基于GPT-2和互信息的语言单位信息量对韵律特征的影响

郝韵[1]　解焱陆[1]　林炳怀[2]　张劲松[1]
[1]北京语言大学信息科学学院，北京100083
[2]腾讯科技，北京100083
haoyun7725@163.com　xieyanlu@blcu.edu.cn　binghuailin@tencent.com
jinsong.zhang@blcu.edu.cn

摘要

基于信息论的言语产出研究发现携带信息量越大的语言单位，其语音信号越容易被强化。目前的相关研究主要通过自信息的方式衡量语言单位信息量，但该方法难以对长距离的上下文语境进行建模。本研究引入基于预训练语言模型GPT-2和文本-拼音互信息的语言单位信息量衡量方式，考察汉语的单词、韵母和声调信息量对语音产出的韵律特征的影响。研究结果显示汉语中单词和韵母信息量更大时，其韵律特征倾向于被增强，证明了我们提出的方法是有效的。其中信息量效应在音长特征上相比音高和音强特征更显著。

关键词： GPT-2；信息量；韵律；音长；互信息

Prosodic Effects of Speech Unit's Information Based on GPT-2 and Mutual Information

Yun HAO[1]　**Yanlu XIE**[1]　**Binghuai LIN**[2]　**Jinsong ZHANG**[1]
[1]School of Information Science, Beijing Language and Culture University, Beijing, 100083, China
[2]Smart Platform Product Department, Tencent Technology Co., Ltd, Beijing, 100083, China
haoyun7725@163.com　xieyanlu@blcu.edu.cn　binghuailin@tencent.com
jinsong.zhang@blcu.edu.cn

Abstract

Research has shown that linguistic units carrying more information tend to be realized with enhanced speech signals. Most previous studies measure the information that a linguistic unit carries with its surprisal. However, such measurement lacks the ability to model long-distance contextual effects. The current study proposes novel measures of linguistic unit' s information by incorporating the GPT-2 pre-trained language model and mutual information (MI) between text and its phonemic transcription. We examine the prosodic effects of word surprisal and MI-based information of final and tones in Mandarin Chinese. Results show that more information of both words and finals enhance prosodic prominence, proving the validity of our proposed measurements. Besides, the effects of information are more notable on duration feature compared with pitch and intensity feature.

第二十一届中国计算语言学大会论文集，第46页-第55页，南昌，中国，2022年10月14日至16日。

Keywords: GPT-2 , information , prosody , duration , mutual information

1 引言

在语言交流过程中，各语言单位（如词、词素、音素等）所携带的信息量会影响我们感知与产出的难易程度。早在1929年，Zipf (1929)就发现音素的频率与其语音复杂度之间存在着反比关系：音素的频率越高，其语音复杂度就越低。后来的心理语言学与实验语音学领域的研究都提供了类似的证据：Howes and Solomon (1951) 发现辨认单词所需要的视觉呈现时间与根据语料库计算的词频有关：词频越高，被试辨认出所呈现的单词的时间越短。Lieberman (1963)通过填空任务的正确率衡量了英语句子中词的可预测性，并通过语音产出实验发现可预测性更强的词在时长上更短、音高和音强更弱。

真正意义上把语言和信息理论相结合的研究始于Shannon (1949)的信息论。信息论将语言交流视为信息传输的系统：在语言信息传输过程中，说话人将想要传递的信息编码成语音信号，而听者基于噪声下的语音信号对信息进行解码。当信息率为均匀分布且接近信道容量时，可以达到信息传输的最小冗余(Genzel and Charniak, 2002)。在信息论的基础上，Jaeger and Levy (2006)提出了语言传递的均匀信息密度假设(Uniform Information Density Hypothesis)，认为说话人致力于维持一段话语中每单位的Shannon自信息服从均匀分布。Aylett and Turk (2004)也提出了类似思想的平稳信号冗余度假设(Smooth Signal Redundancy Hypothesis)，不同之处在于他们的理论更加强调在语音产出中韵律突显特征对信息量的调节作用：当文本的信息量局部过小或过大时，通过音长、音高或音强的方式弱化或强化语音信号以实现整体上平稳的信息量分布。

韵律突显(prosodic prominence)指一段话语中的某个语音单位在声学或感知上相对突出的特性(Terken and Hermes, 2000; Aylett and Turk, 2004)。韵律突显的声学特性一般体现在时长、音高、音强或其他频谱特征上(Terken and Hermes, 2000)。韵律突显的主要功能便是在话语中突出更加重要、信息量更大的语言单位(Calhoun, 2007)。随着大规模语音语料库的出现，许多研究开始定量探究基于统计的语言单位信息量与韵律特征的关系。Jurafsky et al. (2001) 基于Switchboard 语料库发现英语中单词的频率及Bigram 概率对元音时长有显著的影响。Van Son et al. (2004) 基于荷兰语、芬兰语和俄语语料库发现根据单词中音素的条件概率计算的音素信息量越大，该音素的时长、音强及频谱特征就越容易被加强。

对信息量的韵律效应的研究早期主要集中在印欧语系语言，近年来逐渐拓展到了其他语言如Kaqchikel Mayan语(Tang and Bennett, 2018),日语(Shaw and Kawahara, 2019; Hashimoto, 2021)及包含各语系语言的跨语言比较(Pimentel et al., 2021)。对于汉语中语言单位的信息量，早在20世纪60 年代就有中国科学院声学研究所对汉语的单词出现频率、声韵母及声调的出现频率、声韵母结合概率等进行了统计分析(张家騄, 2010)。关于汉语中信息量与韵律的关系，周韧 (2007)主张句法组合中信息量大的成分将得到重音，而信息量小的成分得不到重音，但并未进行定量的统计分析。Tang and Shaw (2021)基于语料库和Bigram 语言模型发现汉语中词的信息量对时长、音高和音强均有显著的影响，但他们仅探究了词层级的信息量效应，对更细粒度的语言层级(如音素、声母或韵母等)的信息量没有涉及。

传统方法多使用N-gram概率的方法衡量单词或音素等语言单元的信息量，但此类方法无法对长距离的上下文语义关系进行建模(Daland and Zuraw, 2018)。为了解决此问题，本研究提出

两种改进的语言单位信息量的计算方法：一种是引入预训练语言模型计算字词的信息量，另一种是引入文本- 拼音互信息的方法计算音位层级的信息量。相比传统的N-gram 语言模型，基于大数据训练的预训练语言模型具有更强的泛化能力，且模型结构中的深度注意力机制可以学习到长距离的上下文语义依赖关系。文本- 拼音互信息方法建立在文本-音位-文本传输模型(Zhang et al., 2008; Zhang et al., 2010)的基础上。在音位的功能负载研究中，该方法可以量化特定音位对辨别语义的贡献程度(Zhang et al., 2010; Wu et al., 2014; Chen et al., 2016; Zhang et al., 2021)。该方法的另一个优势是可以在同样标准下量化不同语音范畴的信息量，包括声韵母、声调、韵律边界等(Wu et al., 2014; Chen et al., 2016)。因此，我们提出使用特定音位信息丢失时文本- 拼音互信息的损失来量化韵母与声调的信息量。基于语音语料库的实验结果证明了汉语的单词和韵母信息量更大时，其韵律特征倾向于被增强，且本研究引入的方法对韵律参数有更好的回归效果。

2 相关工作

2.1 基于自信息的语言单位信息量

目前的信息量的语音产出效应相关研究主要采用基于自信息的方式衡量语言单位的信息量，如式(1)所示。

$$SI_{unit_i} = -\log_2 P(unit_i|context) \tag{1}$$

其中$context$表示该语言单位出现的环境条件。单词自信息的研究中一般计算单词给定前n个词条件下的自信息(Jurafsky et al., 2001; Bell et al., 2009; Tang and Shaw, 2021)。音素自信息的研究中，有些考虑单词中某音素在给定所有前接音素的条件概率(Van Son et al., 2004; Priva, 2015)；也有研究仅考虑给定前一个音素条件下的音素概率(Malisz et al., 2018; Shaw and Kawahara, 2019)。以上方法虽然可以反映单词或音素在给定局部环境条件下的可预测性，但无法对话题、新旧信息等更长距离的上下文依赖关系进行有效的建模(Daland and Zuraw, 2018)。

2.2 文本-拼音互信息理论

文本-拼音互信息理论在Zhang et al. (2008)中首次被提出，以用于为汉语语音识别设计音素集。后来该理论被应用于音系学相关研究，用来计算音位的功能负载(Zhang et al., 2010; Wu et al., 2014; Zhang et al., 2021)。基于该理论的功能负载衡量了音位在受到上下文语境影响的条件下对语言信息传递的重要程度，对我们将要研究的语音单位信息量具有重要启示意义。计算互信息的文本-音位-文本传输模型如图(1)所示。

其中W表示原始文本，即说话人想要传达的信息；F表示W对应的拼音形式；$\hat{W}$表示对F解码得到的所有文本的集合。如果信息编码与解码过程是无损的，应该满足$W = \hat{W}$。然而语言传输过程中可能由于噪声、同义词等因素而产生信息损失。W编码为F需要依赖该语言的音素词典Φ，而F 解码为W需要依赖词典Φ和语言模型LM。W与其拼音形式F之间的互信息定义为式(2)：

$$MI(W;F) = H(W) - H(W|F) \tag{2}$$

互信息量化了一个随机变量在已知另一个随机变量的情况下减少的不确定性。$MI(W;F)$表示根据音素序列F解码出原始文本W的可能性。文本-拼音互信息越大，说明

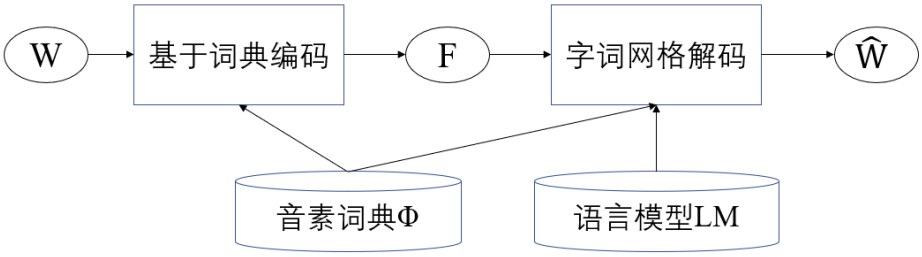

Figure 1: 文本-音位-文本传输模型

越容易从拼音还原出正确的文本内容。根据Shannon-McMillan-Breiman 定理，如果(W, F) 同时是平稳的(stationary)并且是各态经历的(ergodic)，公式(2)可以推导为(3)：

$$MI(W;F) = \lim_{n\to\infty} -\frac{1}{n}\log\sum_{i=1}^{m} P(W_i') \tag{3}$$

其中W_i'表示所有拼音形式为F的文本，即$\hat{W}$中的元素。我们可以根据语言模型计算得到$P(W_i')$。式(3)表明发音为F的文本概率和越大，文本-拼音互信息越小。

3 本文方法

3.1 预训练语言模型GPT-2

随着深度学习及预训练语言模型的兴起，语言研究者们开始关注其对人类语言处理表现上的预测能力。GPT-2是由OpenAI提出的第二代基于Transformer结构的大规模预训练语言模型(Radford et al., 2019)。其模型结构为Transformer的解码器部分，训练目标是对于一段给定文本预测下一个单词的概率分布。GPT-2在生成类似人类创作的文本任务上表现突出，并且在多项预测人类语言处理任务(如眼动数据)上表现优于其他预训练语言模型。例如，Wilcox et al. (2020) 比较了N-gram, LSTM, RNNGs 和GPT-2 模型在预测句子加工时长及眼动表现上的效果，发现GPT-2模型的表现最佳。Hao et al. (2020) 也发现GPT-2 模型在预测阅读句子的眼动数据上表现优于XLM、Transformer-XL等其他预训练语言模型。基于以上背景，本研究尝试将GPT-2预训练语言模型应用于估计汉语单词及声韵母、声调层级信息量，进而探究信息量对语音产出中韵律特征的影响。作为自回归语言模型，GPT-2可以基于给定的上文输入预测单词出现的概率。我们将基于式(4) 计算句子中第t 个单词w_t的自信息，并基于式(5)计算长度为N的句子信息量以用于后续计算文本-拼音的互信息。

$$SI(w_t) = -\log_2 P(w_t|w_1, ..., w_{t-1}) \tag{4}$$

$$SI(s) = -\sum_{t=1}^{N} \log_2 P(w_t|w_1, ..., w_{t-1}) \tag{5}$$

3.2 基于文本-拼音互信息的信息量

我们基于文本-拼音互信息理论提出一种计算语音单位信息量的方法。该方法基于这样的假

设：某个语音单位对信息传递的贡献度可以被假设为当该语音单位在传输过程中丢失时（即听话人没有听到该声音），文本-拼音互信息的减少程度。即某语音单位基于互信息的信息量定义为公式(6):

$$MIloss(p) = \frac{MI(W;F) - MI(W;F')}{MI(W;F)} \tag{6}$$

其中p可以表示任何语音单位，包括声韵母、声调等。F表示文本W的规范发音，而F'表示p丢失时的发音。式(6)量化了p丢失的情况下文本-拼音互信息减少的程度，互信息损失越大，说明该语音丢失造成的混淆程度越大，即该语音越重要。图(2)展示了同样的文本内容“你好”在三种不同的语音编码情况下的信息传递过程，其中F表示拼音形式的发音，括号中为声调。当发音为规范发音“ni(3) hao(3)”时可能解码得到包括“你好”、“拟郝”...等的文本序列集合。而当“你”的韵母或声调信息丢失时，解码文本集合$\hat{W}$扩大，可能得到其他发音的文本如“女好”、“哪好”或“尼好”、“逆好”等。在韵母或声调的发音丢失的情况下，由于文本数量增加，文本- 拼音互信息减少。如果增加的文本概率小，那么互信息减少的程度小，说明该语音单位的信息贡献较小；反之如果增加的文本概率大，那么该语音的信息贡献大。

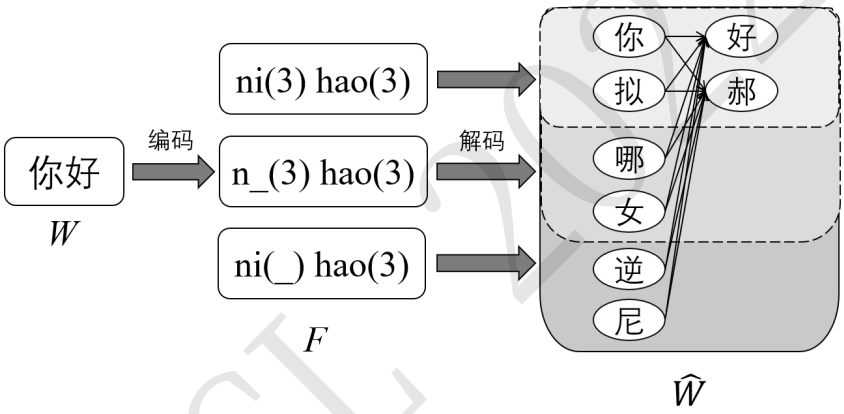

Figure 2: 示例“你好”在拼音信息无损/损失韵母/损失声调情况下的编码与解码过程

计算基于互信息的语音单位信息量的具体过程如下：首先，一个句子被转录成拼音，并进一步转录成音位序列。接下来基于音素词典，原始音位序列被解码成所有可能的文本序列。对于句子中的每一个音节，得到该音节中某个语音单位p信息丢失后的音位序列对应的所有文本序列。最后通过语言模型获得所有文本序列的概率，并根据式(6)计算丢失p后的互信息损失。

4 实验设置

4.1 语音语料

本实验的语音语料来自北京语言大学汉语中介语语料库(曹文and 张劲松, 2009)中的母语部分。语料文本包含选自对外汉语教材的301句话，发音人包含12个中国母语者（6个男性，6 个女性）。我们的实验选取了其中的189个字数为8个字以内的句子。语料库中包含对音节及声韵母边界的标注。我们使用Praat软件在边界标注的基础上提取了每个音节韵母段的时长、音高最大值和最小值、音强最大值。去除未能成功提取音高的音节之后，本实验的最终数据为12459个音

节。在后续分析中，我们对所有语音数据进行了发音人归一化处理，并将时长和音高数据取对数以使它们更接近正态分布。

4.2 语言模型

本研究使用的中文预训练GPT-2模型为Du (2019)训练并发布在Github和Huggingface上的GPT2-chinese-cluecorpussmall模型，模型结构包含12 个层，12个注意力头和768个隐藏层节点。模型的训练数据为CLUECorpusSmall(Xu et al., 2020)，包含新闻语料、社区互动语料、维基百科语料和评论数据语料四个部分，总数据量超过14G、50亿字。我们同时训练了先行研究中常用的Bigram语言模型以便与GPT-2模型的结果进行对比：使用KenLM工具包(Heafield, 2011)进行训练，使用modified Kneser-Ney方法(Heafield et al., 2013)进行平滑处理；训练语料为CLUECorpusSmall中的评论数据部分，包含2.3G左右文本、约10亿字。

4.3 统计方法

本研究采用线性混合模型的方法，基于R统计软件的lmerTest库对相关变量进行回归分析。我们对每个因变量基于Bigram 和GPT-2 语言模型的信息量分别进行了线性混合模型的回归。每个回归模型中包含了除信息量之外其他可能影响韵律特征的控制变量，并包含了不同发音人的随机效应。模型的因变量、控制变量和信息量变量如下所示。

- **因变量** 韵母段时长，音高最大值，音高范围，音强最大值。
- **控制变量** 当前音节声母/韵母/声调，前接/后接音节声调，后接音节声母，单词内前/后音节个数，标点符号分割的短句内前/后音节个数，句子内前/后音节个数，句子语速（音节/每秒），由THULAC工具包(孙茂松et al., 2016)得到的词性标注。
- **信息量变量** 单词自信息(Bigram/GPT-2)，韵母合并后的互信息损失(Bigram/GPT-2)，声调合并后的互信息损失(Bigram/GPT-2)。对互信息损失变量取对数，并对所有信息量变量都进行了归一化处理。

在回归过程中，首先对每个因变量建立只对控制变量进行回归的基线模型；再通过向后剔除方法去掉不显著的控制变量；最后在筛选控制变量后的基线模型上分别加入基于Bigram模型的信息量和基于GPT-2 模型的信息量，即对每个因变量最终得到基线、Bigram和GPT-2 三个回归模型。对各模型进行方差膨胀系数检验发现VIF<2.5，说明各模型均不存在多重共线性。

5 实验结果

5.1 信息量对韵律特征的影响

由线性混合模型统计得到各信息量衡量方法分别对4种韵律特征的固定效应及显著性如表5.1所示。其中固定效应β值表示信息量对韵律特征影响的大小和方向，p 值表示影响的显著性，加粗表示$p < 0.05$，即固定效应显著。

对韵母词长的统计结果显示，两种单词自信息量均有正向的显著影响（Bigram：$\beta = 0.037$，$p = 0.01$，GPT-2：$\beta = 0.038$，$p = 0.04$），即单词信息量越大，韵母时长越长。这与前人对其他语言中信息量的语音效应的结论相符，也与Tang and Shaw (2021)对汉语的研究结论相一致。我们提出的基于Bigram 和GPT-2 的韵母互信息损失对时长都有有正向的显著影响（Bigram:

信息量	语言模型	音长		音高最大值		音高范围		音强最大值	
		β值	p值	β值	p值	β值	p值	β值	p值
单词自信息	Bigram	0.027	**0.01**	0.012	0.18	0.022	**0.04**	0.003	0.81
	GPT-2	0.038	**0.04**	0.019	0.1	0.033	0.07	0.038	0.82
韵母互信息损失	Bigram	0.062	**<0.001**	-0.008	0.55	0.027	0.06	0.014	0.31
	GPT-2	0.05	**<0.001**	0.011	0.23	0.009	0.44	0.024	0.05
声调互信息损失	Bigram	0.012	0.33	-0.037	**<0.01**	-0.019	0.27	-0.003	0.87
	GPT-2	-0.013	0.14	-0.034	**<0.001**	-0.044	**<0.001**	-0.018	0.25

注：显著性水平$p < 0.001$ ***，$p < 0.01$ **，$p < 0.05$ *，$p < 0.10$.

Table 1: 信息量变量对韵律特征的固定效应及显著性

$\beta = 0.062$，$p < .001$，GPT-2: $\beta = 0.050$，$p < 0.001$)；但声调互信息损失对时长影响不显著(Bigram: $\beta = 0.012$，$p = 0.33$，GPT2: $\beta = -0.013$，$p = 0.14$)。

对于音高特征，信息量效应则多数不显著。基于Bigram和GPT-2的声调互信息损失对音高最大值有显著的负向影响（Bigram: $\beta = -0.037$，$p < 0.01$，GPT-2: $\beta = -0.034$，$p < 0.001$)，这与我们的预期相反。基于Bigram的单词自信息对音高范围由有显著的正向影响（$\beta = 0.022$，$p = 0.04$)，基于GPT-2的单词自信息对音高范围也有接近显著的正向影响（$\beta = 0.033$，$p = 0.07$)，该结果的趋势与前人的发现Tang and Shaw (2021)一致。

对于音强最大值，结果显示只有基于GPT-2 的韵母互信息损失有接近显著的正向影响($\beta = 0.024$，$p = 0.05$)。以上实验结果表明，在三种韵律特征中时长最容易受到信息量效应的影响，而音高和音强较少受到信息量效应的影响。单词自信息和韵母互信息损失对韵律特征的影响都是正向的，这与前人的研究结果及我们的预期相同。声调互信息损失对韵律参数的影响多数不显著，只对音高呈现负向的效应，与我们的预测不符。前人研究很少涉及超音段单位的信息量，我们提出的声调互信息损失是对超音段单位信息量与语音产出之间关系的初步探索，还需要未来进一步探究和讨论。

5.2 对数似然值比较

为了比较加入基于Bigram和GPT-2语言模型的各信息量变量对韵律特征回归的贡献程度，我们引入了对数似然值的变化($\Delta logLikelihood$)，表示加入某变量后与基线模型相比对数似然值的提升（5.2)。对数似然值越大，说明模型对韵律特征的拟合效果越好。正的$\Delta logLikelihood$表明该信息量变量对韵律特征的回归结果有提升。表中加粗显示了效果更优的语言模型。

表(5.2)的结果显示多数信息量变量都可以提升对韵律参数的拟合效果，尤其是对音长参数的拟合效果。其中，基于GPT-2的单词自信息对所有韵律参数的拟合都有帮助，且全部优于基于Bigram 的单词自信息，这说明我们提出的基于预训练语言模型的单词自信息是有效的。基于文本-拼音互信息的韵母和声调互信息损失也有部分可以显著提升模型的拟合效果，说明了我们提出的基于互信息的信息量的有效性。在基于文本- 拼音互信息的韵母和声调信息量中，可以看到韵母互信息损失对韵律参数的贡献较大，优于声调互信息的损失；其中基于Bigram的韵母互信息损失对所有韵律参数的拟合都有帮助；基于GPT-2 的韵母互信息损失在解释音强最大值

信息量	语言模型	音长	音高最大值	音高范围	音强最大值
单词自信息	Bigram	29.49	-3.18	0.67	-6.42
	GPT-2	**47.64**	**0.25**	**7.73**	**4.76**
韵母互信息损失	Bigram	**115.5**	**66.7**	**74.17**	70.13
	GPT-2	41.77	-2.99	-3.06	**74.62**
声调互信息损失	Bigram	**37.57**	**6.74**	-1.31	-1.51
	GPT-2	16.44	1.7	**3.14**	**2.67**

Table 2: 各信息量贡献的对数似然值$\Delta logLikelihood$

时优于Bigram 模型，在解释其他韵律参数时效果弱于Bigram 模型。声调互信息损失在音长参数上有较明显的效果，且基于Bigram模型的信息量优于GPT-2，对其余韵律特征的回归贡献则较小。

6 总结与讨论

基于信息传递效率的语言研究认为语音的韵律突显与语言文本的语境信息量呈正相关(Lieberman, 1963; Aylett and Turk, 2004)，且该现象的语音实现存在跨语言的差异(Malisz et al., 2018)。本研究通过基于语料库的实验探究了汉语中语言单位（单词、韵母和声调）的信息量对韵律声学特征（音长、音高和音强）的影响。为了更好地对语境信息进行建模，我们提出了基于预训练语言模型GPT-2 和文本- 拼音互信息的语言单位信息量的衡量方式。在我们提出的两种方法中，基于GPT-2估计的单词信息量相比Bigram模型对韵律参数的拟合有明显提升，这说明了相对于传统方法，预训练语言模型得到的单词信息量可以更好地解释人类语言产出的有关现象；基于文本-拼音互信息的韵母信息量对韵律特征尤其是时长也有显著的正向影响，说明了我们提出方法的有效性。我们还考虑了汉语声调信息量对韵律特征的影响，但其效应普遍不显著，在部分特征上结果与预期不符。对超音段层级信息量的语音效应相关定量研究目前较少，只有一些关于重音可预测性与相关声学参数的讨论(Athanasopoulou et al., 2017)。本研究对于声调信息量的研究是对该方向的初步探索，目前的结果受限于语音语料及语言模型性能等因素的影响，还需未来进一步的探索与考察。

我们的研究结果支持了平稳信号冗余度假设(Aylett and Turk, 2004)，即韵律突显在语音信号中起到了平滑语言的信息量的作用，并且发现时长是汉语中主要体现信息量效应的韵律特征。先行研究对基于N-gram统计的信息量与言语产出/感知的关系研究已有足够相关实验证据及理论，但对深度学习模型训练得到的单词表示与人类语言能力之间关系的探究还在起步中。我们的实验中，基于GPT-2 模型计算的信息量与基于Bigram模型的信息量对韵律特征的效应在显著性上得到了相似的结论，但对各因变量的对数似然值贡献上存在一定差异。在未来的研究中，我们将尝试进一步探索汉语中的声调范畴与语音信息量在韵律表现上的交互作用，并通过引入其他预训练语言模型和进行模型微调等改进信息量的计算方式，继续探索语言单位信息量与人类语言产出/理解的关系。

致谢

本工作得到中央高校基本科研业务专项资金（20YJ040002）、北京语言大学梧桐创新平

台(19PT04)、以及语言资源高精尖中心项目“面向智能语音教学的汉语中介语语音多模态语料库研究”(KYR17005)的资助，张劲松是本文的通讯作者。

参考文献

曹文and 张劲松. 2009. 面向计算机辅助正音的汉语中介语语音语料库的创制与标注. 语言文字应用, (4):10.

孙茂松, 陈新雄, 张开旭, 郭志芃, and 刘知远. 2016. Thulac：一个高效的中文词法分析工具包. https://github.com/thunlp/THULAC-Python.

周韧. 2007. 信息量原则与汉语句法组合的韵律模式. 中国语文, (3):15.

张家騄. 2010. 汉语人机语音通信基础. 上海科学技术出版社, 上海.

Angeliki Athanasopoulou, Irene Vogel, and Hossep Dolatian. 2017. Acoustic properties of canonical and non-canonical stress in french, turkish, armenian and brazilian portuguese. In *INTERSPEECH*, pages 1398–1402.

Matthew Aylett and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, 47(1):31–56.

Alan Bell, Jason M Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language*, 60(1):92–111.

Sasha Calhoun. 2007. *Information structure and the prosodic structure of English: A probabilistic relationship*. Ph.D. thesis, University of Edinburgh.

Yue Chen, Yanlu Xie, Bin Wu, and Jinsong Zhang. 2016. A study on functional load of chinese prosodic boundaries under reduction of syllable information. In *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 1–5. IEEE.

Robert Daland and Kie Zuraw. 2018. Loci and locality of informational effects on phonetic implementation. *Linguistics Vanguard*, 4(s2).

Zeyao Du. 2019. Gpt2-chinese: Tools for training gpt2 model in chinese language. https://github.com/Morizeyao/GPT2-Chinese.

Dmitriy Genzel and Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 199–206.

Yiding Hao, Simon Mendelsohn, Rachel Sterneck, Randi Martinez, and Robert Frank. 2020. Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. *arXiv preprint arXiv:2009.03954*.

Daiki Hashimoto. 2021. Probabilistic reduction and mental accumulation in japanese: Frequency, contextual predictability, and average predictability. *Journal of Phonetics*, 87:101061.

Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H Clark, and Philipp Koehn. 2013. Scalable modified kneser-ney language model estimation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 690–696.

Kenneth Heafield. 2011. KenLM: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 187–197, Edinburgh, Scotland, July. Association for Computational Linguistics.

Davis H Howes and Richard L Solomon. 1951. Visual duration threshold as a function of word-probability. *Journal of experimental psychology*, 41(6):401.

T Jaeger and Roger Levy. 2006. Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, 19.

Daniel Jurafsky, Alan Bell, Michelle Gregory, and William D Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language*, 45:229–254.

Philip Lieberman. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and speech*, 6(3):172–187.

Zofia Malisz, Erika Brandt, Bernd Möbius, Yoon Mi Oh, and Bistra Andreeva. 2018. Dimensions of segmental variability: Interaction of prosody and surprisal in six languages. *Frontiers in Communication*, 3:25.

Tiago Pimentel, Clara Meister, Elizabeth Salesky, Simone Teufel, Damián Blasi, and Ryan Cotterell. 2021. A surprisal–duration trade-off across and within the world's languages. *arXiv preprint arXiv:2109.15000*.

Uriel Cohen Priva. 2015. Informativity affects consonant duration and deletion rates. *Laboratory phonology*, 6(2):243–278.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Claude E Shannon. 1949. Communication theory of secrecy systems. *The Bell system technical journal*, 28(4):656–715.

Jason A Shaw and Shigeto Kawahara. 2019. Effects of surprisal and entropy on vowel duration in japanese. *Language and speech*, 62(1):80–114.

Kevin Tang and Ryan Bennett. 2018. Contextual predictability influences word and morpheme duration in a morphologically complex language (kaqchikel mayan). *The Journal of the Acoustical Society of America*, 144(2):997–1017.

Kevin Tang and Jason A Shaw. 2021. Prosody leaks into the memories of words. *Cognition*, 210:104601.

Jacques Terken and Dik Hermes. 2000. The perception of prosodic prominence. In *Prosody: Theory and experiment*, pages 89–127. Springer.

Rob Van Son, Olga Bolotova, Louis CW Pols, and Mietta Lennes. 2004. Frequency effects on vowel reduction in three typologically different languages (dutch, finish, russian). In *Interspeech*. Citeseer.

Ethan Gotlieb Wilcox, Jon Gauthier, Jennifer Hu, Peng Qian, and Roger Levy. 2020. On the predictive power of neural language models for human real-time comprehension behavior. *arXiv preprint arXiv:2006.01912*.

Bin Wu, Jinsong Zhang, and Yanlu Xie. 2014. A clustering analysis of chinese consonants based on functional load. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pages 1–4. IEEE.

Liang Xu, Xuanwei Zhang, and Qianqian Dong. 2020. Cluecorpus2020: A large-scale chinese corpus for pre-training language model. *ArXiv*, abs/2003.01355.

Jinsong Zhang, Xinhui Hu, and Satoshi Nakamura. 2008. Using mutual information criterion to design an efficient phoneme set for chinese speech recognition. *IEICE TRANSACTIONS on Information and Systems*, 91(3):508–513.

Jinsong Zhang, Wei Li, Yuxia Hou, Wen Cao, and Ziyu Xiong. 2010. A study on functional loads of phonetic contrasts under context based on mutual information of chinese text and phonemes. In *2010 7th International Symposium on Chinese Spoken Language Processing*, pages 194–198. IEEE.

Yuqing Zhang, Zhu Li, Bin Wu, Yanlu Xie, Binghuai Lin, and Jinsong Zhang. 2021. Relationships between perceptual distinctiveness, articulatory complexity and functional load in speech communication. *Proc. Interspeech 2021*, pages 1733–1737.

George Kingsley Zipf. 1929. Relative frequency as a determinant of phonetic change. *Harvard studies in classical philology*, 40:1–95.