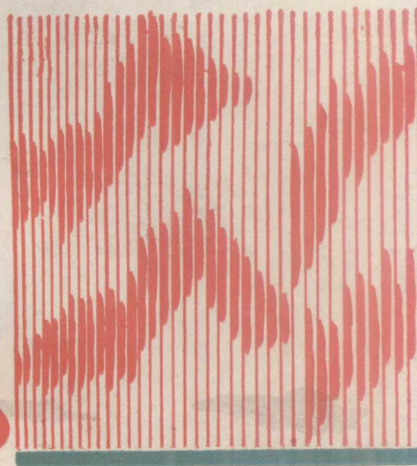




丁宁 译 张家骥 校



嗓音鉴别的 理论与实践

群众出版社

嗓音鉴别的理论与实践

(美) 国家研究理事会

国家科学院

国家工程研究院

医学研究院

丁 宁 译 张家骅 校

群 众 出 版 社

一九八九年·北京

知央已分野内

会事理民理察理(集)

理理理理理理

理理理理理理

理理理理理理

理理理理理理

嗓音鉴别的理论与实践

〔美〕国家研究理事会等著 丁宁 译

群众出版社出版发行 新华书店经销

北京通县印刷厂印刷

787×1092毫米 32开本 5.5印张 113千字

1989年5月第1版 1989年10月第1次印刷

ISBN7-5014-0386-4/H·3 定价：2.15元

印数：00001——1000册

说 明

本研究项目业经美国国家科学研究委员会理事会批准，委员会成员是从国家科学院理事会、国家工程研究院和医学研究院推选出来的。负责该报告的委员会成员，是根据他们的专长挑选的，同时也考虑了一定的平衡。

本报告是根据审批委员会所认可的程序，经著者以外的一个小组审阅的。报告审批委员会由国家科学院、国家工程研究院和医学研究院院士组成。

前 言

美国联邦调查局 (FBI) 于1976年3月要求国家科学院承担利用噪音的声谱来鉴别说话人的工作。联邦调查局注意到15年来噪音鉴别技术得到的广泛应用, 以及所进行的几种与此相关的科学实验。此外, 各级法院对于是否接受以声谱为证据作了规定。从事声谱分析与实验的工作者已经建立了一个专业性组织。与此相关的, 关于技术的可靠性和证据结果的可接受性也一直存在着广泛的争议。

国家科学研究委员会在1976年7月指定声谱评价委员会负责并根据联邦调查局的要求进行研究。这个委员会隶属于国家科学研究委员会的行为与社会科学联合会。委员会的活动主要是全体委员、研究主任和顾问之间的非正式的坦率的讨论。有关委员会的详细活动和简介见附录E。

委员会第一次会议期间, 讨论了联邦调查局所要求的技术手段, 并同意进行下列五项工作以完成它所负的责任。

1. 研究有关说话人的言语特征及其声谱表现形式, 和它们在鉴别噪音工作中的应用。
2. 利用语声中所包含的信息, 分析鉴别噪音中所出现的误差率, 并研究影响误差率的因素。
3. 通过说话人语声中的信息, 为鉴别噪音提出新的或改进的方法。
4. 说明对噪音鉴别人员的训练方法, 并寻求训练和考查

噪音鉴别人员的新方法。

5.搜集有关的科技、法律文献，描述现有的评价噪音鉴别技术的数据，并准备写出评论和文献目录。

委员会决定不审查某些尽管在它们自己的领域中是重要的，但与委员会的基本任务无关的那些题目（见附录D）。下列问题与本委员会工作无关：录音磁带是原件，还是已被篡改；（被洗掉的）录音磁带包含什么信息，怎样使信息复原；在录音制作中是否会侵犯个人隐私；从录音中怎样测定说话人是被迫讲话以及说话人讲的是谎话还是实话。

在进行本报告的讨论时，委员会力求区分有关事实的概念和有关价值的概念。通过准确的用词来无歧义地表明这种区别。在谈到事实时，我们用“准确与误差”这两个词，这两个词通常具有统计学的性质，它们是可以客观地测量和估价的。在谈到价值判断时，我们使用“可靠性和可接受性”。所谓价值判断是指所给的准确与误差的信息对特定的应用是否满意。

噪音鉴别中被判断为可接受的准确率或误差率取决于噪音鉴别中正确或不正确的判断。有关可靠性或可接受性的判定，必须由司法或立法机构做出。这些部门是决定这类事件的责任与权力机构。执行噪音鉴别职能的人不能出于法庭目的来确定可靠性或可接受性。他们只能在现有知识允许的范围内，正确决定这些噪音样本是否能够匹配，详细说明他们测定的有关准确率和误差率。同样，委员会作为一个技术评价机构，而不是司法或立法机构，不能对用于法庭目的的噪音鉴别的可靠性或可接受性做出价值判断。

委员会衷心感谢为我们的工作做出贡献的人们。

Donlas L. Hogen研究主任在这项研究中做出了重要努力,他担任了事务和行政管理,在委员会的整个活动中给予了明智的支持。Fansto Poza凭他长期从事言语处理和嗓音鉴别的经验,对委员会的工作提供了非常宝贵的资料和建议。我们还从其他几位顾问那里得到了很多帮助,在法律方面有Panl. F. Rothstein和Christopher. Smell; 在做判断的科学问题方面有John. A. Swets。我们还得到了行为与社会科学联合会的编辑们的大力帮助: Engenia Gvohman给我们提供了很有帮助的综述; Christine lintz. Moshane编辑报告和处理出版细节。在整个研究过程中,行政秘书Waldena Banks很有效、很耐心地作了许多工作。

声谱评价委员会主席

Richard. H. Bolt.

提 要

这份报告对有关嗓音鉴别的技术、法律和科学诸方面在当前的实践以及将来可能的改进作了统一的讨论。在今天的实践中，一个鉴别人听录制的嗓音和看用图象表示嗓音某些特征的嗓音图，努力把一个未辨明身份的人的嗓音样本与一个或更多个已辨明身份的嗓音样本进行匹配。用于声音传递、录音、播放与分析的技术的发展，起先并不是为了嗓音鉴别的目的。当前的做法是基于有关语音的一些有限知识基础之上的，在很大程度上被认为是一门经验性的艺术。在这种技术中，鉴别人要获得熟练的技能需要经过广泛地训练和体验。对某一对给定的样本，鉴别人从几个供选择的报告中选择一个最典型的，根据已确定的标准，指出或者没有结果，或者样本可以彼此匹配，或者样本不能匹配。

证人靠视听方法，通过嗓音鉴别提供证据，首例出现在1966年的法庭案件中，到写本报告时已有一百多例了。对嗓音鉴别证据，司法部门有各种各样的反映，有承认的，也有否认的。从某种程度上说，考虑到在不同的法庭和实验条件下嗓音鉴别的相对精确性，则不同的法律观点反映了不同的技术观点。

在语音学和声学的科学研究中，已经得到相当多的与言语信息有关的语音资料，可是有关鉴别说话人的资料却很少。现有的嗓音鉴别建立在这样一种假设之上，即：说话人

本身的变异小于不同说话人之间的变异，而且它们的性质也不同。但是，这种假设在目前还没有科学理论和充分的数据来证实。鉴别结果中可能的错误看法主要来自各类专业人员的判断和不完整的实验结果，而不是这些结果在法庭应用中客观的数据表示。

委员会认为，鉴于目前嗓音鉴别中技术的不准确性如此之大，所以要求法庭在应用该技术时要极为谨慎。委员会不主张或反对把嗓音鉴别的视听方法用于法庭，但是建议，如果使用这种证据，一定要把这种方法的局限性清楚地、全面地向事实裁决人，或者法官，或者陪审团加以说明。因为这种方法有可能在某种程度上继续在法庭上应用，所以委员会建议使用某些现有的方法和不久的将来能够改进方法的知识和技术。出于同样原因，委员会指出，在更远的将来，利用新的知识，沿着本报告提出的研究方向，使这种方法得到更进一步地改进。

目 录

前 言

提 要

第一章	噪音鉴别的技术与实践	(1)
第二章	噪音鉴别的科学概况	(11)
第三章	噪音鉴别的法庭概况	(39)
第四章	发现、结论与建议	(67)

附 录

A.	噪音鉴别的现行操作程序	(80)
B.	噪音鉴别的科学问题	(89)
C.	有关噪音鉴别的法律文献	(143)
D.	声谱评价委员会	(147)
E.	委员会成员和工作人员简历	(158)

第一章 嗓音鉴别的技术与实践

早在1660年就有个证人根据被告的嗓音辨认出了被告。¹

直到1937年才对嗓音鉴别进行科学研究。²

电话通讯的发展使得嗓音鉴别可以不受空间和距离的限制。

录音技术的发展使得嗓音鉴别有可能在任何时间进行。

40年代，一种新仪器——声谱仪——为嗓音鉴别增添了视觉手段。³

1966年，法庭第一次认可了以语音声谱为基础的嗓音鉴别的证据。⁴

背 景

图1所示的三张言语声谱（亦称语图）样本是用声谱仪制作的。这种仪器可以对一个信号进行频率分析，并将该信号的频谱显示为随时间变化的图形。在这三张样本中，水平轴线表示时间，这些样本显示的是一段历时两秒钟的言语片段。垂直轴线表示频率，样本显示的频率范围大约100Hz至4000Hz（赫兹或周/秒）。任何一处的相对灰度表示该频率在该时刻的相对声压级。

这几张声谱样本中的相对灰度所表示的声压级范围大约为24分贝。

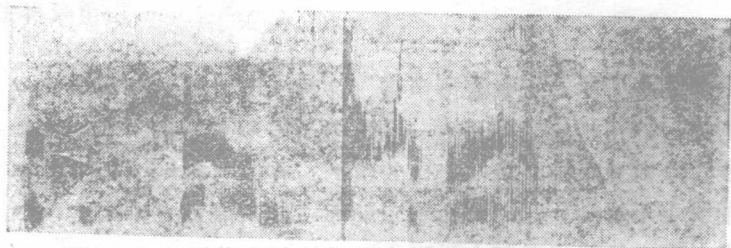
这三张声谱样本，是一个男性说话人所说的 “This is a



说话人A：1天



说话人A：4天



说话人B：1天

图1 三张声谱样本

Sound spectrogram.”上两语图是说话人A的，下一语图则是说话人B的。从这些声谱样本中可以看出，同一个人关于同一言语材料的两次发言，在声谱上可产生明显的差别。而同一言语材料当说话人不同时会出现更大的差别。某

些案例中还可能出现相反的情况，两个不同发音人的声谱样本，看起来要比同一说话人的两个样本更相似。如何测定这些差别，如何对同一人和不同人的声谱图进行统计性处理，是本报告要叙述的基本问题。

象本文中这种言语声谱的视觉显示方式，允许以不同于听音的方式进行检查，眼睛可以不受约束地在图上沿着时间和频率两维空间反复浏览，因而可以发现图象显示的物理特征的细节，并加以检查。与此相反，听音提供了一种自然的、长期训练的能力，能提取说话人的言语意义、感情色彩的细微差别、方言以及身份等信息。当然不论是声谱，还是听到的声音里都不包含有原始言语中不存在的噪音信息。

在法庭上使用时，言语声谱常被称为“声纹”，这个术语含有与指纹相似的意思。但实际上，二者是根本不同的，指纹是解剖特征，皮肤上纹路的直接图象显示。某个人的指纹模式在其一生中基本保持不变，还从来没有发现过一个案例其手指不同而指纹是相同的。⁵

与指纹相反，言语声谱是噪音产生的复杂过程的间接显示。⁶说话过程包括解剖关系、生理和心理的影响。这些影响可使同一说话人不同时间说的话发生变化，并且随着年龄的增长而发生显著变化。另外，检测、传递、记录和绘制声纹用的物理系统在较大程度上会影响最后的图形。鉴于声谱与声纹这两种证据之间存在着根本区别，所以本篇报告所指的是以声谱作为噪音图，而不是声纹。

此外，委员会把这篇报告的题目叫噪音鉴别而不叫说话人鉴别。原则上，两个说话人的噪音差别完全可能小于测量精度的范围，以致于无法区别。声学分析，不论是靠听语音

还是看嗓音图，都会得出有关嗓音的直接信息，但这些信息只是有关说话人特征的间接的推论的信息。准确鉴别说话人，既要根据指定的嗓音与它所产生的嗓音图之间匹配的可能性，又要依靠嗓音与说话人之间匹配的可能性而定。

借助嗓音图来进行嗓音鉴别是一项专门技术。这种技术在法律与侦察领域已众所周知。在过去的20年里，嗓音鉴别的实践已经作为一种新的职业活动而出现了。它包括看嗓音图的视觉检验和听录制的言语的听觉检验。实践与技术已应用于实验室研究和涉及各种犯罪，包括绑架、谋杀、敲诈和贩毒等法庭程序。法院对建立在嗓音图基础上的证据是否可能接受持有异议。如果真是这样，那也要看是在什么情况下。这些和其他一些法律方面的问题将在第三章讨论。

设备与程序

下面简单介绍一下当前由嗓音鉴别联合会 (IAVI) 所推荐的实践程序，IAVI是当前嗓音鉴别实践唯一的正式组织机构。这个联合会所推荐的程序更详细地描述包括在附录A里。IAVI正在继续改进它所推荐的在嗓音鉴别中进行的培训和专业程序。

目前进行的嗓音鉴别实践通常包括一个或更多个已知嗓音与一个未知嗓音的听视比较。听觉检验包括听辨已知与未知的嗓音录音，以便观察大体的相似点和差别，筛选出用途不大的声样，为进一步的研究提供有用的录音。视觉检验包括分析和对比嗓音图上的嗓音声学模式。鉴别人要力求比较相同的言语要素的声学模式，例如，不同嗓音图中所出现的相同的音素、音节或词。另外，检验人在同一言语上下文

中力求进行言语要素的对比。通常是这样：当比较的要素被同一个人在同一句子或词组中说出，或者这一言语要素单独被人说出时，这一要素的前后没有任何言语。

用于目前实践的主要设备包括磁带、录音机、放音设备和绘制噪音图的语图仪。把被录制的已知噪音样本与未知噪音样本进行对比。有时检查员使用一种特殊的放音设备，它可以向前或向后高速播放，把两个样本进行详细比较。有时还用一种特殊的滤波器，它可以滤去不需要的声音，如电线的嗡嗡声或频率范围内那些不利于言语分析的噪声。

检查员或其他获准人员，采用为获得言语取样而发展的程序，录制已知的噪音样本。这些采样要能代表已知的噪音。检查员或其他获准人员还必须力求再现未知话音所在的物理环境。噪音样本一录下来，便要立即加上标记并妥善保管，以保证每个样本都与产生它的噪音准确地一致。接着，检查员用语图仪为听觉分析准备噪音图。将已知与未知的噪音取样输入语图仪，检查员凭借经验和有关对比、灰度、图象渐变等一般原理，调整语图仪，做出质量令人满意的语图，力求消除噪声，增加对噪音鉴别有特殊意义的高频信息。每个噪音图要标上已知或“未知”字样以示区别。还要在噪音图上标明案件名称、制图日期和序号。

检查员在检验准备噪音图时，要在噪音图上直接标记与言语图形对应的每一个音节和词。这一信息可以用正常的单词拼写，也可用音标，或者两种方法都用。专业检查员鉴别时要精确，因为可靠的噪音鉴别要求每一口语的词和音节与噪音图上的模式应相符合。

实际上，检验通常是视图与听音同时进行。鉴别人把注

注意力集中在已知与未知嗓音样本中所显示出的特殊的音段，并且重复播放直到做出判断为止。如果已知嗓音不止一个，那么就把未知音样与每一个已知嗓音进行反复对比。有时有几个未知嗓音，则须确定这些嗓音是否属于同一个人。

在进行这一程序时，鉴别人找出特殊模式，可能是鉴别工作中十分重要的线索。尤其要力求解决由于说话人之间的变异性而引起的嗓音图上的差别，以指出不同人的嗓音；还须解决由于说话人本身的变异性而引起的嗓音图上的差别，指出同一个人讲同一个词时的不同发音。在分析这些线索和对嗓音鉴别做出决定时，对于在检验中听音和视图部分在数量上各占的比重并没有规定。

鉴别人可以使用尽可能多的有用音样。通常鉴别人对一个嗓音鉴别所需要的时间是根据鉴别人认为必要的时间来决定的。在直接完成两个嗓音对比之后，鉴别人作出下列决定中的一种决定：

- 1.肯定鉴别
- 2.可能鉴别
- 3.无结果
- 4.可能排除
- 5.肯定排除

目前，对嗓音鉴别人的训练程序规定得还不够严谨，因为这一程序要考虑到被训练人的许多方面，比如被训练人的背景，以前所受过的训练和训练中指导者的影响。两门介绍性的训练课程在目前来说是通用的。有些课程是在言语科学方面，在合适的指导人指导之下进行嗓音鉴别分析工作，并且受训人至少参加两年（见附件A）。嗓音鉴别人的专业团

体认为某人已完成上述训练活动，并得到指导人的推荐，证明他在噪音鉴别方面已达到了熟练程度，通过考核之后，才能加入他们的行列。

目前的状况

目前，噪音鉴别技术是一种被相当广泛运用的实用方法。但是缺乏对噪音鉴别的科学问题进行说明的理论基础。噪音鉴别领域的实践者和科学家承认这种理论与实践之间所存在的不一致性。

主要的科学问题是讲话人的变异性。即使声学测量能够显示一个人用同一方式两次说出的同一个词，但对于一个人言语中变异性的统计记录人们知道的很少。实际上，对于不同人讲同一个词时声音产生的变异知道的就更少了。如果没有这种变异性产生原因方面的知识，那么对噪音鉴别的影响就很难预料、掌握和控制。噪音域 (Voice Popwlation) 这些没有答案的有关噪音总体的统计上的有效表示是一个严重的缺陷。因为噪音鉴别最基本的假设就是说话人本身的变异性小于说话人之间的变异性。实际上，变异性是构成噪音鉴别理论与运用的基础。

委员会认为变异的问题能够用纯统计学方法进行处理，不用考虑言语的物理来源与它们的变异性。统计的方法包括说话人有代表性的取样记录的数据积累，单词，语音，生物学的与文化的影响，说话时的环境和其他统计上的重要变量。这种变量可以通过使用多变量分析的方法，从数据本身得到确定。委员会相信，直接研究解剖、生理和文化方面对语音的影响及其变异，将会更有效更充分地了解噪音鉴别，或许能