平其实都是在音位水平的不同研究问题进行探讨,因此我们只在3.2部分统一阐述了音位、音节的加工机制与神经机制)、3.3部分增加了有关加工机制的描述。我们还添加了对相关水平上神经机制的探讨。在3.3部分有关句子的加工中,我们倾向于认为在听觉句子的加工中,类似于读者在理解阅读过程中对刚刚阅读过的单词保留一种不确定性一样(Levy et al., 2009),我们同样采用一种独立于具体词汇语音位置的编码机制。有趣的是,Gwilliams 等人(2022)通过脑磁技术在神经机制层面的研究也支持类似的结果,发现大脑可以快速解码语音特征,语音位置顺序信息是独立存储的。

相关内容已添加到修改稿中,并使用红色字体标示。

意见 2: 是否可以扩充 TISK 模型的介绍。如同作者所言,该模型对于音位位置转置效应具有较好的解释力度,是否可以扩充这部分的内容,提供多一些实证证据的介绍。

回应: 在修改稿中,我们重新阐述了关于 TISK 模型的构成以及工作原理,并加入了对其加工过程的介绍。除此之外,我们通过两个使用拼音文字系统实验材料的实例,介绍了 TISK 模型对这些研究的支持。

意见 3: 可以增加基于 TISK 模型下对于中文口语识别的解释。如同作者所言, 中文在该领域的研究相对较少, 因此如果作者可以多增加这部分的内容, 相信对于相关领域的学者会有不小的启发。

回应: 在新的修改稿中, 为了突出中文口语识别的特殊性, 我们增加了第四部分“**语音位置编码汉语口语识别中语音位置编码的特殊性**”。该部分强调了汉语口语识别中的三点特殊之处, 包括汉字词形音分离特性、汉字词语音加工单元、以及汉字词语音加工单元与语义的关系。除此之外, 我们增加了有关 **TISK 模型在中文口语识别中的解释**, 并通过实例对中文口语识别过程进行了预测。这部分内容我们放在了修改稿第 4 部分最后一段。

第二轮

审稿人 1 意见:

文章比较系统的梳理和比较了语音编码中的位置效应, 并提出了目前编码模型中存在的一些问题和见解。整体上为语音的位置编码理论进行较好的总结, 但是文章中存在个别用词和细节需要检查修正, 例如, 引言部分的最后一句话存在重复“的”字(“展望未来关于汉字词语音位置编码在口语识别中的的特殊性”), 以及第 4 部分的首句“综合上面的综述可见”中的“综述”这个词改为“上文”或者“内容”可能更恰当。

回应: 谢谢审稿人的意见。我们已针对审稿专家提出的两个细节进行了修正, 去掉了多余的字, 并把“综合上面的综述可见”改为了“综合以上内容可见”。除此之外, 我们仔细检查了论文的内容, 修正了一些用词和细节。

审稿人 2 意见: 作者已经按照审稿人的意见做了较为充分的修改, 建议接收。

编委 1 意见: 稿件已达到发表水平, 一个小建议: “研究发现, 人类大脑能够持续并行地处理和编码最近听到的三个语音声音”这里要指明声音单元。

回应: 谢谢编委的意见。我们在正文中增加了对声音单元的描述, 指明了 Gwilliams 文章中的声音单元 (Phoneme 音位), 更改如下: “研究发现, 人类大脑能够持续并行地处理和编码最近听到的三个语音单元 (在 Gwilliams 等人的文章中为音位单元)。”

编委 2 意见: 同意发表。

主编意见: 同意发表。