

# 技术文献

## 增强型语音通话服务 (EVS) 编解码器

目前，由于音频带宽较窄及非语音信号处理水平较差等限制因素，通话服务往往无法提供高品质的声音体验。然而，语音和音频编码技术取得的最新进展将有助于大幅提升通话服务质量，通过提供全频带音频传输实现更贴近原声的声音体验，并改善语言清晰度及聆听舒适度。

最新通过标准化的增强型语音通话服务 (EVS) 编解码器是首个提供超宽带音频带宽，且在9.6kbps比特率下仍能极大改善语音通话质量的3GPP编解码器。同时，在处理音乐以及混合内容等信号方面，EVS的性能可与最新的音频编解码器相媲美。EVS的关键技术是在处理语音信号和音乐信号的专业编码模型之间进行灵活切换。这一编解码器由运营商、终端设备、基础设施和芯片提供商以及语音与音频编码方面的专家联合开发，其中包括爱立信、Fraunhofer集成电路研究所、华为技术有限公司、诺基亚公司、日本电信电话公司 (NTT)、日本NTT DOCOMO公司、法国电信 (ORANGE)、日本松下公司、高通公司、三星电子公司、VoiceAge公司及中兴通讯股份有限公司等。

本文旨在简要介绍EVS技术的通信系统蓝图。在强调主要设计制约因素和特征的同时，也包含了简要的技术见解，并展示与探讨了在标准化过程中，尤其是选择和测定阶段所进行的音质评价测试结果。

Fraunhofer Institute for  
Integrated Circuits IIS  
(Fraunhofer集成电路研究所)

所长：Albert Heuberger (博士兼教授)  
Am Wolfsmantel 33  
91058 Erlangen  
[www.iis.fraunhofer.de](http://www.iis.fraunhofer.de)

联系方式  
Matthias Rose  
电话：+49 9131 776-6175  
[matthias.rose@iis.fraunhofer.de](mailto:matthias.rose@iis.fraunhofer.de)

联系方式 (中国)  
Toni Fiedler  
电话：+86 138 1165 4675  
[china@iis.fraunhofer.de](mailto:china@iis.fraunhofer.de)

联系方式 (日本)  
Fahim Nawabi  
电话：+81 90-4077-7609  
[fahim.nawabi@iis.fraunhofer.de](mailto:fahim.nawabi@iis.fraunhofer.de)

联系方式 (美国)  
Fraunhofer美国公司  
数字媒体技术部\*  
电话：+1 408 573 9900  
[codecs@dmf.fraunhofer.org](mailto:codecs@dmf.fraunhofer.org)

\*Fraunhofer美国数字媒体技术部隶属于Fraunhofer美国分部，旨在推广和支持Fraunhofer集成电路研究所在美国推出的产品。

## 通信系统

语音通话质量与电视上播放的电影原声相比，标准手机通话的语音通常都非常模糊。这主要是受现有电话系统音频带宽的限制。图1显示了传统通信/广播电视系统的不同音频带宽容量及人耳听觉系统所能听到的声音频率范围。



图1：广播电视和通信系统的音频质量

老式电话系统主要支持窄带音频信号，音频带宽的频率仅到3.4kHz。高清语音服务(WB)则可支持宽带音频信号，音频带宽的频率达到7kHz。考虑到人耳听觉能力，实现高保真音质需要带宽频率支持到20kHz，而高清语音服务技术仍无法实现。因此，在高清语音技术之后又进一步出现了全高清语音技术，可提供超宽带和全频带的语音通话质量。超宽带技术的音频频谱为16kHz，而全频带的频率分量高达20kHz。

目前，固话服务可实现窄带或宽带语音质量。这些系统主要采用比特率为64kbit/s的G.711 [1] 或G.722 [2] 编解码技术。在移动通信环境下，窄带为默认质量标准；而现在出现了越来越多的宽带服务。用于窄带或宽带移动通信服务的编解码器通常采用AMR-NB [3] 和AMR-WB [4] 编解码技术，并且一般在12kbit/s比特率下运行。一些移动网络甚至可支持如23.85kbit/s等更高的AMR-WB速率，尽管与默认速率相比，在语音质量方面的改善仍较为有限。移动通信编解码器在处理语音信号方面进行了极大的优化，其结果导致在处理音乐等信号方面的能力则不尽如人意。

目前，用于远程或专用视频会议的通信系统已能实现全高清语音通话质量。针对这些系统的现行标准编解码器为AAC-(E)LD音频编解码器 [5]。这类编解码器可在24kbit/s至64kbit/s的比特率范围内运行，能够传输语音和音乐信号。

AAC-(E)LD音频编解码器也被应用于OTT服务。典型的OTT应用包括Skype、Facetime等，针对这些应用的IP数据包传输处理不涉及运营商的网络管理。

3GPP EVS编解码器 [6,7] 解决了移动电话和固话系统中存在的两大主要问题，即音频带宽较窄及非语音信号处理能力较差。同时，这一编解码器能在移动服务的常用比特率下运行。这为适用于各类网络通信质量方面的用户体验新标准奠定了基础，包括固话服务、移动电话服务和OTT服务等。在以下的章节中，本文将概述EVS设计可大幅提升服

务质量的主要技术要素。需要注意的是，除音频编码器之外，还有很多因素也会对端到端服务的用户体验产生重大影响。这些因素包括音频前端处理（包括回声消除、噪声抑制、自动增益控制、风噪声过滤和混响消除等）以及导致延迟抖动及数据丢包等的网络行为。

## EVS编解码器介绍

### 设计目标

EVS编解码器于2014年9月在3GPP通过标准化评估 [6]，具有多项功能实现前所未有的多功能性及通讯高效性。这一编解码器主要面向VoLTE服务，可实现3GPP设定的以下目标：

1. 提高窄带（EVS-NB）和宽带（EVS-WB）语音服务的质量和编码效率；
2. 引入超宽带（EVS-SWB）语音服务，提高通信质量；
3. 提高会话应用程序中混合内容和音乐信号的质量；
4. 具备防止数据丢包和延迟抖动的能力；
5. 针对AMR-WB编解码器 [20] 向下兼容。

如上文所述，本文将主要围绕设计目标2和3展开讨论。为确保信息完整性，本文后面部分也将讨论如何改善目标1中提出的传统窄带和宽带服务。除上述指出的质量改善外，EVS编解码器还具备通信系统所需的一系列功能，如语音活动检测（VAD）、非连续传输（DTX）、舒适噪声生成（CNG）或抖动缓冲管理（JBM）等。该编解码器可在5.9kbit/s至128 kbit/s宽比特率范围内运行，从而提供可适用于所有网络的比特率。参考 [6] 中介绍了EVS标准化中所有的设计限制因素。

### 技术概述

#### 编码模式

总体来看，音频编码可被分为以下两种模式：

- 语音编码：模拟人类声道
- 感知编码：利用人类听觉系统感知的限制因素

正如参考 [8] 中所述，AMR-NB和AMR-WB等高效的语音编码体系通常拥有三大主要组成部分，包括：（1）模拟人类声道的短期线性预测（LP）滤波器；（2）模拟声带激励信号周期的长期预测（LTP）滤波器；及（3）用于对语音信号非预测性内容进行编码的创新型代码本。

AAC [9] 等感知编码体系主要基于以下三大步骤：（1）进行时间/频率转换；（2）通过后续量化阶段减少不相关性，在这一阶段可使用心理声学模型信息来控制量化误差；（3）减少冗余，即在编码阶段，通过熵编码的方式使用代码表处理量化频谱系数和相关边信息。这就形成了适应于输入信号数据及人类感知特点的受源代码控制的编解码器。

一般来说，在低比特率情况下，语音编码方式可更好地处理纯粹、清晰的语音信号，而感知编码方式可更好地处理音乐等一般声音内容，并实现可感知的、明晰的声音质量。

首个结合了这两大编码方式的编解码器为语音/音频统合编解码器 (USAC) [8]。USAC 算法延迟超过了100毫秒，这是双向通信应用不可接受的。然而，得益于USAC卓越的编码性能，这一统合的编码方法已经被采用并且取得了进一步的优化，以满足EVS极具挑战性的要求。

### 低延迟语音/音频切换编码

在32毫秒低算法延迟下，EVS编解码器是首个在语音和音频压缩之间部署基于内容且具备即时切换功能的移动通信编解码器，极大地改善了针对音乐信号等一般内容的编码性能。

该语音编码器是改进型代数码激励线性预测 (ACELP)，还采用了适合不同语音类别的线性预测模式。对于音频信号编码，则采用频域 (MDCT) 编码方式，并特别关注低延迟/低比特率情况下的频域编码效率，从而在语音处理器和音频处理器之间实现无缝可靠的切换。图2展示了EVS编解码器和解码器的高级框图。

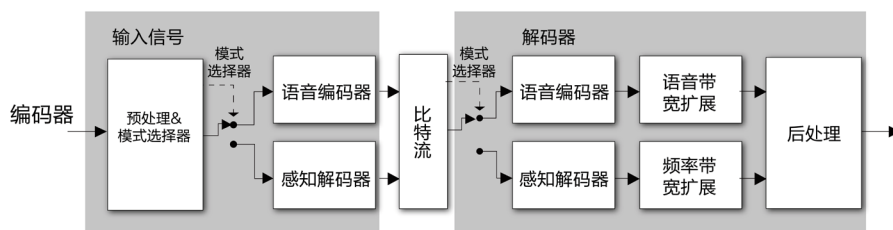


图2：EVS编解码器结构

### 超宽带编码及更先进的技术

EVS可支持超宽带甚至全频带服务，从而克服当前通话中声音不清晰的问题。从技术上来看，EVS可通过扩展带宽实现这一目标。根据是否采用语音或音频模式，可部署时域带宽扩展 (TBE) 技术或一体化频域解决方案。后者可提供多个子模式，如可处理传统音乐信号的谐波模型编码模式等。EVS是首个可提供通过不同方式优化的带宽扩展方式的编解码器，而这样的带宽扩展方式通常基于源代码控制来进行使用或切换。基于专用的内容优化方式，即便在比特率非常低的情况下也能够提供非常自然、清晰的音质。

## 性能评估

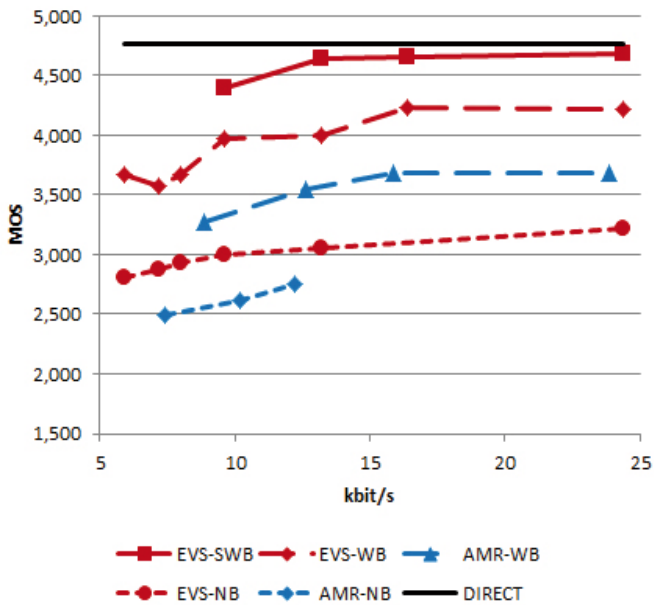


图3：3GPP EVS编解码器针对多带宽清晰语音的鉴定测试结果

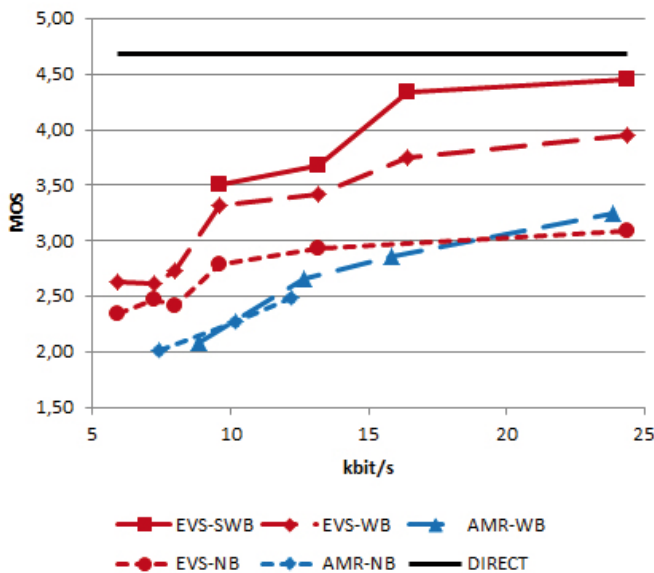


图4：3GPP EVS编解码器针对多带宽混合内容和音乐信号的鉴定测试结果

目前，3GPP已进行了大量测试来验证EVS在不同情况下以及处理不同内容 [11] 时所表现出的性能，其中包括根据P.800 DCR方法 [12] 进行的多频带测试。图3和图4分别显示了（DOMS评分表）中清晰语音（英语）及混合内容和音乐信号的音质级别。结果显示了在不同比特率下移动蜂窝业务常用的窄带、宽带和超宽带下的不同音质。下面将对这些结果进行讨论：

- 对于（超宽带模式下运行的）EVS编解码器，比特率为9.6kbit/s时纯语音音质评价已非常高，大大超过了AMR-WB在23.85kbits/s比特率下的音质，同时这一

数值将随比特率的增加而增加。从13.2kbits/s开始，EVS-SWB的语音音质已经接近“直接来源”（原音）的音质。

- （在超宽带模式下运行的）EVS编解码器在处理混合内容和音乐信号时的性能大大优于AMR-WB编解码器。在相同比特率下，其得分基本都比AMR-WB编解码器高出1.2个平均意见得分（MOS）。在比特率为24.4kbit/s的情况下，EVS编解码器在处理混合内容和音乐信号时所呈现的音质可接近“直接来源”（原音）的音质。
- 甚至在宽带模式（当前仅支持宽带模式的设备或服务）下，EVS编解码器在比特率为9.6kbit/s时所能呈现的音质超过了AMR-WB编解码器在比特率为23.85kbits/s时所能实现的音质。此外，在同一比特率（24.4kbit/s）下，EVS编解码器在处理纯语音以及音乐信号时所呈现的音质要远远好于AMR-WB编解码器。
- 初看上去，在同等比特率下，即便使用双倍音频带宽，AMR-WB编解码器在处理混合内容和音乐信号时呈现的音质也比不上AMR-NB编解码器，这让人觉得出乎意料。但EVS编解码器已经解决了这一问题。
- 在处理窄带输入信号时，EVS编解码器处理纯语音及混合内容和音乐信号时实现的音质要优于AMR-NB编解码器。这一模式在连接至如固话网络等窄带网络时非常有用。

众所周知，由于语言和所选内容的不同，测试的结果及结果分析也各有不同。然而，在3GPP进行筛选的阶段，EVS编解码器已经过10种语言、6种不同背景噪声及各种音乐内容的测试，展现出了其卓越的性能并大大改善了早期标准。这些测试结果以及之后对EVS编解码器进行的进一步性能鉴定结果已在3GPP发布的技术报告（TR）26.953 [11]中公布。

## 应用

自第四代移动网络标准LTE问世以来，蜂窝电话网络开始向基于IP的传输方式转变。LTE标准以发展成熟的GSM和UMTS标准为基础，可提供全IP架构和实现低延迟。LTE需要部署全IP语音业务或VoLTE业务，而这也促使所有语音服务采用IP网络进行传输，最终淘汰基于GSM、UMTS和CDMA等网络的传统转换服务。

在全高清语音技术的推动下，服务提供商可摆脱这些传统服务的限制，包括音频带宽方面的较大限制及需要使用以处理语音信号为主的编解码器等。由于VoLTE可在管理有序的网络中提供优质服务（QoS），EVS有望在音质、稳定性和服务可用性方面超越Skype或Viber等OTT服务。因此，移动运营商将可能在语音服务领域收复失地。

得益于杰出的防错能力 [10]，EVS非常适用于Wi-Fi语音等最佳网络，而且未来还可用于3G/电路交换系统。

## 结语

凭借无法比拟的语音和音频质量等多项新特性，符合3GPP最新标准的EVS编解码器可作为一种效率最高、功能最全的编解码器适用于各种网络（尤其是蜂窝LTE网络和Wi-Fi语音网络等），以实现高质量通信。即使在移动通信服务中，EVS编解码器也能呈现接近原音的音质，从而为用户带来全新的体验。因此，即将推出的EVS编解码器可为移动通信运营商及其客户带来长远的好处。

## 参考资料

- [1] ITU-T Rec. G.711, "Pulse code modulation of (PCM) of voice frequencies"
- [2] ITU-T Rec. G.722, "7 kHz audio-coding within 64 kbit/s"
- [3] K. Järvinen. "Standardisation of the Adaptive Multirate Codec," Proc. EUSIPCO, Sept. 2000.
- [4] B. Bessette, et al., "The adaptive multi-rate wideband speech codec (AMR-WB)," IEEE Trans. on Speech and Audio Processing, vol. 10, no. 8, pp. 620-636, Nov. 2002.
- [5] M. Schnell, et al., "MPEG-4 Enhanced Low Delay AAC - a new standard for high quality communication", 125th AES Convention, Oct. 2008.
- [6] S. Bruhn, et al., "Standardization of the new EVS Codec", Proc. ICASSP, Apr. 2015.
- [7] Martin Dietz et al., "Overview of the EVS Codec Architecture," Proc. ICASSP, April 2015.
- [8] M. Neuendorf, et al.: "Unified Speech and Audio Coding Scheme for high quality at low bitrates", Proc. ICASSP, Apr. 2009
- [9] M. Bosi, et al., "ISO/IEC MPEG-2 Advanced Audio Coding", paper 4382, 101st AES Convention, Nov. 1996.
- [10] V. Atti, et al., "Improved error resilience for VOLTE and VOIP with 3GPP EVS channel aware coding", Proc. ICASSP, Apr. 2015.
- [11] 3GPP TR 26.952, "Universal Mobile Telecomm-unications System (UMTS); LTE; Codec for Enhanced Voice Services (EVS); Performance characterization," <http://www.3gpp.org/DynaReport/26952.htm>
- [12] ITU-T Rec. P.800, "Methods for Subjective Determination of Transmission Quality," Aug. 1996.



本文档的内容“按原样”提供，不带有任何明示或暗示的担保或条件，包括但不限于适销性、针对特定用途的默认担保等。

本文档内容的版权归FRAUNHOFER-GESELLSCHAFT所有，并可在不另行通知的情况下随时进行修改和/或更新。因此，不允许转售本文全部或部分内容或将其用于商业用途，或将其用于其他产品。

版权所有©FRAUNHOFER-GESELLSCHAFT，2015年7月。

## 关于Fraunhofer IIS

媒体世界日新月异，而Fraunhofer IIS始终主导着最先进音频技术的发展。在过去25年的时间里，Fraunhofer集成电路研究所（IIS）音频和多媒体部门始终专注于数字音频技术的研发。作为mp3技术的发明者和AAC（高级音频编码）技术的联合开发者，以及未来广播电视领域的音频技术开拓者，Fraunhofer IIS将音频技术创新变成了现实。

目前，Fraunhofer IIS最引人关注的全新技术包括：提供虚拟环绕音效的Fraunhofer Cingo、打造车载3D音频的Fraunhofer Symphoria、为电话 / 网络通话带来CD般音质的AAC-ELD，以及让电视观众能够随意调整电视对白音量的对白增强技术。

Fraunhofer IIS技术已应用于全球超过80亿台设备中，并向超过1000家公司授权使用其音频编解码软件和特定应用的自定义软件。如今，Fraunhofer IIS的mp3和AAC音频编解码技术在移动多媒体系统中无处不在。

Fraunhofer IIS总部位于德国埃尔兰根，是弗劳恩霍夫应用研究促进协会（Fraunhofer-Gesellschaft）的一个分支部门。Fraunhofer-Gesellschaft由66个研究院组成，在全球拥有超过24000名员工，是目前欧洲最大的研究机构。

欲了解更多信息，请联系Matthias Rose ( [matthias.rose@iis.fraunhofer.de](mailto:matthias.rose@iis.fraunhofer.de) )，或访问以下网站：  
[www.iis.fraunhofer.de/audiochina](http://www.iis.fraunhofer.de/audiochina).