

语音编码

维基百科，自由的百科全书

语音编码（Speech coding），是一种包含说话的数位音讯信号资料压缩应用。语音是人类交流资讯最自然、最有效、最方便的手段，因此大多数通信系统都要包含语音传输的功能。随着科技的发展，虽然影像、数据等非语音信息在资讯传递中所占的比例越来越大，但是有效地传递语音信号仍是众多通信系统必备的功能之一。

语音通信最早可以追溯到1876年贝尔发明的电话机，该装置首次利用声电、电声转换技术实现了远距离的语音讯号传输。早期的语音通讯装置采用的是模拟语音通讯，但它在可靠性、抗干扰能力、快速交换等方面存在很多问题。自从1960年代集成电路出现以后，通讯通道和通讯终端开始从模拟系统演变为全数位系统。与模拟通讯相比，数位通讯在可靠性和抗干扰能力等方面具有很大优势，但不足之处是占用的频带较宽。为了节省数位语音信号传输所需要的带宽，语音编码技术开始得到深入研究。

语音编码属于信源编码范畴，自从1930年代末提出脉冲编码调变（PCM）原理以及声码器（Vocoder）概念后，语音编码一直沿着这两个方向发展，它们也可以称为语音讯号的波形编码与声码化编码或称为非参数编码与参数编码，参数编码有时也称为模型编码。

随着电脑的发展以及大规模、超大规模集成电路的出现，语音编码技术取得了很多突破性的进展，研究出了很多实用的语音编码方案，应用在不同的语音通讯系统中，并在不断地研究、改进中得以完善，逐渐形成国际性或者地区性的语音编码标准。

目录

语音编码的应用

语音编码的分类

按编码分类

按编码速率分类

按编码语音带宽分类

按编码语音带宽分类

语音编码的硬件与系统

语音编码的发展趋势

参考资料

语音编码的应用

几乎语音编码领域的所有研究都是受应用驱动的，由于数字化语音的传输和存储，在可靠性、抗干扰、易保密等方面都远远胜过模拟语音，从1950年代以来，在通讯系统中数位语音所占的百分比不断增加。目前，在整合服务数位网络(Integrated Services Digital Network，ISDN)、卫星通讯、行动通讯、微波接力通讯等系统无一例外采用的都是数字化语音传输和存储。近几年，被称为“最后一个数字化电子行业”的无线对讲机领域也升始研究应用数字化语音编码方案了。

行动通讯系统和互联网(internet)是语音编码技术的两个重要应用领域。行动通讯的发展日新月异，除第一代行动通信系统采用模拟式语音传输外，从第二代移动通信系统开始，全部采用数位语音编码方式。目前广泛使用的“全球通”等第二代数位蜂窝移动通信系统采用了 13 kb/s 的 RPE - LTP (Regular Pulse Excitation - Long Term Prediction)语音编码技术。北美第二代数位移动电话(CDMA)采用的是

Qualcomm (高通)公司的 QCELP(Qualcomm Code Excited Linear Prediction) 语音编码预测。QCELP 语音编码预测不仅可以工作在 4/4.8/8/9.6 kb/s 等固定速率上, 还可以变速率地工作在 0.8~9.6 kb/s 之间。在第三代行动通讯系统中, 变速率语音编码技术得到了广泛应用。TD-SCDMA 标准采用了 AMR (Adaptive Multi-rate)语音编码技术, WCDMA标准主要采用了 AMR 语音编码技术, cdma2000 主要采用了 QCELP 或 EVRC (Enhanced Variable Rate Codec) 语音编码技术。近些年, 由于成本低廉, 网络电话得到深入研究, 目前世界多个标准组织和工业实体提出了很多语音编码方案。其中包括国际电信联盟的 G.711 (速率为64 kb/s), G.723.1 (速率为 5.3 kb/s 或 6.3 kb/s), G.729A (速率为 8 kb/s) 编码方案。GIPS (Global IP Sound)公司、Skype 公司等业界企业也有自己的编码方案, 如 iLBC、SILK 等编码算法等。考虑到语音数据封包包在网络环境中传输的特殊性, 以及不同通讯网络的结构性等因素, 嵌入式多速率语音编码算法也在近几年成为一个重要的研究领域。随着网络带宽的不断增加, 终端处理能力的不断增强, 用户对话品质要求的提高, 宽频、超宽频、全带宽语音编码技术得到了广泛的研究, 很多传统的窄频语音编码技术得到深入研究并扩充到宽频编码版本上。

除了行动通讯系统和互联网外, 语音压缩编码技术还广泛应用于保密通讯、卫星通信以及水下通信等领域。在这些应用领域中, 传输带宽通常较窄, 超低速率语音压缩编码算法的研究成为一个焦点。

语音编码的分类

针对不同的应用场合、不同的语音来源以及对语音编码目的的不同, 可作如下的分类:编码、编码速率、编码语音带宽以及编码讯号环境。

按编码分类

按编码方法分类, 语音编码可以分为波形编码、参数编码以及混合编码。

- **波形编码:** 语音信号的波形编码图使重建语音信号的波形维持原语音信号的波形形状。该编码算法通常将语音信号作为一般的波形信号来处理, 其有适应能力强、话音质量好的优点, 但需要的编码速率高。脉冲编码调变(PCM)、自适应增量调制(ADM)、Adaptive(ADPCM)、自适应预测编码(APC)、自适应子带编码(ASBC)、自适应变换编码(ATC) 等都属于该类编码器。它们分别在 16 ~ 64 kb/s 的编码速率上给出高的编码质量, 当速率进一步降低时, 其性能下降较快。
- **参数编码:** 语音信号的特征参数通过对语音信号特征参数的提取及编码, 力求使重建语音讯号具有尽可能高的清晰度, 即保持原语音的语义, 而重建讯号的波形可能同原始语音讯号有较大的差别。此类编码器的优点是编码速率低, 可以达到 2.4 kb/s 甚至更低, 主要问题是合成语音品质较差, 自然度较低, 连熟人都一定不能听出讲话的人是谁。另外, 该类编码器对环境讯杂比较敏感, 需要在安静的环境才能给出较高的清晰度, 对通道误码也比较敏感。通道声码器、共振峰声码器以及目前广泛使用的线性预测声码器都是典型的参数编码。
- **混合编码:** 混合编码采用参数编码和波形编码的混合编码形式。该类算法的共同特点是: 先进行线性预测分析, 去掉语音信号的短时相关性, 然后利用合成分析法及感觉加权均方误差最小原则分析出合适的替代冗余讯号的激励源, 最后对LP参数和激励讯号源进行编码和传送。由于混合编码方案的激励模型和误差计算与时域波形相连系, 使合成语音具有较强的追踪输入语音变化的性能, 从而能够改善合成语音的质量以及抗噪声的能力。

按编码速率分类

按语音编码速率分类是根据语音讯号压缩编码后进行存储或传输所需要的数据速率进行分类的一种方法。例如, 按照编码后的数据速率, 可以将窄频语音编码分为5类:

- **高速率语音编码:**编码速率为32kb/s以上
- **中高速率语音编码:**编码速率为16 ~ 32kb/s

- 中速率语音编码:编码速率为4.8 ~ 16 kb/s
- 低速率语音编码:编码速率为1.2 ~ 4.8 kb/s
- 超低速率或极低速率语音编码, 其编码速率低于1.2 kb/s

目前, 语音编码的国际标准主要集中在中速率以上, 低速率以及超低速率语音编码算法的研究仍然是一个热门领域。

按编码语音带宽分类

通常将人耳可以听到的频率在 20 Hz~20 kHz 的声波称为音频讯号,标准的话路带宽语音信号是300~3400 Hz, 若加上少量保护频带, 一般意义上说的标准话路频带宽度应该是4 KHz, 根据取样定理, 带宽为 4 KHz 的语音讯号的取样频率应为 8 kHz, 量化精度一般在 16 位左右。通常所说的窄频语音编码指的就是这种频带范围内的语音讯号编码。

随着网络传输带宽的不断增加, 语音编码的另一研究领域也开始引起广泛重视, 这就是宽频语音编码。宽频语音编码的频率范围通常在 50~7000 Hz, 加上保护带宽, 通常为 8 kHz, 取样频率一般为 16 kHz。国际电信联盟公布的自适应多速率带宽(AMR-WB)语音编码器的标准为 G.722.2 协议, 其频率带宽就在 50~7000 Hz, 取样频率为16 kHz, 属于带宽语音编码的范畴。

在多媒体领域中, 主要处理音频讯号, 包括音乐、语音、风声、雨声、鸟鸣声、机器声等。音频编码主要是针对频率范围较宽的音频讯号进行的编码, 主要应用于数位广播和数字电视广播、消费电子产品等。典型的编码有 MPEG-1 / MPEG-2 的 layer 1、2、3 和 MPEG-4 AAC 的音频编码等。

按编码语音带宽分类

语音讯号具有短时平稳性, 通常假定语音讯号在 5~30 ms 之间的时段内, 某些物理特征参数是可以视为不变的, 所以语音讯号的参数编码及混合编码通常以语音帧为单位进行。按照被编码的语音讯号所在环境可以将编码算法分为时域编码和频域编码。

- 时域编码: 时域编码是指对语音的时域讯号进行编码, 如 PCM 编码等。
- 频域编码: 频域编码是指对语音的频域讯号进行编码, 次频带编码(SBC)就是一种频域编码方法。

语音编码的硬件与系统

绝大多数语音编码系统需要时时工作, 对于系统的硬件环境提出了较高的要求。随着语音编码算法的日益复杂, 许多算法的运算量达到几十个 Million Instructions Per Second (MIPS)。但相应的, 近 20 年来, 随着微电子行业的迅速发展, 数字信号处理器(DSP) 技术也产生了巨大的突破, 这就为语音编码算法的实用化铺平了道路。在DSP领域最成功的企业是美国的德州仪器公司(Texas Instruments, TI 公司), 自从 1982 年推出 TMS320 系列的第一款定点 DSP TMS3210 以来, 其 DSP 已经发展了若干系列, 处理能力最高可达 GHz。除 TI 公司外, 美国 AD 公司、Motorola、NEC 等公司的 DSP 也各有特点。具体选择哪一种硬件平台来实现语音编码, 要根据应用的实际需求综合考虑, 灵活决定。

根据 DSP 所能处理的数据格式, 可以将其分为整数DSP和浮点数DSP。

- 整数DSP: 整数DSP速度较快且价格便宜, 但程式撰写困难, 精度较低。
- 浮点数DSP: 浮点数DSP精度高, 但其价格通常较贵。

在为某种语音编码算法选择合适的 DSP 时, 需要综合考虑运算量、存储量以及性价比等众多因素。在语音编码领域, TI的TMS320C54X 及 TMS320C55X 系列DSP得到了广泛应用, 这两款DSP均为整数 DSP。

语音编码的发展趋势

经过近几十年的发展，语音编码领域涌现了很多的成果。目前在语音编码方面有几个热门的领域：低速率语音编码、频带语音编码、变速率语音编码以及嵌入式语音编码。

ITU 已制定的语音编码标准主要集中在中速率以上。如何在低速率甚至超低速率上获得较高的语音品质，仍然是语音编码领域的焦点。在低速率语音编码领域，目前研究较多的模型包括混合激励线性预测模型、正弦激励线性预测模型、正弦变换编码、多频激励编码以及波形插值编码等。

除了致力于制定更低速率的语音编码标准外，ITU正在讨论已制定的语音编码标准近进行全带宽的拓展，使其能够适应语音的应用，例如 AMR 从窄频带 AMR-NB 到宽频 AMR-WB 再到超宽频 AMR-WB+，G.729 全频带扩展到 G.729.1 等。ISO/MPEG 从标准化组织目前正在制定下一代音频编 / 解码标准 USAC (Unified Speech and Audio Coding)，即语音 / 乐音联合编解码器。USAC 可以对任意比例混合的语音 / 音乐讯号进行编 / 解码，同时，无论是语音成分还是音乐成分，其编码性能至少不亚于当前最好的专业语音编码器或乐音编码器的编码性能。USAC 标准具有广泛的应用前景，可以广泛地应用于全球数字广播，第三、四代行动通讯等领域。

随着行动通讯的发展，尤其是第三代行动通讯的发展，对于语音压缩编码算法提出了更高的要求，不但要求编码速率较低以增加系统容量，而且要求合成语音音值较高以确保通话品质。用传统的压缩编码方式，很难同时满足这两个要求，在这种情形下，提出了变速率语音压缩编码的方法。在行动通讯系统中采用变速率语音压缩编码，可以根据需要动态地调整编码速率，在合成语音质量和系统容量中取得灵活的折衷，最大限度地发挥系统的效能。1990年代，随着 CDMA 行动通讯系统的快速发展，各种通讯组织相继推出了不同的变速率语音编码技术标准，主要包括 QCELP、EVRC、AMR 和 SMV 4 种算法标准。

随着网络技术的不断发展，以网络为载体的语音讯号传输得到了广泛研究。在现实的网络条件下，由于网络拥塞、过延时和缓冲器溢出等原因，决定了网络传输中必然经常出现封包丢失等问题，这对语音讯号的实时传输具有很大影响，传统的编码器因为本身的特点，包丢失会导致合成语音品质严重下降。嵌入式语音编码算法从本质上来说也是一种变速率语音编码算法，其码流的分布为嵌入式结构，核心码流能够保证基本的合成语音质量，外围层的码流不断提高合成语音的品质，收到的比特流越多，合成语音的品质就越好，嵌入式语音编码的这种结构特别适用互联网上的语音传输，因此得到了各国学者的重视与研究。2006年，ITU制定了新的带宽嵌入式语音编 / 解码方案 G.729.1，并制定了新一代宽频嵌入式(EV-VBR)语音编码提案，具有5层嵌入式编码结构，最终由 VOICEAGE 和 Nokia提交的编码器成为基线编码器，经这与其他竞争者的技术融合后，在2008年成为新一代嵌入式语音编码标准 G.718。由于其没有兼容G.729的要求，相应速度上具有比 G.729.1 更高的语音通讯品质。

参考资料

- 李晔, 崔慧娟, 唐昆 等编著, 数字语音编码技术 (Digital Speech Coding Technologies)

取自 ["https://zh.wikipedia.org/w/index.php?title=語音編碼&oldid=65106168"](https://zh.wikipedia.org/w/index.php?title=語音編碼&oldid=65106168)

本页面最后修订于2021年4月8日 (星期四) 01:33。

本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用。（请参阅使用条款）
Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。
维基媒体基金会是按美国国内税收法501(c)(3)登记的非营利慈善机构。