AI赋能的实时音频信源信道联合编码研究

周 通*① 陈宏智① 许佳龙① 孙 鹏① 姜大洁① 刘健康①
①(维沃软件技术有限公司 北京 100015)

摘 要：目前，第三代合作伙伴计划(3rd Generation Partnership Project, 3GPP)的业务与系统技术规范组第4工作组(TSG SA Working Group 4, SA4) 已经开始研究基于人工智能(Artificial Intelligence, AI)的语音编解码器。本文首次以3GPP地球静止轨道(Geostationary Earth Orbit, GEO)语音典型的物理层参数配置与目前SA4正在研究的Descript 音频编解码器(Descript Audio Codec, DAC)作为评估基线，保证了基线结果的公平性，并首次提出了总功率限制的DAC信源信道调制联合优化方案、恒模限制的DAC信源信道调制联合优化方案和DAC信源信道联合编解码方案。在语音质量可接受的前提下，所提方案相比于基线方案有2.5 dB～4 dB的覆盖增益。希望为GEO语音的后续研究提供有益思路。

关键词：GEO语音；信源信道联合编码；联合信源信道编码调制；AI

中图分类号：TN926.1 **文献标识码**：A **文章编号**：1009-5896(2026)00-0001-10

DOI: 10.11999/JEIT260379 **CSTR**: 32379.14.JEIT260379

1 引言

近年来，信源信道联合编码(Joint Source-channel Coding, JSCC)在学术界和工业界都引起了广泛关注。在工业界，基于信源信道联合编码的信道状态信息(Channel State Information, CSI)的压缩反馈传输[1–4]，已经成为了第三代合作伙伴计划第20版本(3rd Generation Partnership Project Release 20，3GPP R20)的一项研究用例。相比于3GPP 从R18开始研究的基于人工智能(Artificial Intelligence, AI)的CSI压缩反馈，JSCC将在物理层进行CSI压缩与信道编码进行联合优化，从而可以在低信噪比区间获得高的余弦相似度平方(CSI压缩反馈性能的关键指标)。学术界则主要研究信源为视频[5–7] 、图片[8–10]、以及音频[11–14] 的信源信道联合编码方案。主要原理是将视频 、图片、以及音频这些人类可以感知的信源与信道编码以及调制进行联合优化，以提升在低信噪比区间的接收端恢复信号与信源的相似度。

尽管学术界在视频、图像、音频的信源信道联合编码领域已积累了丰富的研究成果，但以3GPP为代表的工业界标准组织迄今并没有推进该技术的全面预研与标准化。当前工业界的相关探索，仅局限于扩展现实(Extended Reality, XR)等特定场景下基于数据重要度指示的差异化传输处理[15,16]。其主要原因包括如下两个方面：

第一，一般来讲，视频和图片的编解码算法不属于3GPP的工作范畴。例如，视频和图片的编解码算法协议由动态图像专家组(Moving Picture Experts Group, MPEG)和 联合图像专家组 (Joint Photographic Experts Group, JPEG)等3GPP之外的组织制定的，而非由3GPP来制定。

第二，信源信道联合编码架构方面面临挑战。为了支持信源信道联合编码，有两种典型的架构。一种架构是在应用层进行信源信道联合编码，另一种架构则是在物理层进行信源信道联合编码[1,17]。其中，在物理层进行信源信道联合编码时，需要应用层采用不破坏语义的加密算法。但由于应用层的信源编码加密算法是3GPP之外的组织如MPEG/JPEG直接定义的，且不能保证不破坏语义。因此采用第一种架构时需要新设计不破坏语义的加密算法，带来额外的工作量。或者不使用加密算法，但这又会带来隐私和安全的问题。而在应用层进行信源信道联合编码，应用层获得的物理层的信道信息往往是过时的，因此限制了信源信道联合编码的增益[1]。

音频与视频、图像的情况有所不同。相较于由其他组织主导的视频与图像编解码领域，3GPP在音频编解码标准上具备更多的主导权。3GPP 曾经自研了AMR(Adaptive Multi-Rate) 语音编码，目前3GPP SA4正在研究的面向手机直连地球静止轨道(Geostationary Earth Orbit, GEO)场景的语音编码，目的是制定GEO接入下 IMS 语音业务超低码率编解码技术规范，开展相关标准化研究工作。SA4 超低码率语音编解码(Ultra Low Bit rate Speech Codecs, ULBC)项目的研究包括针对业界公开的Descript 音频编解码器(Descript Audio Codec, DAC)[18] 在典型的物理层参数配置下进行

收稿日期：2026-xx-xx；改回日期：2026-07-08；网络出版：2026-07-22
*通信作者： 周通 tong.zhou@vivo.com

了丢包补偿(Packet loss concealment, PLC)研究和复杂度分析[19]。目前，SA4 ULBC项目研究的编解码预设1～3 kbps的比特率工作区间, 目的是为了匹配高轨卫星到地面的极大的路径损耗。其中，比特率指单位时间内音视频媒体编码输出的二进制比特数据量，是衡量编码输出速率的技术指标。

此外，传统媒体(例如，视频、图片)在应用层进行信源信道联合编码时，通常面临CSI滞后所导致的跨层架构挑战，而GEO语音场景则能够有效规避这一问题。在GEO语音场景中，得益于轨道同步特性，卫星相对于地面基站基本保持静止。这意味着系统几乎不受多普勒频移的影响且传播路径高度固定，使得信道的相干时间较长，从而表现出信道变化较为缓慢的特性。因此，可以假设物理层的CSI能够被准确传递至应用层，而不会出现信道信息过时的问题。与此同时，由于GEO轨道的链路传播距离长，该场景下语音业务的单向空口传输时延达到 248～280 ms[19]，物理层难以进行快速的自适应调整。基于这些特性，将信道编码机制上移至应用层进行设计和优化是更为合理的选择。此外，考虑到较大的往返时延超过了语音包的时延需求。GEO语音场景不采用混合自动重传请求(Hybrid Automatic Repeat Request, HARQ)机制。

基于上述分析，我们认为有必要研究面向GEO语音场景的信源信道联合编码。其中，信源信道联合编码又可以进一步分为不联合调制的JSCC和联合调制的信源信道编码，即联合信源信道编码调制(Joint Source-channel Coding and Modulation, JSCCM)。

目前，学术界对面向音频的JSCC和JSCCM的仿真研究并未考虑3GPP GEO语音典型的物理层传输配置。本文将结合3GPP SA4正在研究的DAC编码器与 GEO语音典型的物理层参数配置进行面向音频的信源信道联合编解码的研究。本文的主要贡献包括：

(1)首次以3GPP GEO语音典型的物理层参数配置和SA4正在研究的DAC编码器作为评估基线，保证了基线结果的公平性。

(2)首次提出总功率限制的DAC信源信道调制联合优化方案，在语音质量可接受的前提下，所提方案相比于基线方案有4 dB的覆盖增益。

(3)首次提出恒模限制的DAC信源信道调制联合优化方案，在语音质量可接受的前提下，所提方案相比于基线方案有4 dB的覆盖增益。

(4)首次提出DAC信源信道联合编解码方案，在语音质量可接受的前提下，所提方案相比于基线方案有2.5 dB的覆盖增益。

本文第2节将介绍基线方案，即应用层DAC与物理层传输方案。第3节提出总功率限制的DAC信源信道调制联合优化方案，恒模限制的DAC信源信道调制联合优化方案，与DAC信源信道联合编解码方案。第4节介绍仿真评估的结果，最后是结论与下一步计划。

2 应用层DAC与物理层传输方案

本节将介绍DAC编码器在应用层对音频信号的处理流程，并阐述现有的3GPP物理层传输方案。

DAC 编解码器是一种可扩展的编解码器，在比特率方面具有可调性。目前已公开了输入采样率分别为16 kHz、24 kHz 和 44.1 kHz的模型权值[18] 。考虑到GEO场景的覆盖性能难以支持高采样率的语音编码，我们进一步研究了输入采样率为8 kHz的低码率场景。DAC编解码器采用自编码器架构，音频信号先被降采样为潜在表征，然后使用残差矢量量化器(Residual Vector Quantizer ,RVQ)进行量化。图1描述了应用层DAC与物理层传输方案的处理流程。

1. 发送端处理流程: 在发送端(例如上行链路中的用户设备)，处理包含以下三个阶段：

第一阶段：应用层DAC编码。在应用层(例如，IP多媒体子系统(IP Multimedia Subsystem，IMS))将原始语音信号经过DAC编码器(包括下采样和RVQ量化器)处理生成为N比特的实时传输协议负载(Real-time Transport Protocol payload, RTP payload)。该编码阶段具体包括如下3个子过程。

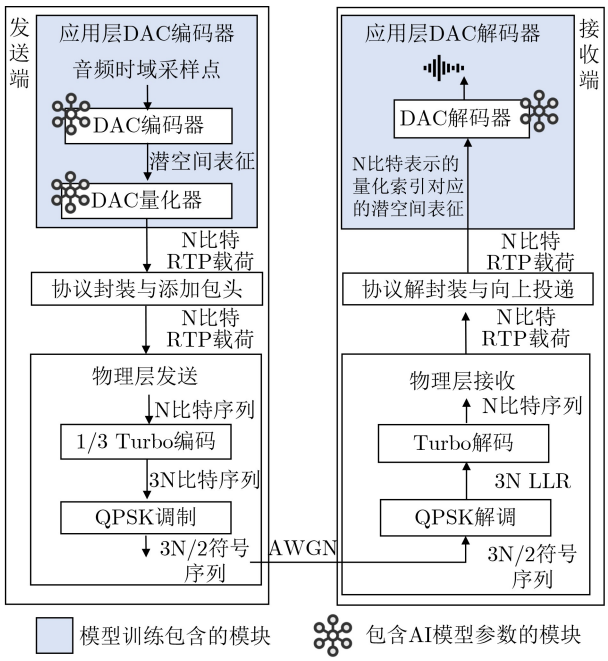

图 1 应用层DAC与物理层传输方案的处理流程

(1)时域采样：将长度为T的音频信号通过采样率s的采样得到$T{\cdot}s$个采样点。假设T=500 ms，s=8 kHz，则共有0.5秒×8 kHz = 4000个采样点，每个采样点的时长为1000 ms÷ 8 k =0.125 ms。

(2)DAC编码器(特征提取与下采样)：将每M个采样点映射为1个语音帧，M可由如下公式计算获得：

$$M = \prod_{k=1}^{N_c} \text{stride}_k \tag{1}$$

其中，N_c是DAC包含的卷积次数，stride_k是第k次卷积的步长，$1 \le k \le N_c$。

当采样率为8 kHz时，可以设置DAC特征提取与下采样模块的N_c=4，步长分别为[2, 4, 4, 8]，则$M = 2\times4\times4\times8 = 256$，每个语音帧的时长为256×0.125 ms = 32 ms。下采样比例可以通过设置不同的卷积步长进行调整。考虑到DAC原始配置是基于最低16 kHz的采样率，而本研究是针对超低码率场景采用的8 kHz采样率，因此，对步长进行了调整。经过特征提取与下采样后可以获得每个语音帧潜空间表征$\boldsymbol{x}_{\text{latent}}^{\text{DownP}}$。

(3) DAC量化器(RVQ量化)：每个语音帧的潜空间表$\boldsymbol{x}_{\text{latent}}^{\text{DownP}}$通过RVQ量化得到$N_{\text{codebook}}$个码本，每个码本由$N_{\text{codes}}$比特表征，每个码本大小为$2^{N_{\text{codes}}}$(索引范围为$0 \sim 2^{N_{\text{codes}}}-1$)。量化后共产生$N = N_{\text{codebook}} \cdot N_{\text{codes}}$个比特序列，同时产生$N$个比特序列对应的潜空间表征$\boldsymbol{x}_{\text{latent}}^{\text{RVQ}}$。应用层DAC编码器输出的码率为

$$R = \frac{N \cdot s}{M} \tag{2}$$

其中，N是DAC编码后的RTP载荷的比特数，s是音频信号的采样率，M是一个语音帧包含的时域采样点个数。当$N_{\text{codebook}} = 8$，$N_{\text{codes}} = 8$，s = 8 kHz，M = 256时，码率为2 kbps。其中，码本个数与码本大小都可以通过不同的设置进行调整，从而改变应用层语音编码器输出的码率。

第二阶段：协议封装与添加包头。包括在应用层添加RTP包头，以及在网络层和数据链路层(MAC、RLC、PDCP层)分别添加包头。

第三阶段：物理层**发送**。在物理层，N比特的RTP负载经过1/3码率的Turbo编码[20]，得到一个 3N 比特长度的序列。在调制之前，可选的，还可以进行速率匹配，然后将比特序列映射为复数符号。如果未进行速率匹配，则调制过程将产生3N/2个复数符号(例如使用QPSK调制[21])。随后将复数符号映射到3N/2个资源单元(Resource Elements, RE)上发送出去。

2. 接收端处理流程与发送端相对应，接收端(如基站与核心网侧)包含三个逆向处理阶段：

第一阶段：物理层接收。基站的物理层首先对接收到的信号进行解调，得到 3N 个对数似然比(Log-Likelihood Ratio, LLR)。这些 LLR 接着经过 Turbo 解码，恢复出 N 比特的RTP载荷。

第二阶段：协议解封装与向上投递。基站将恢复的N比特RTP载荷向上层(MAC、RLC、PDCP层)依次投递，再发送至核心网。在向上投递过程中，每投递到一层，则拆掉对应层的包头。

第三阶段：应用层DAC解码。在IMS中，将接收到的N 比特RTP载荷进行反量化得到对应的潜空间表征$\tilde{\boldsymbol{x}}_{\text{latent}}^{\text{RVQ}}$，并将其送入DAC解码器的输入进行上采样和特征提取，获得恢复的语音信号。

图1为DAC编解码推理结合3GPP各层传输的完整流程。而在DAC编解码器训练时，复用了开源 DAC的训练方式[18]，即在训练过程中不考虑无线传输信道的影响(仅包括图中蓝色模块)。

3　方案设计

上一节介绍了DAC编码器与现有3GPP支持的物理层传输方案。DAC编码器在训练时未考虑无线信道传输带来的影响，因此，是一种信源编码与信道编码分离编码的方案。本节提出3种信源编码与信道编码联合优化的方案，分别为总功率限制的DAC-JSCCM、恒模限制的DAC-JSCCM，以及DAC-JSCC。如图2，DAC- JSCCM是在应用层同时进行DAC信源编码、信道编码和调制，而DAC-JSCC是在应用层同时进行DAC信源编码和信道编码，调制仍然在物理层进行。总功率限制的DAC-JSCCM和恒模限制的DAC-JSCCM区别在于总功率限制的DAC-JSCCM是在总RE上进行功率归一化，而恒模限制的DAC-JSCCM是在每个RE上进行功率归一化。下面的3个小节将分别详细介绍提出的3种方案。

3.1　总功率限制(Total Power Constrained, TPC)的DAC-JSCCM

本小节提出的总功率限制的DAC-JSCCM方案同时对信源编码、信道编码与调制进行联合优化，并且只在总RE上进行了功率限制，因此，相比于图1的基线方案1，所提方案可获得最大的AI增益空间。

总功率限制的 DAC-JSCCM方案的流程图如下图2(a)所示。与基线方案1的分离式编解码架构相比，差异主要体现在应用层和物理层。

在发送端，基线方案1中原本在物理层进行的信道编码与调制被移至应用层，从而实现了信源信道编码与调制的联合优化。

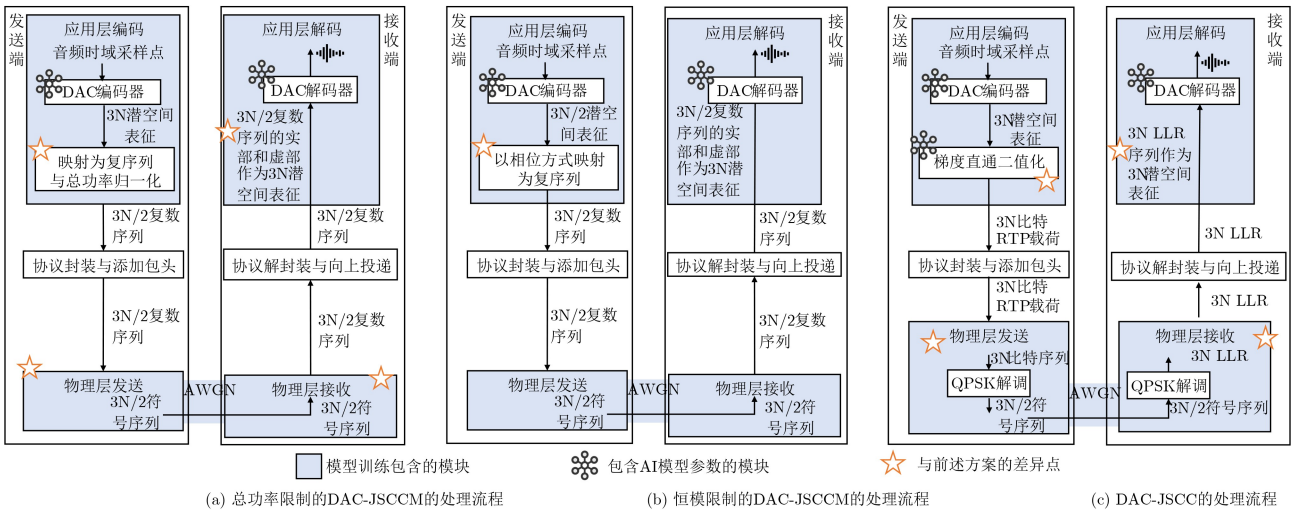

图 2 提出3种方案的处理流程

第一阶段：应用层基于DAC的信源信道联合编码与调制。发送端的原始语音信号首先送入应用层的 DAC 编码器进行处理。在此阶段，保留了与基线方案1相同的DAC编码器模型结构，用于提取隐空间特征序列。区别在于去除了DAC量化器，直接将隐空间特征序列映射为调制符号。该过程包含两个操作：

(1)连续特征至复数序列映射：假设DAC的特征提取与下采样模块输出的每个语音帧的潜空间表征为 $\boldsymbol{x}_{\text{latent}}^{\text{TPC}}=(x_1^{\text{TPC}},x_2^{\text{TPC}},\cdots,x_{3\text{N}}^{\text{TPC}})\in \text{R}^{1\times 3\text{N}}$。将每2个实数组成一个复数，得到一组长度为3N/2的复数序列 $\boldsymbol{y}_{\text{latent}}^{\text{TPC}}=(y_1^{\text{TPC}},y_2^{\text{TPC}},\cdots,y_{3\text{N}/2}^{\text{TPC}})\in \text{C}^{1\times 3\text{N}/2}$，复数序列中的复数 $y_l^{\text{TPR}}=x_{2l-1}^{\text{TPR}}+j\cdot x_{2l}^{\text{TPR}},\ 1\le l\le 3\text{N}/2$。

(2)全局功率归一化(TPC)：为了满足物理射频链路的发射功率约束，将复数序列 $\boldsymbol{y}_{\text{latent}}^{\text{TPC}}$ 进行功率归一化得到调制符号序列 $\boldsymbol{y}_{\text{M}}^{\text{TPC}}=(\overline{y}_1^{\text{TPC}},\overline{y}_2^{\text{TPC}},\cdots,\overline{y}_{3\text{N}/2}^{\text{TPC}})\in \text{C}^{1\times 3N/2}$。假设在所有RE上进行功率归一化，则 y_l 的计算公式如下：

$$\overline{y}_l^{\text{TPC}}=y_l^{\text{TPC}}\Big/\sqrt{\sum_{n=1}^{3\text{N}/2}|y_n^{\text{TPC}}|^2}(1\le l\le 3N/2) \tag{3}$$

第二阶段：协议封装与添加包头。包括在应用层添加RTP包头，以及在网络层和数据链路层(MAC、RLC、PDCP层)添加包头。

第三阶段：直通物理层发送。在物理层，系统将调制符号序列 $\boldsymbol{y}_{\text{M}}^{\text{TPC}}\in \text{C}^{1\times 3\text{N}/2}$ 映射到3N/2个RE上发送出去，而无需在物理层进行图1的信道编码与调制。

在接收端，基线方案 1 中原本在物理层进行的解调与信道解码被移至应用层，从而实现了信源信道解码与解调的联合优化。

第一阶段：直通物理层接收。基站的物理层接收的带噪声复数序列 $\tilde{\boldsymbol{y}}_{\text{M}}^{\text{TPC}}\in \text{C}^{1\times 3\text{N}/2}$：

$$\tilde{\boldsymbol{y}}_{\text{M}}^{\text{TPC}}=\boldsymbol{y}_{M}^{\text{TPC}}+\boldsymbol{z} \tag{4}$$

其中，$\boldsymbol{z}\in \text{C}^{1\times 3\text{N}/2}$ 是高斯白噪声。假设信号功率归一化为1，若给定的信噪比γ以分贝(dB)为单位，则对应的线性信噪比为$10^{\gamma/10}$。由此，可推导出复噪声的总方差为$10^{-\gamma/10}$。在此条件下，噪声实部与虚部均相互独立地服从均值为0、方差为$0.5\cdot 10^{-\gamma/10}$的实高斯分布。

基站在接收到带噪声复数序列 $\tilde{\boldsymbol{y}}_{\text{M}}^{\text{TPC}}\in \text{C}^{1\times 3\text{N}/2}$ 后，无需进行CRC校验，也无需进行如图1所示的在物理层信道译码和解调。

第二阶段：跨层透传与解封装。基站的物理层不进行CRC校验，直接向上层(MAC、RLC、PDCP层)投递。并发送至核心网。在向上投递过程中，每投递一层，则拆掉对应层的包头。

第三阶段：应用层DAC解码。在IMS中，复用现有的DAC解码器对带噪声复数序列 $\tilde{\boldsymbol{y}}_{\text{M}}^{\text{TPC}}\in \text{C}^{1\times 3\text{N}/2}$ 进行处理，获得恢复的语音信号。现有DAC解码器的输入是量化索引对应的潜空间表征，在本方案中，将复数序列的实部和虚部当作3N潜空间表征，作为DAC解码器的输入，并复用DAC解码器的模型结构。

图2(a)为总功率限制的DAC-JSCCM方案的推理和训练流程。相比于基线方案1，总功率限制的DAC-JSCCM方案在应用层进行了信源编码、信道编码以及调制的联合优化。此外，在DAC-JSCCM训练时，考虑了无线传输信道的影响。因此，具有更好的抗噪性。

3.2 恒模限制(Constant-envelope constrained, CEC)的DAC-JSCCM

3.1节提出的总功率限制的DAC-JSCCM是在

所有RE上功率归一化，这将造成某些RE上集中了过多的功率，从而提高了峰值与平均功率比(Peak-to-Average Power Ratio, PAPR)。PAPR较高的系统会导致功率放大器工作在较高的非线性区域，从而导致信号失真、效率低下和干扰增大等问题。为解决上述问题，本小节提出了恒模限制的DAC-JSCCM。恒模限制的DAC-JSCCM在每个RE上进行功率归一化，可以减少由于特定RE功率过高导致的PAPR增大的问题。因为每个RE的功率是均衡的，信号的峰值通常较低，不容易产生极高的功率峰值，从而可以降低PAPR。

如图2(b)所示，恒模限制的DAC-JSCCM方案的流程与3.1节非常类似。唯一的区别在于发送端应用层第一阶段的处理。其核心创新点是将功率约束从总RE功率约束加强为单RE功率约束，以获得与QPSK类似的PAPR。该方案在应用层基于DAC的信源信道联合编码与调制的差异化处理如下：

假设DAC编码器输出的每个语音帧的潜空间表征为$\boldsymbol{x}_{\text{latent}}^{\text{CEC}}=(x_1^{\text{CEC}},x_2^{\text{CEC}},\cdots,x_{3\text{N}/2}^{\text{CEC}})\in\text{R}^{1\times 3\text{N}/2}$。为了满足更严格苛的单 RE 纯恒模约束，该处理摒弃了3.1的实/虚部正交组合机制，转而将实数域的隐性特征直接作为复数的相位。基于潜空间表征$\boldsymbol{x}_{\text{latent}}^{\text{CEC}}\in\text{R}^{1\times 3\text{N}/2}$获得调制符号$\boldsymbol{y}_{\text{M}}^{\text{CEC}}\in\text{C}^{1\times 3\text{N}/2}$的计算公式如下：

$$\boldsymbol{y}_{\text{M}}^{\text{CEC}}=\exp(\text{j}\cdot 2\pi\cdot\boldsymbol{x}_{\text{latent}}^{\text{CEC}}) \tag{5}$$

该函数的数学特性在于，通过非线性的复共轭指数运算，确保输送到任意一个接收端 RE 上的基带符号模长恒久等效为 1。

图2(b)为恒模限制的DAC-JSCCM方案的推理和训练流程。相比于总功率限制DAC-JSCCM方案，本方案将DAC下采样模块输出的潜在表征用于产生复数相位，从而保证了调制符号的恒模特性，降低了PAPR。此外，在恒模限制的DAC-JSCCM训练时，也考虑了无线传输信道的影响，因此具有良好的抗噪性。

3.3 DAC-JSCC

3.1和3.2小节提出的DAC-JSCCM方案在应用层对信源编码、信道编码和调制进行了联合优化。但是在应用层进行调制，将导致网络层以及RAN的SDAP/PDCP/MAC层传输复数序列而非比特序列，对标准协议影响较大。因此，本小节提出DAC-JSCC方案，将调制过程仍然保留在物理层，因此，发送端的网络层以及RAN的SDAP/PDCP/MAC层的传输都可以复用第2节DAC方案的传输方式。相比于3.1和3.2节提出的方案，本小节提出的方案对协议影响更小。

如图2(c)所示，DAC-JSCC(调制分离)方案与3.1、3.2的区别在于，本方案在应用层仅进行了信源信道联合编解码，但未结合调制进行优化。本方案在应用层直接输出0/1 离散比特流，并在物理层恢复了调制解调处理。

在发送端，与3.1、3.2的差异点包括在第一阶段的应用层在DAC编码器后引入了梯度直通二值化，以及在第三阶段的物理层传输中包括了调制。下面将针对这两个有差异的阶段进行描述：

第一阶段：应用层基于DAC的信源信道联合编码与梯度直通二值化(Straight-Through Estimator, STE)。假设DAC编码器输出的每个语音帧的潜空间表征为$\boldsymbol{x}_{\text{latent}}^{\text{JSCC}}=(x_1^{\text{JSCC}},x_2^{\text{JSCC}},\cdots,x_{3\text{N}}^{\text{JSCC}})\in\text{R}^{1\times 3\text{N}}$。在本方案中，应用层输出的是0/1比特，如果直接对潜在表征$\boldsymbol{x}_{\text{latent}}^{\text{JSCC}}$进行硬判决，则会造成梯度无法反向回传的问题。为克服这一障碍，本方案在应用层设计了“梯度直通二值化”模块，通过三步联合映射实现平滑转化：

(1)硬二值化判决。在前向传播时，对$\boldsymbol{x}_{\text{latent}}^{\text{JSCC}}$进行阈值操作得到硬二值化序列$\boldsymbol{x}_{\text{hard}}^{\text{JSCC}}=(x_{\text{hard},1}^{\text{JSCC}},x_{\text{hard},2}^{\text{JSCC}},\cdots,x_{\text{hard},3N}^{\text{JSCC}})\in\text{R}^{1\times 3\text{N}}$。

其中，$x_{\text{hard},m}^{\text{JSCC}}$为0或1，$1\le m\le 3N$。$x_{\text{hard},m}^{\text{JSCC}}$的计算公式如下：

$$\begin{aligned}x_{\text{hard},m}^{\text{JSCC}}&=\text{sign}\left(x_{\text{latent},m}^{\text{JSCC}}\right)\\&=\begin{cases}1, & \text{if}x_m^{\text{JSCC}}>0\\0, & \text{else}\end{cases}(1\le m\le 3\text{N})\end{aligned} \tag{6}$$

在反向传播中，定义了一个掩码条件，只有当x_m^{JSCC}不大于1时(即 $|x_m^{\text{JSCC}}|<=1$)，上游传递的梯度才会被传递。

(2)软二值化平滑：在前向传播中，对$\boldsymbol{x}_{\text{latent}}^{\text{JSCC}}$使用了sigmoid函数得到软二值化序列$\boldsymbol{x}_{\text{soft}}^{\text{JSCC}}=(x_{\text{soft},1}^{\text{JSCC}},x_{\text{soft},2}^{\text{JSCC}},\cdots,x_{\text{soft},3N}^{\text{JSCC}})\in\text{R}^{1\times 3\text{N}}$。$x_{\text{soft},m}^{\text{JSCC}}$的计算公式如下：

$$x_{\text{soft},m}^{\text{JSCC}}=\text{sigmoid}\left(x_{\text{latent},m}^{\text{JSCC}}\right) \tag{7}$$

其中，$x_{\text{soft},m}^{\text{JSCC}}$属于[0, 1]的连续区间。

(3)基于‘stop_gradient’的联合前反向映射：为确保系统向信道输出的严格为0 或 1 的离散态，同时保证有效的梯度回传，将硬判决与软判决序列进行如下等式合并：

$$\boldsymbol{y}^{\text{JSCC}}=\boldsymbol{x}_{\text{soft}}^{\text{JSCC}}+\text{stop\_gradient}(\boldsymbol{x}_{\text{hard}}^{\text{JSCC}}-\boldsymbol{x}_{\text{soft}}^{\text{JSCC}}) \tag{8}$$

其中，在前向传播中，实际输出为

$$\boldsymbol{y}^{\text{JSCC}}=\boldsymbol{x}_{\text{hard}}^{\text{JSCC}}=\text{sign}\left(\boldsymbol{x}_{\text{latent}}^{\text{JSCC}}\right) \tag{9}$$

公式(8)是实现STE的核心：在前向传递时，由于‘stop_gradient’算子不对数值产生影响，公式(8)

等效公式(9)，输出了0/1比特流；而在反向求导时，截断了离散项的不可导梯度，仅保留连续平滑的$\boldsymbol{x}_{\text{soft}}^{\text{JSCC}}$参与对后端的链式求导，从而在保障离散输出的同时实现了网络权重的有效迭代。

第三阶段：物理层调制与传输。在物理层，系统对比特序列$\boldsymbol{y}^{\text{JSCC}} \in \text{R}^{1\times 3\text{N}}$进行调制，例如，通过QPSK调制得到调制后的符号序列$\boldsymbol{y}_{\text{M}}^{\text{JSCC}} \in \text{R}^{1\times 3\text{N}/2}$，再映射到3N/2个RE上发送出去。

在接收端处理流程中，与3.1、3.2的差异包括第一阶段的物理层接收和第三阶段的应用层处理。下面将针对这两个有差异的阶段进行描述：

第一阶段：物理层接收与软解调。基站的物理层接收带噪声复数序列$\tilde{\boldsymbol{y}}_{\text{M}}^{\text{JSCC}} \in \text{C}^{1\times 3\text{N}/2}$：

$$\tilde{\boldsymbol{y}}_{\text{M}}^{\text{JSCC}} = \boldsymbol{y}_{\text{M}}^{\text{JSCC}} + \boldsymbol{z} \tag{10}$$

其中，$\boldsymbol{z} \in \text{C}^{1\times 3\text{N}/2}$是高斯白噪声。基站在接收到带噪声复数序列$\tilde{\boldsymbol{y}}_{\text{M}}^{\text{JSCC}} \in \text{C}^{1\times 3\text{N}/2}$后，进行QPSK解调获得LLR序列$\tilde{\boldsymbol{y}}_{\text{LLR}}^{\text{JSCC}} \in \text{R}^{1\times 3\text{N}}$。

第三阶段：应用层DAC解码。 在IMS中，复用现有的DAC解码器对LLR序列$\tilde{\boldsymbol{y}}_{\text{LLR}}^{\text{JSCC}}$进行处理，获得恢复的语音信号。现有DAC解码器的输入是量化索引对应的潜空间表征，在DAC-JSCC方案中，将3N LLR当作3N潜空间表征，作为DAC解码器的输入，并复用DAC解码器的模型结构。

图2(c)为DAC-JSCC方案的推理和训练流程。相比于DAC-JSCCM方案，本方案将DAC下采样模块输出的潜在表征通过梯度直通二值化产生比特序列，在物理层进行调制，而非如DAC-JSCCM方案在应用层进行调制，因此对协议影响更小。在接收端，基站的物理层需要先解调，再向上投递LLR，而非如DAC-JSCCM方案直接向上投递含噪复数序列。此外，在DAC-JSCC训练时，也考虑了无线传输信道的影响，因此具有良好的抗噪性。

4 仿真评估

4.1 数据集

目前公开的音频数据集包括VCTK[22]，common voice[23]，librispeech[24]等。本次仿真选择了common voice数据集。其中，训练数据集、验证数据集和测试数据集的样本是正交的。三者的样本数分别为1 000 000、16和1 000。

4.2 评估指标

评估指标采用语音质量感知评价(PESQ, Perceptual Evaluation of Speech Quality)。PESQ是国际电信联盟(ITU-T)在标准ITU-T P.863 (2018)[25]中定义的一种客观语音质量评估方法。它通过模拟人类听觉感知机制，对通信系统或编解码器处理后的语音信号质量进行客观评价。PESQ的核心思想是：将参考语音信号与系统处理后的退化语音信号进行比较，模拟人耳的听觉处理过程，从而得到一个语音质量分数。该得分可以直接与主观听音测试的平均意见分(MOS, Mean Opinion Score)对比，因此常被称为PESQ MOS。其中，PESQ = 2.5分通常被视为最低可用阈值，低于此分值的语音质量会被认为不适宜用于专业或商业服务。

4.3 基线方案

本次评估选择了2个基线方案。基线方案1是第2章描述的应用层DAC与物理层信道编码方案，基线方案2与基线方案1的区别在于基线方案2在物理层不进行信道编解码。如图5所示，相比基线方案1，基线方案2在应用层DAC量化器输出可以扩大三倍至3N，在物理层传输时不进行1/3 Turbo编码，在接收端的物理层接收时，将Turbo解码替换为硬判决。在应用层处理中增加了反量化将RTP载荷映射为潜空间表征。其他流程基本相同不再赘述。

图3为基线方案2的推理流程。在基线方案2训练时，复用了开源DAC的训练方式[18]，即在训练过程中不考虑无线传输信道的影响(仅包括图中蓝色模块)。与基线方案1的训练区别在于基线方案2将RVQ量化器输出扩大了3倍，以确保物理层不进行信道编码后，使用的传输RE个数仍然为3N/2，与基线方案1保持一致。

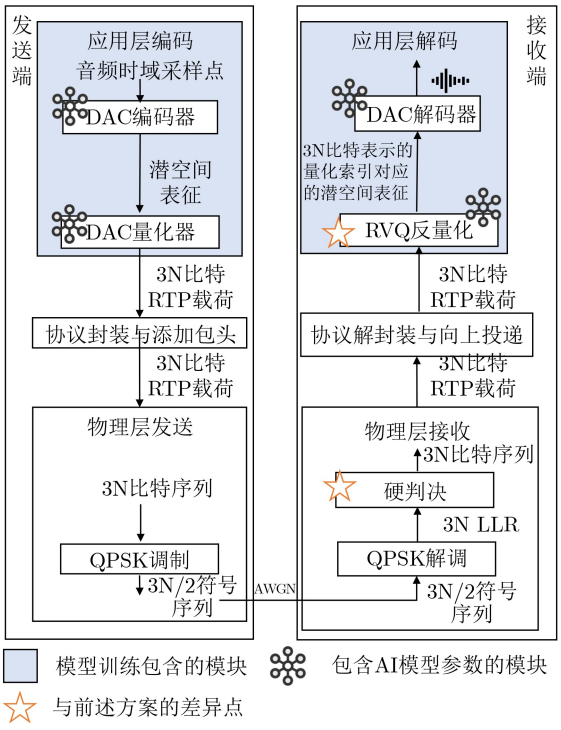

图 3 基线方案2 DAC-物理层无信道编码方案的处理流程

4.4 仿真假设

本次仿真将对比4.3节的基线方案1和基线方案2，以及第3节提出的总功率限制的DAC-JSCCM、恒模限制的DAC-JSCCM、DAC-JSCC。模型训练时的参数配置大部分沿用DAC公开的参数配置[18]，考虑到低码率场景，所有方案假设音频时长为500 ms，采样率为8 kHz，编码后的音频帧数为16，DAC的4次下采样卷积步长为[2, 4, 4, 8]，物理层用于传输的RE个数为96，训练样本数/验证样本数/测试样本数分别为1 000 000/16/100，训练batch size为8，验证batch size为16，训练轮数为1000000，workers个数为6。各方案不同的仿真假设如表1所示。值得注意的是，本文的接收端恢复音频质量对比均建立在统一的带宽资源基准之上，即保证各方案的音频输入完全相同且消耗的物理层资源(RE)总数相等。根据等效资源映射原则，假设传统分离式架构使用1/3码率的信道编码，其对应的信源载荷仅为 N 比特(经编码冗余增至3N比特)。而在无需物理层信道编码的架构可将全部带宽分配给信源层，直接生成3N个比特；进一步地，若应用层输出为复数序列，则总资源等效为占用3N/2个复数符号。

本次仿真暂未考虑步骤2添加的包头的传输，因为包头信息需要进行无损传输，而无法如RTP载荷进行有损传输。

4.5 仿真结果

在基线方案1方案的评估中，首先通过物理层的链路级仿真，得到输入长度为64比特且采用 1/3 Turbo 编码时的 信噪比与误块率表，如表2所示。

再根据表2得到基线方案1的RTP包错误率，假设1个RTP包包含L个语音帧，当信噪比为γ时，1个语音帧(64比特)对应的误块率为p_γ，假设一个语音帧的错误率与误块率相同，即$p_{\text{frame},\gamma}=p_\gamma$。根据现有的传输机制，一个RTP包中只要有一个语音帧传输错误则整个RTP将被丢弃。则一个RTP包的错误概率为$p_{\text{RTP},\gamma}=1-(1-p_\gamma)^L$。以$L$=16，信噪比 γ=1 dB为例，由表2可知误块率p_γ =0.0039，等效为1个语音帧错误的概率$p_{\text{frame},\gamma}=0.0039$，则一个RTP包的错误概率为$p_{\text{RTP},\gamma}=1-(1-0.0039)^{16}\approx 0.06061$。

基于上述假设，图4在输入音频均为500 ms，采样率为8 kHz，且传输RE数均为96的相同前提下，对比了4.3节的基线方案1(DAC + 1/3 Turbo + QPSK)和基线方案2(DAC + QPSK)，以及第3章节提出的总功率限制的DAC-JSCCM，恒模限制的DAC-JSCCM，以及DAC-JSCC。

在实时通信、VoIP或语音增强场景中，PESQ = 2.5分通常被视为最低可用阈值，低于此分值的语音质量会被认为不适宜用于专业或商业服务。因此，在GEO语音通话中，我们选择PESQ = 2.5分作为参考点。由图4可以观察到，当PESQ = 2.5时，总功率限制DAC-JSCCM和恒模限制DAC-JSCCM相比于基线方案1有大概4 dB的覆盖增益。DAC-JSCC方案相比于基线方案1有2.5 dB的覆盖增益。DAC-JSCCM与DAC-JSCC的覆盖增益来源于应用层进行了信源信道联合编码，从而避免了物理层使用Turbo编码而导致的悬崖效应。

基线方案2在高信噪比区域表现与总功率限制DAC-JSCCM类似，但是在信噪比<10 dB区间，PESQ明显下降。基线方案2由于在物理层未使用Turbo编码，因此PESQ随着信噪比缓慢下降，而

表 1　仿真参数设置

参数	基线方案1	基线方案2	总功率限制DAC-JSCCM	恒模限制DAC-JSCCM	DAC-JSCC
应用层输出(N=64)	N比特	3N比特	3N /2复数	3N /2复数	3N比特
物理层调制方式	QPSK	QPSK	无	无	QPSK
物理层信道编码	1/3 Turbo	1/3 无	无	无	无
训练信噪比(dB)	无	无	[–5,10)	[–5,10)	[–5,10)
DAC下采样输出维度	1024	1024	192	96	192
RVQ码本数	8	24	无	无	无
码本大小	256	256	无	无	无

表 2　信噪比与误块率表(64比特长度，1/3 Turbo编码)

信噪比(dB)	误块率
–6	1
–5	1
–4	1
–3	0.925
–2	0.665
–1	0.236
0	0.0464
1	0.0039
2	0.0004
3	0
4	0
5	0
6	0
7	0

未出现PESQ陡降的悬崖效应。尽管在推理阶段，基线方案2与DAC-JSCC的前向处理流程高度相似，但由于DAC-JSCC在训练过程中将信道噪声影响纳入了联合优化，使其网络权重具备了内在的抗噪能力。因此，在低信噪比(SNR)区间内，DAC-JSCC的PESQ得分相较于基线方案2实现了显著的性能增益。

图5为96个RE下，总功率限制的DAC-JSCCM、恒模限制的DAC-JSCCM，以及QPSK调制(对应提出的DAC-JSCC和2个基线方案) 的PAPR互补累积分布函数(Complementary Cumulative Distribution Function, CCDF)对比。从图5中可以观察到，当CCDF$=10^{-3}$时，总功率限制的DAC-JSCCM的PAPR比QPSK调制高出4 dB，信号峰更尖锐，对功放线性度要求更高，更容易产生失真。而恒模限制的DAC-JSCCM与QPSK调制的PARP非常接近。

5 结论与下一步计划

目前，3GPP SA4已经开始研究基于AI的语音codec，本文以目前SA4正在研究的DAC编解码器为基础，提出了总功率限制DAC-JSCCM，恒模限制DAC-JSCCM，以及DAC-JSCCM方案。并结合现有3GPP GEO语音的物理层典型配置进行了仿真。通过仿真观察到当PESQ为2.5(最低可用阈值)时，相比于DAC结合现有物理层传输技术，提出的DAC-JSCCM有4 dB的覆盖增益，提出的DAC-JSCC方案有2.5 dB的覆盖增益。SA4正在研究比DAC更先进的GEO语音编解码，后续我们将结合更优的GEO语音编解码扩展新的研究，并将考虑在衰落信道下进行仿真评估。

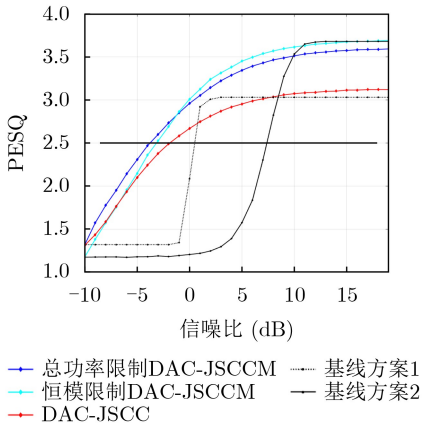

图 4 提出方案与基线方案的信噪比与PESQ对比

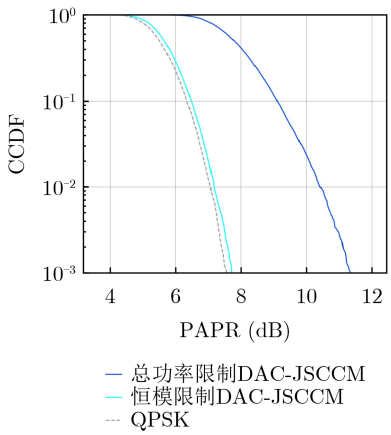

图 5 提出方案1、2与QPSK的PAPR CCDF对比

参 考 文 献

[1] 周通, 姜大洁, 谭俊杰, 等. 语义通信的用例、挑战与标准化影响浅析[J]. 移动通信, 2025, 49(7): 2–11. doi: 10.3969/j.issn.1006-1010.20250522-0001.
ZHOU Tong, JIANG Dajie, TAN Junjie, *et al*. A survey of semantic communication: Use cases, open challenges, and standardization trends[J]. *Mobile Communications*, 2025, 49(7): 2–11. doi: 10.3969/j.issn.1006-1010.20250522-0001.

[2] OPPO. 基于信源信道联合编码的CSI反馈[R]. IMT-2030-Semantic_2024004, 2024. **(查阅网上资料, 未找到本条文献信息, 请确认)**.

[3] OPPO. 信道与无线环境语义通信_基于联合信源信道编码的CSI反馈[R]. IMT-2030-Semantic_2024036, 2024. **(查阅网上资料, 未找到本条文献信息, 请确认)**.

[4] 小米. 面向CSI反馈的信道估计与联合信源信道编码的联合设计[R]. IMT-2030-Semantic_2024056, 2024. **(查阅网上资料, 未找到本条文献信息, 请确认)**.

[5] YANG Ke, WANG Sixian, DAI Jincheng, *et al*. SwinJSCC: Taming swin transformer for deep joint source-channel coding[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2025, 11(1): 90–104. doi: 10.1109/TCCN.2024.3424842.

[6] GUO Jiangyuan, CHEN Wei, SUN Yuxuan, *et al*. VideoQA-SC: Adaptive semantic communication for video question answering[J]. *IEEE Journal on Selected Areas in Communications*, 2025, 43(7): 2462–2477. doi: 10.1109/JSAC.2025.3559160.

[7] TUNG T Y and GÜNDÜZ D. DeepWiVe: Deep-learning-aided wireless video transmission[J]. *IEEE Journal on Selected Areas in Communications*, 2022, 40(9): 2570–2583. doi: 10.1109/JSAC.2022.3191354.

[8] BOURTSOULATZE E, KURKA D B, and GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2019, 5(3): 567–579. doi: 10.1109/TCCN.2019.2919300.

[9] YILMAZ S F, NIU Xueyan, BAI Bo, *et al*. High perceptual quality wireless image delivery with denoising diffusion models[C]. IEEE Conference on Computer Communications Workshops, Vancouver, Canada, 2024: 1–5. doi: 10.1109/INFOCOMWKSHPS61880.2024.10620904.

[10] DAI Jincheng, WANG Sixian, TAN Kailin, *et al*. Nonlinear transform source-channel coding for semantic communications[J]. *IEEE Journal on Selected Areas in Communications*, 2022, 40(8): 2300–2316. doi: 10.1109/JSAC.2022.3180802.

[11] WENG Zhenzi, QIN Zhijin, and LI G Y. Robust semantic communications for speech transmission[C]. 2025 IEEE International Conference on Acoustics, Speech and Signal Processing, Hyderabad, India, 2025: 1–5. doi: 10.1109/ICASSP49660.2025.10889160. **(查阅网上资料,未能确认本条文献修改是否正确,请确认)**.

[12] HAN Tianxiao, YANG Qianqian, SHI Zhiguo, *et al*. Semantic-preserved communication system for highly efficient speech transmission[J]. *IEEE Journal on Selected Areas in Communications*, 2023, 41(1): 245–259. doi: 10.1109/JSAC.2022.3221952.

[13] WENG Zhenzi, QIN Zhijin, TAO Xiaoming, *et al*. Deep learning enabled semantic communications with speech recognition and synthesis[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(9): 6227–6240. doi: 10.1109/TWC.2023.3240969.

[14] CHEN Xiaojiao, WANG Jing, XU Liang, *et al*. A perceptually motivated approach for low-complexity speech semantic communication[J]. *IEEE Internet of Things Journal*, 2024, 11(12): 22054–22065. doi: 10.1109/JIOT.2024.3378779.

[15] 3GPP. 3GPP TS 23.501 System architecture for the 5G System (5GS)[S]. 2025. **(查阅网上资料, 未找到出版信息, 请补充)(查阅网上资料, 未能确认本条文献修改是否正确, 请确认)**.

[16] 3GPP. 3GPP TS 38.323 Packet data convergence protocol (PDCP) specification[S]. 2025. **(查阅网上资料, 未找到出版信息, 请补充)(查阅网上资料, 未能确认本条文献修改是否正确, 请确认)**.

[17] IMT-2030(6G)推进组. 面向6G的语义通信技术研究报告[M]. 北京: IMT-2030(6G)推进组, 2025. **(查阅网上资料, 未找到对应的英文翻译, 请补充)**.

[18] Descript. Descript-audio-codec[EB/OL]. https://github.com/descriptinc/descript-audio-codec, 2026.

[19] 3GPP. 3GPP TR 26.940 Study on ultra low bit rate speech codecs[S]. 2026. **(查阅网上资料, 未找到出版信息, 请补充)(查阅网上资料, 未能确认本条文献修改是否正确, 请确认)**.

[20] 3GPP. 3GPP TS 36. 212 Multiplexing and channel coding[S]. 2025. **(查阅网上资料, 未找到出版信息, 请补充)(查阅网上资料, 未能确认本条文献修改是否正确, 请确认)**.

[21] ETSI. ETSI TS 136 213 V19.1. 0 (2025-10) Physical layer procedures[S]. 2025. **(查阅网上资料, 未找到出版信息, 请补充)(查阅网上资料, 未能确认本条文献修改是否正确, 请确认)**.

[22] OpenDataLab. VCTK[EB/OL]. https://opendatalab.com/OpenDataLab/VCTK, 2026.

[23] Common voice[EB/OL]. https://commonvoice.mozilla.org/zh-CN, 2026.**(查阅网上资料,请补充作者信息)**.

[24] Librispeech[EB/OL]. https://huggingface.co/datasets/openslr/librispeech_asr, 2026. **(查阅网上资料,请补充作者信息)**.

[25] ITU. ITU-T P. 863-2018 Perceptual objective listening quality prediction[S]. 2018. **(查阅网上资料, 未找到出版信息, 请补充)**.

周　通：女，现任vivo通信研究院预研专家，研究方向为通信与AI融合技术.

陈宏智：男，现任vivo通信预研工程师，研究方向为通信与AI融合技术和AI感知辅助通信．.

许佳龙：男，现任vivo通信研究院预研工程师，研究方向为基于AI的联合信源信道编码，基于AI的无线通信物理层技术.

孙　鹏：男，高级工程师，研究方向为massive MIMO，AI等.

姜大洁：男，高级工程师，现任vivo通信研究院预研总监，研究方向为通感一体化、AI与通信结合等.

刘健康：男，现任vivo通信研究院验证专家，研究方向为数据面，IoT-NTN，语音AI Codec等.

责任编辑：廖海贝

Research on AI-Enabled Real-Time Audio Joint Source-Channel Coding

ZHOU Tong[①]　CHEN Hongzhi[①]　XU Jialong[①]　SUN Peng[①]
JIANG Dajie[①]　LIU Jiankang[①]

[①](*vivo Mobile Communication Co. Ltd.*, *Beijing* 10015, *China*)

Abstract: Currently, the 3rd Generation Partnership Project (3GPP) Technical Specification Group Service and System Aspects Working Group 4 (TSG SA Working Group 4, SA4) has begun researching Artificial Intelligence (AI)-based audio codecs. Based on the Descript Audio Codec (DAC) being studied by SA4, this paper presents a joint optimization scheme for DAC source-channel coding and modulation with total power constraints, a joint optimization scheme for DAC source-channel coding and modulation with constant modulus constraints, and a DAC joint source-channel coding scheme. Furthermore, simulation evaluation is conducted using the typical physical layer parameter configuration of 3GPP Geostationary Earth Orbit (GEO) voice scenario. Finally, preliminary research suggestions are made, with the hope of providing useful insights for the

subsequent research on GEO voice.

Objective In recent years, Joint Source-Channel Coding (JSCC) has gained widespread attention in academia and industry. In academia, the sources studied for JSCC include video, image, and audio. However, the industry has not focused on these sources as in academia, but rather on JSCC where channel state information (CSI) is considered as the source. The reason the industry has not considered JSCC for video and images is due to two main factors. First, for video and images, 3GPP generally does not conduct independent research, but instead directly reuses source compression standards developed by external organizations such as the Moving Picture Experts Group (MPEG) and the Joint Photographic Experts Group (JPEG). Second, it is constrained by the challenges of the source-channel coding architecture. For example, in joint source-channel coding at the application layer, the channel information obtained at the application layer is often outdated, which limits the potential gains of source-channel coding. However, the situation is quite different for audio. Audio coding and decoding are designed independently by 3GPP, and currently, 3GPP SA4 is researching AI-based speech coding and decoding. Moreover, in GEO speech scenarios, where satellites remain largely stationary relative to ground stations and the channel changes slowly, the physical layer CSI can be transmitted to the application layer, overcoming the issue of outdated channel information. This makes the research on JSCC for audio more promising within 3GPP, compared to JSCC for video or image sources.

Methods This paper firstly uses the typical physical layer parameter configurations of 3GPP GEO voice and the DAC encoder, currently being studied by SA4, as an evaluation baseline, ensuring the fairness of the proposed scheme's gain evaluation. Based on this, the paper proposes the integration of DAC with channel coding and modulation to improve audio quality in low-bitrate scenarios. First, the total power constrained (TPC) DAC-JSCCM scheme is considered, which offers the maximum potential gain due to the joint optimization of source coding, channel coding, and modulation. Then, considering the high PAPR (Peak-to-Average Power Ratio) impact on the power amplifier, the paper proposes a constant-envelope constrained (CEC) DAC-JSCCM scheme. Finally, since the modulation symbols included in RTP payload require significant changes to the existing protocol, a DAC-JSCC scheme that is more compatible with current protocols is proposed, where bit sequences are included in RTP payload.

Results and Discussions In the evaluation of baseline scheme 1, the link-level simulation at the physical layer is first conducted to obtain the relationship between SNR and BLER for an input length of 64 bits using a 1/3 Turbo code. Then, assuming that the error rate of an audio frame is the same as the BLER, the transmission error rate of an RTP packet containing L audio frames is calculated. According to the existing processing logic, if one audio frame in an RTP packet is transmitted with an error, the entire RTP packet will be discarded. In real-time communication, VoIP, or speech enhancement scenarios, a PESQ score of 2.5 is generally considered the minimum usable threshold, and voice quality below this score is regarded as unsuitable for professional or commercial services. Therefore, in GEO voice calls, we choose PESQ = 2.5 as the reference point. When PESQ = 2.5, the TPC DAC-JSCCM and CEC DAC-JSCCM provide a coverage gain of approximately 4 dB compared to baseline scheme 1 (Fig.4). The DAC-JSCC scheme offers a coverage gain of 2.5 dB compared to baseline scheme 1 (Fig.4). The coverage gains of DAC-JSCCM and DAC-JSCC come from the application layer performing joint source-channel coding, which avoids the cliff effect caused by using Turbo coding at the physical layer. When the Complementary Cumulative Distribution Function (CCDF) of Peak-to-Average Power Ratio (PAPR) is 10^{-3}, the TPC DAC-JSCCM is 4 dB higher than the QPSK modulation, with sharper signal peaks, higher requirements for power amplifier linearity, and is more prone to distortion. In contrast, the CEC DAC-JSCCM is very close to the QPSK modulated PAPR.

Conclusions This paper proposes the Total Power Constrained DAC-JSCCM, Constant Modulus Constrained DAC-JSCCM, and DAC-JSCCM schemes, based on the DAC codec currently being researched by SA4. Simulations are conducted using the typical physical layer configuration of the existing 3GPP GEO voice scenario. The simulation results show that, when the PESQ is 2.5 (the minimum usable threshold), compared to DAC combined with existing physical layer transmission technologies, the proposed DAC-JSCCM provides a coverage gain of 4 dB, while the proposed DAC-JSCC scheme provides a coverage gain of 2.5 dB. SA4 is currently researching more advanced GEO voice codecs beyond DAC, and we will extend our research by combining better GEO voice codecs in the future.

Key words: GEO Voice; Joint Source-channel Coding (JSCC); Joint Source-channel Coding and Modulation (JSCCM); AI