数据处理中的矩阵方法

--读书笔记

单通道语音信噪分离算法研究

硕士研究生：秦炜

学号：202128002627001

学科专业：信号与信息处理

所在单位：声学所东海研究站

完成时间：2022 年 5 月 28 日

单通道语音信噪分离算法研究

摘要：为了评估单通道语音信噪分离的效果，本文分别对六种传统语音增强算法进行了探讨。在理想的高斯白噪声环境下，子空间法增强后的语音信号输出信噪比 SNR 最大，VMD(Variational Mode Decomposition, VMD)增强后的语音信号语谱图保留频率细节部分更多，分段信噪比 segSNR 最高。在八种不同场景不同信噪比复杂环境下，维纳滤波法增强后的语音信号分段信噪比 segSNR 最大，语音感知度最高。

关键词：语言增强；VMD；子空间；维纳滤波法；

目录

1 绪论

在单麦克风场景下，语音增强算法主要通过对带噪语音信号进行时域、频域、时频域、KLT 域以及 SVD 域的系数变换，并根据变换后的系数特征来区分纯净的语音信号和噪声，进而达到信噪分离的目的。在简单场景下输入信噪比较高的情况下，可以直接通过将噪声系数置零来达到语音增强的目的，但对于复杂场景下低信噪比的带噪语音信号简单地进行噪声系数置零可能会导致增强后的语音信号失真。

在八九十年代，所提出大多数语音增强算法是基于频域的。最先谱减法（Spectral Subtraction, SS）思想含噪语音信号能量谱与前几帧估计到的平均能量谱相减，得到估计到的增强语音信号能量谱；但其中噪声能量谱估计得不准确会导致频谱过减而产生“音乐噪声”，降低了语音增强质量。然而，谱减法还存在“相位问题”，将带噪语音信号的相位直接作为增强语音信号的相位可能会造成失真现象。基于统计模型的语音增强算法有维纳滤波法（Wiener Filter, WF）、最小均方误差（Minimum Mean Square Error, MMSE）算法等；MMSE 算法增强后的语音信号中残留噪声量相对较少，其中最小均方误差（Log-MMSE）算法残留噪声更少且舒适度得到了提升。子空间（subspace）方法的原理是将观测信号的向量空间分解为信号子空间和噪声子空间，通过保留信号子空间和消除噪声子空间并从而估计出干净语音。在复杂环境低信噪比情况下，子空间方法仍不能有效地分离出噪声。

由于语音信号非线性、非平稳性等性质，传统的傅里叶变换方法已不再适用。针对语音信号的非平稳、时变等特性，许多学者提出了一些基于时频域方法。小波变换（Wavelet Transform, WT）是时间和频率的局部化分析，通过伸缩平移运算对信号逐步进行多尺度细化，最终达到高频处时间细分，低频处频率细分，能自动适应时频信号分析的要求，从而可聚焦到信号的任意细节。大量实验表明小波变换能够在很大程度上对非平稳信号进行去噪，增强语音信号的信噪比优于其他传统方法增强语音信号的信噪比，但小波去噪的关键取决于阈值的选取，一个合适的小波阈值可以达到较好的语音增强效果，且小波变换不适用于非线性信号。经验模态分解(Empirical mode decomposition, EMD)算法根据信号的局部特性将带噪信号分解为有限个固有模态函数(Intrinsic mode function, IMF)，从而适用于

非线性非平稳信号，大量实验结果表明增强语音信号质量在不同程度上均有一定的提升；但 EMD 算法可能会由于端点检测不准确造成模态混叠现象，造成语音重构信号不准确，从而导致语音质量下降。针对 EMD 存在模态混叠问题，EEMD (Ensemble EMD)、CEEMD (Complementary EEMD)算法可以避免模态混叠现象发生，但无法消除。为了解决 EMD 的模态混叠问题，Konstantin Dragomiretskiy 和 Dominique Zosso 提出了变分模态分解 (Variational Mode Decomposition, VMD)算法，实验表明该算法可以避免端点检测所造成的模态混叠问题，从而增强语音信号质量要明显好于 EMD 增强语音信号质量

本课程论文第 2 章介绍了几种传统的单声道语音增强算法和语音质量评估的主、客观指标。第 3 章为不同语音增强方法对在不同场景下不同信噪比的语音信号进行去噪增强，然后在客观评价的尺度下分别计算分段信噪比和语音感知度。

2 单通道语音增强技术

2.1 语音与噪声特性

2.1.1 语音特性

语音信号是一个典型的随机信号，虽然其二阶统计量的变化随时间变化但是其满足短时平稳性，即在很短的时间内(10~30ms)可以把语音信号看成平稳随机过程。语音信号具有周期性，这是人体本身的发声机制所决定的。实际工程中可通过提取语音信号的基音周期进行一系列的语音信号处理，同时研究语音的声学特征可很好的提升语音可懂度与质量。在工程模型中，声带与声道可分别看作激励源和滤波器，而声带的震动可以是周期同时也可以是非周期的。图 2-1 为工业中通用的语音产生模型。

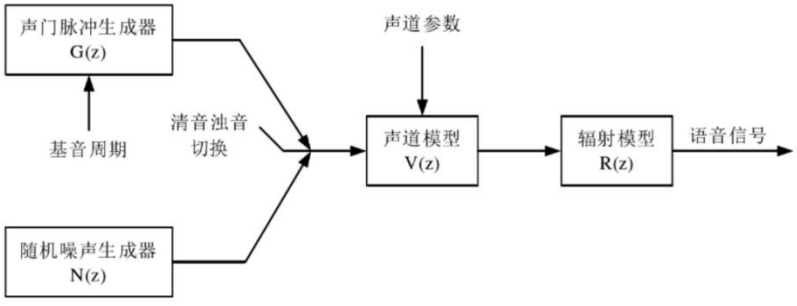

图 2.1

当输出的语音信号 X(z)为浊音时，其模型表达式如下

$$X(z)=G(z)V(z)R(z) \tag{2.1}$$

式（2.1）中 $R(z)=1-z^{-1}$。

当切换为清音时，输出的语音信号类似于白噪声，此时输出的语音信号的表达式为

$$X(z)=N(z)V(z)R(z) \tag{2.2}$$

2.1.2 噪声特性

噪声是自然界最为复杂的信号之一，普遍存在于人们的日常生活中。根据噪声对语音频谱干扰方式的不同，可以将噪声分为乘性噪声和加性噪声。乘性噪声是指噪声和语音在频域是相乘的关系，在时域和语音则是卷积关系。加性噪声是指当噪声对语音的干扰表现为两者信号在时域进行相加，而显然噪声和语音频域中也为相加关系，从能量角度看背景噪声和语音的声强是叠加关系。实际环境中背景噪声可以看成加性噪声，如风扇的声音、汽车引擎声、周围人说话声等。

2.2 语音增强算法

2.2.1 谱减法

谱减法是简单、易于实现，先假设含噪语音信号 y(n) 是纯净语音信号 x(n) 与加性噪声 d(n) 相加得到的，其数学表示形式为

$$y(n)=x(n)+d(n) \tag{2.3}$$

式（2.3）两边做 DTFT 变换，可以得到

$$Y(\omega)=X(\omega)+D(\omega) \tag{2.4}$$

由于人耳对语音相位不敏感，可估计得到语音信号频谱 $\hat{X}(\omega)$ 为

$$\hat{X}(\omega)=\left[\left|Y(\omega)\right|-\left|\hat{D}(\omega)\right|\right]e^{j\phi_y(w)} \tag{2.5}$$

式中：$\hat{D}(\omega)=\left|D(\omega)\right|e^{j\phi_y(w)}$ 为噪声估计值；$\phi_y(\omega)$ 为带噪语音信号的相位谱。

为了得到 y(n) 的功率谱，可将它的共轭函数 $Y^*(\omega)$ 分别与公式（2.4）的

两边相乘，从而得到

$$|Y(\omega)|^2=|X(\omega)|^2+|D(\omega)|^2+2\mathrm{Re}\{X(\omega)D^*(\omega)\} \tag{2.6}$$

假设纯净语音信号 x(n) 与加性噪声 d(n) 不相关，且噪声均值为 0，可得纯净语音信号 x(n) 的估计功率谱为

$$\left|\hat{X}(\omega)\right|^2=|Y(\omega)|^2-\left|\hat{D}(\omega)\right|^2 \tag{2.7}$$

2.2.2 时域维纳滤波

若输入信号 y(n) 和期望信号 d (n) 是联合广义平稳随机过程，那么系统输出对应的误差可以表示为

$$e(n)=d(n)-\hat{d}(n)=d(n)-h^T y \tag{2.8}$$

式（2.8）中 $y=[y(n),y(n-1),y(n-2),\cdots,y(n-M+1)]^T$ 为输入的向量，$h=[h_0,h_1,h_2\cdots,h_{M-1}]^T$ 为滤波器的系数。

为了找到最优的滤波器系数，以最小均方误差的准则进行求解

$$\begin{aligned}J&=E[e^2(n)]=E[(d(n)-h^T y)^2]\\&=E[d^2(n)]-2h^T E[yd(n)]+h^T E[yy^T]h\\&=E[d^2(n)]-2h^T R_{yd}+h^T R_{yy}h\end{aligned} \tag{2.9}$$

式（2.9）中 R_{yd} 为输入信号与期望信号的互相关，R_{yy} 为输入信号的自相关矩阵，为了使代价函数 J 最小，对其进行求导可以得到

$$\frac{\partial J}{\partial h_k}=2E\left[e(n)\frac{\partial e(n)}{\partial h_k}\right]=0,k=0,1,\cdots,M-1 \tag{2.10}$$

由于 $\frac{\partial e(n)}{\partial h_k}=-y(n-k)$，式（2.10）可以简化为

$$\frac{\partial J}{\partial h_k}=-2E\left[e(n)y(n-k)\right]=0,k=0,1,\cdots,M-1 \tag{2.11}$$

利用矩阵求导可以得到

$$\frac{\partial J}{\partial h}=-2R_{yd}+2h^T R_{yy}=0,k=0,1,\cdots,M-1 \tag{2.12}$$

那么滤波器的最优系数为

$$h_{opt} = R_{yy}^{-1} R_{yd} \tag{2.13}$$

式（2.13）称为 Wiener-Hopf 方程的解。

2.2.3 子空间

子空间方法的原理是将观测信号的向量空间分解为信号子空间和噪声子空间，通过消除噪声子空间并保留信号子空间从而估计出干净语音，子空间分解过程是对带噪语音信号做 KLT 变换,然后设置一个门限阈值，利用 KLT 系数的稀疏性，将噪声的 KLT 系数置为 0，之后通过逆 KLT 变换得到增强后的语音。若观测噪声 d(n) 与干净语音 x(n) 无关，则带噪语音可以表示为

$$y(n) = x(n) + d(n) \tag{2.14}$$

假定对干净语音 x (n) 的线性估计为

$$\hat{x}(n) = Hy(n) \tag{2.15}$$

式（2.15）中 H 为 K×K 的滤波器矩阵。

那么滤波误差可以写为

$$\xi = \hat{x}(n) - x(n) = (H - I)x(n) + Hd(n) = \xi_x + \xi_d \tag{2.16}$$

式（2.16）中ξ_x为语音失真，ξ_{d}为残留噪声。

语音失真能量为

$$\xi_x^2 = tr\left(E\left[\xi_x \xi_x^T\right]\right) \tag{2.17}$$

残留噪声能量

$$\xi_d^2 = tr\left(E\left[\xi_d \xi_d^T\right]\right) \tag{2.18}$$

为了得到最优线性滤波器，可以寻求解决如下有约束的优化问题：

$$\min_H \xi_x^2 s.t. \frac{1}{K}\xi_d^2 \le \sigma^2 \tag{2.19}$$

式（2.19）中σ^2为正实数，代表噪声的容忍度。

最优解可以近似写作

$$H_{\mathrm{opt}} = R_x (R_x + \mu R_d)^{-1} \tag{2.20}$$

式（2.20）中 R_x、R_d 分别是干净语音和噪声的协方差矩阵，μ 是拉格朗日乘子。参数 μ 由噪声容忍度确定，其控制着语音失真和残留噪声之间的权衡。

当 $\mu=1$ 时，由于干净语音和噪声的互不相关特性，H_{opt} 变为维纳意义上的最优解 $H_{\text{opt}}=R_xR_y^{-1}$，这意味着，维纳滤波器是子空间算法的一种特殊情况。

对 R_x 进行特征分解

$$R_{\text{x}}=U\Sigma_xU^T \tag{2.21}$$

那么

$$H_{\text{opt}}=R_x(R_x+\mu R_d)^{-1}=U\Sigma_x(\Sigma_x+\mu U^TR_dU)^{-1}U^T \tag{2.22}$$

其中 u_k^T 是 R_x 第 k 个特征向量，于是可以得到次优估计器

$$H_{\text{opt}}=R_x(R_x+\mu R_d)^{-1}=U\Sigma_x(\Sigma_x+\mu\Sigma_d)^{-1}U^T \tag{2.23}$$

因此，子空间算法先将含噪的观测语音变换到 KLT 域，得到 KLT 系数 U^Ty；根据干净语音在 KLT 域的稀疏性对 U^Ty 进行加权处理，得到稀疏化的干净语音 KLT 系数估计值 $U\Sigma_x(\Sigma_x+\mu\Sigma_d)^{-1}U^T$；最后通过逆 KLT 变换得到增强后的语音 $H_{\text{opt}}y$。

2.2.4 小波分析

小波变换与傅里叶变换的不同之处在于：傅里叶变换是无限长三角函数基，而小波变换是有限长会衰减的小波基。这个基函数会伸缩、会平移（其实是两个正交基的分解），然后这个基函数不断和信号做相乘。某一个尺度（宽窄）下乘出来的结果，就可以理解成信号所包含的当前尺度对应频率成分有多少。缩得窄，对应高频；伸得宽，对应低频。

将一个带噪的语音信号通过一个多分辨率的滤波器组，可将其分解为多个子带信号，而这些子信号都具有不同的频率。与此同时，为了解决信号之间的相关性以及能量集中的问题，可采用正交变换进行去除。从而可使得在一些少数的频带上集中信号能量。最后，采用与阈值法、加权法、直接置零法处理其它小波系

数即可达到去噪的效果。小波去噪的效果随着近些年来小波理论的不断发展逐渐提升。

设带有加性噪声的的语音信号为 y(n)，公式（2.1）为其数学表达式，将 y(n) 进行小波变换可得到

$$WT_y(a,b)=WT_x(a,b)+WT_d(a,b) \tag{2.24}$$

式中：a 为尺度因子；b 为时移。a, b 两者均为常数。

令 d(n) 是均值为 0 的平稳随机过程，且具有服从独立同分布的统计特性，记为：$d=(d(0),d(1),\cdots,d(N-1))^T$，因此可得：

$$E\{dd^T\}=\sigma_d^2 I\overset{\Delta}{=}Q \tag{2.25}$$

式（2.25）中 $E\{\bullet\}$ 为均值运算符，Q 为 d 的协方差矩阵。

当小波变换为正交时，小波的变换矩阵也随之变为正交矩阵，设小波变换矩阵为 W 。同时令， y(n) 和 x(n) 所对应的向量为 y 和 x 。 y(n) ， s(n) 和 d(n) 的小波变换分别对应 Y ， S 和 D 向量。

$$Y=Wy,S=Ws,D=Wd \tag{2.26}$$

结合公式（2-24）可得：Y= S+D。令 P 是 D 的协方差矩阵，由于

$$E\{D\}=E\{Wd\}=WE\{d\}=0 \tag{2.27}$$

所以

$$P=E\{DD^T\}=E\{Wdd^TW^T\}=WQW^T \tag{2.28}$$

由于 W 是正交阵，并且 $Q=\sigma_d^2 I$，所以 $P=\sigma_d^2 I$。

通过上述的推导可以看出对于加性噪声信号经过正交小波变换后， x(n) 的相关性在最大程度上被消除了，并且在少数的小波系数上集中着信号的能量。各尺度下的小波变换的模极大值点被保留，并尽可能的减少其他点，最后将处理后的小波系数做逆变换即可达到小波去噪的效果。

2.2.5 变分模态分解

变分模态分解是一种完全非递归、自适应信号处理方法，该方法通过多次迭

代搜索变分模型的最优解来确定每阶分量的带宽和中心频率，从而在频域内能够自适应地实现信号的有效分离。VMD 的整体结构就是变分问题，其约束条件是使每阶模态的估计带宽之和最小，且各模态之和等于输入信号。其对应的约束变分模型构造问题如下：

首先原始信号通过 Hilbert 变换得到解析信号，同时也可以得到其单边频谱，再移动频谱到估计的中心频率，并计算信号梯度平方的 L^2，从而得到各信号的带宽。

则变分问题为：

$$\begin{cases} \min\limits_{\{u_k\},\{w_k\}} \{\sum\limits_k \left\| \partial_t[(\delta(t)+\frac{j}{\pi * t})u_k(t)]e^{-jw_k t} \right\|_2^2 \\ st.\sum\limits_k u_k = f \end{cases} \tag{2.29}$$

式中 f 为待处理信号，$\{u_k\}=\{u_1,u_2,\cdots u_k\}$ 和 $\{\omega_k\}=\{\omega_1,\omega_2,\cdots\omega_k\}$ 分别为分解的 K 个 IMF 模态及其对应中心频率的集合。

为了解决上述变分问题，在式（2.19）的基础上引入二次惩罚参数 α 和拉格朗日乘法算子 $\lambda(t)$，它们的引入可以保证信号在存在高斯白噪声的情况下仍然具有良好的重构精度；同时，二次惩罚参数 α 和拉格朗日乘法算子 $\lambda(t)$ 的组合能够使该算法具有良好的收敛性和对约束条件的严格性。

其增广拉格朗日函数如下：

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \alpha\sum_k \left\| \partial_t[(\partial(t)+\frac{j}{\pi * t})u_k(t)]e^{-j\omega_k t} \right\|_2^2 + \\ \left\| f(t)-\sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t)-\sum_k u_k(t) \right\rangle \end{aligned} \tag{2.30}$$

式中，α 用于确保信号的重构精度，$\lambda(t)$ 使约束条件更具严格性。

因此，这样的变分问题可以用交替方向乘数法（Alternate Direction Method of Multipliers, ADMM）来求解（2.30）式的鞍点,即在频域内不断迭代更新 u_k^{n+1}，ω_k^{n+1}，λ_k^{n+1}。其中，u_k^{n+1} 迭代表达式为

$$u_k^{n+1}=\arg\min_{u_k\in X}\left\{\alpha\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi*t}\right)u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\}+\left\{\left\|f(t)-\sum_i u_i(t)+\frac{\lambda(t)}{2}\right\|_2^2\right\} \tag{2.31}$$

式子 ω_k 和 ω_k^{n+1} 等效，利用 Parseval 傅里叶等距变换，将 u_k^{n+1} 转换至频域，得到各模态的频域更新；同时，将中心频率的取值也转换至频域，得到中心频率的更新方法。频域内模态函数、中心频率和 λ 的更新表达式为：

$$\hat{u}_k^{n+1}(\omega)=\frac{\hat{f}(\omega)-\sum_{i\neq k}\hat{u}_i(\omega)+\frac{\hat{\lambda}(\omega)}{2}}{1+2\alpha(\omega-\omega_k)^2} \tag{2.32}$$

$$\hat{\omega}_k^{n+1}=\frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{2.33}$$

$$\hat{\lambda}^{n+1}(\omega)\leftarrow\hat{\lambda}^n(\omega)+\tau(\hat{f}(\omega)-\sum_k\hat{u}_k^{n+1}(\omega)) \tag{2.34}$$

对于判别精度 ξ>0，若上述更新满足判断表达式，则终止迭代。其判别表达式为：

$$\sum_k\frac{\left\|\hat{u}_k^{n+1}-\hat{u}_k^n\right\|_2^2}{\left\|\hat{u}_k^n\right\|_2^2}<\xi \tag{2.35}$$

最终，VMD 算法通过傅里叶逆变换将频域转化为时域，即能得到待分析信号 f 自适应地分解为相应 K 个窄带的模态分量 IMF $u_k(k\in 1,2,\cdots K)$。

2.3 语音增强算法的性能评估标准

通过评估增强后语音的质量可衡量语音增强算法的有效性和可靠性。语音质量的评估方法主要分为主观评价方法和客观评价方法两种。主观评价指的是将一段语音播放给听众，让听众从语音的舒适度、可懂度以及去噪效果等方面进行评价打分。客观评价测度是通过数学计算的方式从语音的可懂度、感知度以及信噪比等方面进行数字上的计算打分。通过对增强后语音进行主观评价和客观评价，可以从评价结果中得出增强算法的性能并针对相应的问题对算法进行改进和优化。

2.3.1 主观评价标准

主观评价标准主要基于 ITU-TP.835 标准，其中使用最为广泛的主观评价方法为分级判断法。打分标准为 5 分制，即试听者可打分的范围为 1-5 分，其中 1 分为“不满意”，5 分为“非常好”。IEEE 主观方法分技术委员会（Subcommittee on Subjective Methods）规定，将所有试听者的评分进行平均从而得到平均意见得分(Mean Opinion Score, MOS)，MOS 分值所代表的语音质量如表 2.1 所示：

表 2.1 MOS 评分量表

评分	语音质量	失真度
5	非常好	不可察觉
4	好	略可察觉，但不烦人
3	一般	可察觉，轻度烦人
2	差	烦人，但尚可接受
1	很差	很烦人并且难以接受

2.3.2 客观评价标准

语音质量客观评价标准非常多，其中最具有影响力的两种评价标准如下所示：

（1）分段信噪比

分段信噪比(Segment Signal-to-Noise Ratio, segSNR)可以在时域或频域进行计算。语音增强以及语音编码算法的评估可利用时域测量得到。将原始信号和增强后的信号在时间上调整一致，同时将出现的相位差及时进行改正，才能使得在这种测量下有效。分段性噪比的定义如下：

$$serSNR=\frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Nm}^{Nm+N-1}x^2(n)}{\sum_{n=Nm}^{Nm+N-1}(x(n)-\hat{x}(n))^2} \tag{2.36}$$

式中： M 是帧数； N 为帧长（通常为 15~20ms）； x(n) 为原始的语音信号； $\hat{x}(n)$ 为增强后的语音信号。

（2）语音质量感知评价标准

语音质量的感知(Perceptural Evaluation of Speech Quality, PESQ)评价标准是国际电信联盟推荐的主流算法。将非平均扰动值 d_{asym} 和平均扰动值 d_{sym} 进行线性组合即可得到 PESQ 的值：

$$PESQ=a_0+a_1d_{sym}+a_2d_{asym} \tag{2.37}$$

式（2.37）中：$a_0 = 4.5$ ；$a_1 = 0.1$；$a_2 = 0.0309$ 。

利用公式（2.37）计算得到的值在 [-0.5,4.5] 区间，语音感知度的效果与数值成正比。PESQ 可用于可靠地预测在传输信道错误、包丢失或信号延迟变化的情况下编解码器（波形 CELP 型编解码器）的主观语音质量。

3. 语音增强效果及评估

3.1 数据来源

为了验证第 2 章算法的可行性及语音增强的实际效果，本次仿真实验平台为 MATLAB 2021b，实验所用到的语音数据均来自于知名的开源数据库[1]。

3.2 仿真效果及分析

为了验证第 2 章算法的可行性，本小节选取了数据库干净语音中 sp02.wav 作为纯净语音，选取了数据库噪声数据中 white 数据作为白噪声语音，两者相加得到含噪语音。本小节通过用谱减法、维纳滤波法、子空间法、小波变换和变分模态分解等算法对含噪信号进行处理，计算增强后语音的信噪比，可懂度，做出语谱图，说明算法的可行性。

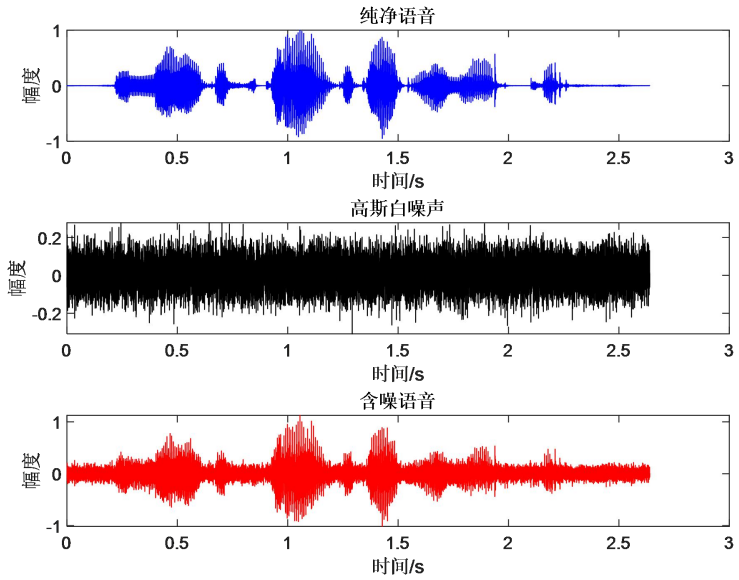

图 3.1 sp02_5db.wav 波形图

[1] http://spib.linse.ufsc.br/noise.html

数据库中噪声采样率为 19.98k，纯净语音数据的采样率为 8k ，首先对噪声数据进行下采样，再加到纯净语音数据 sp02.wav 上，最后做出纯净语音、高斯白噪声（一段）和含噪语音的波形图见图（3.1）。

3.2.1 谱减法

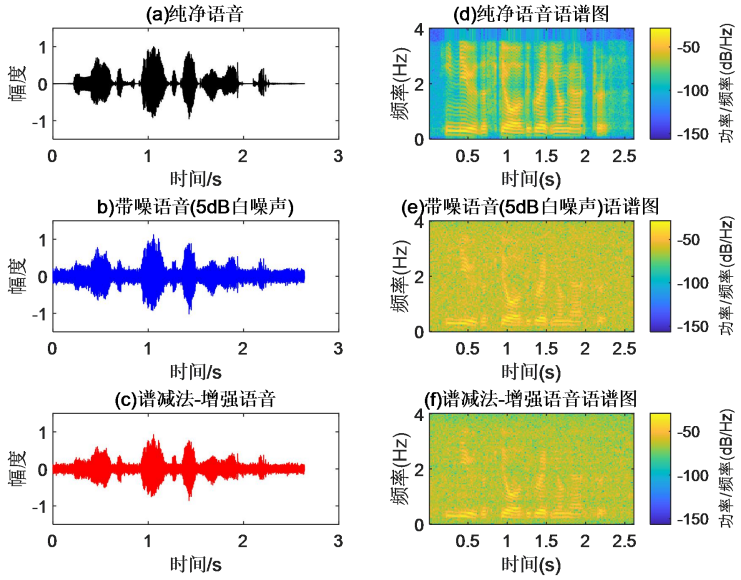

图 3.2 谱减法增强效果图

图 3.2 为信噪比为 5db 的含噪语音经谱减法增强的效果图，从图 3.2 中可以看出，原始干净语音的波形及语谱图都很清楚地显示了语音的特征；中间的图为带噪语音的波形和语谱图，在加噪的语音信号有明显的被污染，而且语谱图中也有加大的粗糙点；最下面的图为经过幅度谱减法增强后的语音信号波形图和语谱图，可以看出噪声被削减一部分，从语谱图中也可以看出原本带噪语音语谱图中部分粗糙点被消减了。增强后语音信噪比 SNR 为 7.3259 db，分段信噪比 segSNR 为 2.7606，可懂度 PESQ 为 1.9626。

多带宽谱减法（Multi-Band Spectral Subtraction, MBSS）是一种对谱减法的改进方法，可以略微提升增强后语音的客观评价指标，详见表 3.1。图 3.3 为信噪比 5db 的含噪语音信号经 MBSS 增强的效果，相比于图 3.2，MBSS 方法增强后的语音信号的语谱图较为干净，去除了含噪语音信号语谱图的粗糙点，但也损失了部分纯净语音信号语谱图的细节部分。MBSS 方法增强语音信号的分段信噪比 segSNR 为 7.3736 db，语音质量感知度 PESQ 为 2.1568，信噪 SN 为 7.6611 db。

表 3.1 谱减法与多带宽谱减法客观指数比较

	SNR/db	segSNR/db	PESQ
SS	7.3572	7.4154	1.9626
MBSS	7.6611	7.3736	2.1568

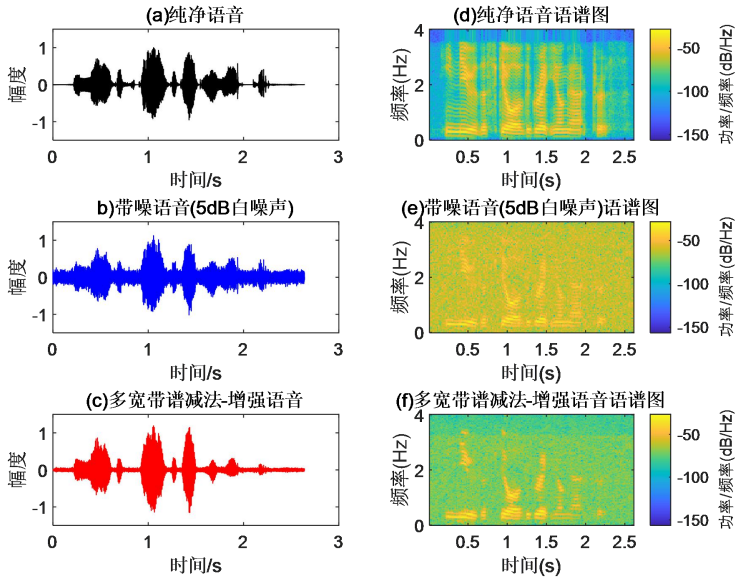

图 3.3 多带宽谱减法增强效果图

3.2.2 维纳滤波法

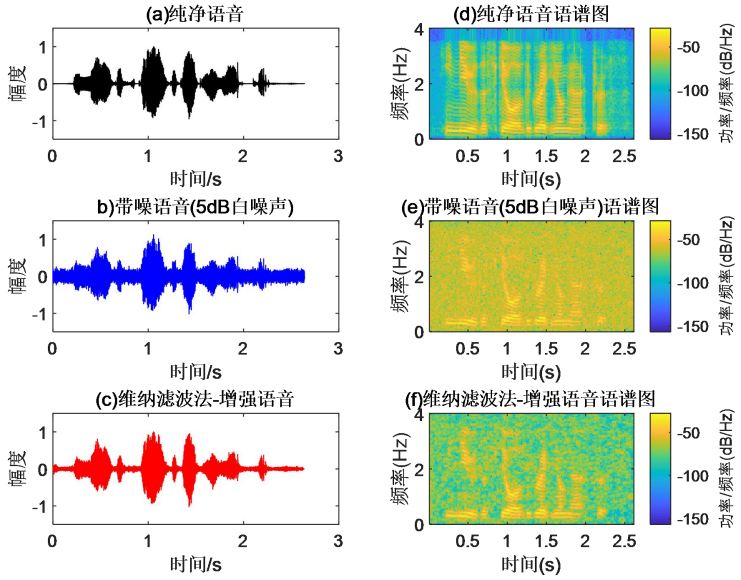

图 3.4 维纳滤波法增强效果图

图 3.4 为信噪比为 5db 的含噪语音经维纳滤波法增强的效果图，从图 3.4 中可以看出，经过维纳滤波增强后的语音信号质量有了一定的改善，但是从图中也可以看出，增强语音信号中依然存在一些噪声，这与维纳滤波器的滤波特性相关。维纳滤波法增强语音信号的分段信噪比 segSNR 为 8.1368 db，语音质量感知度 PESQ 为 2.3146，信噪比 SNR 为 10.2784 db。

3.2.3 子空间法

图 3.5 为信噪比为 5db 的含噪语音经子空间法增强的效果图，从图 3.5 中可以看出，经过子空间法增强后的语音信号质量有了一定的改善，但是从图中也可以看出，增强语音信号相较于纯净语音信号缺失了部分细节。子空间法增强语音信号的分段信噪比 segSNR 为 6.5556 db，语音质量感知度 PESQ 为 2.3612，信噪比 SNR 为 10.6990 db。

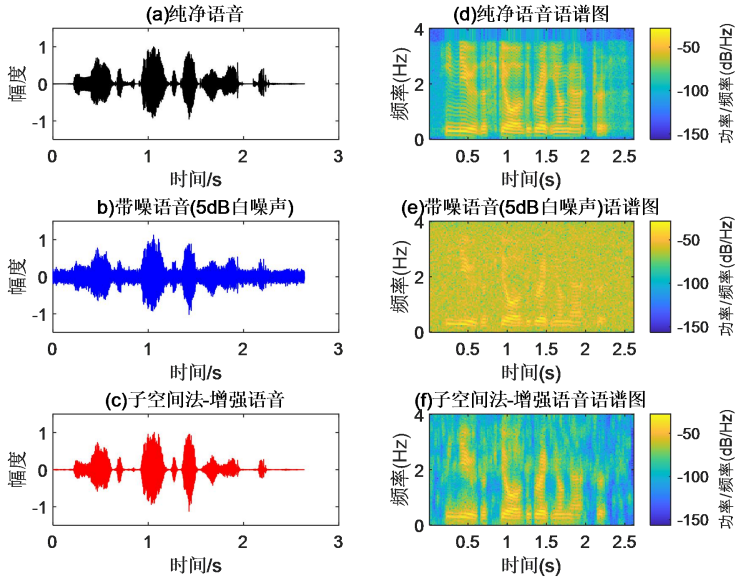

图 3.5 子空间法增强效果图

3.2.4 小波变换

图 3.6 为信噪比为 5db 的含噪语音经小波分析增强的效果图，从图 3.6 中可以看出，经过小波分析增强后的语音信号质量有了一定的改善，但是从图中也可以看出，增强语音信号相较于纯净语音信号缺失了部分高频细节，保留了部分低频细节。小波分析法增强语音信号的分段信噪比 segSNR 为 4.7275 db，语音

质量感知度 PESQ 为 1.8111，信噪比 SNR 为 7.1028 db。

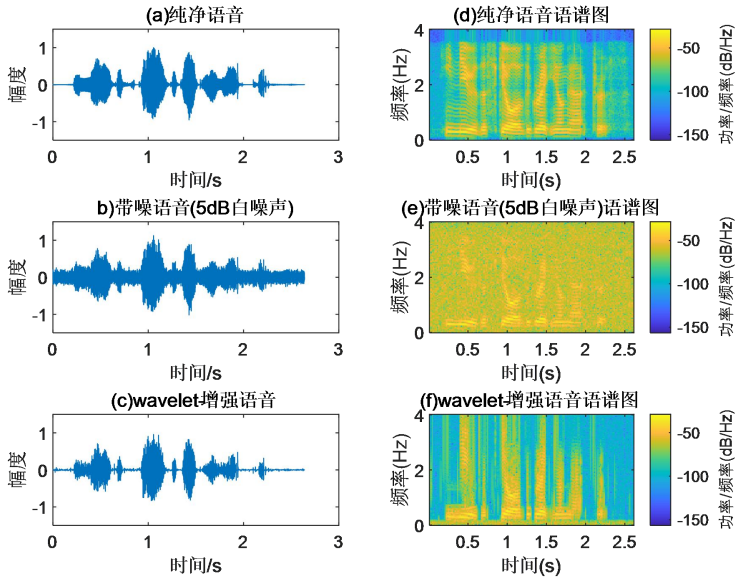

图 3.6 小波滤波法增强效果图

3.2.5 变分模态算法

变分模态分解算法是一种基于频域自适应划分的变分框架算法，受两个超参数模态数 K 、带宽 alpha 影响最大；根据相关文献，本节采用了小波多辨分析的思想优化了 K 值的选取，这人为选取 K 值增强效果有了提高；但小波阈值的选取会间接影响 K 的值的选取，从而影响变分模态分解算法的语音增强效果。作者在此次做了多次尝试（参考了一些文献），均未取得较好（令人满意）的效果；但许多文献结果表明变分模态分解算法语音信号增强效果应当是非常令人满意的。

图 3.7 为信噪比为 5db 的含噪语音经变分模态分解算法增强的效果图，从图 3.7 中可以看出，经过变分模态分解算法增强后的语音信号质量有了一定的改善，但是从图中也可以看出，增强语音信号相较于纯净语音信号保留了较多噪声细节。变分模态分解算法增强语音信号的分段信噪比 segSNR 为 7.2623 db，语音质量感知度 PESQ 为 1.9203，信噪比 SNR 为 6.7627 db。

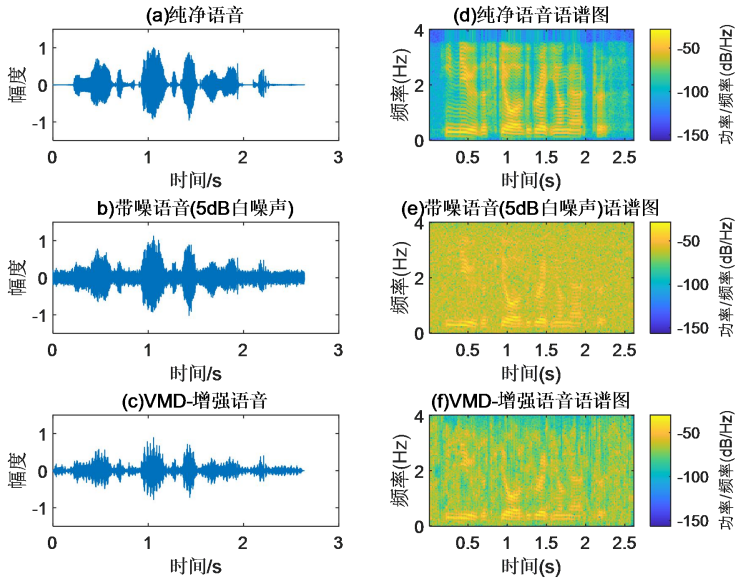

图 3.7 变分模态分解法增强效果图

3.3 客观分析

为了评价语音增强算法的有效性和可靠性，本小节选取了数据库中八种典型场景下不同输入信噪比的 720 段（8*3*30）含噪语音信号进行了实验。这八种场景分别是 airport、babble、car、exhibition、restaurant、station、street 和 train，输入信噪比分别为 0db、5db 和 10db（有一个场景 15db 含噪语音信号下载不了，故而舍去），每种场景每种固定信噪比都有 30 段语音信号。在不同场景不同信噪比的情况下，对上述 6 种增强算法（其中有部分改进过的算法，详见参考文献或程序，由于篇幅问题不在此介绍）增强后的语音信号进行客观的质量评估。利用 2.3 节介绍的客观评价方法，分别计算输出信号的分段信噪比(segSNR)和语音质量感知度(PESQ)，结果见表 3.2 和表 3.3

表 3.2 六种语音增强算法在不同的输入信噪比下输出信号的 segSNR

输入信噪比（db）		0	5	10
airport	MBSS	5.3003	7.0571	9.1652
	wiener	5.3661	7.2095	9.7147
	subspace	3.9478	5.8226	8.4237
	wavelet	3.9903	5.5593	7.7657
	VMD	3.3679	4.7692	6.1557
	Log-MMSE	5.3652	6.9392	9.2905

babble	MBSS	5.3011	7.1478	9.0516
	wiener	4.9394	6.9432	9.4189
	subspace	3.5505	5.6325	8.1537
	wavelet	4.0603	5.6379	7.7051
	VMD	3.7866	5.1014	6.3963
	Log-MMSE	5.0198	6.8032	9.0950
car	MBSS	4.9401	6.8018	8.8341
	wiener	5.4008	7.4111	9.4877
	subspace	3.8118	5.9713	7.9626
	wavelet	3.3730	4.7257	6.7084
	VMD	3.8968	5.2757	6.6087
	Log-MMSE	5.5770	7.3495	9.1705
exhibition	MBSS	5.5264	7.3949	9.2027
	wiener	5.5978	7.5981	9.8227
	subspace	4.3119	6.6113	8.6233
	wavelet	3.9418	5.3298	7.3252
	VMD	4.4751	5.9088	7.3259
	Log-MMSE	5.9699	7.5949	9.4898
restaurant	MBSS	5.4002	7.2760	9.2653
	wiener	5.1502	7.3607	9.6953
	subspace	3.5750	6.1049	8.4493
	wavelet	4.3711	5.9024	8.1708
	VMD	4.0331	5.3286	6.8095
	Log-MMSE	5.2187	7.2600	9.3367
station	MBSS	5.0379	6.8201	8.8713
	wiener	5.3663	7.4664	9.6739
	subspace	3.8437	5.9449	8.2540
	wavelet	3.6436	5.1927	7.1655
	VMD	3.5669	4.9192	6.0764
	Log-MMSE	5.4216	7.2380	9.3269
street	MBSS	5.6045	7.1494	9.2241
	wiener	5.1980	7.0898	9.6156
	subspace	3.5674	5.6670	8.2639
	wavelet	3.9399	5.5339	7.4692
	VMD	4.4713	5.8313	7.1426
	Log-MMSE	5.5294	7.1165	9.4228
train	MBSS	5.6826	7.5175	9.4613
	wiener	5.4893	7.4216	9.7721
	subspace	3.9855	6.0235	8.4815
	wavelet	3.5599	4.7393	6.5138
	VMD	4.9064	6.2788	7.5773
	Log-MMSE	5.7308	7.5215	9.6642

表 3.3 六种语音增强算法在不同的输入信噪比下输出信号的 PESQ

输入信噪比（db）		0	5	10
airport	MBSS	1.8863	2.2431	2.5714
	wiener	1.8261	2.1852	2.5207
	subspace	1.6141	2.0479	2.3869
	wavelet	1.7113	2.0426	2.3385
	VMD	1.6526	1.9155	2.1709
	Log-MMSE	1.8491	2.2268	2.5607
babble	MBSS	1.9157	2.2476	2.5620
	wiener	1.8125	2.1480	2.5020
	subspace	1.5803	1.9761	2.3537
	wavelet	1.7742	2.0489	2.3344
	VMD	1.6509	1.9258	2.1850
	Log-MMSE	1.8711	2.2055	2.5582
car	MBSS	1.8494	2.1760	2.5842
	wiener	1.9131	2.2336	2.6000
	subspace	1.7087	2.1382	2.4712
	wavelet	1.7126	1.9599	2.2573
	VMD	1.6246	1.8601	2.1386
	Log-MMSE	1.9899	2.3293	2.7029
exhibition	MBSS	1.7129	2.1569	2.5296
	wiener	1.7292	2.1154	2.4828
	subspace	1.5452	2.0557	2.3919
	wavelet	1.5524	1.9552	2.2809
	VMD	1.4807	1.8267	2.1139
	Log-MMSE	1.7516	2.1637	2.5537
restaurant	MBSS	1.8523	2.1727	2.5456
	wiener	1.7927	2.1195	2.4954
	subspace	1.5621	1.9779	2.3733
	wavelet	1.7949	2.0456	2.3695
	VMD	1.6771	1.9154	2.2218
	Log-MMSE	1.8514	2.1576	2.5401
station	MBSS	1.8393	2.2156	2.5777
	wiener	1.8499	2.2595	2.5689
	subspace	1.6424	2.1336	2.4577
	wavelet	1.7054	2.0276	2.3104
	VMD	1.5821	1.9108	2.1364
	Log-MMSE	1.8976	2.3348	2.6456
street	MBSS	1.8378	2.1951	2.5459
	wiener	1.8007	2.1454	2.5059
	subspace	1.4692	1.9370	2.3413
	wavelet	1.5998	1.9786	2.2935
	VMD	1.5805	1.8636	2.1641

	Log-MMSE	1.8573	2.2108	2.5495
train	MBSS	1.8268	2.1605	2.5179
	wiener	1.7905	2.1209	2.4596
	subspace	1.4060	1.9315	2.3423
	wavelet	1.3456	1.7240	2.1610
	VMD	1.5395	1.8155	2.0917
	Log-MMSE	1.8800	2.2201	2.5362

从表 3-2、3-3 可以看出多带宽谱减法（MBSS）、维纳滤波法（Wiener）、对数最小均方误差（Log-MMSE）方法能够在去噪效果和语音感知度保持着较好的表现。在低信噪比（0db）情况下，小波变换（wavelet）、变分模态分解（VMD）算法语音增强效果优于子空间法（subspace）；在高信噪比（10db）情况下，小波变换（wavelet）、变分模态分解（VMD）算法语音增强效果差于子空间法（subspace）。除此之外，VMD 算法运算量较大，但性能可提升空间大，本节未对 VMD 优化算法多加探讨。据最新文献结论表明，VMD 算法在极低信噪比情况下语音增强效果令人满意，能够从强干扰噪声中恢复出微弱的有用信号。

去噪效果：在八种不同的场景下，随着输入信噪比的不断增大，六种方法的去噪效果都有不同程度的增强；但是从整体来看 wiener 方法表现得最佳，其中主要表现在以下几个方面：在 car、exhibition、station 等场景下，小波变换、VMD 算法增强后的语音信号分段信噪比 segSNR 较低，而 wiener 方法增强后的语音信号 segSNR 较高，比 MBSS 方法增强后语音信号的 segSNR 高出零点几个 db；在不同场景 0db 情景下，wiener 方法增强后语音信号的分段信噪比 segSNR 都能够保持在 5dB 以上，较为稳定。

（2）语音感知度：在 exhibition 场景下，六种方法所获得增强语音信号都表现出了较差的语音感知度，但是总体来看 wiener 方法所得到的语音信号的感知度最强。在不同的场景下，wiener 方法所得到的语音信号的感知度略好 MBSS 和 Log-MMSE 方法。

4. 结论

本次课程论文对几种常见的单通道语音信噪分离传统方法进行了探讨；在 3.2 节模拟仿真实验中，噪声模拟仿真数据用的是高斯白噪声，其频谱平稳，是一种理想情况。在高斯白噪声情况下，上述六种语音增强算法都能一定程度地抑

制住背景噪声，改善语音质量，但其语音增强效果有所不同。

通过观察不同语音增强算法处理后信号的时域波形图和时频域语谱图，可以更全面地了解算法的局限性。虽然子空间方法增强后语音信号的输出信噪比非常的高，但并不意味着其增强效果最好，在语音客观评价指标中常用分段信噪比 segSNR 代替输出信噪比；因为分段信噪比 segSNR 更加能够表现语音的局部情况，而输出信噪比表现的是语音的全局情况。

由于现实场景中存在着各类噪声，比如粉红噪声、工厂噪声、餐厅噪声，不同噪声的特性有所不同。为了使实验更具有普适性，本文分别用文中六种方法在不同场景不同信噪比对含噪语音信号进行了处理，实验得出：在一定信噪比情况下， wiener 方法语音增强效果最佳，MBSS 和 Log-MMSE 方法及次。

待改进之处：

（1）VMD 算法自适应选取参数 K 和 alpha 问题；由于语音信号频带较宽（>8k），VMD 算法是一种基于频域自适应分解的方法，分解层数太多可能会增大重构误差；许多文献将 VMD 与小波变换结合起来，有着不同的阈值选取方式，但未解决根本问题。

（2）在复杂的声学场景下平衡去噪效果和语音质量；语音增强后的语音不仅需要能够让机器识别的准确性提高，还需满足人耳的听觉特性，所以一味的追求去噪效果指标是不可取的。因此在满足去噪的前提下还需要考虑语音质量问题。

5. 致谢

首先很感谢耿老师“费曼教学法”，通过线下课堂讲解课程内容和布置相关思考题相结合方式，能够引发学生课后带着问题思考，而且耿老师能够实时回复学生的邮件，真的很 nice！

在前一周组会上，导师问及研究生与本科生之前最大的不同在哪里？我想应该是角色的变换，本科生之前是以老师为主体，研究生需以自身为主体；（大部分）本科生思维固在解题，而研究生应发展探索思维。通过这个学期耿老师布置的思考题，我想我已经完成了这种思维上的转变。对于一个给定的问题，文献调研工作是最为重要的，然后由里向外逐个击破。从最初“矩阵求次方”问题到最

后“聚类算法”课堂报告，我也能够很明显地感觉到自身能力的提升。

对于读书笔记，从确定主题为语音信噪分离开始，最初的想法是从压缩感知（字典学习）入手，做了好几天发现效果不太理想且这方面资源比较少，KSVD 在语音信号（一维）上的应用又需要一个比较大的数据集去构建字典库，一个月内可能完不成此般工作量；再加之现今北京疫情日益严峻，国科大要求提前结束春季学期，使得课程教学开启了双倍模式。通过后续相关文献的阅读，我发现目前比较新的语音增强算法（除去深度学习）有基于时频分析的经验模态分解（EMD）、VMD 等，于是把目光转向了 VMD 算法，同时也学习了小波变换（基变换、内积、小波子空间）；不过花费了大量的时间也没有达到文献所谈令人满意的效果。除此之外，还对谱减法、最小均方误差方法进行了改进，均取得了较好的效果。

同时也感谢自己近半个月的努力、坚持，最终才能完成一份令自己较为满意的读书笔记。这个学期给我最大的感悟是：纸上得来终觉浅，绝知此事要躬行；在此与君共勉。

最后希望能够早日见到耿老师的新书！祝老师家庭美满，身体健康，万事如意，科研顺利！

6. 参考文献

[1] 吉慧芳,贾海蓉,王雁.改进相位谱补偿的语音增强方法[J].计算机工程与应用,2019,55(08):48-52.

[2] Thimmaraja G. Yadava, H. S. Jayanna, Speech enhancement by combining spectral subtraction and minimum mean square error- spectrum power estimator based on zero crossing[J]. International Journal of Speech Technology, 2019, 22(3): 639-648.

[3] Yadava T G, Jayanna H S. Speech enhancement by combining spectral subtraction and minimum mean square error-spectrum power estimator based on zero crossing[J]. International Journal of Speech Technology, 2019, 22(3): 639-648.

[4] Yan L, Addabbo P, Zhang Y, et al. A sparse learning approach to the detection of

multiple noise-like jammers[J]. IEEE Transactions on Aerospace and Electronic Systems, 2020, 56(6): 4367-4383.

[5] Dragomiretskiy K, Zosso D. Variational mode decomposition[J]. IEEE transactions on signal processing, 2013, 62(3): 531-544.

[6] 郭欣. 基于 K-SVD 稀疏表示的语音增强算法研究[D].太原理工大学,2016.

[7] 谢文华. Spearman 相关系数的变量筛选方法[D].北京工业大学,2015.

[8] 路敬祎,马雯萍,叶东,姜春雷.基于 VMD 的声音信号增强算法研究[J].机械工程学报,2018,54(10):10-15.

[9] 陶帅. 基于变分模态分解的语音增强算法的研究[D].哈尔滨理工大学,2021.

[10] 陆振宇,卢亚敏,夏志巍,黄现云.基于变分模态分解和小波分析的语音信号去噪方法[J].现代电子技术,2018,41(13):47-51.

[11] 贾海蓉,张雪英,牛晓薇.用小波包改进子空间的语音增强方法[J].太原理工大学学报,2011,42(02):117-120.

[12] 李定文. 优化的变分模态分解算法在信号去噪中的应用[D].东北石油大学,2021.