语音增强

——理论与实践

SPEECH ENHANCEMENT

Theory and Practice

Philipos C.Loizou 著

高 毅
肖 莉
邓 方
吴绍炜 译

电子科技大学出版社

语音增强

——理论与实践

Speech Enhancement: Theory and Practice

Philipos C. Loizou 著

高 毅 肖 莉 邓 方 吴绍炜 译

电子科技大学出版社

图书在版编目（CIP）数据

语音增强：理论与实践 /(美) 罗艾洲(P.C.Loizou) 著；高毅等译. -- 成都：电子科技大学出版社, 2012.12

书名原文: Speech Enhancement: Theory and Practice

ISBN 978-7-5647-1293-8

Ⅰ. ①语… Ⅱ. ①罗… ②高… Ⅲ. ①语声信号处理－数字信号处理－研究 Ⅳ. ①TN912.3

中国版本图书馆 CIP 数据核字(2012)第 260091 号

图书著作权登记：图进字（21-2013-30）号

语音增强——理论与实践

Speech Enhancement: Theory and Practice

Philipos C. Loizou 著

高 毅 肖 莉 邓 方 吴绍炜 译

出　　版：电子科技大学出版社（成都市一环路东一段 159 号电子信息产业大厦　邮编：610051）
策划编辑：张　鹏
责任编辑：张　鹏　李燕芩
主　　页：www.uestcp.com.cn
电子邮箱：uestcp@uestcp.com.cn
发　　行：新华书店经销
印　　刷：四川川印印刷有限公司
成品尺寸：185mm×260mm　　印张 33.25　字数 802 千字
版　　次：2012 年 12 月第一版
印　　次：2012 年 12 月第一次印刷
书　　号：ISBN 978-7-5647-1293-8
定　　价：79.00 元

前　言

本书内容来源于我在德州大学达拉斯分校(University of Texas-Dallas)所讲授的语音信号处理课程（我从 1999 年秋开始讲授该课程），同时也是笔者在该领域长期研究工作的结晶。目前，该领域除了少量的适合专家阅读的一些书籍以外，并没有一本语音增强方面的教程，因此我在研究生课程中讲授语音增强的基本原理的时候感到十分不便。对于那些希望涉足该领域的学生和语音方面的学者而言，相信他们也会因为很难找到一篇指导性的综述或者介绍性的论文而感到沮丧（最近的一篇综述性的论文由 Lim 和 Oppenheim 于 1979 年发表在 IEEE 会刊上）。于是这成为写作该书的最初动因。我对该领域的兴趣来源于我对噪声抑制算法的研究[1]，这些算法可以帮助听障人士（人工耳蜗植入者）在噪声环境下更好的交流。开发这些噪声抑制算法的关键之处，在于对现有的语音增强算法的局限以及潜力有基本的理解，我相信本书将提供这方面的知识。

本书总共分为十一章，第一章（引言）中对各章节的内容做了概述。全书内容分为三个部分。第一部分介绍了数字信号处理以及语音信号的基础知识，为理解语音增强算法做铺垫。第二部分介绍过去 20 年中所提出的各类语音增强算法。第三部分介绍评估语音增强算法性能的方法和手段。

书中正文部分专门设计了许多的范例以及图片，以帮助读者理解其中的理论。本书附带的光盘包含了一个语音库，很适合用于评估经算法处理后的语音质量和可懂度。主要的语音增强算法也以 MATLAB 代码的形式随光盘提供。笔者一直认为，利用 MATLAB 开发算法代码，以及利用通用的语音数据库对新的语音增强算法进行评估，对推动该领域的发展是十分关键和必要的。附录 C 对光盘的内容进行了详细的介绍。

本书可以用作语音增强的研究生课程的一学期教材。该课程的先修课程包括数字信号处理以及概率论基础，随机变量与线性代数。本书也可以作为语音信号处理课程的补充教材，可以选择第四章到第八章，以及第九章和第十章的部分章节来学习。

在这里我要衷心感谢我的许多同事和研究生同学，他们对本书的撰写做出了不同程度的贡献。我要感谢 Patrick Wolfe、Kuldip Paliwal、Peter Assmann、John Hansen 和 Jim Kaiser 教授，他们审阅了本书的内容并且提供了宝贵的意见。同时要感谢我的研究生们，他们开发的很多 MATLAB 算法代码都被包含在随书附带的光盘中，尤其要感谢 Sundarrajan Rangachari、Sunil Kamath、Yang Lu、Ning Li 和 Yi Hu，特别是 Yi Hu，他对本项目做出了极大的贡献，包括提出新的语音增强算法（许多算法在书中做了介绍）并用 MATLAB 实现，同时授权我在本书中包含这些代码。我还要感谢 John Hansen 教授，他授权我使用他的关于语音质量客观评价的 MATLAB 代码；感谢 David Pearce，他授权我使用 AURORA 数据库中的噪声录音文件。感谢 Chaitanya Mamidipally 和 Jessica Dagley 在录制带噪语音数

---

1 该工作得到美国国立卫生研究院（NIH）美国耳聋及其他交流障碍研究所（NIDCD）资助

据库过程中提供的帮助。我要感谢 Taylor&Francis 出版集团的策划 B.J. Clark，他对本书的出版给予了支持与鼓励。

最后，我要把最深的感谢献给我的妻子 Demetria，在本书形成的过程中，她的理解与支持一直伴随我的左右。

Philipos C. Loizou

德克萨斯州立大学达拉斯分校电子工程学院

译 者 序

2010 年，摩托罗拉系统（中国）有限公司成都研发中心数字信号处理（DSP）部门倡导并发起了以兴趣小组为单位的读书活动。作为语音信号处理的研发人员和爱好者，我们几位同事希望在语音增强这一重要而且活跃的技术领域做一些学习和探索。Philipos C. Loizou 博士的《语音增强——原理与实践》无疑是该领域难得的一本教材。在阅读的同时，我们发现国内很少有针对语音增强的专著，因此萌生了将此书译为中文的想法。

这一想法得到了作者本人的支持以及部门经理的鼓励，因此在此后一年多的时间里，在白天繁重的研发工作之余，我们坚持了对该书的翻译工作，无数个节假日和深夜均伴随此书渡过。此间公司项目一度十分繁忙，翻译的工作量也超乎预料，但是学习的乐趣仍然让我们累并快乐着。

原著是 Philipos C. Loizou 博士长期从事语音信号处理研究和教学工作的结晶。我们认为本书的主要特点在于：第一，与其他专著不同，本书由浅入深地介绍语音增强所需的基础知识，适合作为教材或自学参考书；第二，本书全面介绍了单通道语音增强的四大类算法以及噪声估计算法，其中的许多理论和方法也是学习多通道语音增强算法的基础。第三，详细介绍了语音质量/可懂度评价的各类方法，并做了大量工作对主流算法进行比较测试，具有很强的参考性。第四，原书附带光盘包括各类主要算法的 MATLAB 代码实现，并提供一个通用的语音数据库用于对算法进行测评，极大地方便了研发与测试。

在本书的翻译过程中，原书作者 Philipos C. Loizou 博士为我们寄来手稿，并一直关心我们的翻译进度以及遇到的困难，十分及时地解答我们提出的问题，同时介绍他的学生，现为威斯康星大学密尔沃基分校（University of Wisconsin-Milwaukee）助理教授的 Yi Hu 博士与我们就很多术语的中文译法进行了反复探讨。Yi Hu 博士对原书内容十分熟悉，并对其形成具有重要贡献。Loizou 博士和 Yi Hu 博士的帮助使得我们对原书内容具有更准确的理解。译者邓方的导师四川大学教授何培宇博士，成都理工大学副教授陆从德博士为本

书提出了很多建设性意见。中科院自动化所江巍博士审阅了书稿第四章，指出了翻译中的一些疏忽和错误，并探讨了一些术语的译法。在此我们一并表示真诚的感谢。

高毅主要负责第一章至第六章、第九章以及前言和附录等的翻译，肖莉和吴绍炜负责第七章、第八章的翻译，邓方负责第十章、第十一章的翻译。高毅对本书进行了统稿。由于译者水平有限，书中难免存在错误和不当之处，敬请读者批评指正。

2012 年 9 月 29 日从作者的夫人 Demetria Loizou 女士处惊闻 Loizou 教授因病不幸逝世的消息，悲痛不已，仅以此译作缅怀这位老师和朋友。

译　者

2012 年 8 月

目　录

第1章 引 言

语音增强（speech enhancement）主要关心如何改善人耳对带噪语音在某些特定方面的感知，而这些感知往往受到噪声影响。大多数应用中，语音增强的目标是要提高这些受损语音的质量（quality）以及可懂度（intelligibility）。人们往往希望改善语音质量以减轻听觉上的疲劳，特别是需要长时间聆听大噪声环境中的语音时（例如生产车间）。语音增强算法能在某种程度上减轻或者抑制背景噪声，因此有时也被称为噪声抑制（noise suppression）算法。

在很多场合我们都需要对语音进行增强，包括在嘈杂的环境中说话，或者语音受到通信信道噪声的影响。例如，在通过蜂窝式移动电话进行语音通讯时，发送端语音往往带有汽车噪声或者餐馆中的嘈杂声等背景噪声。蜂窝电话标准中，声码器会被用来对语音进行压缩编码，语音增强算法可以用作声码器前端的预处理器，以改善接收端的语音质量（例如文献[1]）。如果蜂窝电话通过语音识别系统来进行语音拨号，那么识别精度很可能受到噪声的影响，因此，在进行识别之前，可将带噪语音通过语音增强算法进行预处理。在空对地通讯的过程中，驾驶舱里极大的噪声将会对飞行员的声音形成严重干扰，这时我们需要利用语音增强算法来改进语音质量及可懂度。在这个例子以及其他很多类似的军用通讯系统中，对语音可懂度的要求通常高于对语音质量的要求。在电话会议系统中，在某一个终端采集到的噪声将被传送到其他所有接收端，如果这个终端所在的房间会产生回响的话，情况会更糟。因此如果能在将音频信号广播到其他接收端之前对带噪语音进行增强，毫无疑问会改善整个系统性能。最后，那些需要使用助听设备（或者人工耳蜗）的听障人士在噪声中交流会感到十分困难。而语音增强算法可以在带噪语音信号被放大之前对之进行预处理，在一定程度上“净化”带噪信号。

前面这些例子说明，语音增强的目标与具体的应用相关。理想情况下，我们希望语音增强算法既能改善语音质量，又能提高可懂度。实际上，大多数算法只是改善了语音的质量。在减少背景噪声的同时，也会引入语音的失真，进而损伤了语音的可懂度。因此，语音增强的主要挑战就在于设计一个高效的算法，在不明显引入信号失真的前提下，对其中的噪声进行有效抑制。

语音增强的具体解决方案与很多因素密切相关，包括具体的应用场景、噪声源或者干扰信号的特性、噪声与纯净信号之间的关系（如果有的话）、麦克风或者传感器的数量。干扰信号可能是类似噪声（例如风扇的噪声），也可能是类似语音的信号，例如：多人说话环境（比如饭馆）中的其他竞争说话人发出的语音信号。声学噪声对于纯净信号而言有可能是加性的，也可能是卷积性的，比如房间里产生了严重回声的情况。再者，噪声与纯净的语音信号之间可能与统计相关或者无关；麦克风的数量也能具有不同的语音增强算法效果。一般来说，麦克风数量越多，越容易对语音进行增强。当至少有一个麦克风靠近噪声源的时候，还可以利用一些自适应技术来消噪。

本书将主要关注受统计无关（且独立）的加性噪声污染的语音信号的增强方法。书中提到的这些增强算法不局限于某几种特定类型的噪声，对很多类型的噪声源均普遍适用（详见后文）。另外，本书假设用于增强的带噪信号——包括纯净语音和加性噪声——均来自于单个麦克风。这是最具挑战性的语音增强应用场景之一，因为没有来自其它麦克风的参考信号。

1.1 了解敌人：噪声

在设计出可以抵御噪声的算法之前，有必要首先了解现实生活中可能遇到的各类噪声的特性以及不同噪声之间的区别，包括时域和频域特征以及噪声声强的范围等等。

1.1.1 噪声源

无论身处何处，我们都被噪声所包围。例如不管在大街上（例如汽车驶过，街道施工），办公室（例如电脑风扇，通风设备噪声），餐馆（例如附近餐桌的谈话声），还是百货商场（例如电话铃声，销售员讲话），噪声都无处不在。正如这些例子所示，日常生活中的噪声存在不同的形式。

噪声可以是平稳的，即不随时间而改变，如电脑风扇噪声。噪声也可以是非平稳的，比如餐馆里面的背景噪声，很多人说话的同时还夹杂着厨房里传出的声音。这种餐馆噪声的频域（以及时域）特征随着周围餐桌上人们的对话以及服务员与顾客交流内容的改变而改变。显然，对不断变化的（非平稳）噪声进行抑制的难度要远远大于平稳噪声。

不同噪声的另一个区别在于它们的频谱形状，特别是噪声能量在频域的分布。例如，风噪声（wind noise）的主要能量都集中在低频段，一般在 500Hz 以下。而餐馆噪声（restaurant noise）的主要能量却分布在更宽的频带内。图 1.1 到图 1.3 展示了汽车噪声、火车噪声以及餐馆噪声的时域波形的例子（这些噪声源取材于本书附带光盘中的 NOIZEUS 语料库[2]），以及相应的长时平均幅度谱。以上三个噪声源中，汽车噪声比较平稳，而火车噪声和餐馆噪声却不平稳。从图 1.1 到图 1.3 所示可以看到，这三个噪声源在频域的区别比在时域更明显：汽车噪声的大部分能量集中在低频段，具有类似低通的特性；另一方面，火车噪声看起来则具有更大的带宽，因为其能量分布在更宽的频率范围。

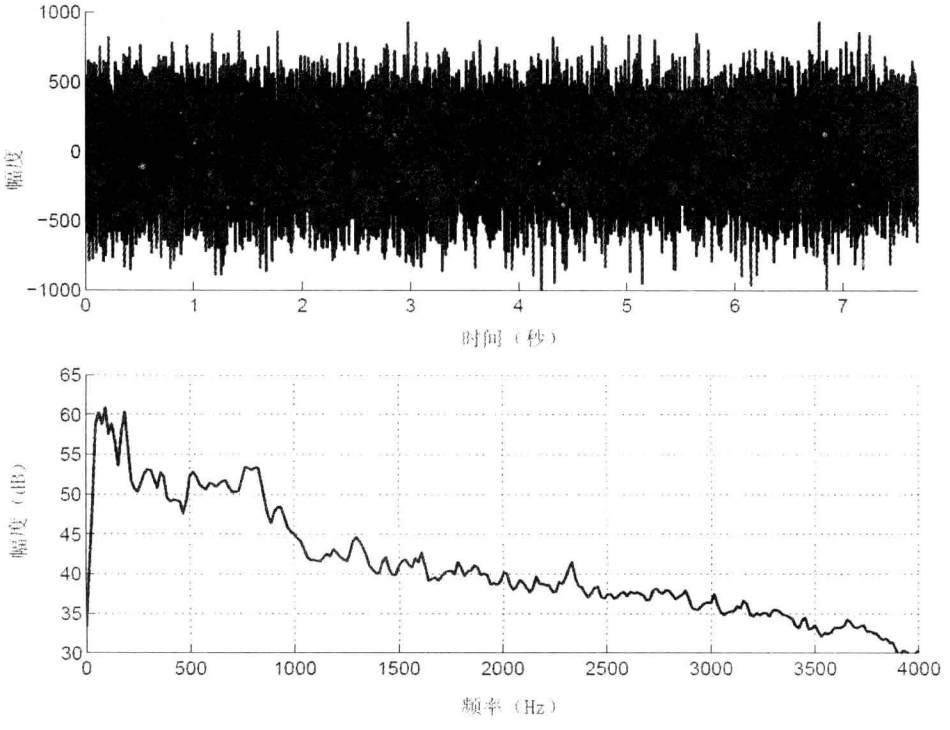

图 1.1　上图显示了一段汽车噪声样本，下图为其长时平均频谱

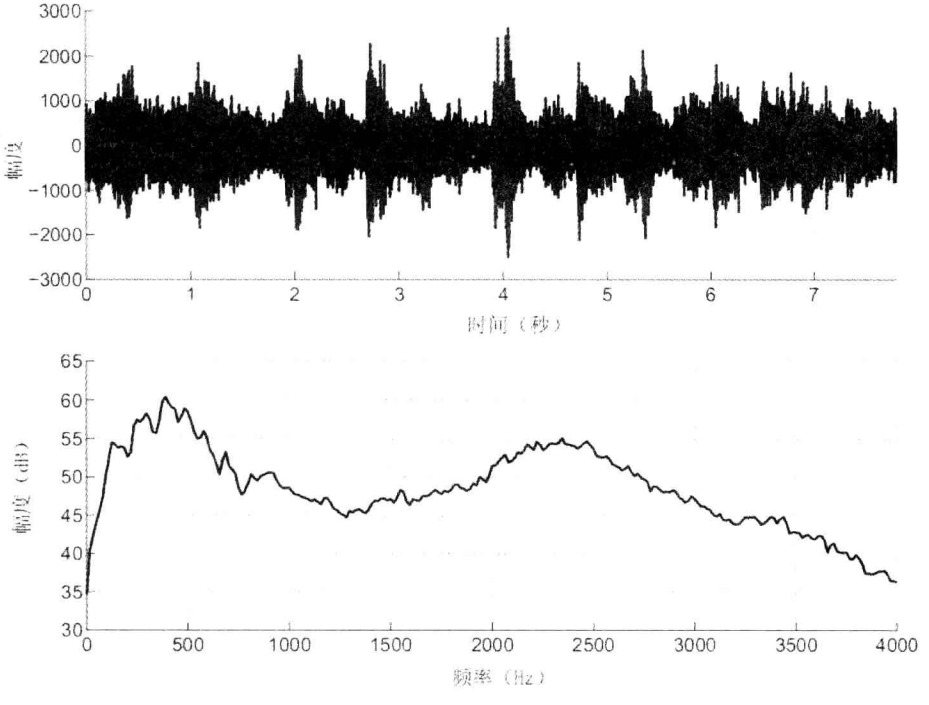

图 1.2　上图为一段火车噪声样本，下图为其长时平均频谱

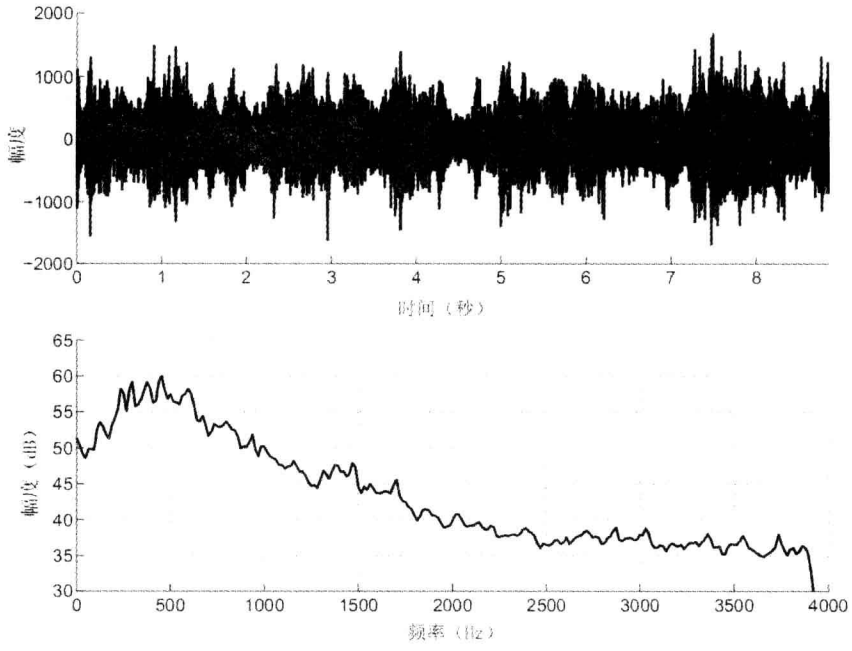

图 1.3　上图为一段餐馆噪声样本，下图为其长时平均频谱

1.1.2　噪声和语音声强

设计语音增强算法的关键知识还包括了解现实世界中的语音和噪声的声强级。借此，我们就可以估计信噪比（SNR）的范围。这一点是很重要的，因为语音增强算法需要在一定的信噪比范围内进行有效的噪声抑制以及改进语音质量。

一些对语音和噪声强度的综合分析测量工作已经由 Pearsons 等人完成[3]。他们考虑了日常生活中能遇到的各种环境，包括教室、城市与郊区房屋（屋内和屋外）、医院（护士站和病房）、百货店、火车与飞机环境。语音和噪声声强级通过声级计（sound level meters）来测量，测试结果通过 dB SPL（decibel sound pressure level，分贝声压级）来表示（dB SPL 是对 0.0002dynes/cm^2 的相对声压，而 0.0002dynes/cm^2 是人耳可以听到的最小声压）。由于说话人与聆听者之间的距离会影响声强级，因此测试中麦克风被放置在远近不同的位置。人们面对面交流的典型距离是 1 米，距离每增加一倍，声强级减少 6dB[4]。在一些特定的场合（例如在乘坐火车或者飞机等交通工具时），人与人之间交流的距离可能减少到 0.4 米[3]。

如图 1.4 所示总结了各种环境下语音和噪声的平均声强。在教室、医院、住所内，以及百货店里面的噪声强度是最低的。在这些环境中，声强级范围大约在 50～55dB SPL，对应的语音强度范围在 60～70dB SPL。这表明在这样的环境中，实际信噪比的范围大约在 5～15dB。在火车和飞机上，噪声强度很高，大约在 70～75dB SPL，相应的语音强度也差不多在这个水平，因此这两种环境下的信噪比在 0dB 左右。在教室中，对于前排最靠近教师的学生来说，实际信噪比是最理想的，但是对于坐在后排的学生来讲就很不舒服了，因为这