# 文本相关的说话人确认:分类器，数据库和RSR2015

计111（10111939）陈楚楠

摘要：RSR2015数据库，旨在评估文本相关识别在不同时间和词汇的限制已经收集和发布的人类语言技术部门在新加坡Infocomm研究所。说英语的人的语音将会被记录，依据多种口音的多样性的平衡。超过151小时的语音数据被移动设备记录。说话人范围包括300名17至42岁的参与者(143名女性和157名男性)。使得RSR2015数据库成为最大的公开的针对文本相关确认的数据库。我们提供评估协议为每个数据库的三个部分，连同两个说话人确认系统的成果:HiLAM系统，基于三层声架构的i-vector/PLDA系统。因此我们提供一个相关的评估方案，以及参考RSR2015数据库性能研究社区的演示方案。HiLAM在大多数场景下比使用了最新技术的i-vector性能好。

关键词：文本相关，RSR2015数据库，说话人确认

## 1介绍

说话人确认过程是指通过比较两个语音样本接受或拒绝身份，依靠两个语音样例:一个用作参考的身份，另一个是在测试期间收集的其他的人的说法。这个通用的定义下，声明人是可以无偿提供的任何话语的，没有时间，质量，记录条件和词汇的内容的约束的语音样本。说话人确认的性能受到许多可能的可变性的说话方式，其中在词汇内容和通道的变化是最不利的。

普遍认为，当测试话语与给定的材料相同，尤其是当其比较小时，能实现更好的精度。在这方面，有两种方法显示有效地解决词汇变化的问题。第一种方法包括进行后验分析弥补词汇不匹配的语音样本，训练和测试之间的话语，而第二个方法考虑对于合作的说话人，词汇的变化很容易降低。首先，尽管第一个方法较高的灵活性，它会受到来自两方面的缺点。一方面，词法分析的计算成本增加确认任务。另一方面，词法的补偿可能会受到大量的词汇不匹配，因为它不可能保证初始的训练和测试词汇一样。第二种方法认为合作的说话人可以要求发音一个预定义的句子或短语在训练和测试阶段。这个过程称为文本相关说话人确认与没有约束输入的文本无关说话人确认相反。换句话说，文本相关的说话人确认可以定义为一个说话人确认任务，其中在测试阶段使用的词汇发音是说话人在注册时的一个子集。

相比于无法控制的环境因素所导致的信道变化，词汇的变化是相对更容易，如果我们可以假定合作的说话人。依据文本相关假设，通过要求测试话语的词汇内容匹配训练材料，我们达到更高的精度和更短的时间训练和测试阶段。因此，文本相关说话人确认非常适合，人体工程学的约束要求精度高，记录持续时间短的商业应用。

在缺乏词汇的约束的训练/测试话语时，允许文本无关技术涵盖广泛，应用的范围包括，法医确认、说话人集群和说话人挖掘。此外，研究文本无关任务是国家标准与技术研究所(NIST)强烈支持举办的国际基准事件，并且在上下文有大量的数据可用。由于这些原因，大量科学论坛关注文本无关说话人确认，而轻视了文本相关的说话人确认的商业潜力。无论如何，文本相关说话人识别可以视为文本无关的说话人确认的一个补充，当确认话语也有类似的持续时间和词典，可以旨在弥补当前的不足和获得更灵活的系统的性能。从历史观点上来看，这篇陈述支持一系列的改进过的文本相关的说话人确认，作为文本无关的说话人确认的子情况。

最近通过集中加强使用开发数据，在文本无关的说话人确认的精度和鲁棒性方面的有了突破。这些改进一直由几十年来一直提供越来越具有挑战性的数据的NIST和LDC强烈支持。虽然文本无关说话人确认正在被越来越多地使用，去训练数据健壮的系统，而文本相关说话人确认苦于缺乏研究数据。这就导致在文本相关这种子情况下，适应现有的方法存在不少的困难。例如，在《使用文本相关说话人节点因素确认》这篇论文发表9年之后，没有与文本相关的说话人确认的论文被发表。这部分可以解释为开发这样的系统缺乏支持文本相关任务的数据库。

数据的缺乏在不同方面上影响着文本相关的说话人识别。现有数据库的限制不允许适当的词汇变化的影响，成为约束的选择说话人的条件。此外，在第二节的现有数据库的概述显示出性别的不平衡，然而在大多数时自动系统的性能是要求能区分性别的不同。最后，自动确认系统的改进要求大量的试验允许统计上显著的性能的措施。

在文本相关的说话人确认下，演讲的词汇内容数据尤为重要，有很多方法可以限制训练和测试话语的词汇。有不同的确认协议，我们可能需要解决词汇约束各级如音素、音节、单词或句子。多项研究表明，保留确认话语中的词汇序列可能导致错误率相对下降50%。因此，选择一个特定的协议从应用程序的观点是至关重要的，因为它会强烈影响精度。然而，很少有研究比较不同词汇的影响进行约束，一定程度上是由于缺乏数据库可以支持一个公平的比较研究。在本文中，我们展示了在IIR的Human Language Technology Department发布的RSR2015数据库，以解决的一些现有的语料库的局限性。这些作为部署健壮的说话人识别的智能根据HOME2015计划的努力记录下来，也是RSR2015数据库的名称的由来。RSR2015数据库的目的是支持文本相关的说话人识别，并允许进行比较确认系统存在不同词汇和时间约束。涉及143名女性和157名男性演讲者总计151小时的录音，RSR2015数据库是最大的公开可用的文本相关的在扬声器和词汇的变化的说话人确认数据库。数据库被安排成三部分，以适用不同的场景。所有三个部分都被在类似条件下记录，以保证任务之间有一个公平的比较。获得后可以在六个不同移动设备包括智能手机和平板电脑商业上可用。RSR2015数据库的第一部分被Larcher等人描述。

剩余的这篇文章中，我们首先概述现有的文本相关的说话人确认数据库。我们通过给他们的描述的文献中的主要特点、优点和缺点，总结了23个数据库。RSR2015将在第三章进行详细的描述。在下面几章中，我们提出现实的评估协议和性能的措施，允许一个在RSR2015数据库的公平的比较系统。在第四章中，我们给出一个调查分类用于文本相关的说话人识别，在描述两个先进的评估RSR2015数据库的系统之前。第五章描述了协议和三个部分RSR2015数据库的两个系统的性能表现。第六章提供实用信息关于如何获得这个数据库。在第七章我们还将讨论一些研究方向和观点关于文本相关的说话人确认。

## 2文本相关说话人识别数据库

在本章中，我们提出一个数据库可供开发和评价文本相关说话人确认的调查。尽管给定的数据库可能不是最详尽的列表，但它构成最大的包含我们尽可能多的文献知识。说话人识别技术互补的信息资源可以在Campbell & Reynolds和一项给出了多模式生物数据库Faundez-Zanuy等地调查中找到。值得一提的是，有一些文献报道结果来自不公开可用的数据库。

我们的目的是提供一些上下文驱动的RSR2015数据库而不是给出一个详尽的描述现有的数据库。事实上，文本相关的说话人确认数据库设计用于各种目的，另外它的多样的协议使得其很难有全集的一个公平的比较。在剩下的这一章中，我们讨论的一些现有的和文本相关的说话人确认的数据库的主要特征。

### 2.1　人口统计

在评估说话人确认系统的性能时，人口统计是至关重要的。在说话人确认中，开发和评估自动系统是由现有的语料库驱动，它由精心挑选而来的人口记录的数据库。为特定的应用程序，选择人口尽可能代表的目标人群，而数据库为通用设计研究目的往往覆盖人口最大的可能。在剩余的这篇文章中，我们考虑的是人口的性别和年龄，它们通常被视为影响说话人确认引擎的两个主要条件。出于这个原因，说话人的集合应该代表目标人群的性别和年龄。此外，人群需要足够大，才能在性能上通过自动说话人确认系统的大量的试验改善，以确保结果在统计上是有意义的。

在实践中，人口的规模，还要考虑技术和经济问题等因素。有趣的是，一个数据库的一大部分，包括文本相关的语音材料，是多模数据库，也就是在表1中列出的24个数据库，其中有14个是多模的。收集多模式的数据库的好处是双重的。首先它允许研究比较和组合形式的身份确认。其次，数据池成本和复杂性的负载，也与生物样本的收集相同。从表1可以看到，在过去的20年中，科学界通过巨大的努力已经收集了维持生物识别技术的研究的数据。然而，这些数据库中的说话人的数量仍然有限，只有7/24个表中的条目数超过200。

另一个限制是不平衡的性别表示，可以从到图1观察到。如果19个数据库的性别信息是可用的，则有8个可以是性别平衡的，因为达到了45%，然而有7个的包含少于30%的女说话人。这种不均衡特别损害作为说话人确认系统的性能，在区别男性和女性的说话人。此外，年龄并不总是可用的信息(至少在表1中列出)。当年龄差距很小说话人之间的分歧已被证明是更加困难。

### 2.2　词汇的变化

说话人的确认系统的性能是强烈依赖于条件作为输入提供的语言材料。例如，许多研究已经进行了估计演讲持续时间的影响的研究。其他作品表明，分歧率取决于演讲内容，用于注册和测试，主要的结论，在固定的时间，不同的话语可能不会对演讲者确认同样有用。在文本相关的说话人确认中，注册和测试话语都是固定的，词汇的内容尤其重要，因为它会影响系统的精度。因此，选择的词汇内容时，应研究对文本相关说话人确认系统的影响。

#### 2.2.1人确认的主流协议

在过去的20年中，NIST所提供的大型数据库及其相关协议已成为事实上文本相关说话人确认技术的标准评价。没有的在文本相关说话人确认的情况下，使跨系统比较困难，增加协议在文献报道的数量的标准。然而，两个主要流反映在现有的数据库。在 Yoho， M2VTS， Verivox 和 Biosec，词法的培训和测试内容被强烈限制下只能使用位数，而数据库如SmartKom cccvpr2c2005 - 10000，或MIT-MDSVC允许更广泛的词汇覆盖使用固定短语。这两种类型的协议是由数据库如Polyvar， OGI Speaker verification， XM2VTS， Ahumada， PolyCost，BANCA， BioMet， MyIdea， Valid， M3， BiosecureID， MBioID， BioSecure，或提供不同的数字字符串一起固定短语的UNMC-VIER。大多数情况下，词汇的变化范围是有限的几个固定的句子和固定数字字符串。例如，在表1列出的24个列表，10个数据库包含少于10个不同的句子。

为了增加词汇覆盖率，一些数据库如PolyCost，Banca，MyIdea或MIT-SDSVC包括词汇的内容不同的说话人。试验中的冒充者宣称目标说话人所使用的文本是由要求每个主题定一些其他学科的内容。在这种协议，说话人交互的冒充者试验的可能性大大限制冒充者和目标不都说同样的演讲内容的情况。