Hilbert-huang 变换在信号处理中的应用研究

摘 要

随着现代科学技术的蓬勃发展，数字语音信号处理作为一个跨学科、综合性的研究领域已成为当今的一个研究热点。经研究表明，语音信号是一个复杂的非线性、非平稳随机过程，这使得基于线性平稳线性系统理论发展起来的传统语音信号处理技术性能难以进一步提高。近年来发展起来并逐步完善的非线性、非平稳信号处理方法为语音信号处理技术的发展带来了新的生机。实际中，语音常常受到环境噪声的干扰而使通话质量下降，使得语音处理系统不能正常工作。在这种情况下，必须采用信号处理方法进行语音的检测和语音增强，抑制背景噪声，以提高语音通信质量。语音增强不仅可以提高语音的清晰度，改善听觉质量，而且在许多语音编码和识别系统中，通过增强处理可以大大改善系统在含噪条件下的性能。语音增强是语音信号处理的一个重要分支，该技术已广泛应用于无线电话、电话会议、场景录音和军事窃听等领域。语音检测和增强技术无论在日常生活中，还是在军事领域，或者对语音信号处理技术都很有应用价值。

关键词：经验模态分解；Hilbert—Huang 变换；语音检测；语音增强

Abstract

With the vigorous development of modern science and technology, digital voice signal processing as an interdisciplinary, comprehensive field of study has become a hot spot of research today. The study showed that speech signal is a complex nonlinear and non-stationary random process, this makes based on linear smooth linear system theory developed speech signal processing technology of traditional to further improve the performance. Developed in recent years to the nonlinear and gradually improve the non-stationary signal processing method, in speech signal processing technology development brings new vitality. The actual speech often affected by environmental noise interference and make calls the quality descend, makes the speech processing system didn't work properly. In this case, must use signal processing method for voice detection and speech enhancement, inhibiting the background noise, in order to improve the quality of speech communication. Speech enhancement can not only improve the clarity of the hearing, speech to improve the quality, and is in many speech coding and recognition system, through increased handling can be greatly improved system in the performance under the condition with noise. Speech enhancement of the speech signal processing is an important branch, this technique has been widely used in wireless telephone, conference calls, scene recording and military eavesdropping, etc. Voice detection and enhancement technology both in daily life, or in the military field, or to the speech signal processing techniques are applied value.

Keywords: empirical mode decomposition; Midlands transform; Hilbert -- Voice detection; Speech enhancement

引 言

HHTL是一种独特的完全自适应时频分析方法，它既适合于非线性、非平稳信号的分析，也适合于线性、平稳信号的分析，并且对于线性、平稳信号的分析比其他的时频分析方法更好地反映了信号的物理意义。以一种全新的非线性信号分析方法——Hilbert-Huang变换(Hilbert-Huang transform，HHT)为基础，将其应用于电力系统暂态信号方面的研究。该方法利用经验模态分解法EMD(empirical mode decomposition)将暂态信号分解成有限个固有模式函数IMF(intrinsicmodefunction)，从而根据Hilbert变换求出每一个固有模式函数瞬时频率和瞬时幅值，对电力系统暂态信号进行分析。仿真和实例分析结果均表明，该方法具有很强的自适应能力，应用于电力系统暂态信号分析方面具有可行性、有效性。

Hilbert-Huang 变换,它以瞬时频率为基本量，以固有模态信号为基本信号，与以往的时频分析方法相比有明显的区别。希尔伯特一黄变换(HHT) 是 1998 年由 NASA 的 Noren EHuang 等人提出，作为一个崭新的时频分析方法，能够进行非线性、非平稳信号的线性化和平稳化处理，被认为是近年来对以傅立叶变换为基础的线性和稳态谱分析的一个重大突破。已在海洋工程、地震工程、土木工程损伤识别等领域得以成功的应用。

以一种全新的非线性信号分析方法——Hilbert—Huang 变换(Hilbert—Huang transform，HHT)为基础，将其应用于变换在信号处理方面的研究。该方法利用经验模态分解法 EMD(empirical mode decomposition)将暂态信号分解成有限个固有模式函数 IMF(intrinsic mode function)，从而根据 Hilbert 变换求出每一个固有模式函数瞬时频率和瞬时幅值，对电力系统暂态信号进行分析。仿真和实例分析结果均表明，该方法具有很强的自适应能力，应用于变换在信号处理暂态信号分析方面具有可行性、有效性。

一、绪论

本课题研究的目的是 Hilbert—Huang 变换在信号处理中的应用研究。

本课题的研究无论在军用还是在民用方面都有着重要的意义。在军用上，如何有效地截获对方的信息和高效地传输信息是电子战和电子对抗的一个重要组成部分。在民用上随着移动通信的普及必然对传输信道的质量、效率和容量提出更高的要求。随着现代科学的蓬勃发展，人类社会愈来愈显示出信息社会的特点。通信或信息交换已成为人类社会存在的必要条件，正如衣食住行对人类是必要的一样。语音作为语言的声学表现，是人类交流信息最自然、最有效、最方便的手段之一，是人们最重要的交际工具。

然而，人们在语音通信过程中不可避免地会受到来自周围环境和传输媒介引入的噪声、通信设备内部电噪声、乃至其他讲话者的干扰。这些干扰最终将使接收者接收到的语音是被噪声污染过的语音。例如，汽车、街道、机场中的电话，常受到强背景噪声的干扰，严重影响通话质量。而且环境噪声的污染使得许多语音处理系统的性能急剧恶化。语音识别虽然已取得重大进展，正步入实用阶段。但目前的语音识别系统大都是在安静环境中工作的，在噪声环境中尤其是强噪声环境，语音识别系统的识别率受到严重影响。低速率语音编码，特别是参数编码，也遇到类似问题。由于语音生成模型是低速率编码的基础，当模型参数的提取受到混杂在语音中背景噪声严重干扰时，重建语音的质量将急剧恶化，甚至变得完全不可懂。在上述情况下，必须加入语音检测和语音增强系统，以提高语音通信质量，或者作为预处理器，以提高语音处理系统的抗干扰能力，维持系统性能。因此，研究语音检测和语音增强技术在实际中有重要价值。目前，语音检测和语音增强已在语音处理系统、通信、多媒体技术、数字化家电等领域得到了越来越广泛的应用。如何在复杂声学环境中准确地检测出语音信号，并实现语音的增强，目前仍是一个很有挑战性的课题。因为语音信号是非线性、非平稳的复杂的随机信号，而本课题面对的是(尤其是短波信道)很难预测的随机复杂噪声，这就给语音的检测和增强造成了更大的困难。新的问题呼唤新的理论，希尔伯特一黄变换(Hilbert—Huang Transform，HHT)是上世纪末 Huang 等人首次提出的一种新的信号分析理论。它的主要创新是固有模态(Intrinsic Mode Function，IMF)概念的提出和经验筛法(EmpiricalModeDecomposition，EMD)的引入。通过 EMD 将信号分解成 IMF，一般为有限数目的

。对每个 IMF 进行 Hiibert 变换就可以获得有意义的瞬时频率从而给出频率变化的精确表达。信号最终可以被表示为时频平面上的能量分布，称为 Hiibert 谱。进而还可以得到信号的边际谱。HIIT 是基于信号局部特征的和自适应的，因而是高效的，它特别适用于分析大量频率随时间变化的非线性、非平稳信号。但 HHT 的第一篇公开文献直到 1998 年才发表，因而这一理论出现的时间还很短暂其完善和发展还有诸多工作要做，HHT 是现阶段一个全新的研究课题。HHT 信号分析具有重要的理论价值和广阔的应用前景已在一些实际工程领域中获得了有效地应用。

本课题的主要研究目的是，深入研究 HIIT 变换理论，对其中的算法做进一步的改进。并将此理论引入到话带信号分析中来，为语音信号处理提供新的手段和方法。为噪声环境中的语音信号检测和增强开辟新的有效途径。

二、Hilbert-huang 变换及信号处理算法

(一)Hilbert-huang 语音检测

语音流检测对于语音信号的编码和识别都有很重要的意义。在不同的场合对于检测有不同的称呼，对于不同的应用有不同的算法。在语音识别中准确的端点检测可以大大地提高识别率。在语音传输中，语音检测可以实现对信号的多速率编码，减少信道的负载。在语音截收中可极大地减少话务处理量。在语音增强中可以实现自适应的语音增强。

(二)、语音识别中的端点检测

在语音识别系统中将语音检测称为语音端点检测(End Point Detection：EPD)。它需要准确地找出语音的起始和终止点的位置。总体上来讲语音识别中的端点检测方法大致有模型匹配法和门限法两大类。

1)模型匹配方法

Mel 倒谱系数(MFCC)是语音识别器常用的特征，是受人的听觉系统研究成果推动而导出的声学特征。文献[1]用 MFCC 作为分类特征来建立噪声模型，计算每帧信号对噪声模型的似然得分，将得分与门限比较，把每帧信号初步分为语音和噪声，最后有一个根据多帧信号的分类情况平滑和判断的过程。整个过程分三步：训练、检测和自适应。其中，训练是指估计噪声模型的参数，每个特征分量为单

高斯建模。自适应是为了解决训练环境和测试环境的不匹配，动态调整模型的参数。实验结果指出在强噪声环境下，该方法优于全能量和基频方法。文献[2]将LDA(linear Discriminant Analysis)

用于 MFCC，使得 MFCC 如同一个单系数用于端点检测。LDA 的目的在于对于端点检测的分类问题，找到一个线性函数来最大化类间差异，最小化类内差异。经 LDA 线性变换的 MFCC 特征结合能量特征获得好于能量方法的效果，尤其在噪声环境下(15dB 以下)。HMM 模型方法是较常见的模型匹配方法。文献[3]提出 HMM 模型方法，分别用一个 HMM 模型对背景噪声和语音建模，取得了较好的效果，但是单个的 HMM 模型不足以描述所有可能出现在端点附近的音素，所以文献[4]提出了为语音建立多个 HMM 模型的方法。首先提取特征，然后用一数帧长的滑动窗，沿时间轴移动。根据检测目的，假设每个滑动窗被分为一前一后的噪声和语音两部分或是语音和噪声两部分。对噪声部分计算在噪声的 HMM 模型下的前向概率，对语音部分，分别计算在多个语音 HMM 模型下的前向概率并从中挑选最大的概率作为语音部分的概率，将两部分的概率相乘作为联合概率，所要求的端点即是联合概率最大的地方。这种方法需要事先用 Baum—Welch 算法训练模型。实验结果表明当噪声的谱特性和语音相差越大时，如低频噪声，检测性能越好。

2)门限比较法

a)子带能量方法

人的语音在 200Hz 到 1000Hz 内能量分布非常密集稳定，所以采用一个带通滤波器对语音进行处理后，把 200Hz 到 1000Hz 的能量作为端点检测的特征。对于不同的应用场合其信噪比变化较大，估计背景噪声动态确定门限。整个过程包括背景噪声估计，检测语音开始，确信开始，检测语音结束和确信结束。首先根据统计预先确定上下限，然后根据开始五帧的平均能量来估计背景噪声并由此确定门限(比如是背景噪声的 1．5 倍加一个常量)。如果该结果大于预先确定的上限，取上限为门限，以防止门限过大检测不到语音，如果该结果小于预先确定的下限，则取预先确定的下限为门限，如果计算结果在预先设定的上。则取预先确定的下限之间，则取该结果为门限，这样可以防止背景噪声检测不准确带来错误。如果发现有连续 n，帧能量大于门限，则认为语音开始，否则重新开始估计背景噪声。对语音终点的检测和语音开始的检测相似，发现连续″，帧能量低于门限时标记结束，如果在随后的 m 帧中没有连续 n。帧能量大于门限则确信语音结束。

b)基于熵的方法

由于语音和大部分噪声在频域的分布上有很大的差异，可以在频域进行端点

检测。文献[5]首先通过快速傅立叶变换来得到每一帧信号的频谱，其中每个频谱向量的各个系数表明了该帧信号在该频率点的大小分布。然后，计算每一帧的每个频谱分量在每帧的总能量中所占的比例，将其作为代表信号能量集中在某频率点的概率大小，即计算熵所需的概率密度函数通过下式得到：

(三)、语音通信中的语音激活检测

在通信中语音检测常常被称为语音激活检测(Voice Activity DetectiOn：VAD)。在移动通信、卫星通信等应用场合，通常采用不连续传输(Di scontinuouS Transmi SSiOff：DTX)的工作模式。DTX 工作模式是指当语音存在时，以正常的速率进行传输；而在无话时，则不传输任何信息或以较低的速率传输背景的信息。语音激活检测是一种检测输入信号是否为语音的技术。它主要应用在语音通信领域，如可变速率语音编码、不连续传输和数字语音插空等。据统计在电话交谈中，有话区间所占的比例大约 40％。这样，提高通信系统效率的一种直接方法就是采用语音激活检测技术。采用 VAD 不仅能够提高通信效率，而且还可以降低手持设备的平均功耗，在移动通信、微波、卫星等通信方式中节省带宽。

随着通信技术的不断发展，VAD 检测算法也不断的涌现出来。语音激活检测是语音通信中重要组成部分，一般情况下，VAD 算法基本原则假定如下：

1)语音信号是非平稳信号,在较短的时间(20～30ms)内频谱不会有很大的变化。

2)相当长的时间内，背景噪声频谱是平稳的，并随时间缓慢地变化。语音信号的电平通常高于背景噪声的电平。

3)在基于变 Bit 率语音编码传输通讯系统中，VAD 占有很重要的位置，它可以增加用户、减少设备的消耗等。因此对 VAD 算法研究受到了越来越大重视。

目前，在通信中应用了两个标准，一个是国际电信组织提出的 ITU—T G.729AnnexB,最近 ITU—T 的 16 工作组对其进行了改进提出了 FUZZY VAD(FVAD)算法；另一个是欧洲通信标准委员会 ETSI 提出的用于 3G 移动通信系统的 AMR-1 和 AMR-2。在研究主动的语音通信中，VAD 可做的相对“粗糙”，个别词语的脱节可通过语义、上下文等“校正”或“补正”。

1)FVAD 算法描述

ITU—T G.729AnnexB 主要是为固定电话和多媒体传输制定的标准。它的采样为 8kHz，帧长为 10ms。采用以下几个参数作为判断标准：

- 整个能量带宽差分 AE，
- 低带能量差分 aE,
- 过零率差分△ZC
- 谱失真丛

利用 6 个模糊规则进行判断有 / 无语音。

（四）短波通信中的语音流检测

目前短波通信中的语音流检测(SpeechStreamDetectiOil{SSD)主要应用于军事通信中，其有其独特的要求，如下：

1)它不但包括双方通话,还包括第三方听话,尤其是后者是获得信息的重要来源。

2)短波信道本身带来的异常复杂的噪声移各种调制各异的信号；

3)缺少了平常对话中的主动性,对于说话的内容、对话的时间等没有事先的知识，这一切对于第三者来说都是随机的。

基于这些特点，我们把这种语言检测称为语音流检测 SSD。短波通信与电话和卫星移动通讯存在着很大差别。在短波通讯中，存在各种各样的调制信号，加密经号等。另外由于传播途经的不同，短波信号更容易受到干扰。因此以上两种方法都不能达到很好的效果。

(五) Hilbert-huang 语音增强

语音增强的主要目标是从带噪语音信号中提取尽可能纯净的原始语音。然而，由于干扰通常都是随机的，从带噪语音中提取完全纯净的语音几乎不可能。在此情况下，语音增强的主要目的就是通过对带噪语音进行处理，以消除背景噪声，改善语音质量，提高语音的清晰度、可懂度，提高语音处理系统的性能。这些目的往往不能兼得，通常要根据语音处理系统的具体需要而定。语音增强方法的研究始于 20 世纪 70 年代中期。随着数字信号处理理论的成熟，语音增强发展成为语音信号处理领域的一个重要分支。1978 年，LJm 和 Oppenhim 提出了语音增强的维纳滤波方法。近些年来，随着 VLSI 技术的发展和高速 DSP 芯片的出现，语音增强方法逐步走向实用，同时新的语音增强方法又相继涌现。语音增强不但与语音信号处理理论有关，而且涉及到人的听觉感知和语音学。噪声来源众多，随应用场合而异，它们的特性也各不相同。即使在实验室仿真条件下，也难以找到一种通用的语音增强算法，能适用于各种噪声环境。所以必须针对不同的噪声，采取不同的语音增强对策。

三、基于语音生成模型的增强研究

维纳滤波只能保证在平稳条件下最小均方误差意义的最优估计，同时采用维纳滤波并没有完全利用语音的生成模型，卡尔曼滤波器可以弥补这些缺陷。

卡尔曼滤波通过引入卡尔曼新息，并将要解决的滤波与预测混合问题转化为纯滤波和纯预测两个独立的问题，适合于非平稳条件下的虽小均方误差意义下的最优估计。

(一)基于神经网络的语音增强

语音增强在一定意义上是区分背景噪声和语音的问题，因此可以利用神经网络技术实现语音增强。经过多年的发展，已经提出了一系列应用于语音增强的神经网络方法。上世纪 80 年代中期 Tamura 和 waibel 等人就利用四层全连接 BP 网从各种平稳和非平稳噪声中提取语音。神经网络在语音增强中的应用主要有以下两个方面。

1)时域滤波法

时域滤波的方法基于测试语音和噪声环境的分布和训练时相同且分布保持不变的假设，需要利用带噪语音和干净目标语音分别进行训练，得到合适的预测神经元模型。为得到语音的最大似然估计，在扩展的卡尔曼滤波过程中，使用训练得到预测神经元模型，将噪声抑制 a

2)变换域滤波法

使用带噪语音和干净的目标语音在变换域中对神经网络进行训练。SNR 或其它一些测度可以作为网络的输入，这种方法的前提是 SNR 估计是正确的，即语音、噪声的统计分布是特定的。利用训练得到的神经元，构造可以对语音和噪声进行分类的分类器，实现语音增强。

(二)、基于 HMM 的语音增强

为了更好地描述信号的非平稳性，可以采用基于状态空间的变换方法，对不同类别的语音和噪声建立不同的模型。目前有两种转换方法，一种是构造分类器，利用分类器对当前信号进行最佳逼近；另一种就是隐马尔科夫模型(HMM)。HMM 的各个状态可以对语音、噪声信号所不同的区域进行充分建模，另外，由于要准确地将噪声估计出来，必须保证在只有噪声的情况下 HblM 也可以正确进行分类。此时，利用 HMM 可以对状态转移概率进行建模，将可能为噪声的信号部分滤除就可以做到语音增强了。