

---

北京化工大学

信息科学与技术学院

电子信息综合课程设计报告

专    业：通信工程

所在班级：通信 2201

学生姓名：范俊鹏

学生学号：2022040343

导师姓名：万洪杰

# 课程设计报告

## 一、题目背景和意义

人类的交流主要通过语言，在现代通信技术中，语音是最重要以及最基本的通信媒介。由于语音的形成过程和发音器官的运动密切相关，这种物理运动比起声音振动的速度要缓慢得多，因此语音信号在 10-30ms 的短时间间隔内可假定为平稳过程。随着现代科技的发展，通信过程中的信息量呈爆炸式增长，但通信所使用的带宽资源并不是无限的，这驱动了低速率语音压缩编码技术的产生。在诸如卫星通信和保密通信等传输带宽有限的场景下，低码率语音编码技术至今仍然是现代通信系统的重要组成部分。<sup>[1]</sup>

线性预测编码（LPC）是基于分析和模拟人的发音器官而建立的描述模型，是进行语音信号分析最有效的分析方法之一。其基本思想是将语音信号用一组过去时刻语音采样的线性组合来逼近，并根据实际采样与线性预测之差的平方和最小的原则来决定预测参数集。LPC 技术提供了一组简洁的语音模型参数，能够较精确地表征语音信号的频谱幅度，通过利用语音信号的冗余性和人耳的听觉掩蔽效应，该技术可以在保持一定语音质量的同时，实现极低比特率的语音传输与高效的数据压缩。这种参数编码方式为语音信号处理开辟了新道路，是当今许多高级编码算法的基础。<sup>[2]</sup>

本课程设计的目标是手动实现从端点检测到语音重合成的完整链路，验证全极点模型在语音建模中的有效性。通过构建一个完整的语音编解码系统，研究内容包括利用双门限法判别有声段、分析清浊音帧的自相关特性、基于中心削波法进行基音周期检测，以及采用 Durbin 算法求解 AR 系数并转换为反射系数。最终结合矢量量化码本与发声模型，旨在低码率条件下实现语音的重构，从而深入理解语音信号处理的底层逻辑与物理特性。

## 二、国内外研究现状

语音编码技术经历了从传统信号处理方法向深度学习方法的全面转变。在早期研究中，脉冲编码调制(PCM)参数编码技术，如 LPC-10 算法与码激励线性预测(CELP)，通过模拟人类发音机制提取声学特征，在极低比特率下实现了可懂语音的传输，成为现代通信的重要基石。为了在有限码率下更有效地压缩特征参数，矢量量化(VQ)技术被引入语音处理领域，其中 LBG 算法通过迭代聚类训练寻找最佳码本，成为特征压缩与数据去冗余的经典手段。<sup>[3]</sup>

随着人工智能技术的迅猛发展，深度神经网络模型在语音处理任务中的应用日益广泛，其性能普遍超越了传统算法。Google 提出的 WaveNet 开启了神经声码器的新纪元，通过堆叠空洞卷积实现了对语音波形的高保真逐点生成，但其极高的计算复杂度限制了在端侧设备的实时部署。为提升运算效率，研究者随后开发了 WaveRNN 等模型，利用循环神经网络（RNN）和矩阵稀疏化技术降低了算力开销。

近年来，将传统语音产生模型与深度学习相结合的混合式编解码构架成为主流研究方向。其中，LPCNet 模型具有显著的代表性，它将传统线性预测技术与递归神经网络相结合，将频谱包络建模的负担从神经合成网络中移除，使其大部分能力可以用于建模激励信号。这种设计使得 LPCNet 能够以低于 3 GFLOPS 的复杂度实现高质量的宽带语音合成，即便在计算资源有限的手机或嵌入式系统中也能实现实时运行，有效解决了传统参数编码器音质不佳与端到端神经编码器复杂度过高之间的矛盾。<sup>[4]</sup>

### 三、主要内容与待解决的问题

#### 3.1 主要内容

本课题旨在构建一个基于线性预测理论的完整语音编解码系统，实现从原始语音信号分析到参数化重构信号的全链路处理。研究工作严格遵循从底层算法自研到系统集成的逻辑，具体内容涵盖以下核心模块：

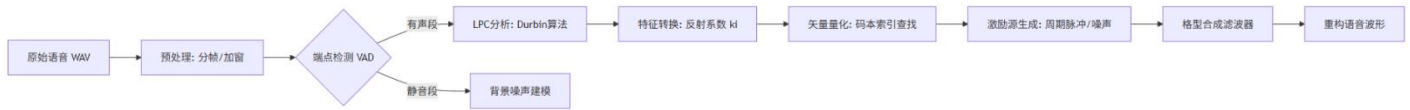
首先是前端分析模块。利用短时能量与短时平均过零率的双门限检测法，对输入的 WAV 文件进行端点定位，准确判别有声段。针对提取出的有声帧，通过中心削波预处理消除声道共振峰的干扰，利用短时自相关函数提取清、浊音帧的时域特征，并计算其基音周期。

其次是线性预测分析模块。基于语音生成的“激励-声道”模型，手动实现 Levinson-Durbin 递推算法，将短时平稳的语音帧建模为全极点滤波器。通过求解线性预测（AR）系数并将其转换为反射系数，实现对声道特性的参数化表征。<sup>[5]</sup>

再次是特征量化模块。为实现高效的低码率传输，通过收集多个语音语料库样本，提取每一帧有声段的反射系数作为特征向量。利用 LBG 聚类算法在特征空间内进行迭代训练，最终生成用于矢量量化的特征码本。

最后是语音重合成模块。根据判别的清浊音标志和基音周期，构建相应的激励源（浊音为周期脉冲串，清音为伪随机白噪声）。结合码本索引还原反射系数，驱动带记

忆的格型合成滤波器重构波形，从而完成完整的编解码闭环过程。



### 3.2 待解决的关键问题

在系统设计与算法实现过程中，需要重点解决以下关键技术问题：

第一，算法的稳定性与鲁棒性。在线性预测分析中，反射系数的模值若接近或超过 1 会导致滤波器不稳定。因此，如何在 Durbin 递推过程中通过合理的截断处理确保格型滤波器的稳定性，是保证合成语音质量的前提。

第二，合成语音的自然度优化。传统 LPC 合成往往存在明显的“电音感”。如何通过优化清浊音判别逻辑、改进激励源的增益平衡以及消除帧间转换时的相位不连续性，是降低重构语音失真、提升音质的关键。

第三，矢量量化效率与失真的平衡。LBG 算法在训练过程中易陷入局部最优。需要研究如何通过合理的分裂因子和迭代门限设置，使生成的码本在极低维度和较小规模下，仍能最大程度减少量化失真，以保证在窄带及超低码率条件下的可懂度。

## 四、设计与实施方案

### 4.1 预处理方案

本设计首先利用 Python 标准库 wave 读取标准 PCM 编码的 WAV 文件，并将量化后的信号归一化至 $[-1, 1]$ 区间，以消除录制设备增益不同的影响。根据语音信号在 10-30ms 内具有短时平稳性的特征，系统将信号划分为 25ms 的长帧，并设置 10ms 的帧移以保证帧间平滑。为减少分帧造成的频谱泄露和吉布斯效应，每帧信号需乘以汉明窗。汉明窗函数的数学表达式为

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1$$

其较宽的主瓣和较低的旁瓣特性有利于后续的自相关及频谱包络分析。

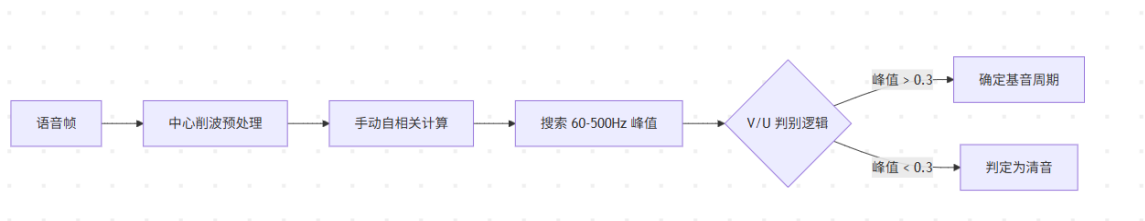
### 4.2 端点检测算法实现

系统采用基于短时能量和短时平均过零率的双门限前端检测法。算法的核心逻辑是通过对语音信号前 10 帧环境背景噪声的统计，自动设定能量高门限(ITU)、能量低

门限(ITL)和过零率门限(IZCT)。检测过程分为三个阶段：首先利用高能量门限 ITU 初步锁定有声段的核心区域；随后利用低能量门限 ITL 向核心区两端进行搜索，以包含能量较低的语音起止部分；最后利用 IZCT 门限结合过零率特征，进一步修正起始点和结束点，从而确保高过零率的清音（如摩擦音）不被误判为静音。

#### 4.3 基音检测与清浊音判决逻辑

基音周期估计采用自相关法，并引入中心削波技术作为预处理手段。由于声道共振峰的干扰常导致自相关函数在基音周期位置的峰值不够突出，设计中通过中心削波函数滤除单帧内幅值低于特定门限（通常为最大幅值的 65%左右）的信号成分，仅保留能表征周期性的高幅值脉冲。在寻找自相关函数最大值时，系统将搜索范围限定在对应人类发音物理限制的 60Hz 至 500Hz 频率区间。清浊音(V/U)的判别准则结合了帧能量和归一化自相关峰值：若峰值超过 0.3 且能量高于噪声门限，则判定为浊音，否则判定为清音或静音。<sup>[6][7][8]</sup>



#### 4.4 线性预测（LPC）分析方案

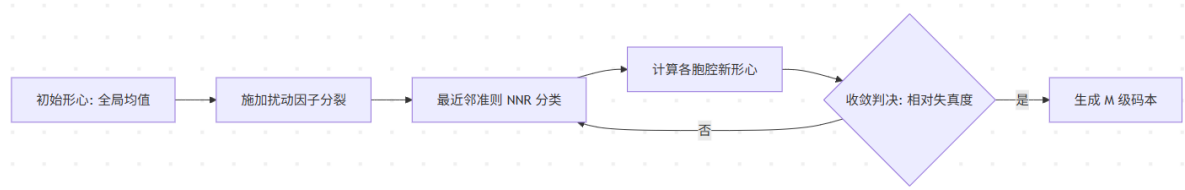
线性预测分析基于语音生成的“源-滤波器”模型。系统通过计算短时加窗信号的自相关函数，并利用 Levinson-Durbin 递推算法求解线性预测方程组。该算法通过递归方式计算每一阶的 AR 系数  $a_i$ ，并利用最小均方误差准则使预测残差能量达到最小。为保证重合成阶段滤波器的稳定性，算法在递推过程中会监控反射系数  $k_i$ ，并引入稳定性截断逻辑，确保其模值严格小于 1。

$$s(n) = - \sum_{i=1}^p a_i s(n-i) + e(n)$$

$$k_i = \frac{R(i) - \sum_{j=1}^{i-1} a_{i-1}(j) R(i-j)}{E_{i-1}}$$

#### 4.5 矢量量化与码本设计方案

为了实现高效的数据压缩，本设计采用矢量量化(VQ)技术对声道参数进行表征。首先将收集到的多组语音样本提取出的 AR 系数转换为反射系数，构建高维特征空间。码本训练采用经典的 LBG 分裂算法：从全体特征向量的全局平均值（形心）开始，通过施加微小扰动因子实现码本规模的倍增分裂（ $1 \rightarrow 2 \rightarrow 4 \rightarrow \dots \rightarrow M$ ）。在每次分裂后，根据最近邻准则（NNR）利用欧氏距离对特征矢量进行聚类划分，并反复计算各胞腔的新形心，直至相对失真度低于预设阈值。最终生成的离散码本将作为编解码器的索引基准。[9][10]



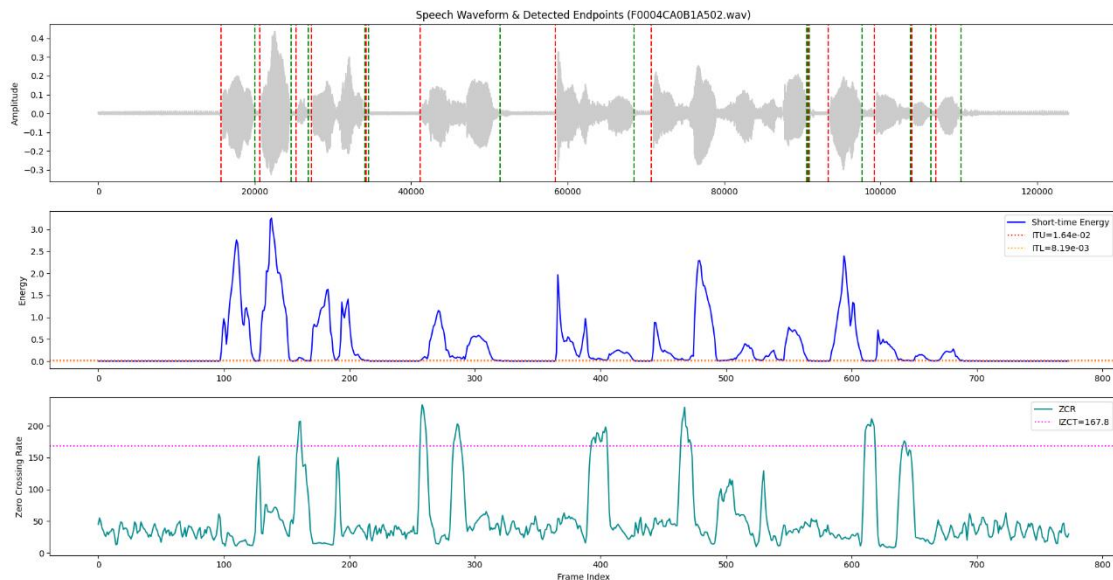
#### 4.6 语音重合成发声模型构建

语音重合成过程遵循线性预测发声模型。解码端根据接收到的清浊音标志、基音周期和码本索引还原反射系数。激励源生成逻辑为：浊音帧采用与基音周期同步的单位脉冲序列，清音帧采用伪随机白噪声。激励信号经增益调整后驱动带记忆的格型合成滤波器。格型结构在物理意义上对应于无损声管模型，相比于直接形式滤波器，它对参数量化误差更不敏感，且通过内部状态量的传递，能有效解决帧间相位不连续导致的“爆破音”问题，使合成语音更接近自然音质。

## 五、实验结果分析

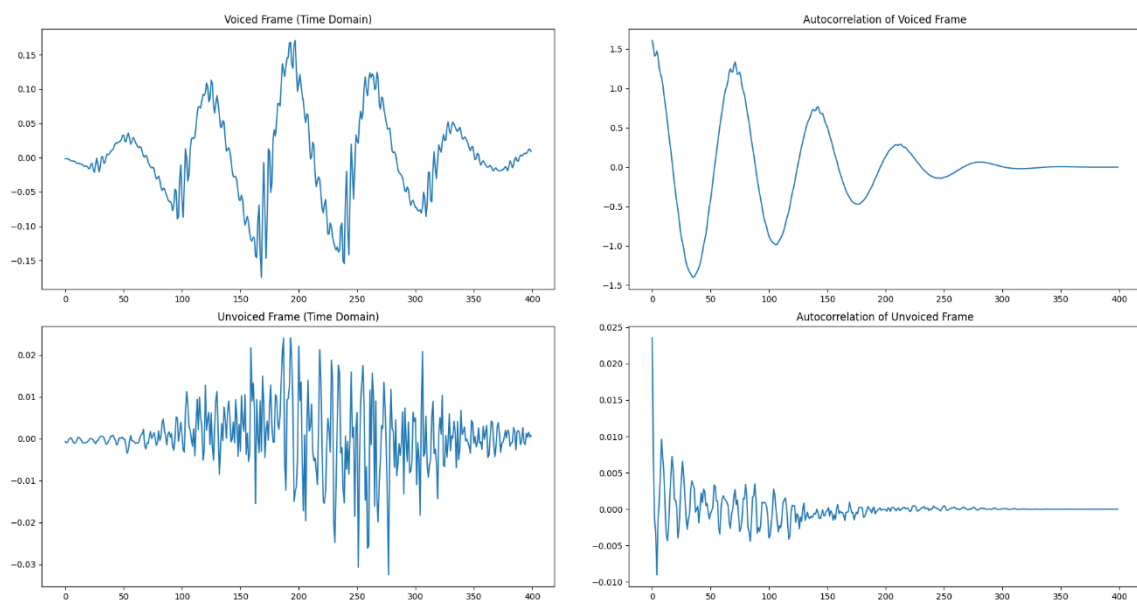
### 5.1 端点检测结果分析

在针对 WAV 语音文件的端点检测实验中，通过手动实现的双门限法成功判别了有声段与静音段。实验结果显示，利用前 10 帧环境背景噪声统计出的能量高门限(ITU)和低门限(ITL)能有效地锁定语音核心区，而过零率门限(IZCT)则成功修正了清音（如摩擦音）的起止点，避免了漏检。在输出的波形标注图中，红虚线与绿虚线准确标注了每个语音段的起始与结束位置，验证了状态机逻辑在处理多段静音（不少于 3 段）复杂语音时的鲁棒性。



## 5.2 自相关函数与基音轨迹分析

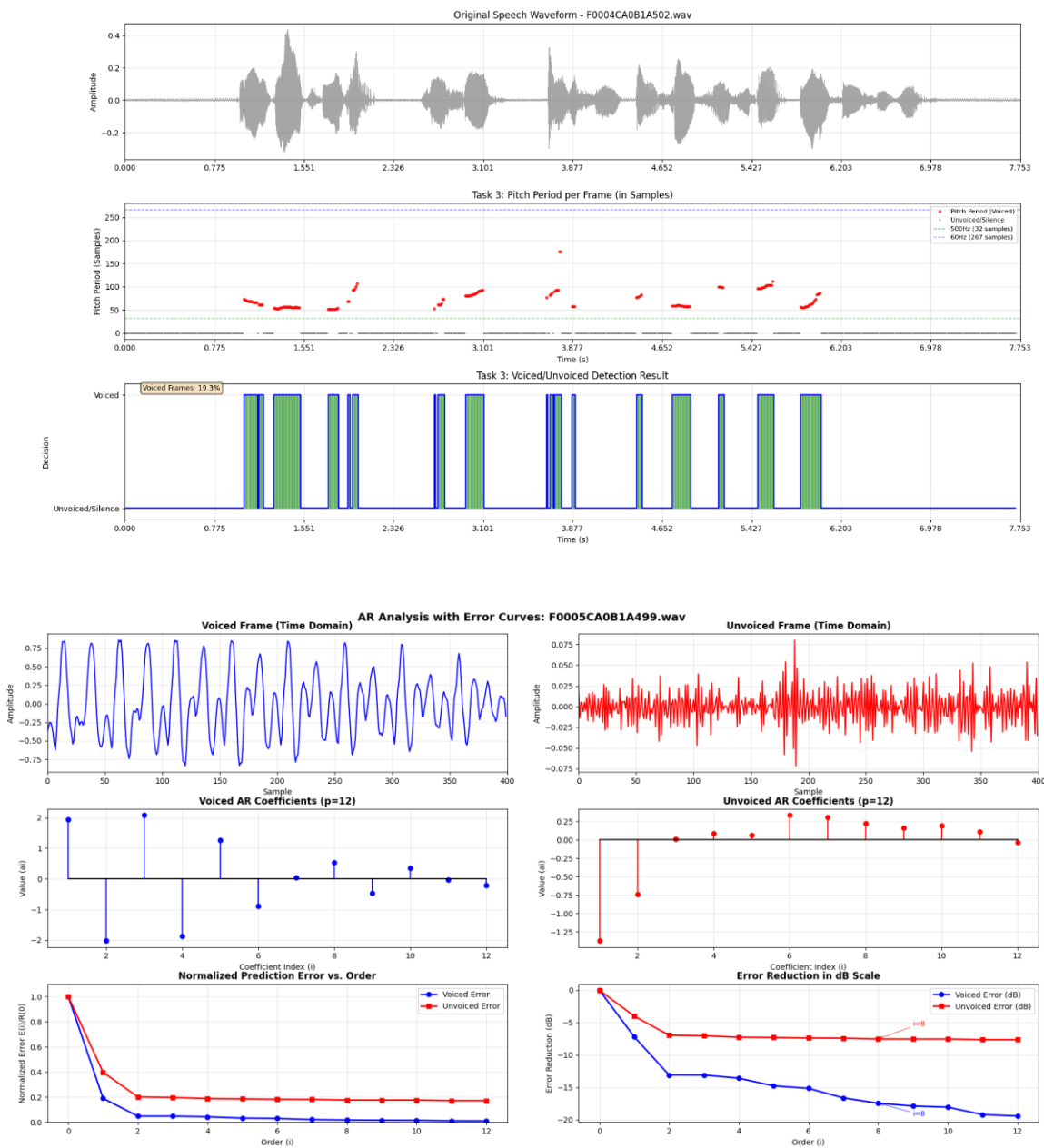
对清/浊音典型帧的自相关分析表明，浊音信号的自相关函数具有显著的周期性峰值，其第一个最大峰值点（除零点外）对应于基音周期的采样点数。相比之下，清音信号的自相关曲线则类似于白噪声，无明显周期性峰值。在引入中心削波预处理后，



有效地去除了声道共振峰的干扰，使基音峰值更加尖锐突出。全语音段的基音轨迹点迹图显示，基音频率稳定在 60Hz 至 500Hz 之间，且清浊音判决结果与原始语谱的能量分布高度吻合。

### 5.3 AR 系数与预测误差分析

线性预测阶数  $p$  的选择对重构质量至关重要。实验分析显示，预测误差能量  $E(i)$  随阶数  $i$  的增加而下降。在  $i$  小于 8 阶时，误差下降幅度极大；当阶数处于 10-12 阶时，下降趋势放缓并趋于稳定。这验证了本设计选择 12 阶 AR 系数的合理性，即在保证较低计算复杂度的同时，已能较精确地提取声道的频谱包络信息。同时，浊音帧的预测误差明显低于清音帧，说明全极点模型对浊音的建模能力更强。

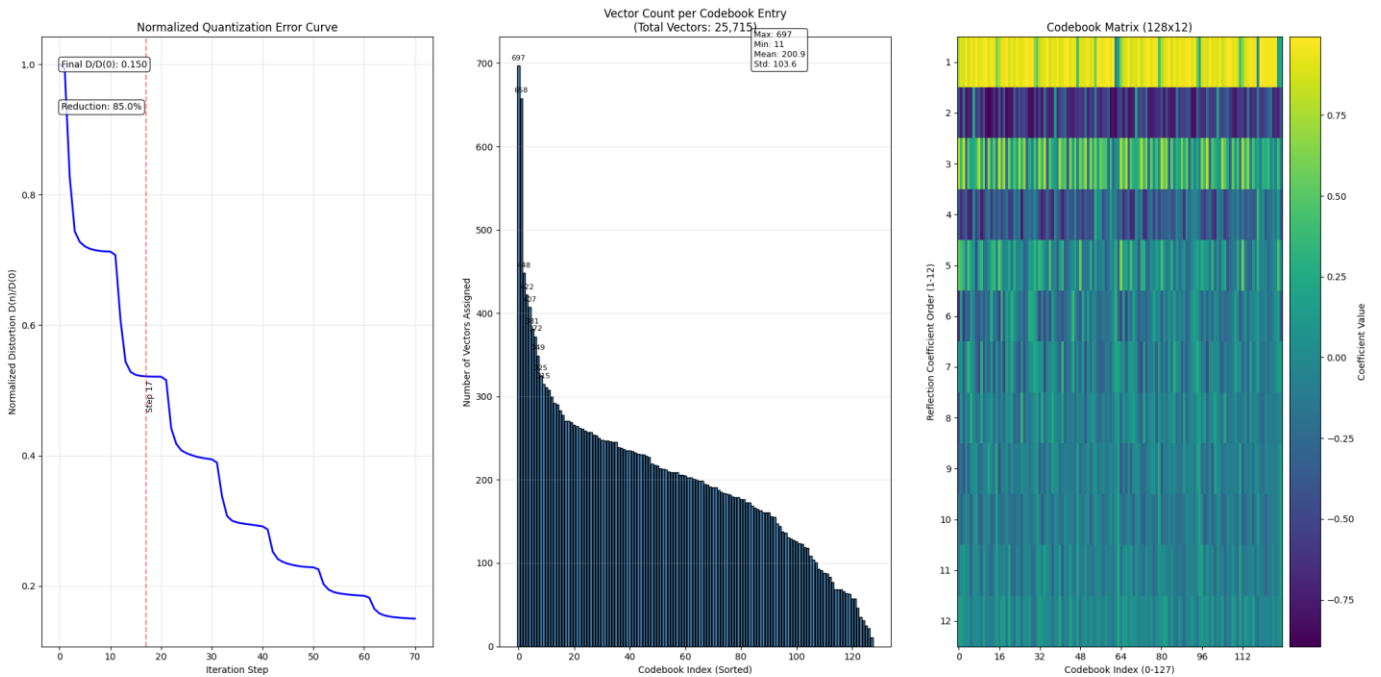




## 5.4 码本分布与矢量量化性能分析

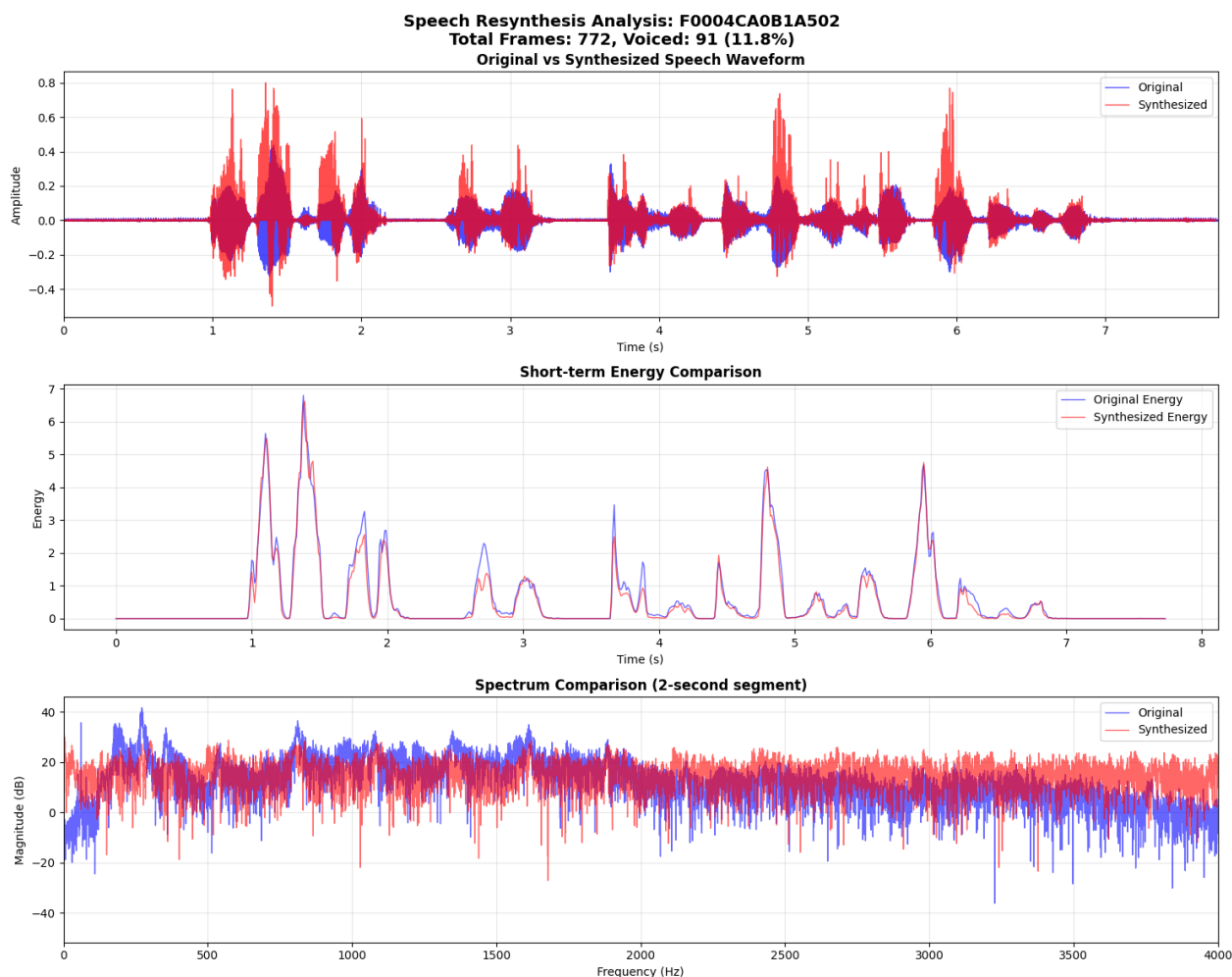
通过 LBG 算法训练生成的反射系数码本展示了良好的聚类特性。在特征空间散点图中，训练得到的 16 个或更多码矢量（形心）均匀分布在反射系数的高维特征空间内。PSNR（峰值信噪比）测试结果表明，随着迭代次数增加，量化失真平稳下降，最终码本能够以极低的码率（如每帧通过索引传输）表征声道物理特征。

LPC-VQ Analysis Results (Total Features: 25,715)



## 5.5 语音重合成音质评价

最终的重合成实验对比了原始波形与合成波形的差异。利用格型合成滤波器重构的语音保留了原始语音的易懂度，通过带记忆的状态传递机制，有效消除了帧间相位不连续产生的爆破音。虽然传统 LPC 合成在听感上仍有一定“电音感”，但结合深度学习的研究（如 LPCNet）表明，将线性预测与神经网络结合可将计算复杂度降低至 3 GFLOPS 以下，并显著提升合成语音的自然度，这为低码率语音编解码器的进一步优化指明了方向。



## 参考文献

- [1] 王晶, 徐亮, 陈晓娇, 等. 基于神经网络的低码率语音编码技术研究综述[J]. 信号处理, 2024, 40(12): 2261–2280. DOI:10.12466/xhc1.2024.12.014.
- [2] 史水平, 李世作. 线性预测编码(LPC)技术及其在音频文件上的应用[J]. 现代电子技术, 2004, 27(4): 21–23. DOI:10.3969/j.issn.1004-373X.2004.04.007.
- [3] J. -M. Valin and J. Skoglund, "LPCNET: Improving Neural Speech Synthesis through Linear Prediction," ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 5891–5895, doi: 10.1109/ICASSP.2019.8682804. keywords: {Complexity theory;Speech synthesis;Sparse matrices;Logic gates;Predictive models;Neural

networks;Cepstrum;neural audio synthesis;parametric  
coding;WaveRNN},

- [4] 王诗瑶. 基于深度学习和线性预测的低速率语音编码算法研究和实现[D]. 陕西:西安电子科技大学, 2022.
- [5] 孙震, 张江鑫. 关于线性预测滤波器阶数的分析研究[J]. 杭州电子科技大学学报, 2010, 30(5):153-156. DOI:10.3969/j.issn.1001-9146.2010.05.039.
- [6] 于俊凤, 曹俊兴. 基于组合参数的清浊音判决方法[J]. 太原理工大学学报, 2004, 35(4):467-469. DOI:10.3969/j.issn.1007-9432.2004.04.025.
- [7] 刘丽娟, 冯宏伟. 基音周期的一种检测方法[J]. 科学技术与工程, 2006, 6(17):2774-2776, 2781. DOI:10.3969/j.issn.1671-1815.2006.17.040.
- [8] 宗源, 曾毓敏, 张梦, 等. 基于趋势分析的 AMDF 基音检测改进算法[J]. 振动与冲击, 2014(20):35-39. DOI:10.13465/j.cnki.jvs.2014.20.008.
- [9] 孙燕. 自适应模糊聚类 LBG 矢量量化算法[J]. 计算机工程与应用, 2014(23):203-205. DOI:10.3778/j.issn.1002-8331.1301-0257.
- [10] 杨裕亮, 杨爽, 匡镜明. 线谱对矢量量化中的码本设计[J]. 北京理工大学学报, 2000, 20(6):725-728. DOI:10.3969/j.issn.1001-0645.2000.06.016.