

普通话态度语音的感知实验研究

卢 琰 维罗尼克·欧贝吉 阿尔贝·里亚尔 顾文涛

(西安外国语大学 西方语言文化学院,陕西 西安 710128;法国国家科研中心 GIPSA 实验室 法国 Grenoble 38041
法国国家科研中心 LIMSI 实验室,法国 Orsay 91403;南京师范大学 语言信息科技研究中心,江苏 南京 210097)

摘 要:在言语交际中,说话人的情绪(emotions)属于不自主控制的潜在的心理状态,而说话人表达的态度(attitudes)或者称社会情感(social affects)则是说话人自主控制的。本文通过听辨实验,研究了中国人对普通话 19 类态度语音的感知特性。实验采用条件受控的语音语料,考察了态度类别、话语长度、听辨人性别这三个因素对态度辨识率的影响,以及不同态度类别在听感上的混淆模式与聚类关系。实验结果表明,普通话态度语音的辨识率总体上高于机遇水平,其中“陈述”和“失望”的辨识率最高,“自信”的辨识率最低。态度类别的主效应、态度类别与话语长度的交互效应,是影响态度辨识率的主要因素。基于听辨混淆模式进一步做层次聚类分析后发现,19 种态度从语音感知上可大体聚类为 7 组,同组内的态度在认知上表现出一定的关联性。

关键词:普通话 态度 语音感知

中图分类号:H014 **文献标识码:**A **文章编号:**1008-9853(2013)03-169-06

一、引 言

人们在言语交际中表现出的情感,包含了对外部信息认知加工的两个不同层面(Aubergé & Gestalt, 2002):一是不自主控制的情感(involuntarily controlled expressions),即“情绪”(emotions);二是自主控制的情感(voluntarily controlled expressions),即“态度”(attitudes)或称“社会情感”(social affects)。我们认为,情绪与态度是两种截然不同的心理状态。情绪与生俱来,是对外部刺激的本能反应;而态度与语言、文化有非常密切的联系,在不同的社会文化环境中会表现出一定的差异,某些态度的表达方式需要后天的学习。

近年来,一些心理学家和语言学家对态度的分类进行了研究(Wichmann, 2002; de Moraes et al., 2010; Gu and Fujisaki, 2013)。在参考已有研究的基础上,本文定义了生活中常见的 19 种态度,如表 1 所列,并且将这些态度分为以下两大类:

(1) 体现说话者对所言之事的看法或观点的态度。如果说话者没有表达任何特别的态度,也被视为一种中性态度,例如简单的“陈述”和“询问”。

(2) 体现交际双方社会关系的态度,例如“礼貌”和“权威”;或者体现特定交际背景和对象的态度,例如“亲密”、“儿向”和“引诱”。这里“儿向”是指儿向语(即成人与婴幼儿对话时采用的语体)呈现的

收稿日期:2013-06-08

作者简介:卢琰,女,1979 年生,陕西宝鸡人,西安外国语大学西方语言文化学院讲师,法国格勒诺布尔市 GIPSA 实验室博士生。维罗尼克·欧贝吉,女,1960 年生,法国人,法国国家科学研究中心 GIPSA 实验室研究员。阿尔贝·里亚尔,男,1973 年生,法国人,法国国家科学研究中心 LIMSI 实验室研究员。顾文涛,男,1972 年生,江苏扬州人,南京师范大学文学院特聘教授、博士生导师,日本东京大学客座研究员。

基金项目:国家社会科学基金青年项目“普通话与粤语的韵律比较及二语习得韵律偏误分析研究”(10CYY009);江苏高校哲学社会科学重点研究基地重大项目“语音韵律的跨语言比较及二语韵律的偏误分析与自动评测”(2010JDXM024);“江苏省社会科学基金项目“汉语二语习得的语音韵律偏误研究”(09YYB006)”。

态度。

表 1 态度的分类

分类	态度名称(及英文缩写)
表明观点和看法的态度	陈述(DECL) 询问(QUES) 欣赏(ADMI) 自信(CONF) 发怒(IRRI) 妥协(RESI) 蔑视(CONT) 讽刺(IRON) 怀疑(DOUB) 确信(OBVI) 失望(DISA) 中性的惊讶(吃惊)(NEU-S) 积极的惊讶(惊喜)(POS-S) 消极的惊讶(惊愕)(NEG-S)
体现特定社会关系或交际对象背景的态度	礼貌(POLI) 权威(AUTH) 引诱(SEDU) 儿向(IDS) 亲密(INTI)

继 Martins - Baltar(1977) 和 Fónagy(1991) 之后,许多学者研究了不同语言的态度语音韵律(Fujisaki & Hirose, 1993; Mejvaldova, 2000; Diaferia, 2002; Mac et al., 2010; Gu et al., 2011),有一些学者还将研究视野扩展到态度语音韵律的跨语言跨文化比较(Shochi et al., 2009) 以及二语教学(Shochi et al., 2010)。

本文将从听觉感知的角度,研究中国人对普通话 19 种态度语音的辨识模式,继而考察这些态度在听感上的距离与聚类关系。考虑到除了态度类别本身,听辨结果还可能与话语的语言学特征、以及听辨人的特征有关,本文选择话语长度、听辨人性别也作为考察因素。

二、听辨实验

我们以音节数、字调、句法结构为变量,设计了一个条件受控的语料库,包含了单字词、二字词、三字词、四字短语以及九字长句,共 152 条。这些词和句子在语义层面均为中性(语义上不暗示任何态度或情绪),但是在语用层面,可以在不同语境中用各种态度自然地表达。

录音人是一位 32 岁的中国女性,普通话发音标准,而且从事语言教学多年。我们为录音人提供了表 1 所列 19 种态度的定义及细致描述,录音人在录音前经过反复训练。语音的录制在法国格勒诺布尔市 GIPSA 实验室的隔音室中进行,已录制的语音库共有 2888 句(152 句文本 × 19 种态度)。

我们从该语音库中选取了 399 句(21 句文本 × 19 种态度)用于听辨实验。表 2 给出了用于听辨实验的语料构成,其中 NP 表示名词词组,VP 表示动词词组,句法结构的括号里的数字表示词组所含音节数,而声调组合中的 1、2、3、4 表示普通话的四声。在当前实验中,暂不考虑字调和句法结构的影响,只考察音节数即话语长度对态度感知的作用。

表 2 听辨实验的语料构成

	音节数	声调组合	句法结构	数量
词	1	1, 2, 3, 4	—	4
	2	1-1, 1-3, 2-2, 2-4, 3-1, 4-2, 4-4	—	7
短语	4	1-1-1-2 1-1-1-4	NP(4) + VP(0) NP(3) + VP(1) NP(1) + VP(3)	6
长句	9	2-1-1-1-1-1-1-1-2 2-1-1-1-1-1-1-1-4	NP(7) + VP(1) + NP(1) NP(3) + VP(3) + NP(3)	4

参加实验的听辨人共 30 位,男女各半,都是普通话本族语者,无任何听力或智力障碍,年龄区间 20-33 岁,平均年龄 25.2 岁。除了 1 名男性是格勒诺布尔某 IT 公司的电脑程序员外,其余听辨人均均为法国格勒诺布尔大学的在读研究生。

听辨人在实验前,充分熟悉并理解了表 1 所列 19 种态度的定义及细致描述。听辨实验安排在一个安静的房间中,通过一个软件界面操作,听辨人戴耳机逐条聆听每一条语料,判断听到的话语表达的是何种态度,在 19 种态度标签中做单项选择。所有实验语料由程序随机排序,每条语料只播放一次。选择判断的时间不加限制,并且听辨人可以随时要求休息以缓解疲劳。

三、实验结果分析

3.1 辨识率统计分析

我们针对听辨实验的辨识率数据,做完全随机设计的三因素方差分析。三个自变量分别为:态度类别(19 个水平)、话语长度(4 个水平)、听者性别(2 个水平)。每个数据单元的观测数据不少于 60 个。表 3 列出了方差分析的结果,其中效果显著的因素有:(1)“态度类别”对辨识率有极显著的主效应($p < 0.001$),而且 η^2 (效应值)最大,达到 0.693;(2)“态度类别”与“话语长度”之间的交互效应极显著($p < 0.001$),且效应值较大(0.220);(3)“话语长度”、“听者性别”都具有极显著的主效应($p < 0.001$),但是效应值很小,只能解释观察到的方差的很小一部分;(4)“态度类别”与“听者性别”之间的交互效应显著($p < 0.05$,但是接近 1% 水平),效应值也很小。

以上结果说明,态度类别、话语长度、听者性别这三个因素都对辨识率有显著影响,其中作用最大的是态度类别;同时,态度类别与话语长度之间的交互效应较强,表明对于不同长度的话语,辨识率随态度类别的变化规律并不相同。

表 3 方差分析结果(粗体表示该因素有极显著效果 $p < 0.001$)

因素	SumSq	Df	F	p	η^2
态度类别	253.43	18	86.22	0.0000	0.693
话语长度	16.19	3	33.06	0.0000	0.044
听者性别	1.97	1	12.06	0.0005	0.005
态度类别 × 话语长度	80.30	54	9.11	0.0000	0.220
态度类别 × 听者性别	5.57	18	1.90	0.0122	0.015
话语长度 × 听者性别	0.13	3	0.26	0.8574	0.000
态度类别 × 话语长度 × 听者性别	7.87	54	0.89	0.6958	0.021

图 2(a-c)分别给出了以态度类别为变量、以态度类别与话语长度为变量、以及以态度类别与听辨人性别为变量的态度辨识率的统计直方图(同时显示了均值与标准差)。根据图 2 所示,我们具体考察不同态度类别、不同话语长度、以及不同性别听辨人条件下,态度辨识率的变化规律。

(1) 态度类别:如图 2(a)所示,除“自信”以外的 18 种态度,平均辨识率都在机遇水平(即 1/19,如图中虚线所示)以上。其中,“陈述”和“失望”的辨识率最高,超过了 50%。

(2) 话语长度:平均来看,单字词的辨识率最低(18.4%),然后由低到高依次是二字词(26.8%)、九字句(28.0%)与四字短语(28.4%)。进一步做事后检验发现,只有单字与其他话语长度之间辨识率有很显著差异($p < 0.01$),其他话语长度之间辨识率没有显著差异($p > 0.05$)。这是容易理解的,单字信息量太少,所以表达的态度更难判断。但是,表 3 显示“态度类别”与“话语长度”之间有显著并且较

强的交互效应,说明对于不同的态度类别,辨识度随话语长度的变化方式并不相同。这从图 2(b)可以看出。例如,“儿向”(IDS)的辨识结果与总体规律相反,其单字和二字词的辨识度明显高于四字短语和九字句。这很可能是因为对婴幼儿说话一般很短很简单,较长的语句听起来反而不真实,事实上四字短语和九字句更多地被误判为“引诱”(SEDU)。

(3) 听者性别:女性听辨人的平均辨识率为 27.2%,略高于男性的 22.8%。但是,表 3 显示“态度类别”与“听者性别”之间有显著的交互效应,说明对于不同态度类别,辨识度在男女之间的高低关系并不相同。如图 2(c)所示,对于“陈述”(DECL)、“礼貌”(POLI)、“消极的惊讶”(NEG-S),男性听辨人的辨识度反而高于女性听辨人。

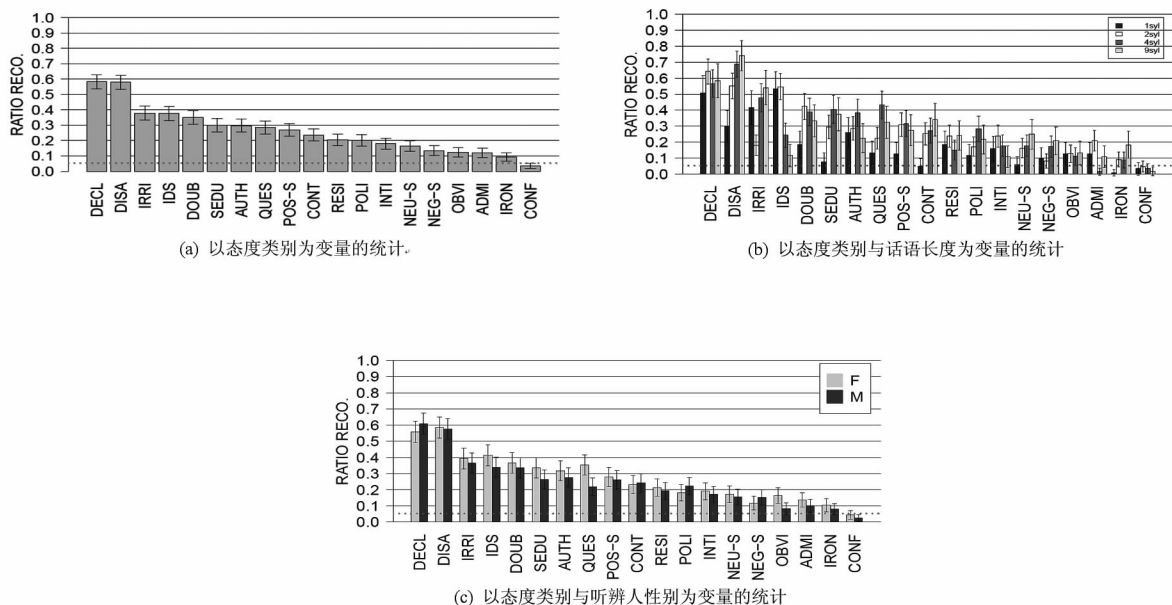


图 2 普通话 19 种态度的辨识度

3.2 态度聚类分析

根据听辨人对不同态度的听辨结果,我们建立了一个混淆矩阵,并据此绘制了图 3,以说明各种态度的辨识度及相互之间的混淆关系。图中只标示了混淆率在机遇水平 2 倍(即 $2/19 \approx 10.5\%$)以上的关系。由图 3 可观察到以下的主要混淆关系:

(1) 半数以上的态度容易被误判为“陈述”(DECL),尤其是“自信”(CONF)(56.0%)和“礼貌”(POLI)(49.2%);相反,“陈述”(DECL)只是较少地被误判为“确信”(OBVI)。(2)“中性的惊讶”(NEU-S)、“积极的惊讶”(POS-S)和“消极的惊讶”(NEG-S),都容易被误判为“怀疑”(DOUB),而后者与“询问”(QUES)也容易相互混淆;同时,“积极的惊讶”(POS-S)、“消极的惊讶”(NEG-S)、“怀疑”(DOUB)、“询问”(QUES)也容易被误判为“中性的惊讶”(NEU-S)。但是,很少有被误判为“积极的惊讶”(POS-S)或“消极的惊讶”(NEG-S)的情形。(3)“讽刺”(IRON)和“蔑视”(CONT)容易相互混淆。(4)“儿向”(IDS)与“引诱”(SEDU)容易相互混淆,后者还可能被误判为礼貌”(POLI)。(5)“妥协”(RESI)和“失望”(DISA)容易相互混淆,但是前者被误判为后者的比率要大得多。(6)“权威”(AUTH)和“发怒”(IRRI)容易相互混淆。

很多态度被误判为“陈述”,这与越南语(Mac et al., 2010)和英语(Diaferia, 2002)态度语音感知实验的结果类似。究其原因,很可能与实验任务的复杂性有关,因为从 19 种态度标签中做单项选择,在认知上本来就有一定难度,特别是在缺乏上下文语境的情况下。所以,当听辨人无法确定感知到哪种态度

时,选择“陈述”是最可行的回答。关于其它态度之间的混淆关系,将在下文做聚类分析后讨论。

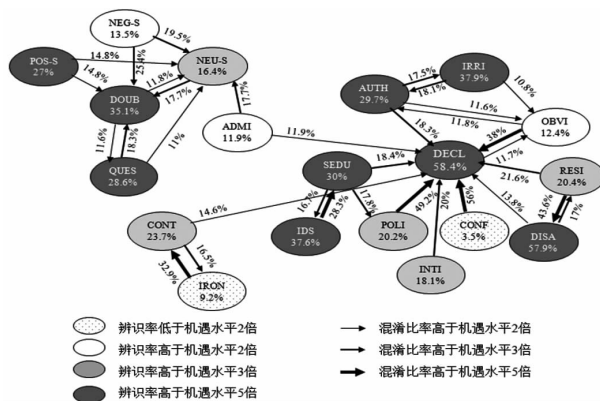


图3 普通话 19 种态度的混淆关系

椭圆形内为态度名称及辨识度,箭头标示混淆关系,箭头上的百分比表示混淆率(图中只标出在机遇水平 2 倍即 10.5% 以上的)。

在混淆矩阵的基础上,我们进一步做层次聚类分析,以考察各种态度之间的听感距离及聚类关系。各种态度的听感距离,可由混淆矩阵各向量间的相关系数确定。设两个行向量的相关系数为 r 。由于态度越相似、听感距离就越短,于是定义这两种态度的听感距离为 $1 - r$ 。计算出听感距离矩阵后,用软件 R 的程序对该矩阵进行层次聚类。聚类分析结果如图 4 所示。结果表明,19 种态度在听感上大体可以划分为以下 7 组:

(1)“欣赏”(ADMI)和“积极的惊讶”(POS-S):这两类均是强烈的积极情感,与积极、主动的评价相联系。

(2)“消极的惊讶”(NEG-S)、“中性的惊讶”(NEU-S)、“询问”(QUES)和“怀疑”(DOUB):这四类态度都和一种无预期的或者与预期不符、不确定的心理状态相关。

(3)“儿向”(IDS)和“引诱”(SEDU):这两类均指向特定背景的交际对象,“儿向”的对象为婴幼儿,“引诱”的对象为具有吸引力的异性。这两类态度的表达都经常使用气嗓音来暗示关心与亲密(Wichmann, 2000; Campbell, 2004)。

(4)“权威”(AUTH)和“发怒”(IRRI):“权威”是一种旨在突显说话者较高社会地位的言语行为,这样的社会身份往往可以对较低地位的人表达怒气或不满。因此,这两种态度在言语交际中经常出现重叠。

(5)“妥协”(RESI)和“失望”(DISA):两者均源于外部世界对说话人的否定,都反映了某种消极、被动的心理状态。

(6)“讽刺”(IRON)和“蔑视”(CONT):都是说话人表达不友好或者对立的姿态。Mac et al. (2010)关于越南语态度感知的研究也得到了这一聚类关系。与“失望”和“妥协”相比,这两类态度虽然也是消极的,但是表达方式是主动的。

(7)其余的态度被归于一组,这些态度有“陈述”(DECL)、“自信”(CONF)、“礼貌”(POLI)、“确信”(OBVI)和“亲密”(INTI)。值得注意的是,我们原本设想“亲密”会与“儿向”和“引诱”这两类态度很接近,但是实验结果表明并非如此。

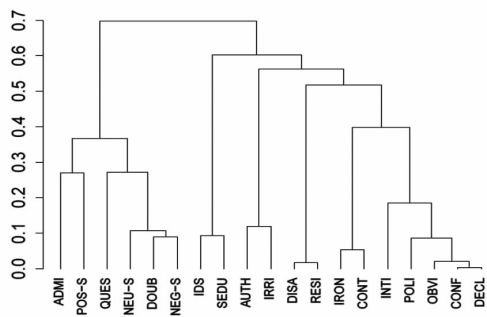


图4 感知态度的聚类分析

四、结 论

态度(或称社会情感)与情绪不同,态度是说话者自主控制的,有些需要在特定的社会文化环境中习得。本文通过听辨实验,研究了普通话 19 种态度语音的感知特性,深入分析了各种态度之间的听辨混淆模式和聚类关系。

实验发现,除“自信”外,其余 18 种态度的辨识率均高于机遇水平。方差分析表明,态度类别的主效应、态度类别与话语长度的交互效应,对态度的辨识有极显著影响并有很高的效应值。此外,辨识率受话语长度的影响,体现在单字的辨识率与较长语句相比显著更低;辨识率受听辨人性别的影响,体现在女性的辨识率在总体上略高于男性。基于各种态度之间的听辨混淆关系做进一步的聚类分析,发现 19 种态度可以大致划分成 7 组,同组内的各种态度在认知上表现出一定的关联性。

在下一阶段的工作中,我们将开展跨语言跨文化的感知实验,对比中、外听辨人对普通话态度语音的感知差异,同时,进一步将感知实验扩展到多模态,考察态度信息的多种表达方式之间的相互关系。

参考文献:

- Aubergé, V. (2002). A gestalt morphology of prosody directed by functions. *Proc. Speech Prosody*, pp. 151 – 154, Aix – en – Provence, France.
- Campbell, N. (2004). Perception of affect in speech – towards an automatic processing of paralinguistic information in *Proc. INTERSPEECH*, pp. 881 – 884, Jeju Island, Korea.
- Daneš, F. (1994). Involvement with language and in language. *Journal of Pragmatics*, 22: 251 – 264.
- Diaferia, M. L. (2002). *Les Attitudes de l'Anglais: Premiers Indices Prosodiques*. Master thesis, National Polytechnique Institut of Grenoble, France.
- Fónagy, Y. (1991). *La Vive Voix*. Paris: Payot.
- Fujisaki, H. and Hirose, K. (1993). Analysis and perception of intonation expressing paralinguistic information in spoken Japanese. *Proc. ESCA Workshop on Prosody*, pp. 254 – 257, Lund, Sweden.
- Gu, W., Zhang, T., and Fujisaki, H. (2011). Prosodic analysis and perception of Mandarin utterances conveying attitudes. *Proc. INTERSPEECH*, pp. 1069 – 1072, Florence, Italy.
- Gu, W. and Fujisaki, H. (2013). Data acquisition and prosodic analysis for Mandarin attitudinal speech. In: F. Shi and G. Peng (eds.), *East Flows the Great River: Festschrift in Honor of Prof. William S – Y. Wang's 80th Birthday*, pp. 483 – 500, Hong Kong: City University of Hong Kong Press.
- Mac, D. K., Aubergé, V., Rilliard, A. and Castelli, E. (2010). How prosodic attitudes can be recognized and confused; Vietnamese multimodal social affects. *Proc. the 2nd International Workshop on Spoken Language Technologies for Under – Resourced Languages*, Penang, Malaysia.
- Martins – Baltar, M. (1977). *De l'énoncé à l'énonciation: une approche des fonctions intonatives*. Paris: Didier.
- Mejvaldova, J. (2000). *Expressions prosodiques de certaines attitudes en thèque et en fran? ais: étude comparative*. Ph. D. thesis, Université Paris 7 – Denis Diderot, Paris.
- de Moraes, J. A., Rilliard, A., Alberto, B. and Shochi, T. (2010). Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. *Proc. Speech Prosody*, paper 340, Chicago, IL.
- Shochi, T., Gagnié, G., Rilliard, A., Erickson, D. and Aubergé, V. (2010). Learning effect of prosodic social affects for Japanese learners of French language. *Proc. Speech Prosody*, paper 155, Chicago, IL.
- Shochi, T., Rilliard, A., Aubergé, V. and Erickson, D. (2009). Intercultural perception of English, French and Japanese social affective prosody. In: S. Hancil (ed.), *The Role of Prosody in Affective Speech (Linguistic Insights vol. 97)*, pp. 31 – 59, Bern: Peter Lang AG.
- Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. *Proc. ISCA Workshop on Speech and Emotion*, pp. 143 – 148, Newcastle, North Ireland.

attitudes could be clustered approximately into seven groups, in each of which the attitudes showed correlated cognitive functions.

责任编辑:虞 澄

Perceptual Study for Mandarin Attitudinal Speech

Yan Lu Véronique Aubergé Albert Rilliard Wentao Gu

Abstract: In contrast to emotions which are involuntarily controlled by the speaker, attitudes (i. e., social affects) are voluntarily controlled and their expressions are usually dependent on language and culture. By conducting a perceptual experiment with Chinese native subjects, the present study examined the perceptual characteristics for Mandarin speech conveying 19 types of attitudes. With a controlled speech corpus, the effects of three factors, i. e., the type of attitude, the length of utterance, and the gender of subject, were investigated. The experiment showed that the 19 attitudes were basically well identified; among them, ‘Declaration’ and ‘Disappointment’ showed the highest rate of identification, while ‘Confidence’ showed the lowest rate of identification. The main effect of the type of attitude, together with the interactive effect between the type of attitude and the length of utterance, plays a major role in the variation of the rate of attitude identification. Based on the patterns of perceptual confusion among attitudes, a hierarchical clustering suggested that all 19 types of attitudes could be clustered approximately into seven groups, in each of which the attitudes showed correlated cognitive functions.

Key words: attitude, speech perception, Mandarin, identification, clustering