



中国科学院心理研究所
Institute of Psychology, CAS



视觉与计算认知实验室
VISUAL AND COMPUTATIONAL COGNITION LABORATORY

《说谎心理学》

第17讲 语音线索

傅小兰

中国科学院心理研究所
中国科学院大学心理学系

2022年6月24日

课程安排

5次课，每次4课时

◆ 6月20日

- 1 说谎与识谎
- 2 谎言识别的过去、现在和未来
- 3 谎言研究的理论与方法
- 4 说谎的认知
- 5 说谎的情绪反应及生理变化

◆ 6月21日

- 6 说谎的意志过程
- 7 说谎的个体和群体差异
- 8 情境与个体偏好

◆ 6月22日

- 9 说谎得逞的影响因素
- 10 生理反应
- 11 基于面部线索的谎言识别

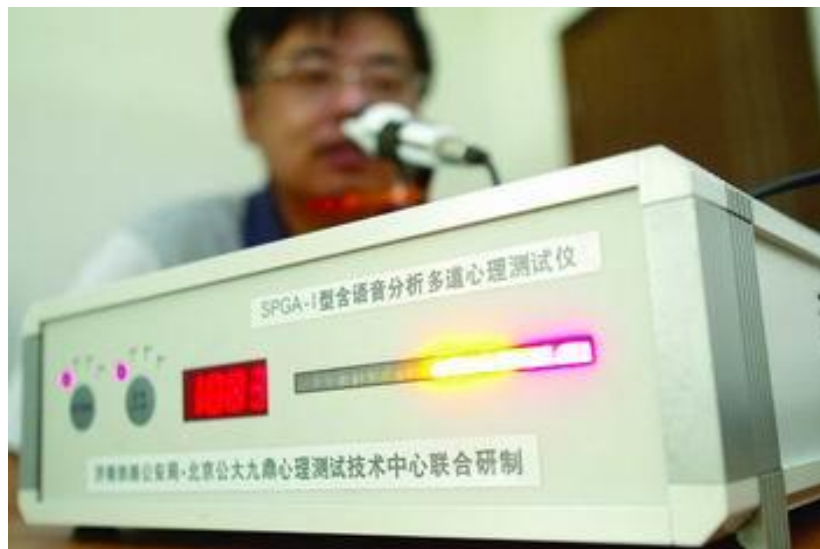
◆ 6月23日

- 12 基于身体姿态的谎言识别
- 13 基于言语内容的谎言识别
- 14 生理信号的自动识别
- 15 微表情
- 16 姿态线索的自动识别

◆ 6月24日

- 17 语音线索
- 18 多模态信息
- 19 谎言识别的个体和群体差异
- 20 谎言识别能力的提升及应用

语音测谎



2006年，济南火车站

问：“你的包里带笔了吗？”

答：“是。”

问：“你的包里带刀了吗？”

答：“不。”

.....

问：“你的包里带钱包了吗？”

答：“不。”

结论：“你在说谎，你的包里有钱包。”

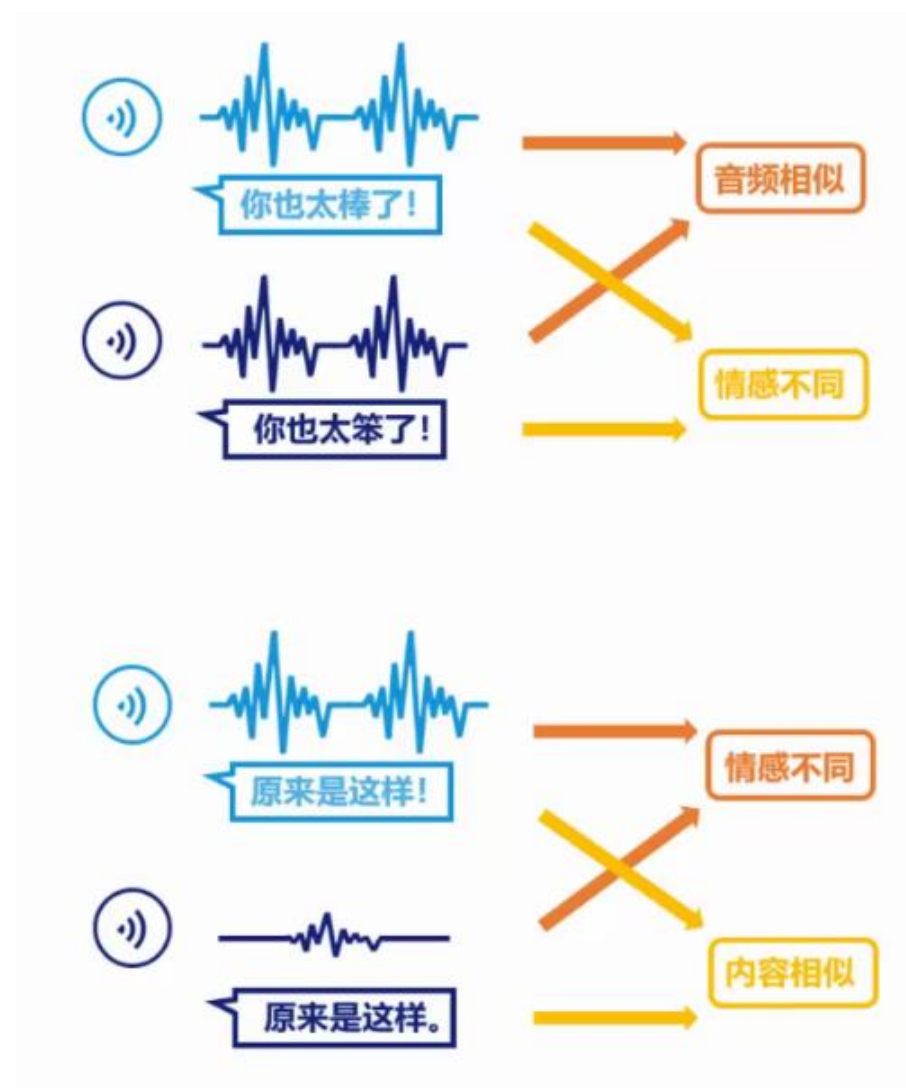
语音测谎的优势

- 传统测谎仪
 - 接触式生理测量仪多采用人的皮肤电反应、呼吸波和脉搏波等参数
 - 测试方法：由测试者依据说谎时相对于诚实作答时生理指标的改变规律来推测被测试者是否在说谎
- 语音测谎
 - 语音信号采集方便且隐蔽，不会给被测试者带来额外压力
 - 不易受到其他生理参数的干扰，更具客观性和有效性
 - 可以在不知情的情况下采集数据
- 应用保障——现有语音分析技术日益成熟

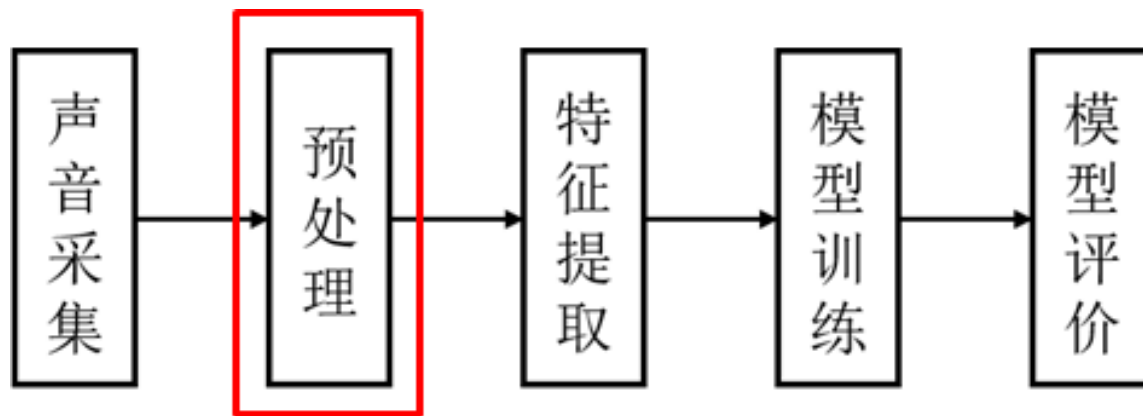


基于语音线索的谎言识别

- 语音测谎一般框架
- 声学特征与谎言相关性分析
- 语音线索自动识别



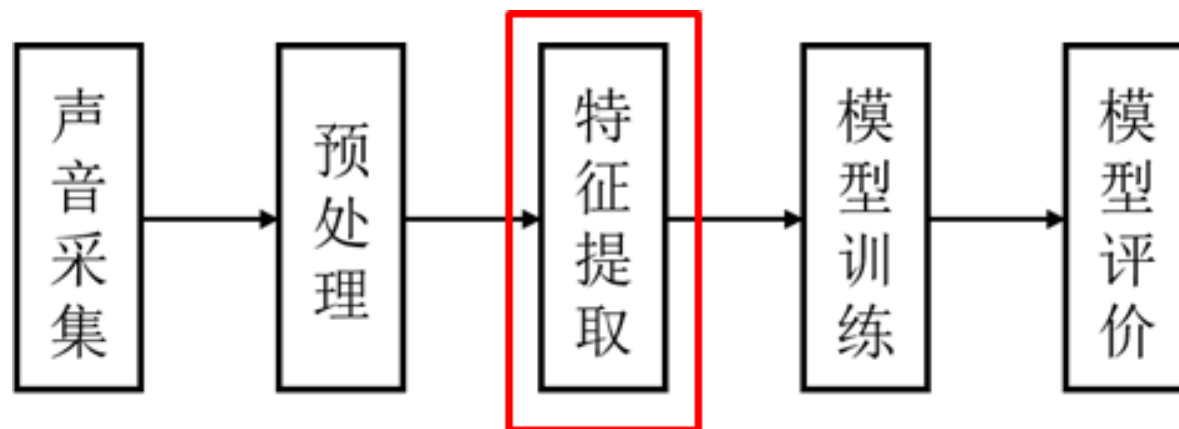
一、语音测谎一般框架



- 预处理模块

- 方便后续语音特征的提取和计算
- 一般包括语音端点检测、预加重、加窗、分帧等基本步骤，也会涉及计算基频、能量等基本语音参数
- 基频的准确度直接影响后续一些特征提取的准确度

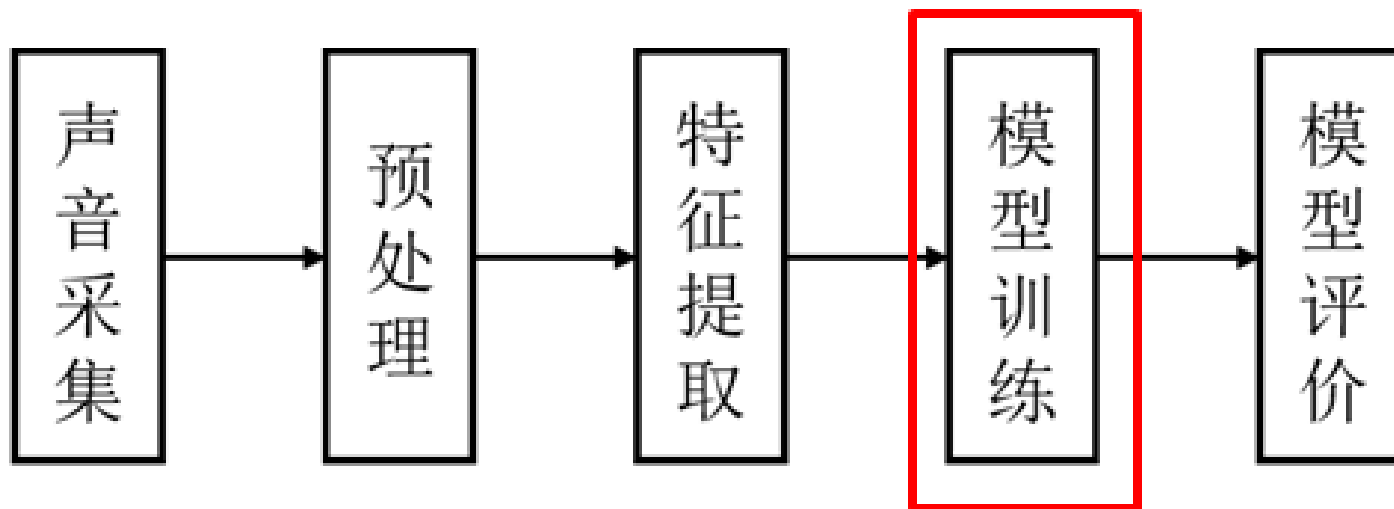
语音测谎一般框架



- 特征提取模块

- 提取常见的语音参数，如基频，频谱等
- 通常也会进行特征选择，主要目的是选择出具有高区分能力的特征参数，将与谎言相关较大的特征送给模型训练模块。减小模型参数，加快模型计算

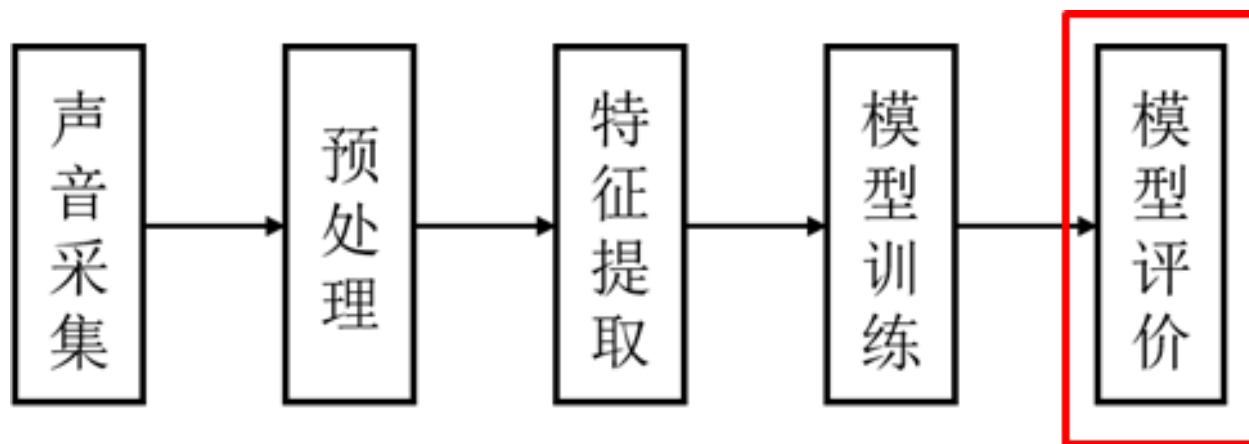
语音测谎一般框架



- 模型训练模块

- 借助一些机器学习算法对提取到的语音特征进行模式学习，挖掘出在什么样的特征模式下，该语音是谎言
- 其输入接口是语音特征向量和样本标签，输出是训练好的统计模型

语音测谎一般框架



- 模型评价模块

- 该模块最终给出模型识别结果的性能指标，如常用的准确率、召回率和置信度等
 - ▣ **准确率**：在识别出的所有谎言中，有多少个是真正的谎言
 - ▣ **召回率**：在送检的所有谎言中，有多少个谎言被识别出来了

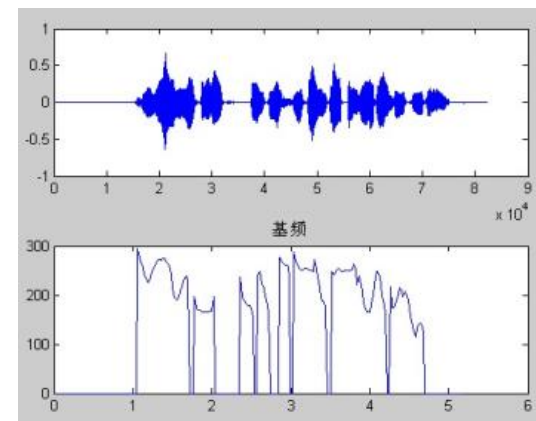
基于语音的谎言识别

- 语音测谎一般框架
- 声学特征与谎言相关性分析
 - ✓ 语音声学特征：韵律特征，音质特征，频谱特征
 - ✓ 语音声学特征提取方法
 - ✓ 特征相关性分析方法
 - ✓ 声学特征与谎言相关性分析
- 语音线索自动识别

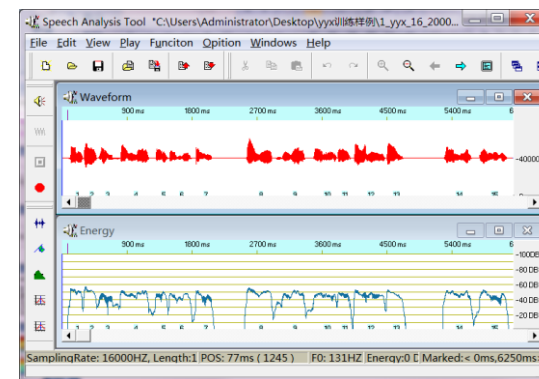
语音声学特征

1、韵律特征

- 指说话人说话时不同的语气、语调等，在声学参数上表现为发音速率、短时能量和基音频率等参数随着时间的变化
- 如语速和能量可以反映出说话者的情绪状态，也可以对说谎状态有所体现
 - 当人信心满满自觉天衣无缝的时候，语速较快，语音的能量也会较高
 - 当谎言被拆穿的时候，会出现结巴、语速较慢、能量也较低的现象



基频曲线



能量曲线

语音声学特征

2、音质 (voice quality) 特征

- 反映发音时声门波形状的变化，也即声音质量的变化，如鼻音和声道肌肉紧张程度等。测谎常用的音质特征包括：基频抖动 Jitter 和振幅抖动 Shimmer

- 基频抖动 Jitter 是描述基频快速反复变化程度的物理量，体现了声音的粗糙程度和嘶哑程度

$$jitter_{absolute} = \sum_{i=2}^N |T_i - T_{i-1}| / (N - 1)$$

$$\text{meanPeriod} = \sum_{i=1}^N T_i / N$$

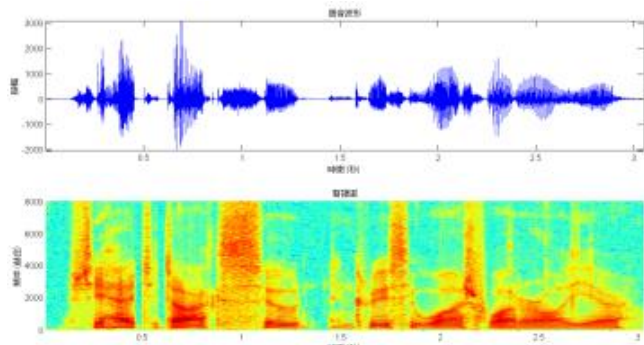
$$\text{jitter} = jitter_{absolute} / \text{meanPeriod}$$

- Jitter 产生的原因包括生理上的原因(带有个人信息的，也是在某人说话中不会发生变化的部分)，情感的突变(基频的变化，声源的变化)，声调的变化，音强的变化等等
- 振幅抖动 Shimmer 反映的是周期间振幅的变化，与基频抖动 Jitter 类似

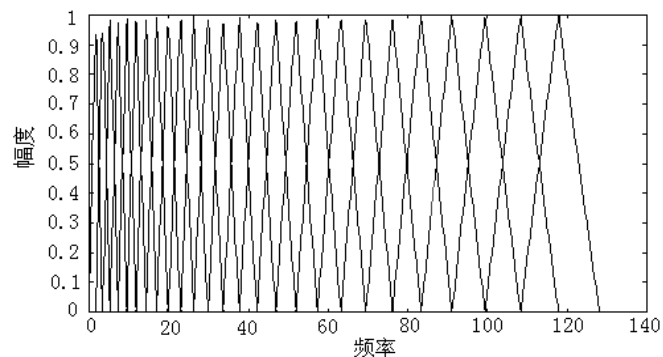
语音声学特征

3、频谱特征

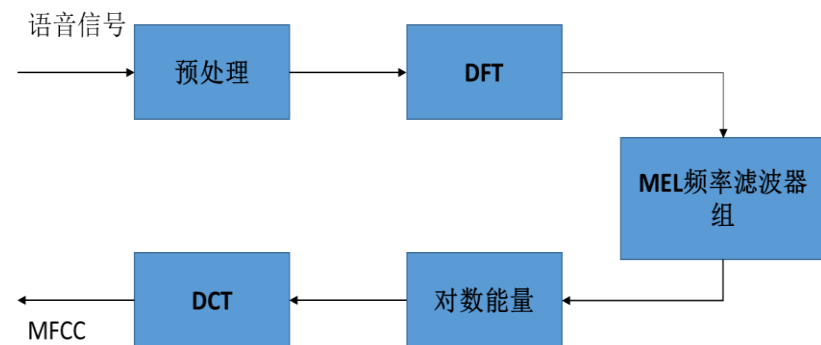
- 将语音的时域信号通过短时傅里叶变换变为频域信号，可以提取频谱特征，探索声道的特征。常用的频谱特征有线性预测倒谱系数 LPCC 和 Mel 频率倒谱系数 MFCC
- 线性预测倒谱系数 LPCC 在所有的频率上线性分布，与人的听觉特性不一致，而且 LPCC 包含了语音高频部分的大部分噪声细节，使其抗噪声性能较差
- Mel 频率倒谱系数 MFCC 参数将人耳的听觉感知特性和语音的产生机制相结合，被证明是在语音识别任务中应用最成功的特征描述之一



语谱图



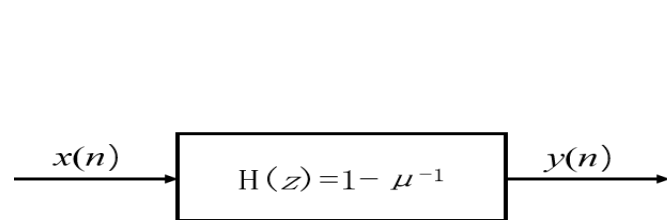
Mel尺度滤波器组



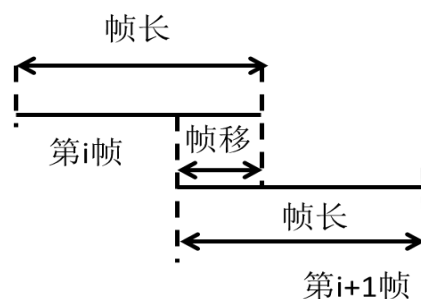
MFCC提取过程

语音特征提取与分析方法

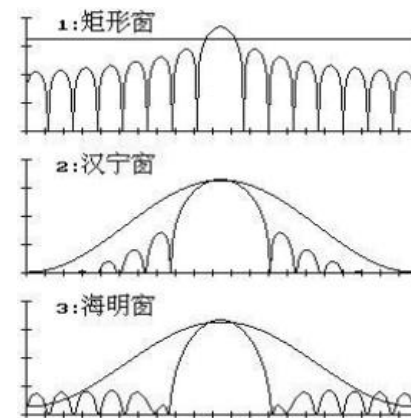
- 预处理能提高信号质量，进而提高特征提取准确度
- 常见的预处理步骤包括预加重、语音分帧、加窗等，部分情况下还需要语音端点检测，用于将较长的语音句切成较短的语句子句，方便后续处理



预加重



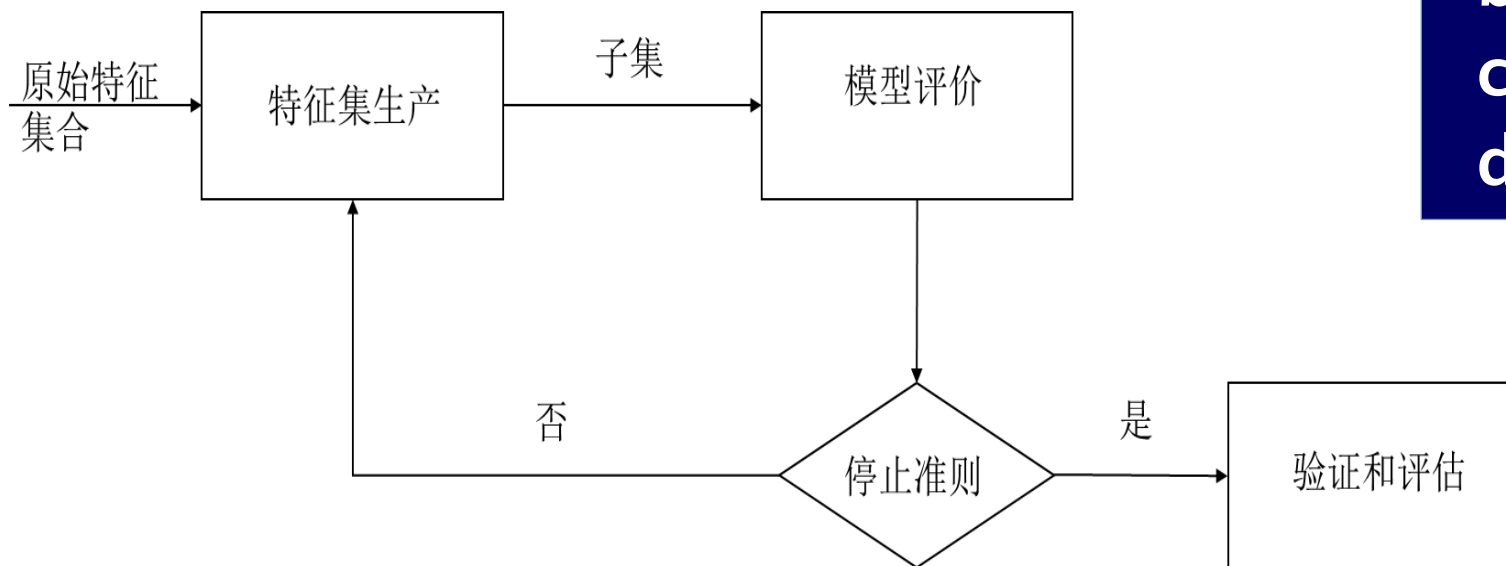
语音分帧



加窗，窗口函数

特征相关性分析方法

- 并不是所有的特征都会与谎言显著相关



特征的特性:

- a. 鉴别性
- b. 鲁棒性
- c. 可表示性
- d. 可计算性

- 由于该方法直接采用分类性能作为评价指标，因此一般能够获得针对特定任务较好的特征子集

声学特征与谎言相关性分析

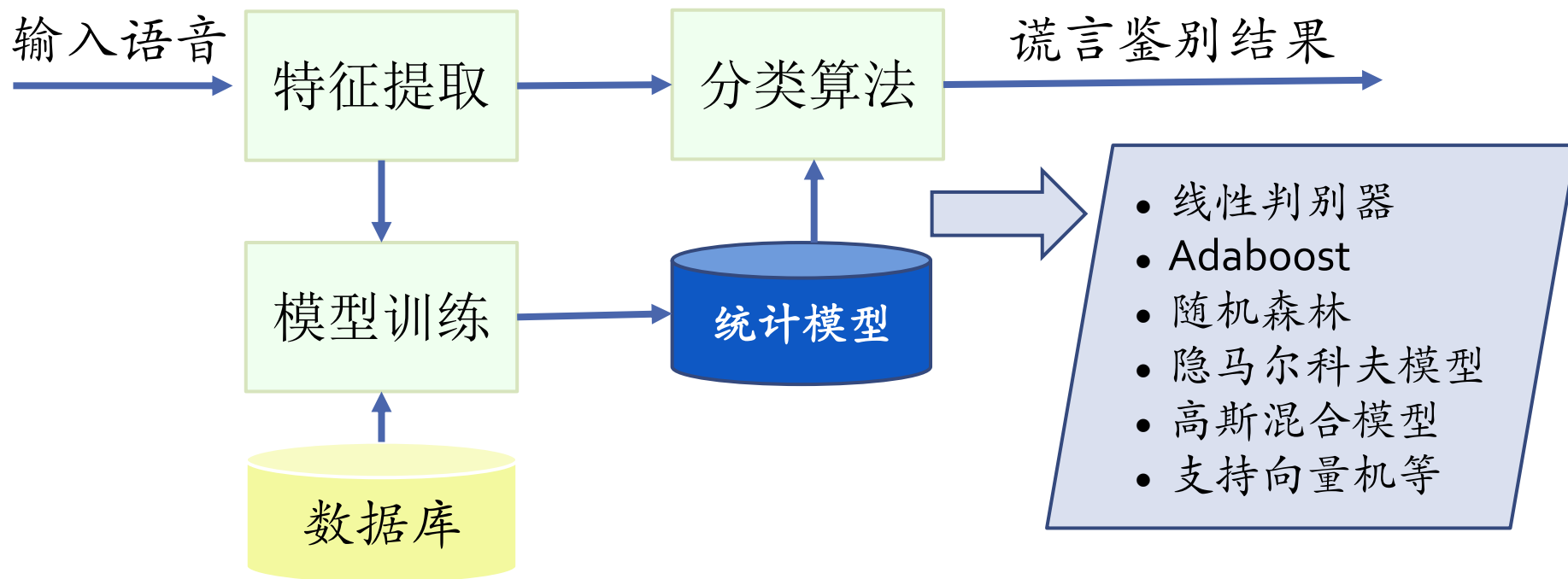
- 人在说谎时，语速、语气等都会产生变化，比如语速变慢、语气变弱、停顿较多等。也有可能说谎者为了掩饰内心的心虚，不自觉地提高音量或提高基频，还可能伴随重复，表达不连续、不完整等
- 由于说谎会引起情绪和生理的变化，因此也会改变语音发声和共振的特点，比如焦虑、害怕会使得喉部肌肉紧张，增加声门下方的空气压力，最终改变声音的音色
- 市场上在售的语音分析式测谎仪，主要以声音压力分析理论（voice stress analysis, VSA）为基础。该理论认为喉部肌肉微颤能够区分人是否紧张，技术上以人在紧张时发出的7-15Hz范围内的语音微抖动为主要检测指标

基于语音的谎言识别

- 语音测谎一般框架
- 声学特征与谎言相关性分析
- 语音线索自动识别
 - ✓ 自动识别方法概述
 - ✓ 语音谎言数据库
 - ✓ 语音谎言自动识别方法

语音谎言自动识别方法概述

- 典型的模式分类问题，因此模式识别领域中的许多算法都可用于语音测谎分析的研究，如线性判别器、迭代算法Adaboost、随机森林、隐马尔科夫模型、高斯混合模型、支持向量机等



语音谎言数据库

1. CSC数据库
2. Idiap Wolf 数据库
3. DyVis 数据库
4. 多模态谎言数据库
5. 苏州大学谎言数据库

语音谎言数据库

1、CSC数据库

- CSC (Columbia-SRI-Colorado Corpus) 数据库，美国哥伦比亚大学构建的谎言识别数据库
- 语料库设计：录音人如果通过谎言成功欺骗了对方，则可以多获得100美元的物质奖励。这一巧妙的设计增加了该谎言语料录制的真实性
- 32名母语为英语的学生，男女各一半
- 15.2h的对话，每个对话大约25-50min，去除掉面试官的语音，最终约有7小时的目标语音

语音谎言数据库

2、Idiap Wolf 数据库

- 多人录制，由瑞士IDIAP研究机构构建
- 数据采集以狼人杀游戏为基本剧本，狼人为说谎者，其余村民等人的语音则认为真实语音
- 通过麦克风和三个同步摄像头，录制音视频数据
- 实验人员分为四组，每组平均8-12人
- 对每个人的角色进行了标注，同时标注了游戏结果
- 该数据集包含了约7个小时的音视频对话数据

语音谎言数据库

3、DyVis 数据库

- 剑桥大学09年录制，起初用于研究语音动态变化
- 数据库由100个18至25岁的标准英语发音的男性参与者对话录音组成，共包含了4个场景：警察访谈，“帮凶”电话对话，段落朗读和句子朗读
- 警察访谈部分可以用来做谎言研究。在录制过程中，警察会询问参与者参与的贩毒时间的问题。采访完全模仿现实生活，也在回答过程中制造一些紧张的气氛，营造更加逼真的效果，获取更真实的数据

语音谎言数据库

4、多模态谎言数据库

- 生理信号、热量和视觉三种模态下的谎言数据库，英文录制
- 参与者为30个高中生和大学生，25男5女，年龄从22到38岁不等
- 录制了三个不同场景
 - 第一个是犯罪场景，每一个参与者被告知有或者没有偷窃了20美元，需要尽力让审查员相信自己没有偷20美元
 - 第二个场景是每个参与者被描述他们最好的一个朋友，或者是真实的或者是谎言，要求自由描述2-3分钟
 - 第三个场景是描述对于堕胎的看法，每个参与者被要求想象自己在参与一个关于堕胎的辩论，自由陈述2-3分钟，可以是真实的看法或者是谎言

语音谎言数据库

5、苏州大学谎言数据库

- 汉语，参与者40人，男女学生各半，均为苏州大学学生
- 录音设计：以自身经历为参考，围绕一个话题，自由准备真话或谎言，录制开始时才告知有面试官提问，若能成功骗过面试官，则有丰厚报酬
- 48Khz 采样，立体声
- 剔除掉表现被动、选择话题为无关紧要的事情、动机不明显的参与者，最后共选了30位参与者的录音进行标注，其中15位男生，15位女生

语音谎言自动识别方法

- 基于LDA和GMM的谎言识别
- 基于非线性光谱分析的谎言识别
- 基于关键子句的谎言识别
- 基于深度学习和迁移学习的汉语谎言识别

基于LDA和GMM的谎言识别

探索不同声学特征在正常语音和谎言语音中的差异

- 参数：MFCC、PLP、基频、过零率等，计算各个参数的分布差异

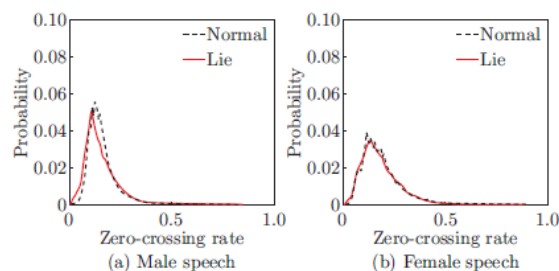


Fig. 6: Distribution of zero-crossing rates between normal and deceptive speech

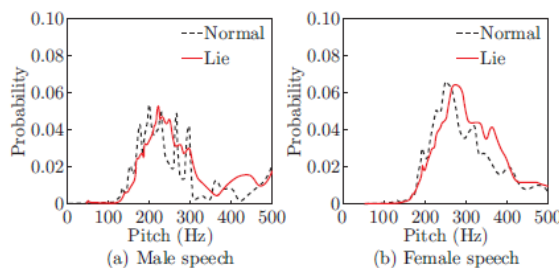


Fig. 7: Distribution of pitch frequencies between normal and deceptive speech

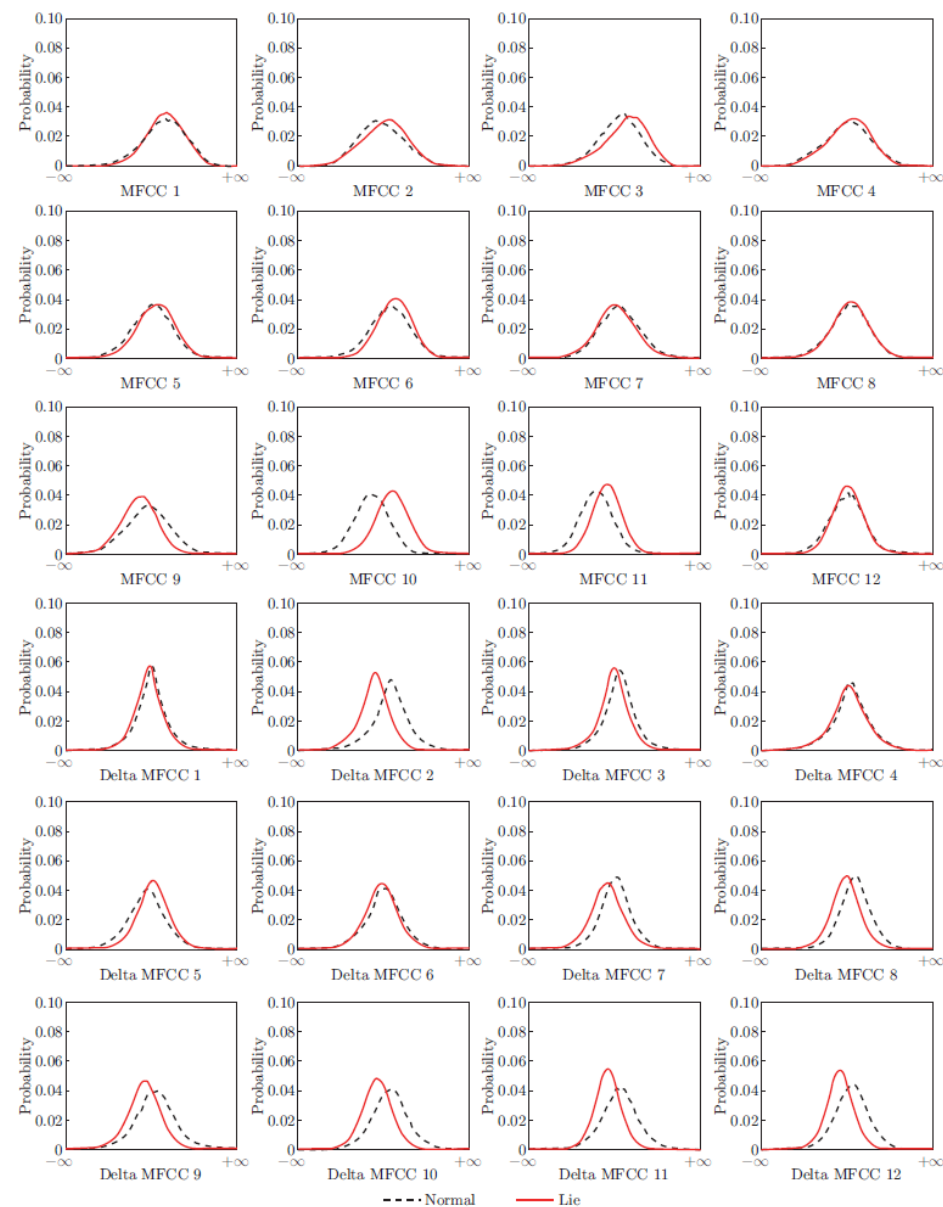
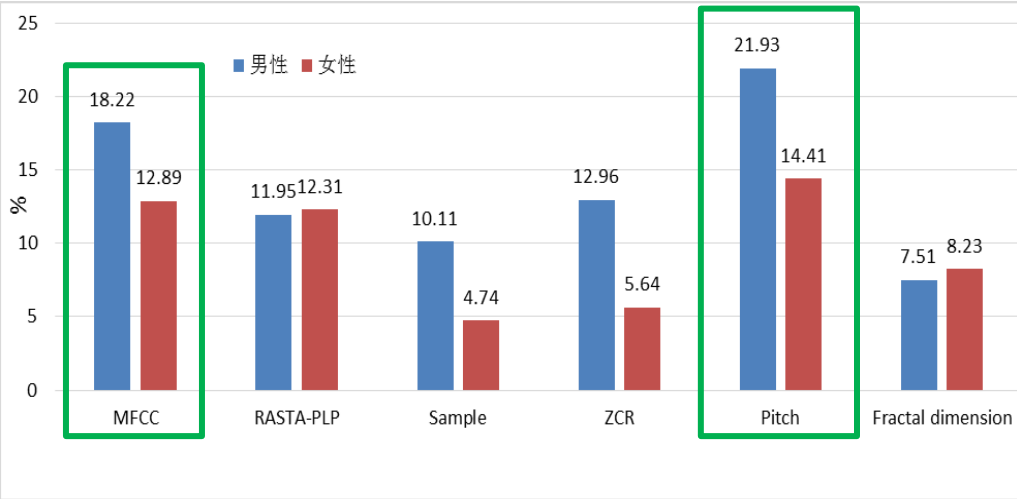


Fig. 1: Distribution of the 12 MFCC and 12 Delta MFCC coefficients between the normal and deceptive speech of males

基于LDA和GMM的谎言识别

探索不同声学特征在正常语音和谎言语音中的差异

- MFCC和基频在正常和谎言情况下差异较大



不同语音特征在正常和谎言的统计差异

表17.3 利用LDA方法在男性(M)和女性(F)上的谎言识别率

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M(%)	43.3	50.2	51.3	61.0	62.2	61.0	63.0	54.6	54.5	62.4	44.0	32.8	49.7	46.5	69.2
F(%)	63.3	61.2	47.0	28.2	40.6	49.1	64.3	58.0	69.0	55.0	49.0	62.4	59.0	78.4	36.3

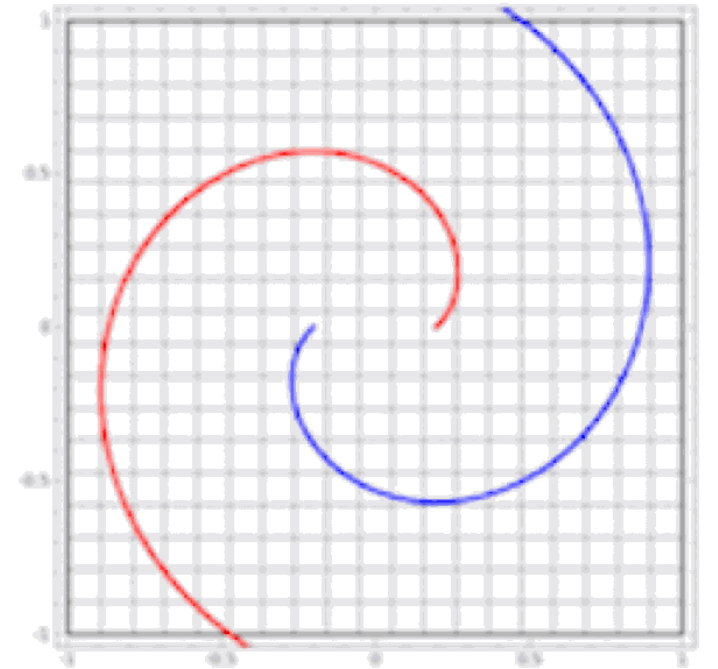
表17.4 利用GMM方法在男性(M)和女性(F)上的谎言识别率

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M(%)	52.6	58.1	48.7	69.2	70.5	63.3	55.9	63.1	59.6	66.6	41.8	40.7	56.8	52.5	78.0
F(%)	71.4	69.0	53.3	33.9	45.5	54.2	68.1	52.3	75.6	59.1	53.2	58.0	64.5	79.9	38.6

LDA的平均准确率分别为53.7%和54.7%。GMM的平均准确率分别为58.5%和58.4%

基于非线性光谱分析的谎言识别

- 谎言语音的形成较为复杂，使得录制的语音数据具有非线性特点
- 非线性光谱分析方法能够使不易区分的特征通过一定非线性变换变得易于区分
- 基于人类语言感知的心理声学隐蔽特性，实验选用2组特征来进行谎言分析，Bark能量特征和知觉意义上的显著能量特征
- 实验语音由演员录制



基于非线性光谱分析的谎言识别

虽然欺骗性话语之间的相似性，以及真实性和欺骗性话语之间的差异性，在时域中不明显，在特征域却被很好地表现出来了

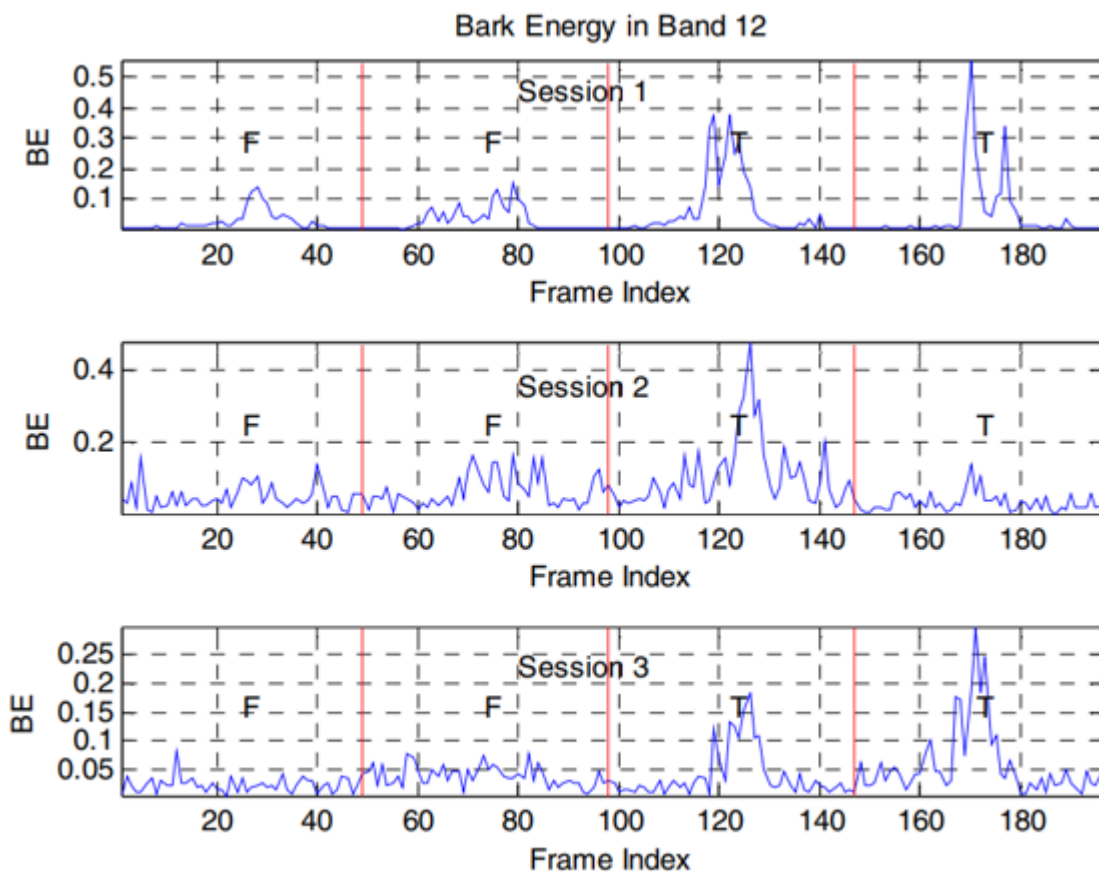


图 17-9 真实的“no”和欺骗的“no”之间的 Bark 能量

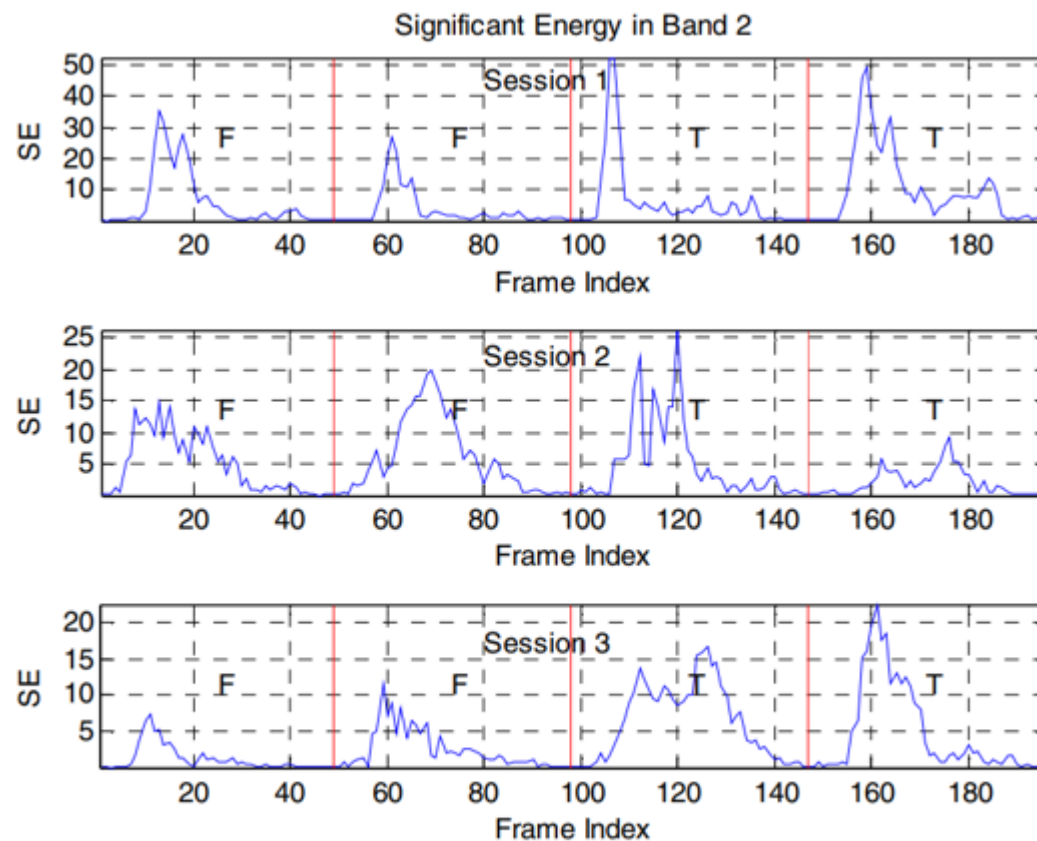


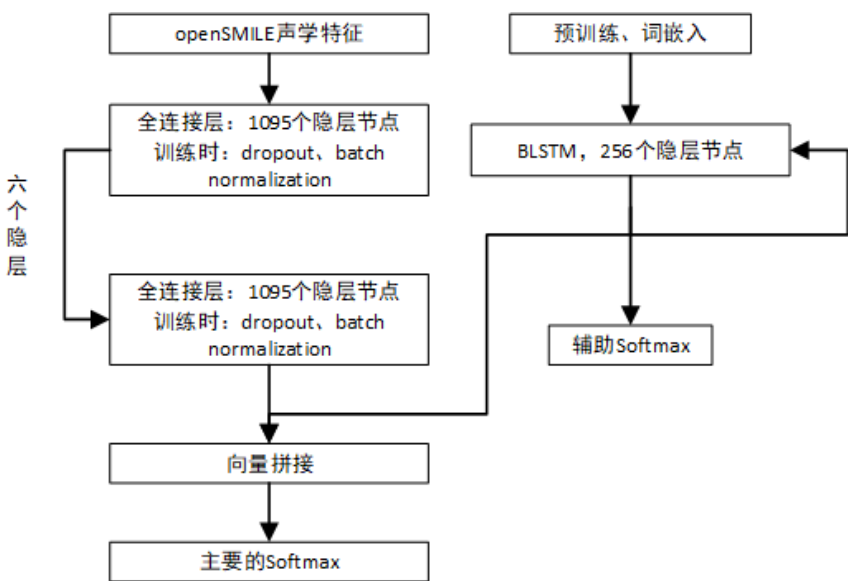
图 17-10 真实“no”和欺骗“no”的显著能量特征

基于关键子句的谎言识别

- 哥伦比亚大学发现一些关键分割句的检测对于正确识别出说谎与否至关重要，一些关键问题的回答对于语音测谎可作为先验知识
- 实验采用特征组合的方法，将声学特征、韵律特征、词法特征组合，利用深度神经网络进行了对比实验
- 实验依据三种类型的声学韵律特征(Interspeech 2013 ComParE 的比赛基线特征、Interspeech 2009的情感比赛基线特征和13维的MFCC特征)和词法特征(n-grams模型和GloVe词向量表示方法)
- 数据采用Columbia X-Cultural Deception (CXD) 数据库的一部分，取其标注较确定的部分。CXD是一个英语数据库，共包括170个对话

基于关键子句的谎言识别

- 在试验中，DNN与BLSTM相比，DNN效果较优
- 在BLSTM中，词法特征优于MFCC特征
- 混合声学词法特征的测谎系统将F1值提高到63.9%



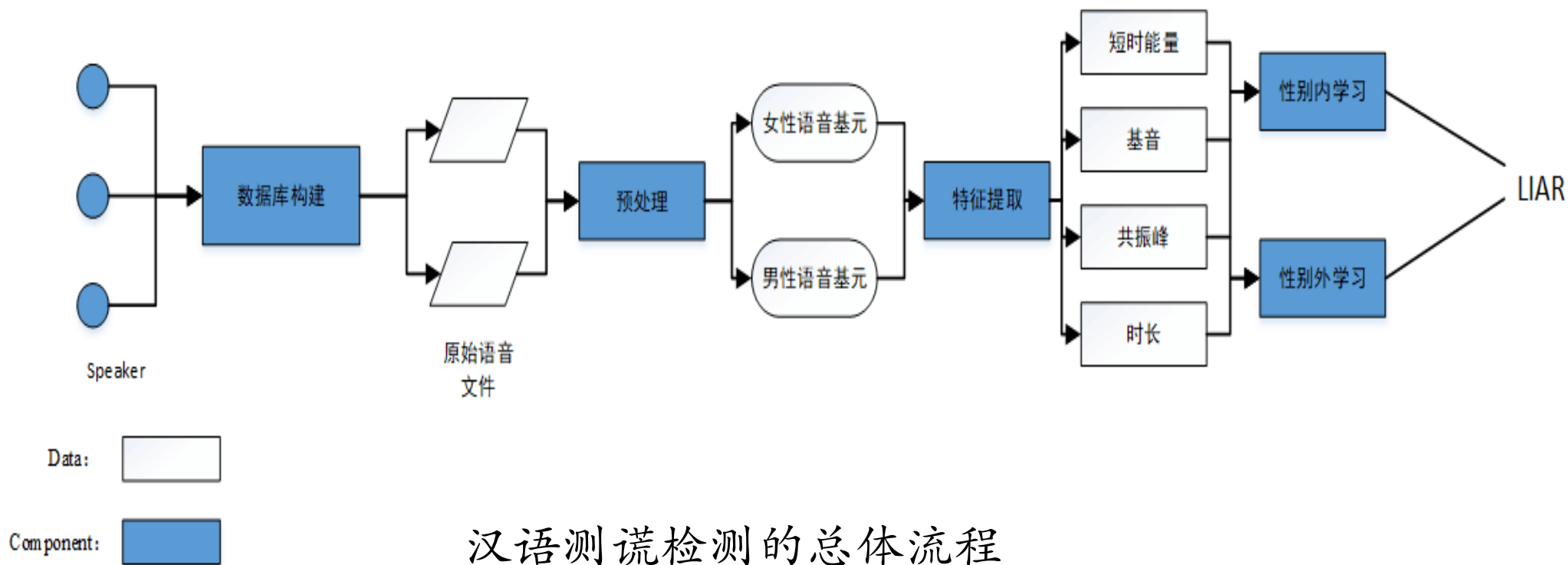
混合声学词法特征的测谎系统图

表17-5 不同特征模型组合的实验结果

Model	Features	Prec.	Recall	F1
LR	Trigrams	58.67	63.95	61.19
RF	OpenSMILE09	72.67	50.44	59.54
RF	OpenSMILE09, Trigrams	76.11	46.99	58.10
DNN	OpenSMILE13	63.65	58.03	60.71
DNN	OpenSMILE09	65.87	59.84	62.71
BLSTM	MFCC	54.19	55.10	54.64
BLSTM	WE	60.46	60.45	60.46
Hybrid	OpenSMILE09, WE	67.32	60.80	63.90

基于深度学习和迁移学习的汉语谎言识别

- 建立汉语谎言数据库，通过迁移学习分析跨性别的语音测谎识别



基于深度学习和迁移学习的汉语谎言识别

- 语音声学特征主要包括短时能量、基频、共振峰和时长等，并提取其统计特征，包括最大值、最小值、均值、中值等
- 多个机器学习算法被用来进行语音测谎分析，包括逻辑回归(LR)、决策树(J48)、多层感知机(MLP)、支持向量机(SVM)、AdaBoost(GBDT)等

不同模型在同性别数据中的测谎性能比较

Method	Male			Female		
	Prec.	Recall	F1	Prec.	Recall	F1
LR	0.5982	0.6622	0.6286	0.6124	0.7321	0.6670
J48	0.5654	0.8469	0.6781	0.6681	0.7464	0.7051
MLP	0.5825	0.6194	0.6004	0.6443	0.7064	0.6739
SVM	0.6144	0.7204	0.6632	0.6315	0.7073	0.6673
GBDT	0.6238	0.7081	0.6633	0.7056	0.6802	0.6927

基于跨性别数据的AdaBoost模型性能比较

Percentage	AdaBoost		TrAdaBoost	
	p_{fm}	p_{mf}	p_{fm}	p_{mf}
0.1	0.557	0.576	0.583	0.656
0.2	0.567	0.605	0.579	0.638
0.3	0.580	0.623	0.575	0.609
0.4	0.603	0.678	0.574	0.614
0.5	0.620	0.643	0.614	0.629

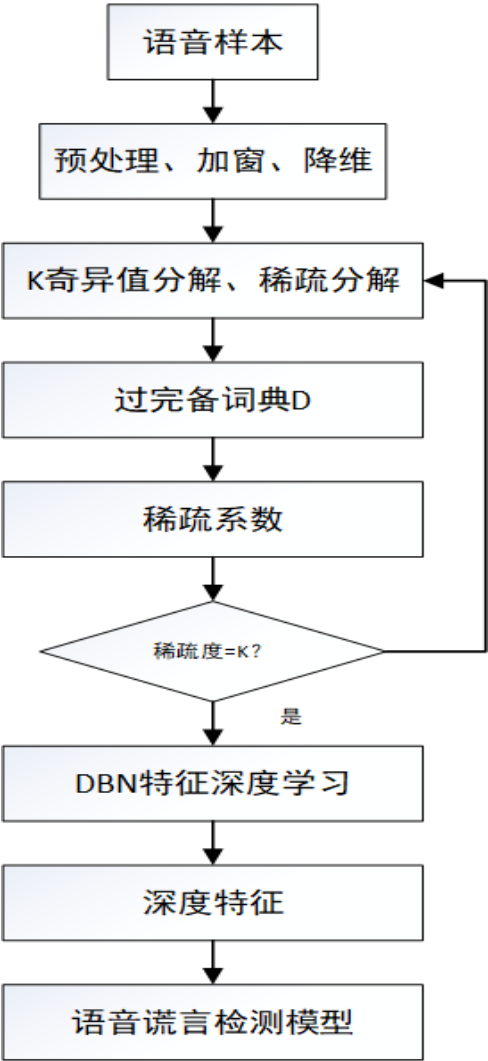
p_{fm}表示女性是源域而男性是目标域

基于深度学习和迁移学习的汉语谎言识别

- 由于声学信号和听觉感知的非线性特性，苏州大学提出了基于K-SVD(奇异值分解)的深度置信网络(Deep Belief Network, DBN)算法，组合了稀疏算法的高维数据线性分解能力和DBN的深度非线性能力

表17.8 不同特征下的谎言识别性能比较

样本数	深度特征	稀疏特征	基本特征
N = 1000	60.12	50.46	37.63
N = 2000	61.82	51.25	37.87
N = 3000	63.25	52.13	38.09
N = 4000	65.34	52.30	38.24
N = 5000	65.89	52.97	38.96
N = 6000	66.02	53.05	39.59
N = 7000	66.86	53.24	40.27
N = 8000	67.23	54.32	41.48
N = 9000	68.67	55.43	42.56
N = 10000	69.83	56.42	45.78



语音自动测谎的展望

语音测谎前景美好，但目前应用仍存在诸多困难



1. **语料库不足。**由于说谎语料较难获取，现有的语音谎言数据库大多不是真实场景的，并且数据量较小
2. **噪音问题。**噪音会影响语音声学特征的采集的准确性
3. **个性化问题。**由于个性、性别、年龄、情绪等因素的不同，每个人发音也会不同，导致了不同人在说谎时的外在表现不同，增加了测谎实用时的难度
4. **反测谎问题。**一些训练有素或者心理素质好的人也会控制表述方式，尽量使自己表现的和说真话时一致

大胆尝试，小心求证！

为谎言识别发展添砖加瓦！

谢谢！ 敬请批评指正！

