

## 自然口语对话语境中回声问句的解码初探

李爱军, 贾媛, 柳雪飞, 张良

中国社会科学院语言研究所

{liaj,jiayuan}@cass.org.cn

### 提要

自然口语交互是信息传递的一个重要手段。其中, 疑问信息的表达和感知(编码和解码), 是人机语音交互理解和翻译中的重要课题。本研究主要探索无疑问句法标记的回声问句在自然口语对话中的编解码机制, 即无疑问句法标记的回声问, 在上下文语境和离境两种情况下疑问信息的解码情况。研究通过对具有上下文语境信息和不具有上下文的离境信息两种情况下回声问句进行感知实验, 初步发现: (1) 回声问句在具有上文语境和离境情况下, 其疑问语气的表达程度不同; (2) 在离境情况下, 绝大部分回声问(约 93%)的疑问程度较低; 在有上下文语境的信息中, 所有的回声问都可以被感知到有疑问语气的存在, 且大部分(约 70%)的疑问语气大于陈述语气; (3) 上下文的语境信息对无标记的回声问句的感知起决定作用。交际双方会将包括上下文的语篇结构信息和语义、语调韵律等特征综合起来, 对疑问信息进行编码和解码。

**关键词** 回声问, 语境, 离境, 感知与解码, 语调韵律

### 1. 引言

在口语语音识别、理解和翻译中, 交际双方的疑问信息的确定是重要的研究课题。在面对面的交互中, 疑问信息可以通过文字、句法、语义、语调、韵律、姿态和面部表情等多模态信息同时传达; 而在电话交互或者移动互联网语音交互应用中, 主要通过对话语境的句法语义和语调韵律信息等来传递。

对于有句法标记的疑问句, 说话人的疑问信息断定相对容易, 但对于没有句法标记的疑问句, 即字面上看起来同陈述句一样, 在口语中也能表达疑问的句子, 其疑问信息的断定就相对困难。在这种情况下, 疑问信息可以通过语调等韵律特征以及上下文语境信息(包括上下文的篇章结构、语义等)来传递, 或者同时使用这些组合特征来传递。

本文关注自然口语对话中没有句法标记的回声问句。回声问指口语交际对话时, 交际一方(A)针对另外一方(B)的问题重复发问。本文的回声问除了这种情况, 还包括针对前面的内容, 重复发问。如我们转写的一段电话订餐场景口语对话, A 和 B 分别表示交际双方:

- A: 你好, 我要订一个土豆丝、鱼香茄子和一碗米饭。能到送到太月园三零九吗?  
B: 太月园三零九?  
A: 没错。  
B: 好的。

第一个问句有疑问语气词“吗”标记疑问, 而 B 重复 A 的问句的一部分内容(或者全部)再问的时候, 字面上是“太月园三零九”, 句法上没有疑问标记, 但对话

人 A 在交际过程中, 非常容易感知到 B 是再次问是否是“太月园三零九”, 从而回答“没错”。但是, 对计算机来说, 要理解 B 的这个没有句法标记的回声问句就比较困难了。本研究的主要目的是探索在口语交互中, 人是如何对回声问疑问信息进行解码的, 是通过上下文信息, 还是通过语音韵律的信息, 或者是两者的结合。研究结果将为机器理解疑问信息提供可使用的特征。

前人研究中对回声话语有着复杂多样的命名与定义[1, 2, 3], 但其中都体现了回声话语的本质, 即对前述话语某种程度的重复。Randolph Quirk et.al.[1]又将回声话语按照类别分为回声问句(echo question)和回声感叹句(echo exclamations)。其他作者[2, 3]又依据不同的话语功能将回声话语分为更加详细的类别。

吕叔湘先生[4]最早提出关于疑问句疑问程度的问题, 李宇明和唐志东[5]在其研究中将疑问句分为高疑问句, 低疑问句和无疑问句三种。林茂灿[6]又在后续的研究中依据疑问强度将回声问句分为 5 类, 用于判断回声问在失去语境环境后的疑问程度变化。

以往对回声问的研究主要集中在语义语用及交际的功能[4, 7, 8, 9, 10, 11]等角度。在交际功能方面, 回声问的表达主要采用疑问句的形式或语气重复对方所说的全部或部分话语, 并要求对方重述或证实所说的内容。并且由于交际功能的不同, 又将回声问分为重述性回声问(recapitulatory echo questions)和解释性回声问(explicatory echo questions)。在具体的交际场合中, 邵敬敏[9]将回声问分为 4 类: 即基本式, 语气式, 提醒式和惊愕式。

从语音上, 语音学学者又将研究范围拓展到回声问的韵律层面上[6, 12, 13], 进一步研究回声问句和陈述句与疑问句在调域、调阶、边界调、音长、音强以及感知方面的差异。

林茂灿[6]对朗读形式的回声问及对话形式的是非问进行了韵律特征分析, 具体阐述了疑问语气和陈述语气与边界调的关系, 并且提出了疑问句在脱离语境的环境下有不同的疑问程度, 其变化主要是因为疑问短语本身的疑问信息有强有弱。在关于回声问边界调的研究中, 林指出, 边界调为重音节时, 疑问句  $F_0$  曲拱的斜率比陈述句大, 或者音阶比陈述句高, 或者同时满足两个条件, 但  $F_0$  曲拱不会因疑问而改变其调型特征。同时, 林也指出回声问的边界调时长比同一内容的陈述句长。

林茂灿的研究还设计了感知实验来研究疑问语气和陈述语气间是否为突变。其结果显示  $F_0$  曲线斜率和音阶不仅能够区分疑问语气和陈述语气, 还可以给出二者之间的续说语气, 进而也证实了疑问可以有不同的强度。

在林的研究中, 值得关注的是其实验主要采用了朗读语料, 并没有设定真实语境, 朗读者的语气带有表演夸张的成分。

胡方[12]的研究, 不同于林的自然对话的录音分析, 采用了设计文本, 由 4 个普通话发音人朗读。分析主要

集中在疑问词位置, 语调重音等方面。其研究结果显示焦点成分在所有录音句中都保持或加强其自身的词调基频拱度曲线。在表示惊讶的回声话语中, 发音人普遍将整个句子的调域提高, 且扩展了疑问词的基频拱度上升幅度借以表示“惊讶”的情绪意义。

Yuan 等[13]以及 Liu 和 Xu [14]在对比中文陈述句和疑问句的韵律信息差中指出, 疑问句的语调曲拱要高于陈述句曲拱, 且在句末音节的调域上, 疑问句明显高于陈述句。然而其实验结果显示边界调并不一定可以直接显示两种句式在韵律层面的差异。

对于上述研究的回顾, 我们发现对于回声问句的研究中, 正如林茂灿所说, 疑问语气与陈述语气是否突变, 实则为感知范畴的问题。

随着语音识别技术的进一步发展, 越来越多的移动终端上都实现了语音输入的功能, 产生了大量的语音识别文本。一个重要的研究就是给这些文字加上不同的句型标点, 增加文本的可读性和用户体验, 同时利于后续自然语言处理模块对于信息的使用。陈萧等[15]的研究中采用了纯文本信息添加中文口语句子句末标点的方法, 建立起全局词汇信息与标点关系的模型, 并且应用多层感知器来建立了标点模型, 进而实现了标点(句号, 问号和叹号)的生成。从言语工程的角度来讲, 在语音信息中总结出结构信息以及语气信息, 将极大的帮助其训练测试模型, 高效且正确的判断自然语言类型, 加强语音识别, 最终实现人机顺畅交流。

因此, 随着移动互联应用的发展, 人机语音交互应用走向实用, 有待我们在朗读口语的基础上, 进一步探索自然口语对话中的疑问句的编码和解码过程。本研究通过感知实验, 初步探索在自然口语对话中回声问的疑问信息与语境和语调韵律的关系。为此, 我们对实际口语对话中的回声问进行 2 个感知实验。实验 1 是在真实对话语境中感知回声问的疑问程度, 实验 2 是在离境情况下感知这些回声问的疑问信息, 即将对话中的回声问句抽出后判断其是否为疑问句。

根据研究目的, 本文提出 2 个假设, 假设 1: 如果在语境中感知的疑问结果优于离境情况, 那么说明语境的上文信息对回声问解码所起的作用更大, 这里上下文信息(context)包括句法语义和韵律的上下文信息, 两者均有可能对解码起作用, 到底哪个作用更大还需要进行进一步分析; 假设 2: 如果离境情况下感知的疑问信息结果与具有上下文语境的情况下不存在差异, 则说明语境信息对回声问感知作用不大, 回声问的解码主要是语调韵律特征起作用。

## 2. 感知实验及结果

### 2.1. 自然口语对话库

本研究采用的语料来自自然口语电话库 CSSAI-CASSIL[16, 17]。该语音库是国内第一个具有言语行为标注的自然口语语音库, 覆盖了 5 个服务应用领域, 包括机场信息咨询、餐厅订餐服务、旅馆服务、出租车服务和旅行社服务等 700 多段对话。

### 2.2. 回声问实验材料

#### 2.2.1. 上下文语境环境对话语料

在上述口语对话库中, 按照对话行为标注挑选 104 组对话, 用于具有上下文语境的感知。由于原始库中对话长度都相对较长, 为了便于感知, 将挑选的对话再进行截

取, 截取的原则是保证意义完整, 并且不少于 6 个话轮。104 组对话含有待听辨的 110 个回声问, 和其对应的陈述句共 215 句, 另外还有 56 句干扰句, 干扰句包括陈述句、各种类型的疑问句、祈使句和感叹句。如: 出租车服务场景截取的一段对话, 其中加粗加黑的句子是待感知的目标句:

**B: 然后就是这个时间就是这个地点呀 和那什  
么都一样 我就要明天早上起来八点一刻**

A: 明天八点一刻

B: 啊 对 还有一个明天下午也是那个

A: 早晨八点一刻

B: 对

A: 然后也是去那个么

B: 对对 全都一样

A: 么零么六么 是吧

B: 么六么 啊

#### 2.2.2. 离境回声问句

将对话中的回声问切分出来, 形成离境回声问句 110 句和与之对应的句子 102 句(包括陈述句 92 句, 是非问句 7 句, 正反问句 2 句以及特指问句 1 句)。另外截取了 38 个不同的祈使句和感叹句作为干扰句。最终选定 324 个句子作为离境刺激。

## 2.3. 被试

我们招募了 16 名(8 男 8 女)大学生被试参加了有语境环境的感知实验, 他们来自北京和天津的高校, 平均年龄 26 岁, 母语为北方官话区, 普通话水平在二级甲等以上。另外招募了 16 名(8 男 8 女)大学生被试参加了离境环境的感知实验, 他们来自天津的高校, 平均年龄 24 岁, 母语为北方官话区, 普通话水平在二级甲等以上。

## 2.4. 实验过程

对于有语境的感知实验, 每个被试要对这 104 组对话进行听辨, 每一组对话都有相应的文字, 听完对话后, 听辨人对其中加黑加粗的目标句进行疑问还是陈述语气的判断, 并给出语气强弱得分, 从强到弱分为 5 级: 5 表示语气最强, 1 表示语气最弱。被试在正式实验前有指导练习, 之后进入正式感知测试。被试播放声音的耳机型号均为森海塞尔 PC163D 被试可以自己操作对话的播放次数。

离境感知实验采用 E-Prime 软件进行。每个刺激播放后, 被试判断是疑问句还是非疑问句, 并用左手和右手食指按“F”和“J”键进行选择, 这两个按键分别表示“是”和“否”, 感知结果得分分别对应为“1”和“0”。软件还记录被试的反应时间。324 个刺激随机分为 3 组播放, 组间休息 5 分钟。

## 2.5. 感知结果与分析

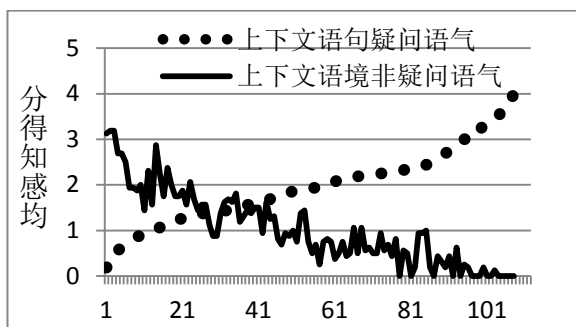
### 2.5.1. 对话语境中回声疑问语气的感知结果

将对话中 108 句目标回声问句的感知结果和得分进行统计分析, 按照这些句子的疑问语气得分进行排序, 结果如图 1 所示。图中横坐标表示刺激排序后编号, 纵坐标表示感知平均得分可以看到, 对话中的回声问句的疑问语气程度不同, 随着疑问语气的不断增强, 陈述语气不

断减弱。回声问的疑问语气如果按照感知 5 级评分，则疑问语气平均分布在 0.2~4.1 之间。

陈述和疑问语气的辨认曲线相交于 30%(32/108) 左右的位置，即有 30%的回声问句，其疑问语气小于陈述语气，疑问语气得分小于 1.5；70%的回声问句的疑问语气大于陈述语气，疑问语气大于 1.5 分。

图 1:在有上下文语境信息的对话中回声问句感知为疑问和陈述语气的辨认曲线



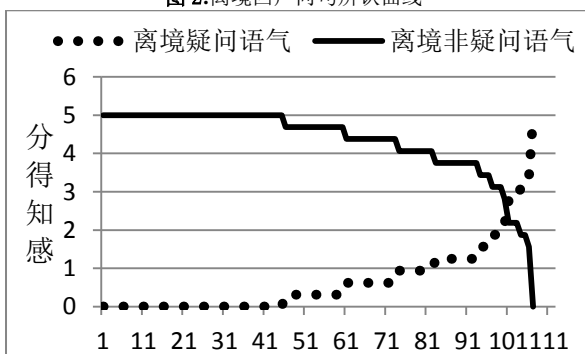
### 2.5.2. 离境回声问感知结果

在引言部分我们介绍了林茂灿对于朗读的离境回声问的研究。这里我们主要考察自然对话中回声问的离境感知情况如何。

图 2 统计了 108 句回声问在离境即孤立句环境下的 16 位被试的听辨结果。首先看到，这些回声问的疑问程度也有差异，分布在 0~16 之间。如果按照 5 度得分规整相当于 0~5 分。这些离境的回声问，随着感知到的疑问程度的增强，感知到陈述语气的程度也在减弱。

感知为陈述和疑问句的辨认曲线相交于 101-102 号刺激之间，也即 93%的回声问被感知为陈述语气的得分高于疑问语气；只有 7%的回声问疑问语气大于陈述语气。在 93%刺激中，有 40%的回声问 100%被感知为陈述句；约一半 50%的回声问能够感知到有疑问语气，但是疑问语气程度低于陈述语气。

图 2:离境回声问句辨认曲线



### 2.5.3. 感知结果对比分析

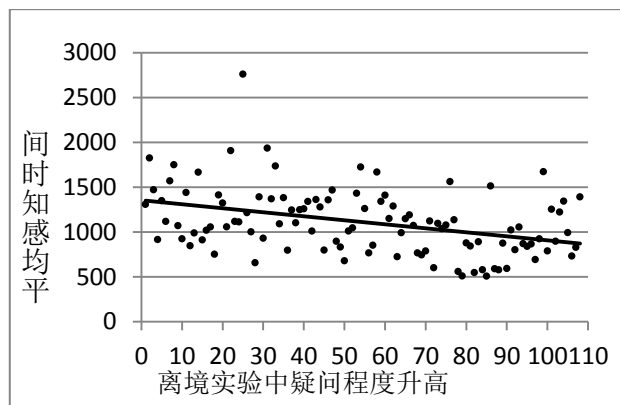
为了对比对话语境和离境下情况下的感知结果，研究采用了卡方检验。结果显示在两个感知实验中，判断为陈述句的结果差异不大 ( $df = 440, p\text{-value} = 0.9014$ )。然而两个实验在疑问句的判断上显示了非常显著的差异 ( $df = 528, p\text{-value} = 2.704e-05$ )，即对话语境和离境对回声问的疑问程度感知有显著的作用。这也从另一个方面证实了图 1 和图 2 的结果，即上下文信息对回声问的判断起到了显著作用。

E-Prime 软件同时也记录了被试对每个刺激的判断反

应时间。其关系如下面图 3 所示。横轴为根据刺激句感知所得等到的疑问程度分布，即从左至右疑问程度加强。纵轴为判断的反应时间，单位为毫秒。

从图 3 中的回归线中 ( $r^2=0.2192$ ) 可以清晰地看到，在离境环境下，疑问得分越高，对其判断所需的反应时间相对越短。反之，在离境环境下，失去了语境的信息的帮助，没有显著的声学疑问信息的刺激句使被试更难以判断其性质，所需要的反应时间也越长。

图 3:离境实验中反应时间与疑问程度感知关系。



## 2.6. 上下文语境信息与疑问解码的关系讨论

对离境和上下文语境的回声问感知结果支持假设 1，即在无句法标记的情况下，上下文语境信息对这类回声问的感知起到了决定的作用。在上下文信息中，句法语义和语调韵律等对回声问的疑问信息解码都有可能起作用。这里我们简单描述一下观察到的一些现象，大量统计分析还在进行中。

从口语语篇结构来看，比如在例 1 中，回声问特征除了重复前一句或者前几句的全部或者部分内容外，后面接着的话轮一般出现应答言语行为，表示肯定或者否定回答，如“对、嗯、没错、不对、对的、不不……”等。这些句法和结构特征，是判断回声问的一些有效标记。

从语调韵律看，上下文韵律特征对比，包括相同说话人的调阶和调域的变化、或者边界调的变化，都可能产生感知上的对比，从而表达疑问信息。

在此，我们分析了同一内容的回声问和对应的陈述句在边界调上的差异。如图 4 所示，分别为阴平、阳平、上声和去声的陈述和回声问边界调时长归一化值。蓝色为陈述句边界调，红色为回声问边界调。由于是不同发音人的句末边界调对比，这里只能看调型的差异，调阶不具有可比性。

在图 4 中我们可以看到，在句尾为非焦点的情况下，回声问的边界调在和陈述句相比，调型的尾部都有抬升趋势。尤其在阳平和上声中更为明显。这也印证了林等的结论，即疑问句的边界调和陈述句相比有显著斜率变化。

我们又比较了句尾词为焦点的情况，去声边界调对比。见图 5 所示。可以发现回声问调阶较高、 $F_0$  斜率变大，调域也相对更宽，其边界调的斜率的绝对值比其陈述句有显著增加。在后续的研究中，我们还将探讨焦点位置对回声问的影响。

因此我们认为，说话者通过上下文语境，结合句法语义和韵律特征共同作用来表达没有标记的回声问，听音人也是通过上下文语境中的句法语义和韵律特征共同



解码获得疑问信息的。

图 4:回声问和陈述句边界调非焦点调型对比 (三角为回声问, 方块为陈述)

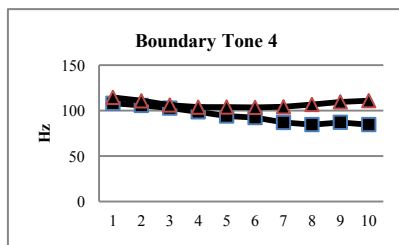
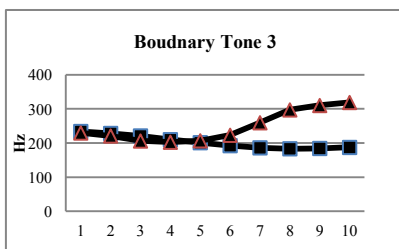
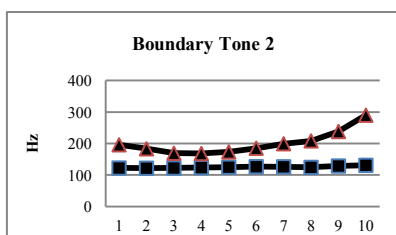
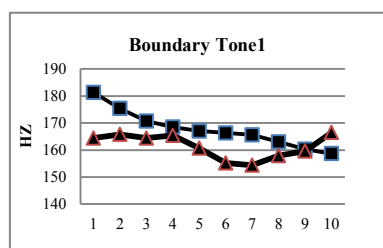
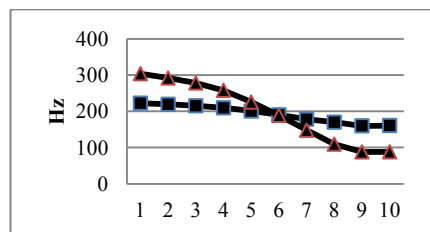


图 5:回声问和陈述句边界调为焦点去声变化 (三角为回声问, 方块为陈述)



### 3. 结论

本研究对自然口语对话中的回声问句的解码进行了初步探索。结果与林茂灿对朗读的句子的 Gating 实验得到的结果有很大的差异[6], 朗读语体的回声问在离境环境下的回声问疑问程度远远高于自然口语对话中回声问离境的疑问程度。说明, 对话语体不同, 疑问信息传递的编解码方式也不同。在自然口语交际中, 疑问信息的传递方式更多依赖语境信息。

通过在上下文语境和离境两种情况中回声问句的疑问信息的感知, 发现 (1) 回声问无论在上下文语境还是在离境情况下, 其疑问语气都有程度的不同; (2) 在离境情况下, 绝大部分 (93%) 的疑问程度较低 (低于非疑问的感知), 而其中还有约 50% 不能被感知为疑问句 (疑问语气为 0)。在上下文语境信息中, 所有的回声问都可以感知到有疑问语气的存在, 大部分 (70%) 的疑问语气大于陈述语气; (3) 上下文的语境信息对无标记的回声问句的感知起决定作用。包括上下文的语篇结构和语义、语调韵律等特征将综合起作用。

要揭示回声问句的疑问信息传递的方式和规律, 还需要进一步的研究。比如, 从图 4 的几个边界调对比实例看到, 尽管边界调在陈述和疑问语气有斜率等特征的变化, 但这种变化与朗读语体的差异, 以及与疑问语气强弱的关系还有待分析, 同时, 也要对回声问的言语行为或者语用功能进行分类研究。

### 4. 致谢

本项目得到国家 973 课题“互联网环境中中文言语感知与表示理论研究”(2013CB329301)、中国社会科学院创新工程项目“语音与言语科学重点实验室”的支持。

### 5. 参考文献

- [1] Quirk, R. et. al. 1985. A Comprehensive Grammar of the English Language. London and New York: Longman Group Limited.
- [2] 陈治安、文旭(Chen, Zhian; Wen, Xu). 2001.论言语交际中的回声话语.解放军外国语学院学报, (4).
- [3] 谢喜爱(Xie, Xiai). 2003.英语交际中的回音言语叹息.安阳师范学院学报, (6).
- [4] 吕叔湘(Lv, Shuxiang). 1942.《中国语法要略》《吕叔湘文集》, 商务印书馆.
- [5] 李宇明, 唐志(Li, Yuming; Tang, Zhi). 1992. 汉族儿童文具系统习得探微.华中师范大学出版社.
- [6] 林茂灿 (Lin, Maocan). 2006.疑问和陈述语气与边界调.中国语文第四期.
- [7] 丁声树 (Ding, Shengshu). 1961.现代汉语语法讲话.北京: 商务印书馆.
- [8] 王志 (Wang, Zhi). 1990.回声问.中国语文.
- [9] 邵敬敏 (Shao, Jingmin). 1992. “回声问”的形式特点和语用特征分析.华东师范大学学报.
- [10] 刘月华, 潘文乐 (Liu, Yuehua; Pan, Wenyu).2001.实用现代汉语语法(增订本).北京: 商务印书馆
- [11] 冉永平 (Ran, Yongping). 2001.元表征结构及其理解.外语与外语教学. (4)
- [12] 胡方 (Hu, Fang). 2005.普通话疑问词韵律的语音学分析.中国语文第三期 pp:269-288.
- [13] Yuan, Jiahong., Shih, Chilin., Kochanski, Gergory. P. (2002). Comparison of Declarative and Interrogative Intonation in Chinese. Speech Prosody,711-714, Aix-en-Provence, France.
- [14] Liu, F. and Xu, Y. 2005. Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation. Phonetica 62: 70-87.
- [15] 陈萧, 柯登峰, 徐波. (Chen, Xiao., Ke, Dengfeng., & Xu, Bo.)2013.基于全局词汇信息的中文口语句子标点生成.第十二届全国人机语音通讯学术会议.
- [16] Zhou, K.; Li, A.; Zong, C.. 2010. Dialogue-Act Analysis with a Conversational Telephone Speech Corpus Recorded in Real Scenarios. OCOOSDA 2010, Nepal.
- [17] Zhou, K.; Li, A.; Yin, Z.; Zong, C.. 2010. CASIA-CASSIL: a Chinese Telephone Conversation Corpus in Real Scenarios with Multi-levelled Annotation. LREC2010.

(原载 ICPLC 2013中国香港 2013年12月)