

语音信源的语义编码传输方法研究



Semantic Coded Speech Transmission

牛凯/NIU Kai, 姚圣时/YAO Shengshi, 戴金晟/DAI Jincheng

(北京邮电大学泛网无线通信教育部重点实验室, 中国 北京 100876)
(The Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI: 10.12142/ZTETJ.202302008

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.tn.20230411.1437.016.html>

网络出版日期: 2023-04-12

收稿日期: 2023-01-25

摘要: 作为一种新兴的通信范式, 语义通信在有效提升端到端传输性能方面展现出巨大的潜力。语音信源的语义编码传输研究方法分为两大类: 基于信号波形的语音语义编码传输和生成式语音语义编码传输。在基于信号波形的语音语义编码传输中, 现有方案的语义信息无法度量, 编码效率低。基于非线性变换的语音语义信源编码方案通过对语音的语义特征进行变分建模, 有效衡量语义特征内容复杂度, 并引入信源信道联合编码, 使语义编码传输更加高效可靠。针对生成式语音语义编码传输方法, 分析了其优势、挑战, 以及研究前景。

关键词: 语音编码; 语义编码传输; 语义通信

Abstract: As an emerging communication paradigm, semantic communications has shown great potential in effectively boosting end-to-end transmission performance. The problem of semantic coded speech transmission is investigated, which can be divided into two main categories: waveform-based and generative semantic speech coded transmission methods. In waveform-based semantic speech coding and transmission, existing solutions cannot quantify semantic information effectively, resulting in low efficiency. The proposed speech semantic coding scheme based on nonlinear transform measures the complexity of semantic features through variational modeling and introduces joint source-channel coding, making semantic coded transmission more efficient and reliable. The advantages, challenges and future research prospects of generative semantic speech coded transmission are summarized.

Keywords: speech coding; semantic coded transmission; semantic communication

过去70年, 通信技术取得了突飞猛进的发展。基于经典信息理论^[1], 先进的编码技术包括信源编码、信道编码, 已经逼近了信源熵/率失真函数以及信道容量。然而, 经典信息论只研究语法信息, 致力于信息传输的有效性和可靠性。从认识论观点看, 信息分为3个层次: 语法、语义和语用^[1]。C. E. SHANNON 和 W. WEAVER^[2]指出了通信的另一层面, 即通信的语义问题以及有效性问题。得益于人工智能技术的发展, 之前未能被解决的信息的语义及语用层次重新得到了研究人员的关注。北京邮电大学张平院士深入分析语义信息特征, 提出语义基 (Seb) 模型^[3], 指出语义信息可以用 Seb 进行表征。张平院士提出“智简 (Intellicise)”理念^[4], 进一步提出模型驱动的语义通信框架, 实现通信系统由传输比特演进为传输经语义处理得到的新特征, 例如 Seb 等。牛凯等深入研究了从经典通信到语义通信的范式转变, 提出了语义通信系统的基本框架^[5-7]。其他团队也对语义通信开展了多项研究工作: 石光明等提出了语义通信的基本框

架^[8], 秦志金等针对语义通信系统探讨了深度学习赋能的语义通信理论、框架和系统模型^[9], 北京交通大学、帝国理工大学团队分析了信源信道联合编码在语义通信中的应用^[10]。本文总结并回顾已有的面向语音信源的语义编码传输方法, 并指出未来的研究方向。

语音编码一直是移动通信中的基础问题, 其目的是获得语音信号的紧凑表征, 从而实现高效传输。波形编码和参量编码是传统语音编码的两大主流方法。波形编码旨在编码端对波形信号进行数字化, 并在解码端尽可能重构波形, 以脉冲编码调制 (PCM)、差分脉冲编码调制 (DPCM) 和自适应差分脉冲编码调制 (ADPCM)^[11]为代表。通过传输声学模型的系数, 参数编码以低比特率成本提供了可比较的重建质量。线性预测编码 (LPC) 引入了信源-滤波器模型, 它的参数通过线性来预测, 并根据语音样本来确定。码激励线性预测 (CELP)^[12]是一种结合波形编码和 LPC 模型的混合编码方法, 适用于窄频带和中频带语音编码。Opus 是近几年性能较好的开源音频编码器^[13], 支持动态可调速率以及从窄带到全带宽的音频编码。其中, 语音编码采用了基于 LPC 的 SILK 和能量约束重叠变换 (CELT) 编码器。

基金项目: 国家自然科学基金 (92067202、62293481、62071058)

现有的通信系统采用传统的音频编码器级联信道编码的方法,但其编码信息与语义无关,且易出现差错传播的情况。从语义信息传输的角度来设计传输方法,可以在语义空间内纠正信道传输引起的偏差。这种方法的针对性更强,最终可以提升端到端的编码传输效率。基于神经网络的语音编码方法已经有了初步研究,引入线性预测和基于变分自编码器的矢量量化(VQ-VAE)^[14],可以将语音特征以低速率传输。这些方法对语音信号的语义编码传输起到了指导作用。传统语音评价指标的不可微的性质导致无法直接用于语义编码模型优化,因此寻找合适的语音信号的语义传输的感知优化函数至关重要。

1 语义通信系统简介

语义通信系统的模型^[5]如图1所示,参照C. E. SHANNON和W. WEAVER的思想^[2],该模型分为Level A技术级通信与Level B语义级通信两个层级,具体由信源、语义知识库、语义发射机、信号发射机、信道、信号接收机、语义接收机、信宿8个部分组成。

其中,语义知识库包括信道环境特征、信源及任务特征,可以对语义发射机/接收机、信号发射机/接收机起到指导作用。语义编码传输问题即如何在语义知识库的指导下进行语义发射机/接收机和信号发射机/接收机的编译码设计问题。

语音信源的语义传输任务通常分为两大类:1)面向听觉的语音信号传输;2)传递内涵文本信息的语音信号传输。两类任务的语义编码优化设计准则也不同:面向语音文本内容的传输以误词率、文本重合度为语义传输性能的主要评价指标;面向人的听觉的语义传输主要考虑的是人的主观感受。针对如何提高人的主观感受(或称为感知编码),研究者提出了客观语音质量评估(PESQ)^[15]、ViSQOL^[16]等语音感知质量评价指标。在语音质量的主观评价方面,除平均主

观意见分(MOS)外,常见的还有多激励隐藏参考基准测试(MUSHRA)^[17]方法。

2 面向语音的语义编码传输方法

语音信源的语义编码传输一般包括两类方法:基于波形的语义编码传输和生成式的语义编码传输。

2.1 基于波形的语义编码传输

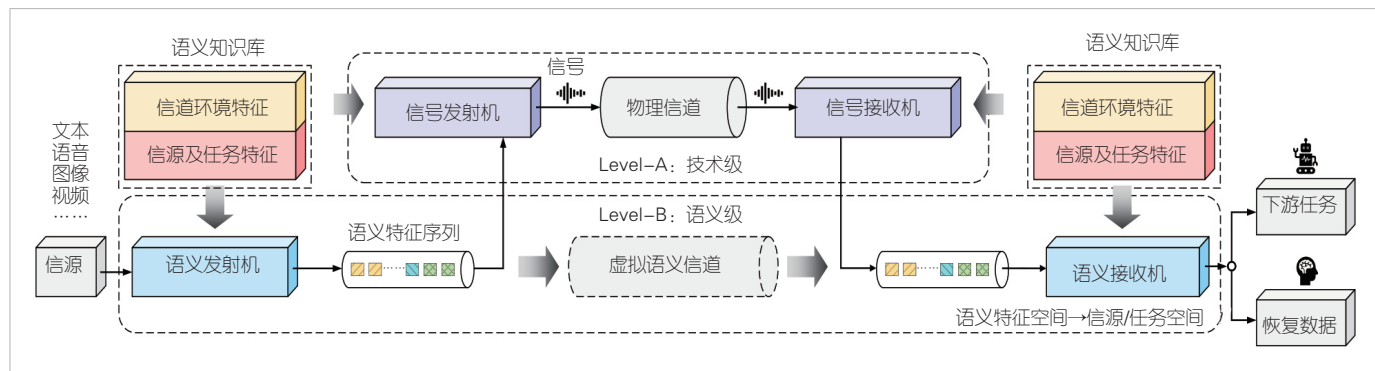
基于时域波形信号的语义编码传输方法直接从语音波形信号中提取语义特征并映射为信道中的传输符号。此类方法属于信源信道联合编码范畴。利用深度神经网络的信源信道联合编码,重建质量随信噪比的降低而柔和降低,不存在悬崖效应。QIN Z. J.等针对语音信源,构建了一种面向语音信源的深度语义通信网络(DeepSC-S)^[18],提高了语音信号的重建质量。该方法主要利用卷积神经网络与“压缩-激励”模块提取语音信源语义特征,实现从信源数据到信道中传输符号的映射。该编码过程与信道条件有关,达到信源信道联合编码的效果。

语音信源向量 $\mathbf{s} \in \mathbb{R}^m$ 经过编码神经网络 f_e 映射,得到语义编码向量 $\mathbf{x} \in \mathbb{R}^k$,满足编码约束 $\mathbf{x} = f_e(\mathbf{s}; \phi_e)$ 。其中, ϕ_e 为DeepSC-S编码器模型参数。经过加性高斯白噪声(AWGN)信道后,解码器输入语义向量为 $\mathbf{y} = \mathbf{x} + \mathbf{n}$ 。其中, $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}_k)$ 是独立同分布的高斯噪声向量, σ_n^2 是噪声方差, $\mathbf{I}_k$ 是单位阵。接收端译码器由卷积神经网络构成,经过解码,得到输出向量 $\hat{\mathbf{x}} = f_d(\mathbf{y}; \phi_d)$ 。最后,整个模型联合进行端到端联合训练,损失函数为:

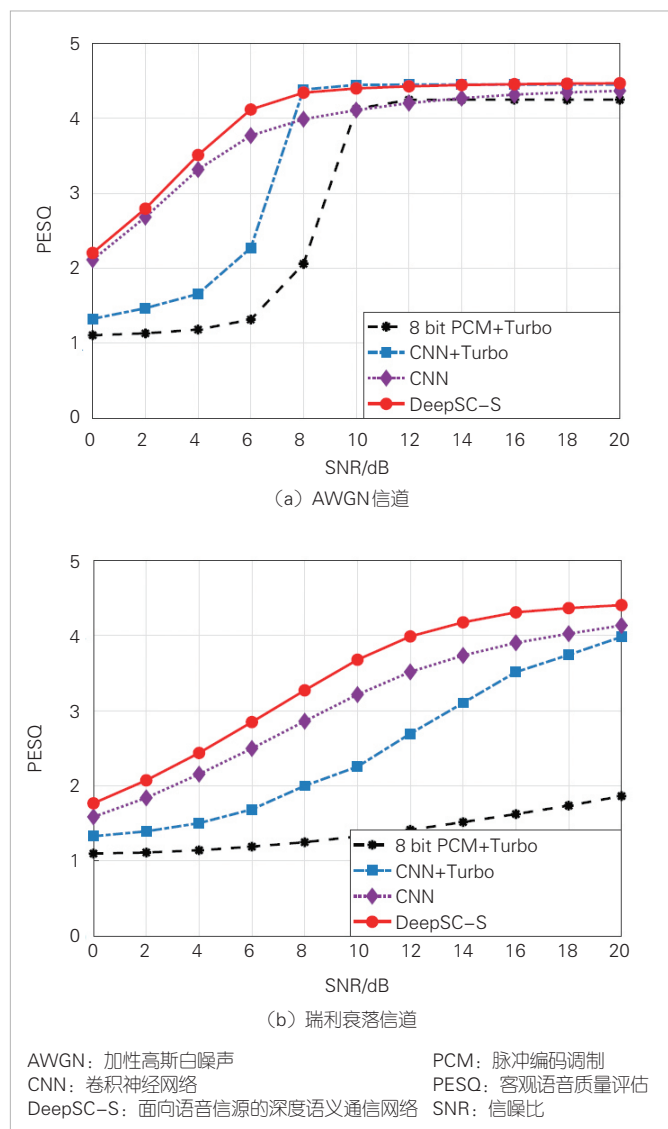
$$(\phi_e, \phi_d) = \arg \min_{\phi_e, \phi_d} \mathbb{E}_{\mathbf{s} \sim p(\mathbf{s})} \mathbb{E}_{\hat{\mathbf{s}} \sim p(\hat{\mathbf{s}}|\mathbf{s})} [d_s(\mathbf{s}, \hat{\mathbf{s}})], \quad (1)$$

其中, $d_s(\mathbf{s}, \hat{\mathbf{s}})$ 为信源和信宿语音信号的失真,通常采用均方误差(MSE)表示。

DeepSC-S的PESQ-SNR曲线如图2所示,在不同信道中



▲图1 语义通信的系统模型



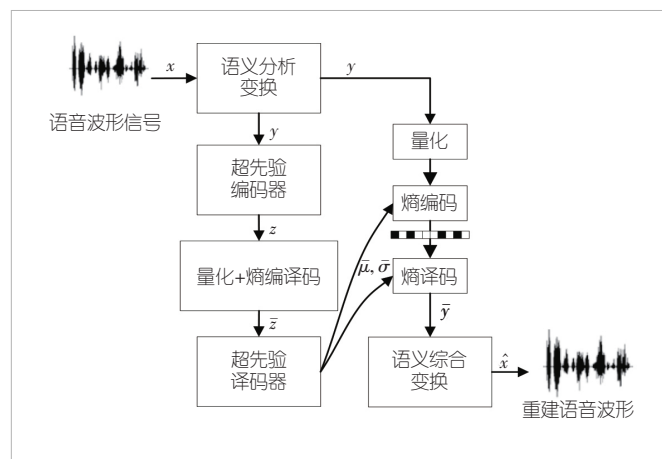
▲图2 DeepSC-S的PESQ-SNR曲线^[8]

的相同信噪比下，DeepSC-S（红色）的性能优于传统分离式编码PCM+Turbo码的性能。

然而，DeepSC-S对不同帧语音信号的编码速率是相同的，且提取的语义特征不明确，无法区分不同语音帧上的内容差异、语义差异，即无法从细粒度上刻画语义信息量的多少，因此难以获得接近于熵编码的压缩增益。这导致传输带宽较高，且编码效率较低。

为了进一步提高语音信号的语义传输效率，我们借鉴了非线性变换编码的思想^[19-21]，设计了基于非线性变换的语音语义信源编码方案^[22]和语义编码传输方案^[23]。下面我们简要介绍两种方案的基本原理。

基于信源变换编码的流程如图3所示。首先，语音经过分帧预加重等预处理后再经波形分析变换就可得到语义特征



▲图3 基于非线性变换的语音信源语义编码流程图

向量 $y = g_{a,\phi_g}(x)$ ，其中 ϕ_g 为波形分析网络的模型参数。其次，一对超先验编解码器将语义特征向量 y 变分建模为简单的高斯分布，从而使用算数编码器根据概率值对语义向量 y 进行熵编码。在收端，波形综合变换网络将恢复的语义特征向量 $\tilde{y}$ 重构为语音波形信号 $\hat{x} = g_{s,\phi_s}(\tilde{y})$ ，其中 ϕ_s 为波形综合网络的模型参数。

在端到端优化时，为解决量化不可导的问题，引入了取值范围为 $\left(-\frac{1}{2}, \frac{1}{2}\right)$ 的均匀噪声，得到语义特征向量的近似量化形式 $\tilde{y}$ ， $\tilde{y} = y + U\left(-\frac{1}{2}, \frac{1}{2}\right)$ 。类似地，超先验向量 z 也可近似量化为 $\tilde{z}$ ，从而得到了熵模型：

$$p(\tilde{y}|\tilde{z}; \psi_h) = \prod_i \left(\mathcal{N}(\mu_i, \sigma_i^2) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right), \quad (2)$$

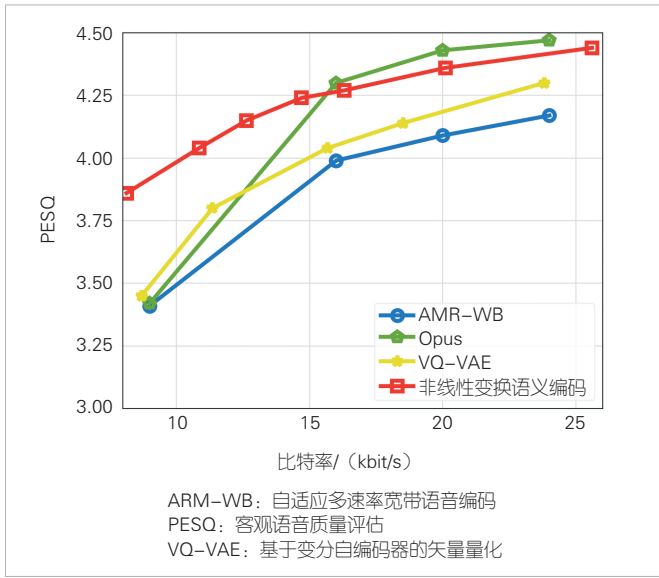
其中， $z = h_{a,\phi_h}(y)$ ， ϕ_h 为超先验编码器的参数集合， $\mathcal{N}(\mu, \sigma^2)$ 表示均值为 μ ，方差为 σ^2 的高斯分布， $*$ 表示卷积。

变换编码的优化问题是一个有损信源编码问题。若语音语义传输是面向人主观听觉感知质量的，那么优化语音的主观感知质量比波形失真更加重要。总的来说，该问题可以转变为最小化真实的联合后验概率与实际概率密度分布的KL散度，即：

$$\mathbb{E}_x D_{\text{KL}}[q_{\tilde{y}, \tilde{z}|x} \| p_{\tilde{y}, \tilde{z}|x}] = \mathbb{E}_x \mathbb{E}_{\tilde{y}, \tilde{z} \sim q_{\tilde{y}, \tilde{z}|x}} \left[-\log p(\tilde{y}|\tilde{z}) - \log p(z) - \log p(x|\tilde{y}) + \log q(\tilde{y}, \tilde{z}|x) \right], \quad (3)$$

其中，等式后大括号内第1、2项刻画了 y 和 z 的编码速率，第3项刻画了失真度量，第4项为求均值后为常数。

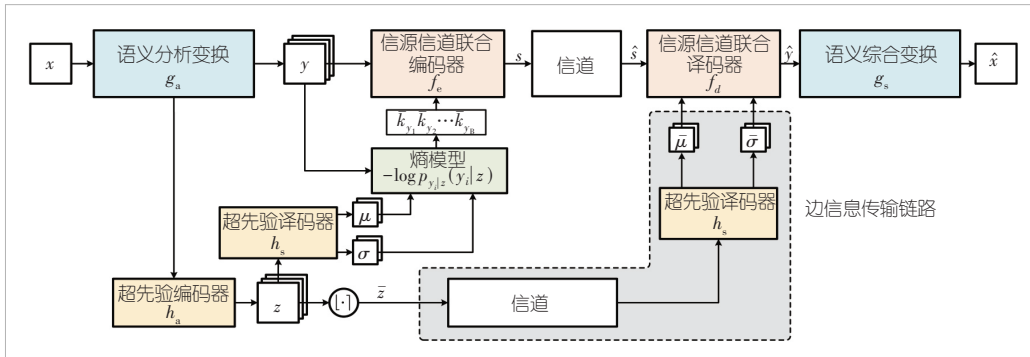
基于非线性变换的语义编码的率失真性能曲线如图4所示。我们可以发现，在低速率时，相较于自适应多速率宽带



▲图4 基于非线性变换的语音语义编码 PESQ 分数(取值为 1.0~4.5)

语音编码 (AMR-WB)^[23]以及基于变分自编码器的矢量量化 (VQ-VAE) 神经网络编码, 基于非线性变换的语义编码的感知质量优势明显; 在高速率时, PESQ 也与 Opus 性能相近。

在非线性变换语义信源编码的基础上, 语音信源的语义-信源信道联合变换编码不用对语义特征向量量化并熵编码, 而是将其映射为信道传输符号, 其流程如图 5 所示。一方面, 语义特征向量 $\mathbf{y}$ 需要通过超先验编解码器将其变分建模为高斯分布。此步骤是为了用语义特征向量的概率分布及概率值来衡量该语音帧的内容复杂度, 从而决定编码所需要的符号数。另一方面, 利用 Transformer^[24] 网络构成的信源信道联合编码器, 可以将 N 帧的语义特征向量 $\mathbf{y}^1, \dots, \mathbf{y}^N$ 映射为信道中传输符号序列 $\mathbf{s} = f_{e,\phi}(\mathbf{y}^1, \dots, \mathbf{y}^N)$ 。其中, ϕ_f 为信源信道联合编码器的参数集合。每个语义特征向量最终的编码长度由熵模型给出。



▲图5 基于非线性变换的语音语义信源信道联合编码系统结构

对于传输问题, 其优化函数转变为端到端的率失真优化问题, 速率此时定义为传输带宽, 训练的损失函数如公式 (4):

$$L = \mathbb{E}_{u \sim p(u)} \left\{ \lambda \left[\underbrace{-\eta \log P(\bar{s}|\bar{z})}_{k_s} - \underbrace{\frac{\log P(\bar{z})}{C_z}}_{k_z} \right] + d_s(u, v) \right\}, \quad (4)$$

其中, $s = g_a(u; \phi_g)$, $z = h_a(s; \phi_h)$, $\bar{s} = Q(s)$ 与 $\bar{z} = Q(z)$ 表示均匀量化, $v = g_s(r; \theta_g)$, k_s 是语义隐向量 $\mathbf{s}$ 的维度, k_z 是边信息 $\mathbf{z}$ 的维度, C_z 是相应的信道容量, 参数 η 控制语义信息编码速率, 拉格朗日因子 λ 控制发送信息速率 $k_s + k_z$ 与端到端的语义失真 d_s 之间的折中关系。

图 5 中基于 Transformer 网络的信源信道联合编/译码器包括 4 层 8 头注意力层, 维度为 1 024。语音数据来自于 TIMIT (英文语音数据集)^[25], 采样率为 16 kHz。图 6 给出了不同信道下基于非线性变换的语义编码传输方法的仿真结果。图 6 (a) 为 AWGN 信道在传输带宽为 4 kHz 的条件下, 不同编码传输方法在 PESQ 评估指标下的端到端率失真曲线图以及主观得分箱线图; 图 6 (b) 为 COST2100 衰落信道^[26]在传输带宽 10 kHz、平均信噪比为 2 dB 的条件下, 不同编码方案的 PESQ 评估指标下的端到端率失真曲线图及 MUSHRA 主观得分箱线图。其中, 红色线为非线性变换语义编码传输方法的曲线, 蓝色线和绿色线为传统分离式编码方法曲线。我们可以发现, 在相同主观/客观感知质量指标下, 相较于传统语音编码的 AMR-WB、Opus 编码器级联 5G LDPC 信道编码, 以及 DeepSC-S, 基于变换编码的语音语义传输所需带宽更少, 且在衰落信道中也表现出更加出色的鲁棒性。

2.2 生成式语音语义编码传输

生成式语音语义编码传输的研究工作目前还较少, HAN T. X. 等研究了上述所说的两种语音语义传输的范式^[27]。

在语音传输范式中, HAN T. X. 等将卷积神经网络级联双向长短期记忆网络 (LSTM) 作为语义编码器, 对语音帧的 40 组频率滤波器系数进行语义特征提取。同时, 他们还借鉴了 FastSpeech2^[28] 的做法, 通过对齐模块预测每帧对应的子单词, 并计算每个音

素的音高、功率等声学特征。在接收侧，文本特征、声学特征向量经对齐级联，由文本-语音综合器恢复语音频谱，最后声码器根据频谱重新生成语音波形，生成效果如图7所示。重建语音的MOS在AWGN信道中的各个信噪比下都优

于DeepSC-S。

生成式语音语义编码传输的优势在于所需带宽少，无须将语音波形信号的全部信息编码，利用人的听觉特性用少量的滤波器组在编码前就可将高采样率的波形信号转化为低采样率的声学特征信号，这样可以降低数据量。但根据数据处理定理，对语音信号的处理会导致部分信息的损失。如果接收侧关注编码语音特征以外的信息，则会导致语义传输的效果变差。另外，生成式的语义编码传输方法通过生成模型来重建语音波形，生成过程的复杂度、编译码器模型的训练难度都将给生成式语义编码传输方法的设计和实际应用带来挑战。

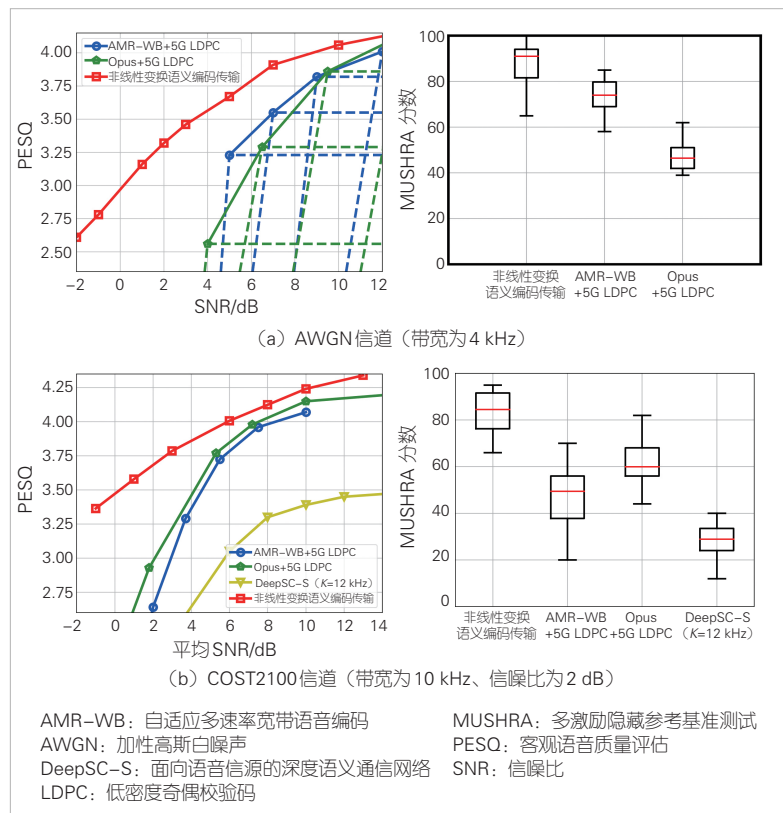
近两年，生成模型正在快速发展，人工智能生成内容(AIGC)的质量及自然度逼近人类。可以预见，基于生成式的语音编码传输也将对语音信号的语义编码设计带来重要影响。

3 结束语

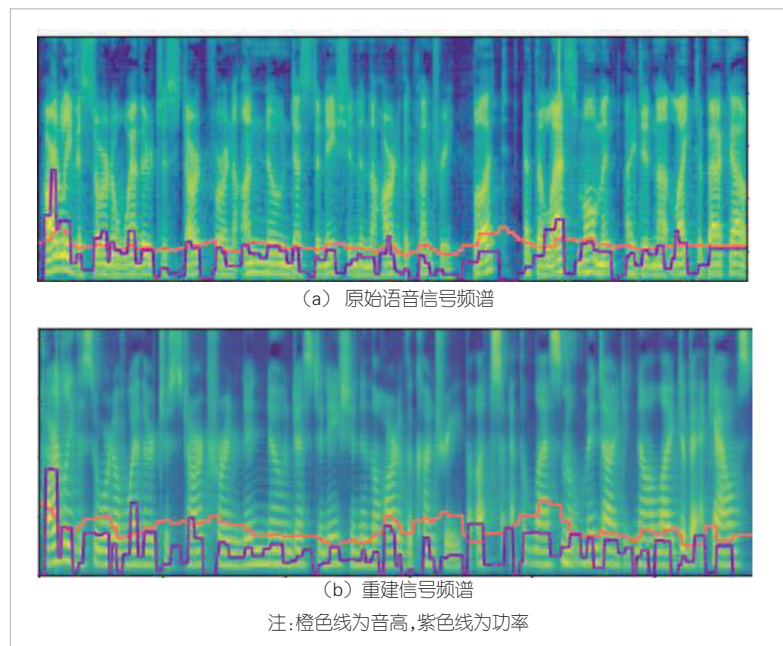
本文中，我们对语音信源的语义编码传输方法进行了整理与总结：首先回顾了现有的传统语音编码技术，分析了语音语义编码传输的优化目标，具体而言，在面向人的感知体验的情境下，优化目标主要为提高人的感知质量；而在面向文本任务的情境下，优化目标主要为提升文本识别正确率。其次，我们将现有的语音语义编码传输方法分为两大类：一类为直接对语音信号样值序列进行编码，一类为对传输语音信号的声学特征进行编码和传输，在接收侧通过生成式声码器重建语音。最后，我们展望了未来生成模型应用于语音语义编码传输的优势及挑战。

参考文献

- [1] SHANNON C E. A mathematical theory of communication [J]. The bell system technical journal, 1948, 27(3): 379-423. DOI: 10.1002/j.1538-7305.1948.tb01338.x
- [2] BURKS A, SHANNON C, WEAVER W. The mathematical theory of communication [M]. USA: The University of Illinois Press, 1951
- [3] ZHANG P, XU W, GAO H. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022, 8(1): 60-73. DOI: 10.1016/j.eng.2021.11.003
- [4] ZHANG P, XU X D, DONG C, et al. Intellicise communication system: model-driven semantic communications [J]. The journal of China universities of posts and telecommunications, 2022, 29(1): 2-12. DOI: 10.19682/j.cnki.1005-8885.2022.2002
- [5] NIU K, DAI J C, YAO S S, et al. A paradigm shift toward semantic communications [J]. IEEE communications magazine, 2022, 60(11): 113-119. DOI: 10.1109/MCOM.001.2200099



▲图6 基于非线性变换的语音语义编码传输重建语音质量性能



▲图7 生成式语音语义编码传输效果^[27]

- [6] DAI J C, ZHANG P, NIU K, et al. Communication beyond transmitting bits: semantics-guided source and channel coding [J]. IEEE wireless communications, 2022, (99): 1–8. DOI: 10.1109/MWC.017.2100705
- [7] 牛凯, 戴金晟, 张平, 等. 面向6G的语义通信 [J]. 移动通信, 2021, 45(4): 85–90. DOI: 10.3969/j.issn.1006-1010.2021.04.014
- [8] SHI G, XIAO Y, LI Y, et al. From semantic communication to semantic-aware networking: model, architecture, and open problems [J]. IEEE communications magazine, 2021, 59(8): 44–50. DOI: 10.1109/MCOM.001.2001239
- [9] XIE H Q, QIN Z J, LI G Y, et al. Deep learning enabled semantic communication systems [J]. IEEE transactions on signal processing, 2021, 69: 2663–2675
- [10] XU J L, TUNG T Y, AI B, et al. Deep joint source-channel coding for semantic communications [EB/OL]. [2023-01-20]. <https://arxiv.org/abs/2211.08747>
- [11] VAUDREUIL G, PARSONS G W. Toll quality voice – 32 kbit/s adaptive differential pulse code modulation (ADPCM) MIME sub-type registration [EB/OL]. (2004-06-01) [2023-01-23]. <https://www.semanticscholar.org/paper/Toll-Quality-Voice-32-kbit-s-Adaptive-Differential-Vaudreuil-Parsons/9236c8615a9b26f7c30a50c61178b7fb45b0652e>
- [12] BCKSTRM T. Speech coding: with code-excited linear prediction [M]. Germany: Springer, 2017
- [13] VALIN J, VOS K, TERRIBERRY T. Definition of the Opus audio codec [EB/OL]. (2012-09-01) [2023-01-25]. <https://www.semanticscholar.org/paper/Definition-of-the-Opus-Audio-Codec-Valin-Vos/ce6784cb8deaf3a05d3395e2624c02e8796a30e5>
- [14] OORD D V A, VINIYALS O, KAVUKCUOGLU K, et al. Neural discrete representation learning [EB/OL]. (2017-11-02) [2023-01-18]. <https://arxiv.org/abs/1711.00937>
- [15] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs [C]//Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal. IEEE, 2002: 749–752. DOI: 10.1109/ICASSP.2001.941023
- [16] CHINEN M, LIM F S C, SKOGLUND J, et al. ViSQOL v3: an open source production ready objective speech and audio metric [C]//Proceedings of 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2020: 1–6. DOI: 10.1109/QoMEX48832.2020.9123150
- [17] SCHOEFFLER M. Towards the next generation of web-based experiments: a case study assessing basic audio quality following the ITU-R recommendation BS. 1534 (MUSHRA) [EB/OL]. [2023-01-22]. <https://www.semanticscholar.org/paper/Towards-the-Next-Generation-of-Web-based-%3A-A-Case-.Schoeffler/908924842e5ec23f9dda290fdc1276606c4abbbd>
- [18] WENG Z Z, QIN Z J. Semantic communication systems for speech transmission [J]. IEEE journal on selected areas in communications, 2021, 39(8): 2434–2444. DOI: 10.1109/JSAC.2021.3087240
- [19] BALLÉ J, CHOU P A, MINNEN D, et al. Nonlinear transform coding [J]. IEEE journal of selected topics in signal processing, 2020, 15(2): 339–353. DOI: 10.1109/JSTSP.2020.3034501
- [20] DAI J C, WANG S X, TAN K L, 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
- [21] YAO S S, XIAO Z X, WANG S X, et al. Variational speech waveform compression to catalyze semantic communications [EB/OL]. [2023-01-22]. <https://arxiv.org/abs/2212.05294>
- [22] XIAO Z X, YAO S S, DAI J C, et al. Wireless deep speech semantic transmission [EB/OL]. (2022-11-04) [2023-01-18]. <https://deepai.org/publication/wireless-deep-speech-semantic-transmission>
- [23] BESSETTE B, SALAMI R, LEFEBVRE R, et al. The adaptive multirate wideband speech codec (AMR-WB) [J]. IEEE transactions on speech and audio processing, 2002, 10(8): 620–636. DOI: 10.1109/TSA.2002.804299
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2023-01-18]. <https://arxiv.org/abs/1706.03762>
- [25] GAROFOLO, JOHN S. TIMIT acoustic phonetic continuous speech corpus [EB/OL]. [2023-01-18]. <https://catalog.ldc.upenn.edu/LDC93S1>
- [26] LIU L F, OESTGES C, POUTANEN J, et al. The COST 2100 MIMO channel model [J]. IEEE wireless communications, 2012, 19(6): 92–99. DOI: 10.1109/MWC.2012.6393523
- [27] HAN T X, YANG Q Q, SHI Z G, et al. Semantic-preserved communication system for highly efficient speech transmission [EB/OL]. (2022-05-25) [2023-01-18]. <https://arxiv.org/abs/2205.12727v1>
- [28] REN Y, HU C X, TAN X, et al. FastSpeech 2: fast and high-quality end-to-end text to speech [EB/OL]. (2022-11-04) [2023-01-18]. <https://arxiv.org/abs/2006.04558>

作者简介



牛凯, 北京邮电大学教授、中国电子学会信息论分会副主任委员; 主要研究方向为信息论与极化码、智能信号处理; 所提极化码高性能编译码算法成为5G标准主流方案, 获中国电子学会科技奖自然科学一等奖; 发表论文200余篇, 申请国家发明专利80项。



姚圣时, 北京邮电大学在读博士研究生; 主要研究方向为信息论、信源信道编码和语义通信。



戴金晟, 北京邮电大学副教授、博士生导师; 从事信息通信、人工智能相关的科研和教学工作, 研究方向为智能语义通信、信源信道编码; 入选“中国科协青年人才托举工程”名单, 曾获中国电子学会科学技术奖自然科学一等奖; 发表论文60余篇, 授权国家发明专利23项, 出版专著1本。