

# 非同时期普通话语音对说话人辨别及自评置信度的影响

王丹琳<sup>1</sup>, 柯青青<sup>2</sup>, 曹洪林<sup>3,4\*</sup>

<sup>1</sup>北京警察学院刑事科学技术系, 北京

<sup>2</sup>长沙市天心区人民法院, 湖南 长沙

<sup>3</sup>中国政法大学证据科学教育部重点实验室, 北京

<sup>4</sup>法大法庭科学技术鉴定研究所, 北京

收稿日期: 2026年4月2日; 录用日期: 2026年5月11日; 发布日期: 2026年5月22日

## 摘要

说话人辨别是听者基于听觉线索判断两段语音是否来源于同一说话人的过程。本研究在同异判断(same-different) AX范式下, 构建1年和16年时间间隔的非同时期普通话语音材料, 控制语料长度和说话人性别, 招募42名男性听音被试完成同异判断及置信度评分, 采用广义线性混合模型进行分析。结果表明, 时间跨度显著降低辨别准确率; 语料长度呈“长句 > 短句 > 词语”的非显著提升趋势; 说话人性别对辨别准确率无显著影响; 自评置信度能够显著正向预测判断准确率。信号检测分析结果显示, 被试整体具备一定感知敏感性且个体差异明显, 判别标准整体接近中性。研究表明, 时间跨度会削弱普通话语音的可辨识性, 自评置信度可作为判断可靠性的有效参考指标。

## 关键词

非同时期, 普通话语音, 说话人辨别, 自评置信度, 信号检测理论

# Effects of Non-Contemporaneous Mandarin Speech on Talker Discrimination and Self-Reported Confidence

Danlin Wang<sup>1</sup>, Qingqing Ke<sup>2</sup>, Honlin Cao<sup>3,4\*</sup>

<sup>1</sup>Department of Forensic Science and Technology, Beijing Police College, Beijing

<sup>2</sup>Changsha Tianxin District People's Court, Changsha Hunan

<sup>3</sup>Key Laboratory of Evidence Science (China University of Political Science and Law), Ministry of Education, Beijing

\*通讯作者。

文章引用: 王丹琳, 柯青青, 曹洪林(2026). 非同时期普通话语音对说话人辨别及自评置信度的影响. *心理学进展*, 16(5), 311-323. DOI: 10.12677/ap.2026.165266

## Abstract

**Talker discrimination refers to the process by which listeners judge whether two speech utterances originate from the same talker based on auditory cues. In this study, non-simultaneous Mandarin speech materials with time intervals of 1 year and 16 years were constructed under the AX same-different judgment paradigm. Speech length and talker gender were controlled, and 42 male listeners were recruited to perform same-different judgments and confidence ratings. Generalized linear mixed models were used for data analysis. The results showed that time span significantly reduced discrimination accuracy; speech length exhibited a non-significant increasing trend of “long sentences > short sentences > words”; talker gender had no significant effect on discrimination accuracy; and self-rated confidence significantly and positively predicted judgment accuracy. Signal detection analysis revealed that listeners generally possessed certain perceptual sensitivity with obvious individual differences, and the overall decision criterion was close to neutral. The study indicates that time span weakens the identifiability of Mandarin speech, and self-rated confidence can serve as an effective indicator for judging the reliability of discrimination.**

## Keywords

**Non-Contemporaneous Speech, Mandarin Speech, Talker Discrimination, Self-Reported Confidence, Signal Detection Theory**

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

语音既承载着语言信息，也包含着与说话人相关的声学特征。人们通过声音识别个体的过程被称为说话人识别(talker recognition)，其主要依赖听者对说话人声音的记忆表征，是一种基于记忆的身份识别方法；而说话人辨别(talker discrimination)则是根据基本听觉参数的匹配来判断两个声音是相似还是不同(Hu et al., 2017)。在实验研究中，说话人辨别研究通常采用同异判断(same-different) AX 任务范式，该实验范式不依赖对具体身份的识别，因此被广泛用于评估听者对说话人语音差异的感知敏感性(Park, 2019)。研究听者对说话人语音差异的感知敏感性，对于理解人类如何从高度可变的听觉输入中提取身份相关信息具有重要意义。

既有研究表明，说话人辨别依赖于听者的辨别能力，还受到说话人的声音特征以及多种情境因素的共同影响。自下而上的声学线索是影响语音感知的重要因素。说话人自身声学特征在感知空间中可分离性的显著影响，声学显著性较高的声音更易被正确匹配，且听者对这些判断的自评置信度也更高，且在噪声或语音变换条件下仍具有稳定优势(Stevenage et al., 2018)。自上而下的语言信息同样显著影响说话人辨别表现。例如语音之间的押韵关系或音段重叠程度可能提高感知相似性，从而既促进正确判断，也可能引发“同一说话人”的反应偏向(Narayan et al., 2017; Quinto et al., 2020; Bradshaw et al., 2025)。此外，口音差异亦被证实会显著干扰说话人辨别，使同一说话人更易被误判为不同个体，或不同说话人被误认

为同一人(Santos et al., 2025)。说话人内部变异,尤其是时间跨度所引入的变化,是影响说话人辨别的重要来源。随着时间推移,说话人的声道特征、音色及发音习惯可能发生变化,从而降低不同时间点语音之间的感知相似性。已有研究发现,非同时期语音样本会显著降低说话人辨别与识别的准确率,且该影响在长时间跨度条件下更为明显(Hollien & Schwartz, 2000)。同时,语音风格、发声方式及噪声条件等因素也会进一步加剧这种不稳定性,例如耳语语音由于缺乏关键声学线索会降低辨别能力,而语音风格不一致或背景噪声的引入同样会削弱判断表现(Bartle & Dellwo, 2015; Smith et al., 2019; Stevenage et al., 2021)。此外,说话风格匹配程度及说话人属性(如年龄和性别)还可能通过引发判断偏向影响辨别过程(Afshan et al., 2022; Vyshnevetska et al., 2025)。除语音与情境因素外,听者个体差异同样在说话人辨别中发挥重要作用。研究表明,不同听者在感知敏感性与判断策略上存在显著差异,准确率范围可从接近随机水平到较高水平不等(Schäfer & Foulkes, 2022)。经验丰富的语音学专家通常在复杂或高变异条件下表现更优,并倾向于采用更为保守的判断策略,而普通听者则更容易将不同说话人误判为同一人(Bartle & Dellwo, 2015)。此外,听者熟悉度及语音时长等因素也会影响辨别表现及其置信度评估(Rose & Duncan, 1995; Hollien & Schwartz, 2000)。这些结果表明,说话人辨别不仅是感知过程,还受到个体经验与决策策略的共同作用。

综上所述,既有研究从声学特征、语言信息、说话人内部稳定性与听者差异等多个层面揭示了影响说话人辨别的重要因素。其中关于时间跨度对说话人辨别的影响存在若干不足,一是现有研究以准确率为主要指标,难以区分辨别准确率下降究竟源于感知敏感性降低还是判别准则变化;二是在材料可比性与信息量控制方面也存在局限,可能混淆身份判断与语言处理;三是自评置信度与客观表现之间的关系尚缺乏系统检验。基于上述研究背景与实践需求,本研究在标准 same-different AX 说话人辨别范式下,构建 1 年与 16 年跨时段普通话语音材料,并在实验设计中控制了语料长度和材料类型,以减少语言内容差异对身份判断的干扰。通过对 42 个青年男性听音被试进行说话人辨别实验,采用广义线性混合模型(Generalized Linear Mixed Models, GLMM)以分离听者与刺激层面的随机变异,并结合信号检测理论同时刻画感知敏感性与判别偏向,进一步检验自评置信度对准确率的元认知预测作用,从而更精细地揭示跨时间语音条件下的决策机制变化。

## 2. 实验方法

### 2.1. 被试

本实验共招募 42 名男性听音被试(18~22 岁),均为北京籍在校大学生,普通话标准,无听力障碍。所有被试与发音者互不熟悉。实验所用语音材料由 20 名男性和 20 名女性发音被试录制,所有发音人普通话标准,无明显方言口音和语言障碍。

### 2.2. 语料

本次实验语料长度设计涵盖词语“长安街”(3 音节)、短句“我来到向往的北京”(8 音节)和长句“五、他去无锡市出差,我到黑龙江检查技术工作”(19 音节),以覆盖普通话语音的基本单元,确保比较不同复杂度语音材料下的辨别表现。

### 2.3. 语音采集与处理

语音数据分为实验组与对照组。实验组发音被试分别在约 16 年间隔的两个时间点完成录音,对照组发音被试在约 1 年间隔的两个时间点完成录音。所有录音均在相对一致的录音环境中进行。

为保证不同条件下语音数据的可比性,所有音频均经过统一预处理,包括重采样、音量归一化及格式标准化处理,最终统一为单声道 PCM wav 格式。

## 2.4. 实验程序

实验运行于 PsychoPy 实验平台(2023.2.3 版本), 采用同异判断(same-different) AX 范式。实验分为词语、短句和长句三个条件, 被试依次完成各条件任务。在每个试次中, 被试需判断两段语音是否来自同一说话人(是/否), 并对判断结果进行自评置信度评分(1~10 分)。实验流程见图 1, 实验用语音对在各条件下均衡设置见表 1。

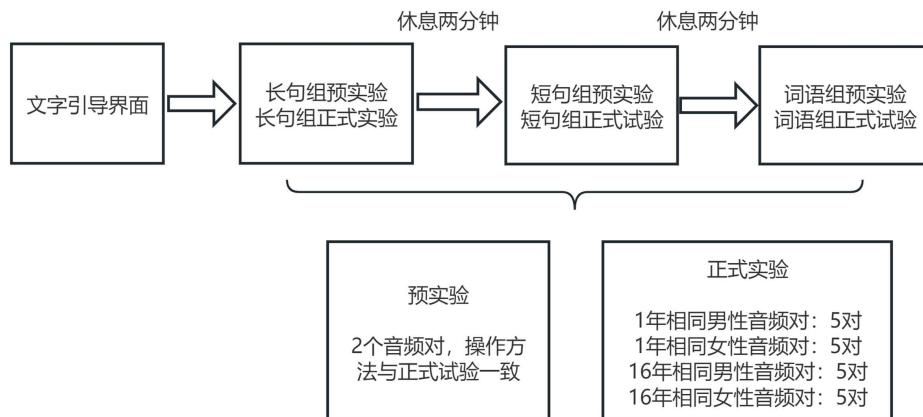


Figure 1. Experimental procedure flowchart  
图 1. 实验流程图

Table 1. Summary of speech pair stimuli  
表 1. 实验用语音对明细表

	时间跨度	性别	同一说话人语音对数	不同说话人语音对数
预实验	不同于正式实验的任意音频对			
正式实验(三组一致)	1 年	男	5 对	5 对
	16 年	男	5 对	5 对
	1 年	女	5 对	5 对
	16 年	女	5 对	5 对

## 3. 实验结果与分析

### 3.1. 数据预处理

Table 2. Data coding scheme  
表 2. 实验数据编码

变量	编号
随机效应: 语音对(Speaker_pair_ID)	1~120
随机效应: 听音被试(Listener_ID)	1~42
固定效应: 语料长度(Length)	词语 = 1; 短句 = 2; 长句 = 3
固定效应: 说话人性别(Speaker Gender)	男性 = 0; 女性 = 1
固定效应: 时间跨度(Interval)	1 年 = 1; 16 年 = 16
连续变量: 自评置信度(Confidence)	1~10

对 42 名听音被试的实验结果进行异常值或缺项的筛查确认所有实验结果均有效。将实验获得的数据统一为一致的数据类型，其中分类变量转换为 R 中的 factor 类型，以便后续计算和分析(见表 2)。

### 3.2. 广义线性混合模型的构建

实验包含多个听音被试对语音对的判断，收集了二元准确率(错误 0 或正确 1)，考虑数据结构中的嵌套和因变量准确率的类型，适合使用逻辑(logistic)回归。GLMM 能够同时处理固定效应，如说话人性别、时间跨度、语料长度、自评置信度和随机效应，如实验参与者个体差异、语音对的内容差异(Suire et al., 2019)。使用 R 语言中 lme4::glmer() 构建模型，对实验中的影响因素进行逐步建模(如图 2)。

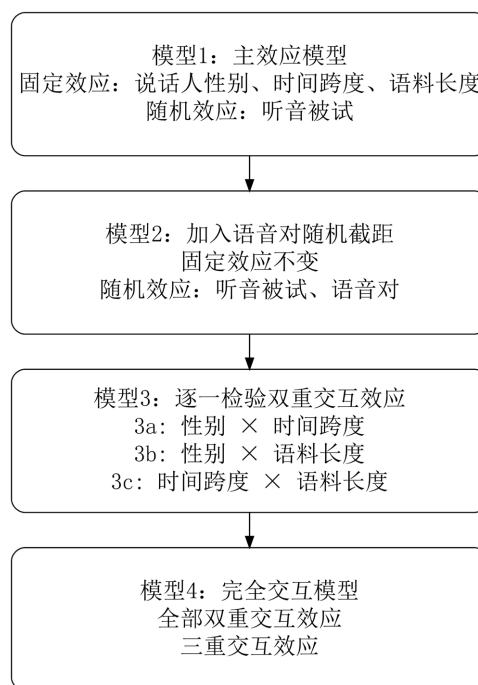


Figure 2. Model construction flowchart  
图 2. 模型构建流程图

### 3.3. 模型的评估和比较

Table 3. Comparison of generalized linear mixed models

表 3. 各广义线性混合模型比较结果

模型	npar	AIC	BIC	logLik	-2LL	$\Delta$ Chisq	$\Delta$ df	$p$
模型 1	6	6359.3	6398.5	-3173.7	6347.3	—	—	—
模型 2	7	5322.7	5368.4	-2654.4	5308.7	1038.6	1	<0.001
模型 3a	8	5322.9	5375.1	-2653.5	5306.9	1.793	1	0.181
模型 3b	9	5325.1	5383.9	-2653.6	5307.1	1.574	2	0.4553
模型 3c	9	5326.2	5384.9	-2654.1	5308.2	0.536	2	0.7649
模型 4	14	5332.4	5423.8	-2652.2	5304.4	0.385	2	0.8251
模型 2 + 时间跨度和语料长度的随机斜率	16	5339.8	5444.2	-2653.9	5307.8	0.952	9	0.9995
模型 2 + 加入说话人性别的随机斜率	19	5344.4	5468.4	-2653.2	5306.4	2.306	12	0.9988

利用 `anova()` 进行似然比检验, 比较上述四级嵌套模型的拟合优度, 判断加入新变量是否显著提升模型拟合。基于赤池信息准则(AIC)和贝叶斯信息准则(BIC)指标, 结果如表 3 所示, 在仅包含被试随机截距的模型基础上, 加入语音对随机效应显著改善模型拟合( $p < 0.001$ ), 说明语音材料差异对判断结果具有重要影响。进一步加入各类交互项及随机斜率后, 模型拟合未显著改善( $p > 0.05$ )。且出现奇异拟合问题。因此, 最终选择包含听音被试与语音对随机截距的模型作为最优模型。

### 3.4. 最优模型分析

最优模型结果(见表 4、表 5)表明, 时间跨度是唯一显著影响说话人辨别准确率的因素( $p = 0.006$ ), 16 年条件下的判断准确率显著低于 1 年条件, 其优势比(odds ratio, OR)约为 0.49。说话人性别与语料长度均未达到显著水平, 但语料长度呈现随长度增加而提升的趋势。随机效应结果显示, 语音对之间的变异大于听音被试之间的个体差异。各交互项均未显著。

Table 4. Random effects estimates

表 4. 随机效应分析

变量	Name	Var	Std. Dev.
语音对	Intercept	1.819	1.349
听音被试	Intercept	0.046	0.213

Table 5. Fixed effects estimates of the main effects model

表 5. 主效应模型的固定效应结果

变量	Est.	SE	z	p
截距	1.115	0.291	3.835	<0.001
说话人性别(女性)	-0.033	0.258	-0.127	0.900
时间跨度(16 年)	-0.706	0.258	-2.731	0.006
语料长度(短句)	0.117	0.316	0.369	0.712
语料长度(长句)	0.359	0.317	1.134	0.257

原始正确率 (分被试听者聚合)

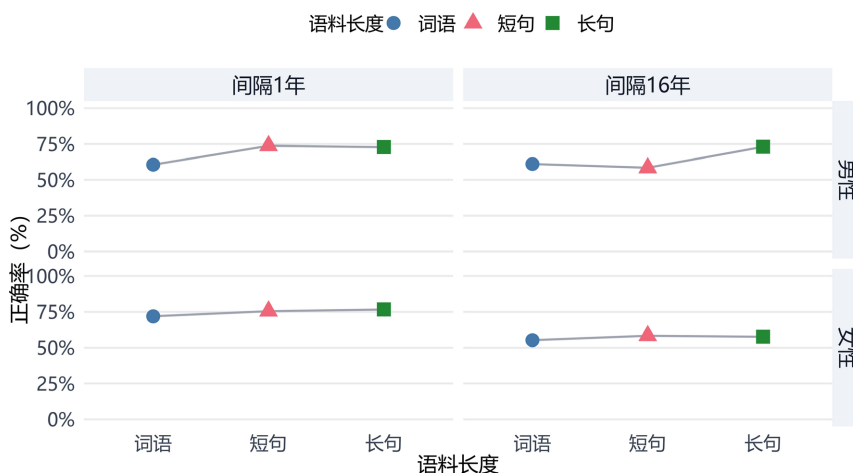


Figure 3. Main effects on accuracy

图 3. 主效应与正确率关系图

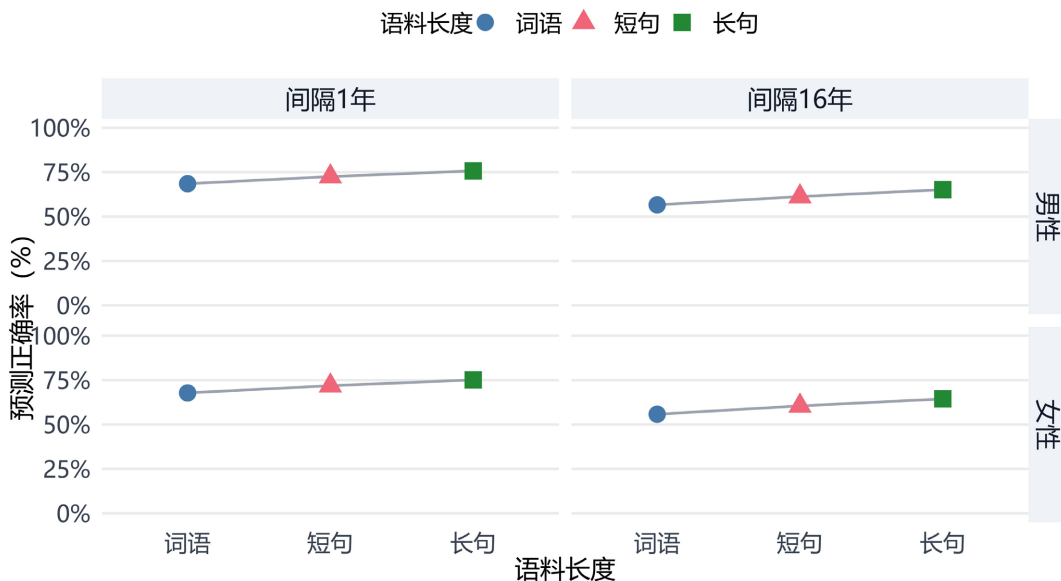
### 3.5. 边际效应分析

在模型 2 的基础上进一步进行了边际效应分析,以直观呈现不同因素条件下的预测准确率。结果见表 6: 边际效应分析结果与主效应模型一致。时间跨度对预测准确率具有显著影响,1 年条件显著高于 16 年条件; 语料长度呈现“长句 > 短句 > 词语”的趋势,但未达到显著水平; 说话人性别及各交互效应均不显著。同时通过可视化方式呈现了不同因素组合下的预测表现(如图 4)。

**Table 6.** Estimated marginal means and pairwise comparisons of main effects  
**表 6.** 三因素主效应的边际均值与组间显著性比较表

变量	emmean	<i>p</i>
男性	0.921	0.899
女性	0.888	
1 年	1.257	0.006
16 年	0.551	
词语	0.746	词语 - 短句 <i>p</i> = 0.928
短句	0.862	词语 - 长句 <i>p</i> = 0.493
长句	1.105	短句 - 长句 <i>p</i> = 0.723

#### 预测正确率的边际效应 (GLMM)



**Figure 4.** Marginal effects of the three factors  
**图 4.** 三因素边际效应趋势图

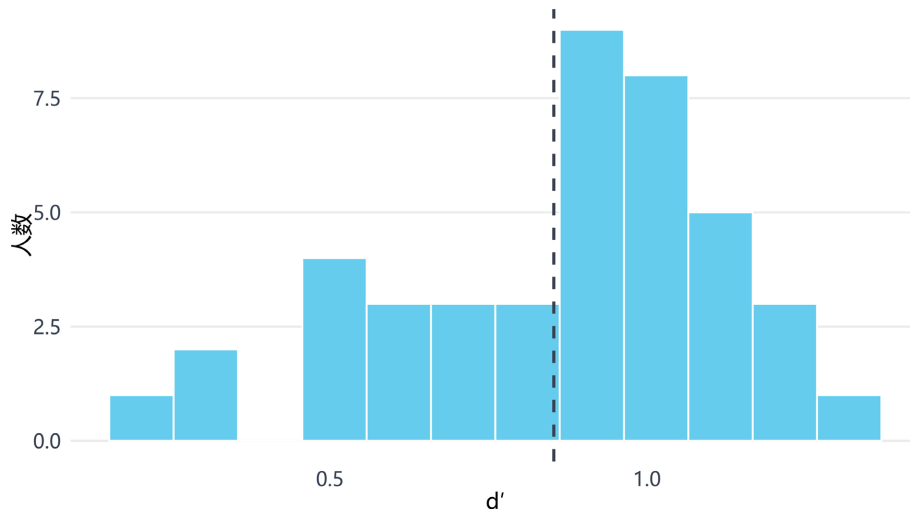
### 3.6. 信号检测理论分析

为了更全面、更深入地理解被试者在说话人辨别任务中的感知能力和决策倾向,本研究进行了信号检测理论(SDT)分析(见表 7)。结果表明,被试整体具有一定感知敏感性(平均  $d' \approx 0.85$ ),表现高于随机水平,但存在明显个体差异(如图 5)。判别标准整体接近中性( $c \approx 0$ ),未表现出显著的反应偏向(如图 6)。

**Table 7.** Descriptive statistics of signal detection measures  
**表 7.** 信号检测指标描述性统计表

变量	Min	Q1	Median	Mean	Q3	Max
d'	0.177	0.672	0.937	0.853	1.023	1.293
c	-0.693	-0.279	-0.089	-0.098	0.017	0.588

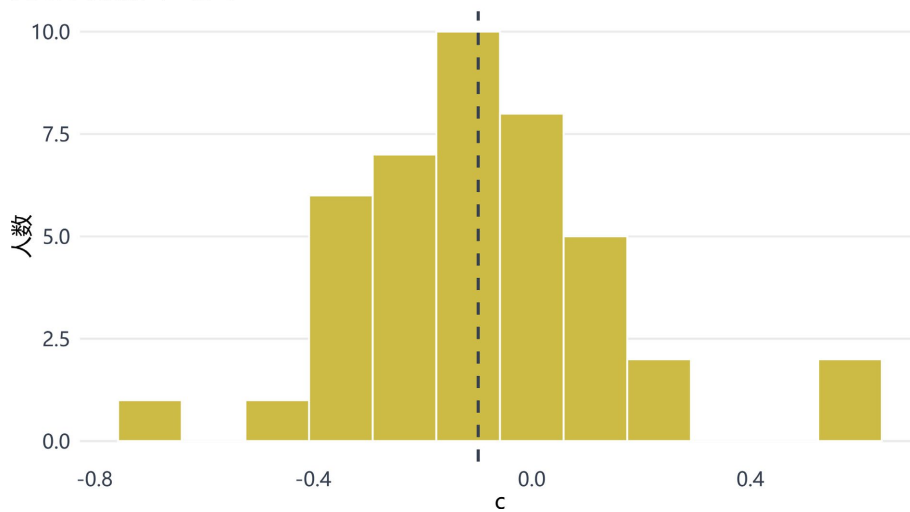
感知敏感性 d' 分布



**Figure 5.** Distribution of d' values

**图 5.** d'值分布直方图

判断准则偏差 c 分布



**Figure 6.** Distribution of criterion c

**图 6.** c 值分布直方图

## 4. 自评置信度分析

### 4.1. 自评置信度的线性混合效应模型

自评置信度是一个连续型变量，因此使用线性混合效应模型(Linear Mixed-effects 模型, LMM)，对听

音被试和语音对的随机效应做了控制, 研究说话人性别、时间跨度、语料长度及其交互作用, 是否会显著影响听音被试自评置信度。

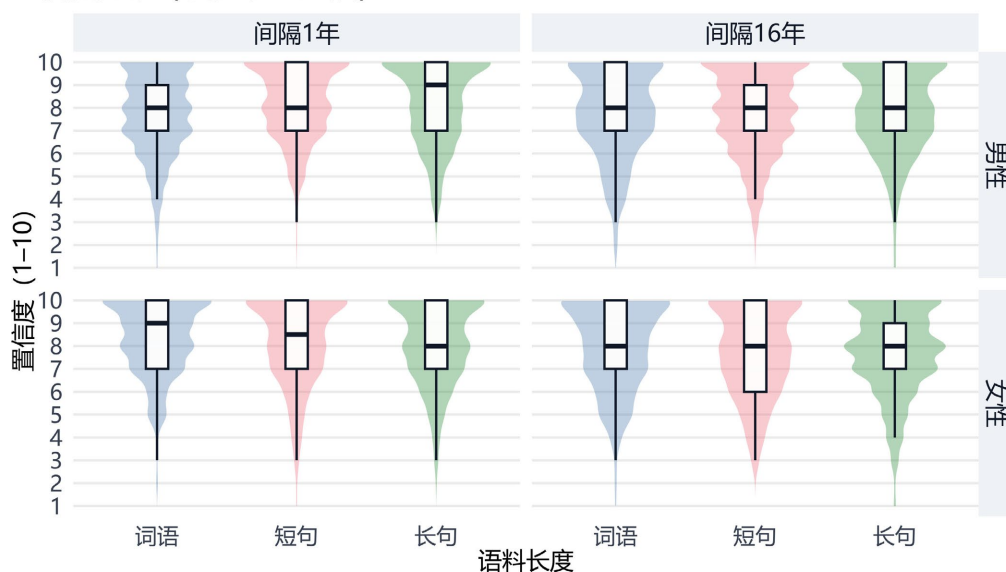
实验结果如表 8 所示。结果表明, 说话人性别与语料长度对自评置信度具有显著影响, 而时间跨度及其交互项均不显著。整体而言, 女性说话人及较长语料条件下的置信度评分更高(见图 7)。

**Table 8.** Linear mixed-effects model results for confidence ratings

**表 8.** 置信度评分的线性混合模型分析结果

变量	<i>Est.</i>	<i>p</i>
截距	7.75	***
说话人性别(女性)	+0.61	0.007
时间跨度(16 年)	+0.04	0.855
语料长度(短句)	+0.54	0.017
语料长度(长句)	+0.51	0.024
说话人性别(女性): 语料长度(短句)	-0.72	0.024
说话人性别(女性): 语料长度(长句)	-0.74	0.020
其余交互作用项	$p > 0.1$	Not significant

置信度分布 (小提琴 + 盒须)



**Figure 7.** Distribution of confidence ratings across conditions

**图 7.** 不同条件下置信度评分分布图

#### 4.2. 自评置信度对准确率的预测作用

本文在已有的主模型基础上加入自评置信度作为一个固定效应变量, 检验模型拟合指标是否有显著改善, 从而自评置信度是否显著提升了对准确率的预测能力。结果见表 9: 加入置信度后, 模型的 AIC 和 BIC 均降低, 对数似然(Log-likelihood)显著提高, 卡方检验显著( $p < 0.001$ ), 说明加入自评置信度后显著改善了模型拟合。尽管自由度仅增加 1, 但模型拟合度的提升幅度较大, 说明自评置信度具有较高的信息量。

**Table 9.** Comparison between the baseline model and the confidence-enhanced model  
**表 9.** 主效应模型与置信度增强模型的对比分析

模型	AIC	BIC	logLik	Chi-sq	Df	<i>p</i>
模型 2	5322.7	5368.4	-2654.4	—	—	—
模型 2 增强	5304.7	5356.9	-2644.3	20.017	1	<0.001

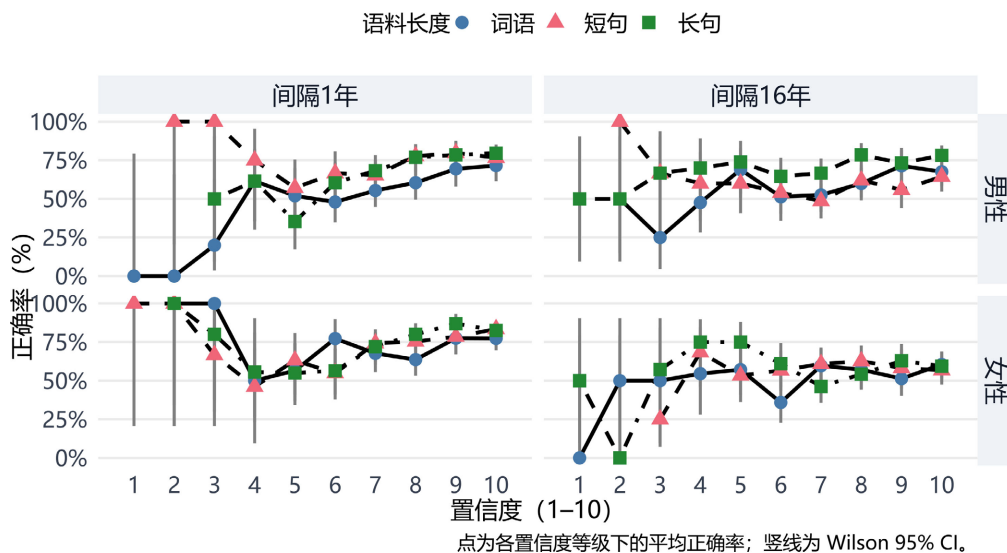
### 4.3. 置信度增强模型分析结果

置信度增强模型对实验结果的分析结果如表 10 所示。时间跨度对说话人辨别准确率具有显著影响，相较于 1 年条件，16 年时间跨度下的判断准确率显著降低。自评置信度对判断准确率具有显著的正向预测作用( $Est. = 0.173, p < 0.001$ )，表明听者在自评置信度较高时更可能作出正确判断。每增加 1 单位标准化信心值，判断正确的优势(odds)约提高 19%。除时间跨度和自评置信度外，其余变量均未对判断准确率产生显著影响。为直观呈现不同实验条件下自评置信度与判断准确率之间的关系，进一步绘制了按置信度等级汇总的准确率变化曲线(见图 8)。

**Table 10.** Results of the confidence-enhanced model  
**表 10.** 置信度增强模型分析结果

变量	<i>Est.</i>	<i>SE</i>	<i>z</i>	<i>p</i>
截距	1.11	0.288	3.855	<0.001
说话人性别(女性)	-0.037	0.255	-0.146	0.884
时间跨度(16 年)	-0.678	0.255	-2.655	0.008
语料长度(短句)	0.114	0.313	0.367	0.714
语料长度(长句)	0.357	0.313	1.142	0.253
置信度	0.173	0.039	4.467	<0.001

置信度—正确率校准曲线 (按置信度等级汇总)



**Figure 8.** Effect of confidence on accuracy  
**图 8.** 自评置信度和准确率关系图

## 5. 讨论

### 5.1. 主要发现和解释

本研究基于广义线性混合模型与信号检测理论的联合分析，系统考察了非同时期普通话语音条件下说话人辨别的感知与决策特征。结果显示，时间跨度、自评置信度以及信息量相关因素在说话人辨别过程中发挥了不同层级的作用，揭示了跨时间语音条件下感知表征与决策机制的变化。

(1) 时间跨度对说话人辨别准确率具有显著影响。听音被试在 1 年间隔条件下的表现显著优于 16 年间隔。结果表明，说话人感知特征并非完全稳定，会随时间推移发生可辨识性衰减。这种衰减可能源于说话人声道特征、音色及发音习惯随时间的变化，产生自然漂移，从而导致同一说话人在不同时间点产生的语音在感知空间中的距离增大。对于听者而言，这种变化会削弱跨时间语音之间的相似性判断，使得“同一说话人”的判断证据变得不充分，从而降低辨别准确率。

(2) 自评置信度是说话人辨别准确率的重要正向预测因子。被试在做出高置信度判断时，更有可能做出正确判断。结果表明，说话人辨别不仅是对外部声学线索的感知区分过程，同时还伴随着对自身判断可靠性的主观评估。听者对自身判断的信心并非完全主观或随机，而是与感知加工过程紧密相关，体现了较为有效的信心校准机制。自评置信度可以作为可作为评估辨别决策质量、理解听者决策状态的重要补充指标。

(3) 语料长度对辨别准确率未产生显著主效应，但呈现“长句 > 短句 > 词语”的非显著提升趋势。结果表明，语料长度可能对说话人辨别具有一定促进作用，但在本研究条件下，该作用尚不足以达到统计显著。主要原因可能是：首先，实验采用的三类材料均为相对受控的朗读语料，虽然在音节数和时长上存在差异，但其语音内容、发音方式和语境复杂度总体较为一致，因此不同长度材料所提供的身份线索增益可能有限。其次，与语料长度所提供的信息量增加相比，时间跨度所引入的说话人内部变异是影响辨别表现的更强因素，从而在一定程度上掩盖了长度效应。最后，虽然较长语料通常能够提供更多韵律、共振峰动态变化及整体发声习惯等线索，但这些线索的利用也取决于任务难度、材料差异幅度及样本量等因素。

(4) 说话人性别未对辨别准确率产生显著影响。结果表明在本研究所采用的普通话受控材料和同异判断任务中，性别相关的声学差异并未显著干扰听者的说话人同异判断。其潜在原因可能是本研究的任务目标是判断两个语音样本是否来自同一说话人，而不是识别说话人的性别类别。因此，尽管男女声在基频、共振峰分布等方面存在一定系统差异，听者在完成同异判断时更可能依赖的是个体层面的身份相关线索，而不是性别这一较为粗略的类别线索。此外，实验所用发音人均均为普通话标准、录音条件相对一致的语音材料，这在一定程度上减少了性别差异与其他语音变异因素之间的耦合，也可能削弱了性别对辨别表现的影响。

(5) 随机效应结果表明，语音对之间的变异大于听音被试之间的个体差异。结果表明相较于听者个体能力的差异，不同语音对本身的可辨识程度差异对判断结果影响更大。某些语音对可能由于说话人声音特征较为鲜明、跨时间变化较小而更容易判断，而另一些语音对则可能由于说话人内部变异较大、语音线索不够突出而更难区分。因此，说话人辨别并非完全由听者能力单方面决定，而是听者感知能力与刺激材料特征共同作用的结果。

### 5.2. 不足与未来展望

本研究揭示了非同时期普通话语音的辨别规律，仍存在若干局限。首先，听音被试样本相对单一，仅包括 18~22 岁的北京籍男性大学生，样本在性别、年龄及语言背景方面的代表性有限，研究结论的外

推性仍需进一步验证。其次, 语音材料类型有限, 实验语料均为朗读文本, 未涵盖自然对话、情绪性语音及带噪声条件, 限制了结果在更复杂听觉环境中的适用性。再次, 实验设计精细度不足, 时间跨度仅设置 1 年与 16 年两个时间梯度以及三类受控语料, 难以揭示中间跨度的变化规律及与真实语境的对应性。此外, 虽然研究发现自评置信度能够显著预测判断准确率, 提示听者具有一定程度的元认知监控能力, 但当前分析主要停留在总体关联层面, 尚未进一步比较不同条件下置信度与正确率之间的一致性是否存在差异。

未来的研究可以在以下几个方面加以深入: 一是可引入更丰富的语音材料类型和听觉条件, 以检验时间效应在更真实语境下的稳健性; 二是扩大被试群体的多样性, 纳入不同性别、年龄和方言背景的听者, 以评估个体差异因素的调节作用; 三是进一步细化时间跨度梯度, 有助于描绘说话人感知表征随时间演变的函数关系, 从而深化对跨时间语音感知机制的理解; 四是结合元认知校准分析, 考察不同实验条件下听者信心判断与客观准确率之间的一致程度, 从而深化对说话人辨别决策质量的理解。

## 6. 结论

本研究得到以下结论: 较长时间跨度会显著削弱听者对说话人一致性的感知判断, 说明说话人感知表征在跨时间条件下并非完全稳定。同时, 自评置信度能够有效反映感知证据的质量, 并对辨别准确率具有显著预测作用, 提示说话人辨别过程不仅涉及感知区分, 也包含重要的元认知评估成分。相比之下, 语料长度和说话人性别的影响相对有限, 表明时间相关的说话人内部变异在该任务中发挥着更为核心的作用。

## 基金项目

国家社会科学基金项目(25BYY079)。

## 参考文献

- Afshan, A., Kreiman, J., & Alwan, A. (2022). Speaker Discrimination Performance for “Easy” versus “Hard” Voices in Style-Matched and Mismatched Speech. *The Journal of the Acoustical Society of America*, 151, 1393-1403. <https://doi.org/10.1121/10.0009585>
- Bartle, A., & Dellwo, V. (2015). Auditory Speaker Discrimination by Forensic Phoneticians and Naive Listeners in Voiced and Whispered Speech. *The International Journal of Speech, Language and the Law*, 22, 229-248. <https://doi.org/10.1558/ijssl.v22i2.23101>
- Bradshaw, L., Chodroff, E., & Dellwo, V. (2025). The Role of Phonetic Overlap for Speaker Discrimination. *The Journal of the Acoustical Society of America*, 157, 3572-3589. <https://doi.org/10.1121/10.0036562>
- Hollien, H., & Schwartz, R. (2000). Aural-Perceptual Speaker Identification: Problems with Noncontemporary Samples. *The International Journal of Speech, Language and the Law*, 7, 199-211. <https://doi.org/10.1558/sll.2000.7.2.199>
- Hu, X., Wang, X., Gu, Y. et al. (2017). Phonological Experience Modulates Voice Discrimination: Evidence from Functional Brain Networks Analysis. *Brain and Language*, 173, 67-75. <https://doi.org/10.1016/j.bandl.2017.06.001>
- Narayan, C. R., Mak, L., & Bialystok, E. (2017). Words Get in the Way: Linguistic Effects on Talker Discrimination. *Cognitive Science*, 41, 1361-1376. <https://doi.org/10.1111/cogs.12396>
- Park, S. J. (2019). *Towards Understanding Voice Discrimination Abilities of Humans and Machines*. Doctoral Dissertation, University of California.
- Quinto, A., Abu El Adas, S., & Levi, S. V. (2020). Re-Examining the Effect of Top-Down Linguistic Information on Speaker-Voice Discrimination. *Cognitive Science*, 44, e12902. <https://doi.org/10.1111/cogs.12902>
- Rose, P., & Duncan, S. (1995). Naive Auditory Identification and Discrimination of Similar Voices by Familiar Listeners. *The International Journal of Speech, Language and the Law*, 2, 1-17. <https://doi.org/10.1558/ijssl.v2i1.1>
- Santos, S. C., Kapadia, A., & Feinberg, D. R. (2025). Hearing People Speak in Different Accents Biases Voice Discrimination. *Scientific Reports*, 15, Article No. 30775. <https://doi.org/10.1038/s41598-025-13117-w>
- Schäfer, S., & Foulkes, P. (2022). The Impact of Voice Recognition Skills on Earwitness Testimony. In *Proceedings of the*

---

*Voice Identity (VoiceID) Conference 2022* (p. 42). University of Zurich.

- Smith, H. M. J., Baguley, T. S., Robson, J. et al. (2019). Forensic Voice Discrimination by Lay Listeners: The Effect of Speech Type and Background Noise on Performance. *Applied Cognitive Psychology, 33*, 272-287. <https://doi.org/10.1002/acp.3478>
- Stevenage, S. V., Neil, G. J., Parsons, B. et al. (2018). A Sound Effect: Exploration of the Distinctiveness Advantage in Voice Recognition. *Applied Cognitive Psychology, 32*, 526-536. <https://doi.org/10.1002/acp.3424>
- Stevenage, S. V., Tomlin, R., Neil, G. J. et al. (2021). May I Speak Freely? The Difficulty in Vocal Identity Processing Across Free and Scripted Speech. *Journal of Nonverbal Behavior, 45*, 149-163. <https://doi.org/10.1007/s10919-020-00348-w>
- Suire, A., Raymond, M., & Barkat-Defradas, M. (2019). Male Vocal Quality and Its Relation to Females' Preferences. *Evolutionary Psychology, 17*, 1-12. <https://doi.org/10.1177/1474704919874675>
- Vyshnevetska, V., Giroud, N., Ramon, M., & Dellwo, V. (2025). Listeners are Biased Towards Voices of Young Speakers and Female Speakers When Discriminating Voices. *Cognitive Research: Principles and Implications, 10*, Article No. 28. <https://doi.org/10.1186/s41235-025-00636-3>