

团 体 标 准

T/AI 109.10—2024

信息技术 智能媒体编码 第 10 部分： 实时通信语音编码

Information technology – Intelligent media coding –

Part 10: Real-time Communication (RTC) Speech Coding

2024 – 10 – 14 发布

2024 – 10 – 14 实施

中关村视听产业技术创新联盟 发布



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目 次

前言	II
引言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 串行流语法规则	3
6 约定	5
6.1 概述	5
6.2 算术运算符	5
6.3 逻辑运算符	5
6.4 关系运算符	5
6.5 位运算符	5
6.6 赋值	6
6.7 助记符	6
7 实时通信语音编解码框架	6
8 实时通信语音解码	7
8.1 概述	7
8.2 AVS3P10 位流解析	8
8.3 宽带特征向量信息和宽带特征向量	20
8.4 解码神经网络	28
8.5 MCR 立体声解码	38
8.6 带宽扩展解码	40
附录 A（规范性） ARSTF 语法及语义	51
附录 B（规范性） AVS3P10 解码神经网络信息	54
附录 C（资料性） AVS3P10 参考编码器	55
附录 D（资料性） 丢包隐藏	74

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是T/AI 109《信息技术 智能媒体编码》的第10部分：实时通信语音编码。T/AI 109已经发布了如下部分：

- 第2部分：视频；
- 第3部分：沉浸式音频；
- 第4部分：视频符合性测试；
- 第6部分：智能媒体格式。

本文件由数字音视频编解码技术标准工作组提出。

本文件由中关村视听产业技术创新联盟归口。

本文件起草单位：清华大学、腾讯科技（深圳）有限公司、中国电子技术标准化研究院、华为技术有限公司、北京理工大学、北京全景声信息科技有限公司、清华大学天津电子信息研究院、广东博华超高清创新中心有限公司。

本文件主要起草人：窦维蓓、肖玮、张亚军、商世东、李婧欣、朱博成、王喆、王晶、许舒敏、潘兴德、张伟民、高文、黄铁军、赵海英、崔晓冉、汪邦虎、李志永、袁俊崧。

引 言

T/AI 109 旨在规定智能媒体的编码、封装与传输规范，旨在为异构网络中的智能媒体数据提供编码、封装和传输服务，拟由 10 个部分构成。

——第1部分：系统。目的在于确立沉浸媒体虚拟现实数据的系统信令。

——第2部分：视频。目的在于确立智能媒体高效视频压缩方法的解码过程。

——第3部分：沉浸式音频。目的在于确立适用沉浸式音频高效压缩方法的解码过程。

——第4部分：视频符合性测试。目的在于确立测试验证编码位流和解码器是否满足T/AI 109所规定的要求。

——第5部分：参考软件。目的在于确立对智能媒体高效视频压缩方法和沉浸式音频高效压缩方法的参考实现。

——第6部分：智能媒体格式。目的在于确立异构网络中智能媒体编码数据的存储格式和传输信令。

——第7部分：音频封装与传输。目的在于确立异构网络中的沉浸式音频数据提供封装和传输。

——第10部分：实时通信语音编码。目的在于确立面向实时通信的语音编码位流和解码过程。

本文件的发布机构提请注意，声明符合本文件时，可能涉及到如下专利：

序号	专利申请号/专利号	专利名称	专利申请人/专利权人	标准条款涉及专利的（章、条编号）
1	202210681816.X	音频编码及解码方法、装置、设备、介质及程序产品	腾讯科技（深圳）有限公司	8.1 至 8.4, 8.6 附录 A、B
2	202210681365.X	音频处理方法、装置、设备、存储介质及计算机程序产品	腾讯科技（深圳）有限公司	8.1 至 8.4, 8.6 附录 A、B
3	202210676984.X	音频解码、编码方法、装置、电子设备及存储介质	腾讯科技（深圳）有限公司	8.1 至 8.4, 8.6 附录 A、B
4	200810106460.7	立体声信号编解码方法、装置及编解码系统	华为技术有限公司、清华大学	8.5

本文件的发布机构对于该专利的真实性、有效性和范围无任何立场。

该专利持有人已向本文件的发布机构保证，他愿意同任何申请人在合理且无歧视的条款和条件下，就专利授权许可进行谈判。该专利持有人的声明已在本文件的发布机构备案，相关信息可以通过以下联系方式获得：

专利持有人：腾讯科技（深圳）有限公司

地址：深圳市南山区高新区科技中一路腾讯大厦 35 层

专利持有人：华为技术有限公司

地址：深圳市龙岗区坂田华为基地

专利持有人：清华大学

地址：北京市海淀区清华大学电子工程馆 4-102

请注意除上述专利外，本文件的某些内容仍可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

联系人：黄铁军（数字音视频编解码技术标准工作组秘书长）

通讯地址：北京大学理科 2 号楼 2641 室

邮政编码：100871

T/AI 109.10—2024

电子邮件: tjhuang@pku.edu.cn

电话: +8610-62756172

传真: +8610-62751638

网址: <http://www.avso.org.cn>

T/AI 109.10—2024

信息技术 智能媒体编码 第10部分：实时通信语音编码

1 范围

本文件规定了面向实时通信的语音编码位流和解码过程。

本文件适用于在线会议、协同办公、社交、互动娱乐等领域低码率高质量语音编解码器的开发和应用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注明日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 5271.1 信息技术 词汇 第1部分：基本术语

GB/T 33475.3 信息技术 高效多媒体编码 第3部分：音频

GB/T 41867 信息技术 人工智能 术语

3 术语和定义

GB/T 5271.1、GB/T 33475.3 和 GB/T 41867中界定的以及下列术语和定义适用于本文件。

3.1

位流 bitstream

符合本文件规范的实时通信语音编码表示的一组二进制数据流。

3.2

编码 encoding

读入音频采样流，并产生一个符合本文件规定的有效位流的过程。

3.3

编码器 encoder

执行编码处理的实体。

3.4

采样频率 sampling frequency

每秒从连续信号中提取离散信号的采样个数。

注：单位为赫兹（Hz）。

3.5

解码 decoding

本文件规定的一种数据处理，即读入编码位流并输出音频采样流的过程。

3.6

解码器 decoder

执行解码处理的实体。

3.7

熵编码 entropy coding

在数据压缩中，按熵原理进行的一种变长无损编码，用以减少数据统计特性上的冗余。

3.8

声道 channel

声音在录制或播放时在不同空间位置采集或回放的相互独立的音频信号。

3.9

正交镜像滤波器组 quadrature mirror filterbank

将信号分成子带信号实现分频效果的滤波器组。

3.10

深度神经网络 deep neural network

由多层神经元组成的网络，通过权值可调的加权连接，接收输入数据并产生输出。

3.11

神经网络通道 neural network channel

神经网络提取特征时使用的卷积核的数量。

3.12

带宽扩展 bandwidth extension

基于低频成分和边信息生成高频成分的编解码方法。

3.13

丢包隐藏 packet loss concealment

基于已恢复语音帧预测丢失帧的解码方法。

4 缩略语

下列缩略语适用于本文件。

ARSTF:	AVS3 实时通信语音传输格式	(AVS3 RTC Speech Transport Format)
AVS:	数字音视频编解码技术标准	(Audio Video Coding Standard)
AVS3:	第三代 AVS 标准	(The Third-generation AVS Standards)
AVS3P10	AVS3 第 10 部分	(AVS3 Part 10)
BWE:	带宽扩展	(Bandwidth Extension)
CNN:	卷积神经网络	(Convolutional Neural Network)
CRC:	循环冗余校验	(Cyclic Redundancy Check)
IMDCT:	修正离散余弦反变换	(Inverse Modified Discrete Cosine Transform)
MCR:	极大相关旋转	(Maximum Correlation Rotation)
MDCT:	修正离散余弦变换	(Modified Discrete Cosine Transform)
PLC:	丢包隐藏	(Packet Loss Concealment)
PReLU:	参数修正线性单元	(Parametric Rectified Linear Unit)
QMF:	正交镜像滤波器组	(Quadrature Mirror Filterbank)
RTC:	实时通信	(Real-Time Communication)
RU:	残差单元	(Residual Unit)
SQ:	标量量化	(Scalar Quantization)
VBR:	变长位率	(Variable Bit Rate)

5 串行流语法规则

位流中的每一个数据项用黑体。通过名字、按位的长度及其类型和传输顺序的助记符来描述。

位流中被解码的数据元素所导致的操作依赖于该数据的值及以前解码的数据元素。下面的语法结构表示数据元素以标准类型出现时的情形。

注1: 如无特殊说明, 本文件中的“位”指二进制位。

注2: 本文件语法用“C”代码规定, 变量或表达式为非零值时等价于条件为真, 变量或表达式为零值时等价于条件为非真。

```
if(condition){
  data_element
  ...
} else{
  data_element
  ...
}
```

若条件为真, 在数据流中产生第一组数据元素, 若条件为非真, 在数据流中产生第二组数据元素。

```
for(expr1;expr2 ;expr3){
  data_element
  ...
}
```

expr1是指定循环初始状态表达式, 通常它指定了计数器的初始状态, expr2是指定的每次循环前的测试条件。条件为非真时循环终止, expr3是每次循环结束时执行的表达式, 一般是增加计数器。

注3: 本结构的最通常用法为

```
for(i=0;i<n;i++){
```

```

    data_element
    ...
}

```

数据元素组产生 n 次。数据元素组内的条件结构可能依赖循环控制变量 i 的值。第一次出现时被置为‘0’，第二次增加到‘1’，如此往复。

```

switch(expr){
    case constcase1:
        data_element1
        break
    case constcase2:
        data_element2
        break
    ...
    case constcasen:
        data_elementn
        break
    default:
        data_elementdefault
        break
}

```

根据表达式 $expr$ 的值，产生对应的数据元素。 $expr$ 的值为 $constcase1$ 时产生数据元素 $data_element1$ ， $expr$ 的值为 $constcase2$ 时产生数据元素 $data_element2$ ，以此类推， $expr$ 的值为 $constcasen$ 时产生数据元素 $data_elementn$ 。当 $expr$ 的值不等于 $constcase1$ ， $constcase2$ ， \dots ， $constcasen$ 中的任何一个值时，产生数据元素 $data_elementdefault$

本结构的一类变体是在 $case$ 后不出现 $break$ ，如：

```

switch(expr){
    case constcase1:
        data_element1
    case constcase2:
        data_element2
        break
    ...
    case constcasen:
        data_elementn
        break
    default:
        data_elementdefault
        break
}

```

$expr$ 的值 $constcasex$ 时，从对应的 $case\ constcasex$ 开始产生数据元素，直到 $break$ 出现。
 $expr$ 的值 $constcase1$ 时产生数据元素 $data_element1$ 和 $data_element2$ ， $expr$ 的值为 $constcasen$ 时产生数据元素 $data_elementn$

注4：数据元素组中可能含有嵌套结构。为简便起见，当后面只有一个数据元素时“[]”省略。

$data_element[]$	$data_element$ 是一数组数据，数据元素的个数由上下文而定；
$data_element[n]$	$data_element[n]$ 是数组数据的第 $n+1$ 个元素；
$data_element[m][n]$	$data_element[m][n]$ 是二维数组的第 $m+1$ ， $n+1$ 个元素；
$data_element[l][m][n]$	$data_element[l][m][n]$ 是三维数组的第 $l+1$ ， $m+1$ ， $n+1$ 个元素；
$data_element[m\dots n]$	$data_element[m\dots n]$ 是位 m 到位 n 之间包括的位。

6 约定

6.1 概述

本文件中使用的数学运算符和优先级与C语言使用的类似。但对整型除法和算术移位操作进行了特定的定义。除特别说明外，约定编号和计数从0开始。

6.2 算术运算符

下列算术运算符适用于本文件。

+	加法运算
-	减（二元运算符）或取反（一元前缀运算符）
×	乘法运算
a^b	幂，表示a的b次幂，也可表示上标。
$\frac{a}{b}$	除法运算，不做取整或四舍五入。
\sqrt{x}	平方根
$\sum_{i=a}^b f(i)$	自变量i取由a到b（含b）的所有整数值时，函数f(i)的累加和。
\log_2	以2为底的对数

6.3 逻辑运算符

下列逻辑运算符适合于本文件。

	逻辑或运算
&&	逻辑与运算
!	逻辑非运算

6.4 关系运算符

下列关系运算符适用于本文件。

>	大于
≥	大于或等于
<	小于
≤	小于或等于
==	等于
≠	不等于

6.5 位运算符

下列位运算符适用于本文件。

&	与运算
	或运算
~	取反运算
$a \gg b$	将a以2的补码整数表示的形式向右移b位。仅当b取正数时定义此运算。向右移至最高有效位时，其值与a 移位运算前的最高有效位相等。

$a \ll b$ 将 a 以2的补码整数表示的形式向左移 b 位。仅当 b 取正数时定义此运算。向左移至最低有效位时，其值等于0。

6.6 赋值

下列赋值运算适用于本文件。

=	赋值运算符
$x = a..b$	x 取从 a 至 b （含 b ）的值，其中 x 、 a 和 b 是整数。
++	自加， $x++$ 相当于 $x=x+1$ 。当用于数组下标时，在自加运算前先求变量值。
--	自减， $x--$ 相当于 $x=x-1$ 。当用于数组下标时，在自减运算前先求变量值。
+=	自加指定值，例如， $x+=3$ 相当于 $x=x+3$ ， $x+=(-3)$ 相当于 $x=x+(-3)$ 。
-=	自减指定值，例如， $x-=3$ 相当于 $x=x+(-3)$ ， $x-=(-3)$ 相当于 $x=x-(-3)$ 。

6.7 助记符

下列助记符适用于本文件。

rpchof	多项式余数，高阶在先；
bslbf	位串，左位在前。位串是带单引号的1和0串。如‘1000 0001’。位串内的空格是便于阅读的，无特殊意义。（bitstream left bit first）
uimbsf	无符号整数，最高有效位优先。（unsigned integer, most significant bit first）

7 实时通信语音编解码框架

本文件规定的实时通信语音编码规范 (RTC Speech Coding) 是 AVS3 系列标准中的第 10 部分，以下简称 AVS3P10。AVS3P10 是一种基于子带编码的多模式编码方案，其编码框架见图 1：按声道数划分，支持单声道和立体声两种编码模式；按采样频率划分，支持宽带（16000Hz 采样频率）和超宽带（32000Hz 采样频率）两种模式。

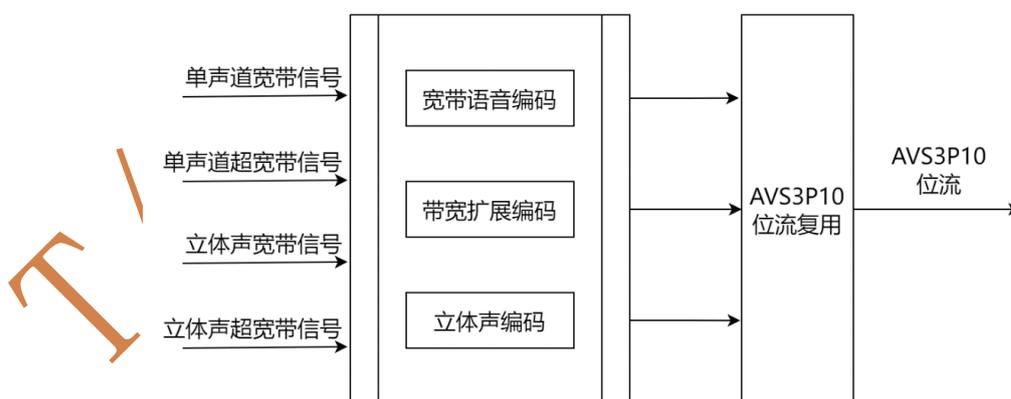


图1 AVS3P10 编码框架

对于单声道宽带信号，AVS3P10 编码直接采用宽带语音编码。对于单声道超宽带信号，AVS3P10 编码将输入信号，划分为低频和高频两个子带；分别采用宽带语音编码和带宽扩展编码。

对于立体声宽带信号，AVS3P10 编码采用宽带语音编码和立体声编码。对于立体声超宽带信号，

AVS3P10 编码采用宽带语音编码、立体声编码和带宽扩展编码。

AVS3P10 编码将编码器提取的信息，进行量化并按照附录 A 中规定的 AVS3P10 位流复用，获得 AVS3P10 编码位流。

AVS3P10 编码的位流封装的传输格式 ARSTF，包含同步字和解码必需的信息，见附录 A。

AVS3P10 编码的参数配置如下：

- a) 支持采样频率为：16000Hz、32000Hz
- b) 支持采样精度为：16bit
- c) 支持编码位率为：
 - 单声道宽带：5.9、7.1、8.0、9.2kbit/s
 - 单声道超宽带：7.9、11.2、12.6、14.5kb/s
 - 立体声宽带：7.6、8.7kb/s
 - 立体声超宽带：11.4、14.5kb/s

AVS3P10 解码框架见图 2。解码器是编码器的逆过程。通过解析位流和解量化，可以获得相应的变量。根据编码模式，使用不同的解码流程。

对于单声道宽带编码模式，经过宽带语音解码，获得单声道宽带重建信号。对于单声道超宽带编码模式，经过宽带语音解码和带宽扩展解码，获得单声道超宽带重建信号。

对于立体声宽带编码模式，经过宽带语音解码和立体声解码，获得立体声宽带重建信号。对于立体声超宽带编码模式，经过宽带语音解码、立体声解码和带宽扩展解码，获得立体声超宽带重建信号。

本文件第 8 章规定了 AVS3P10 解码过程，包括解码器的完整解码流程，以及编码位流解析、宽带特征向量信息和宽带特征向量、解码神经网络、MCR 立体声解码和带宽扩展解码的具体操作步骤。

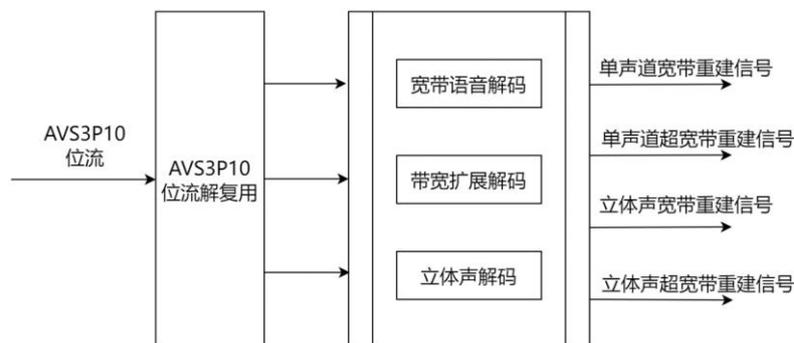


图2 AVS3P10 解码框架

AVS3P10 的编码位流 ARSTF 的语法及语义见附录 A（规范性），AVS3P10 解码神经网络信息见附录 B（规范性）。作为资料性附录，附录 C 提供了 AVS3P10 编码器的参考实现，附录 D 提供了 AVS3P10 解码过程中的一种丢包隐藏方案。

8 实时通信语音解码

8.1 概述

AVS3P10 解码包括单声道宽带解码、单声道超宽带解码、立体声宽带解码、立体声超宽带解码。

AVS3P10 解码器单声道模式流程图见图 3。AVS3P10 解码器单声道模式由解码、逆量化、解码神经网络、带宽扩展解码、QMF 合成滤波器等构成，将位流解码为单声道宽带重建信号或者单声道超宽带

重建信号。

AVS3P10 解码器立体声模式流程图见图 4。AVS3P10 解码器立体声模式由解码、逆量化、解码神经网络、MCR 立体声解码、带宽扩展解码、QMF 合成滤波器等构成，将位流解码为立体声宽带重建信号或者立体声超宽带重建信号。

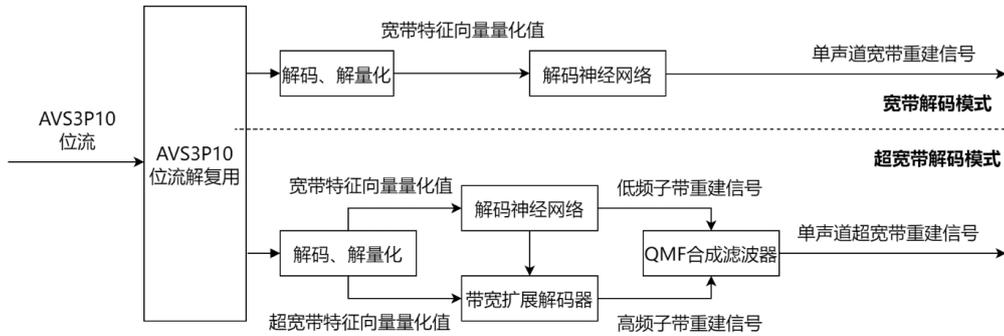


图3 AVS3P10 解码器单声道模式流程图

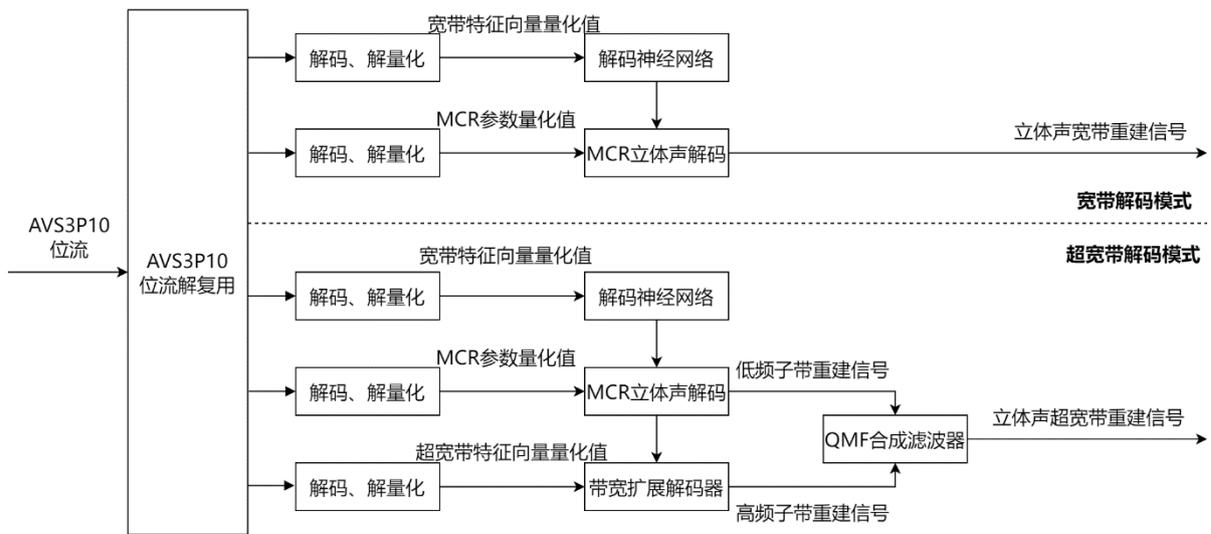


图4 AVS3P10 解码器立体声模式流程图

8.2 AVS3P10 位流解析

8.2.1 语法

AVS3P10 编码位流中的实时通信语音编码数据块 RtcSpeechRawDataBlock() 的解析语法，见表 1。

表1 实时通信语音编码数据块解析语法

语法	位数	助记符
RtcSpeechRawDataBlock() {		
switch(NumChans) {	1	uimsbf
case 0x0: RtcSpeechMonoDec()		
case 0x1: RtcSpeechStereoDec()		
}		

实时通信语音编码单声道解码语法 RtcSpeechMonoDec(), 见表 2。

表2 实时通信语音编码单声道解码语法

语法	位数	助记符
RtcSpeechMonoDec () {		
switch(BandwidthMode) {	1	uimsbf
case 0x0: RtcSpeechMonoWbDec()		
case 0x1: RtcSpeechMonoSwbDec()		
}		
RateMode	2	uimsbf
}		

实时通信语音编码单声道宽带解码语法 RtcSpeechMonoWbDec(), 见表 3。

表3 实时通信语音编码单声道宽带解码语法

语法	位数	助记符
RtcSpeechMonoWbDec () {		
DecodeWbInforBits()		
WbInvQuanCod()		
WbNeuralNet()		
}		

实时通信语音编码单声道超宽带解码语法RtcSpeechMonoSwbDec(), 见表4。

表4 实时通信语音编码单声道超宽带解码语法

语法	位数	助记符
RtcSpeechMonoSwbDec () {		
DecodeWbInforBits()		
DecodeSwbInforBits()		
WbInvQuanCod ()		
SwbInvQuanCod()		
WbNNNetDec()		

表 4 (续)

语法	位数	助记符
SwbBweDec()		
QMFSynthesis()		
}		

实时通信语音编码立体声解码语法 RtcSpeechStereoDec(), 见表 5。

表5 实时通信语音编码立体声解码语法

语法	位数	助记符
RtcSpeechStereoDec () {		
switch(BandwidthMode) {	1	uimsbf
case 0x0: RtcSpeechStereoWbDec()		
case 0x1: RtcSpeechStereoSwbDec()		
}		
RateMode	2	uimsbf
}		

实时通信语音编码立体声宽带解码语法 RtcSpeechStereoWbDec(), 见表 6。

表6 实时通信语音编码立体声宽带解码语法

语法	位数	助记符
RtcSpeechStereoWbDec () {		
DecodeWbInforBits()		
DecodeMcrInforBits()		
WbInvQuanCod()		
McrInvQuanCod()		
WbNeuralNet()		
McrDec()		
}		

实时通信语音编码立体声超宽带解码语法RtcSpeechStereoSwbDec(), 见表7。

表7 实时通信语音编码立体声超宽带解码语法

语法	位数	助记符
RtcSpeechMonoSwbDec() {		
DecodeWbInforBits()		
DecodeSwbInforBits()		
DecodeMcrInforBits ()		
DecodeSwbRInforBits ()		
WbInvQuanCod()		

表7（续）

语法	位数	助记符
SwbInvQuanCod()		
McrInvQuanCod()		
SwbRInvQuanCod()		
WbNNetDec ()		
SwbBweDec()		
McrDec()		
SwbBweRDec()		
QMFSynthesis()		
}		

8.2.2 语义

NumChans	声道数，0 为单声道，1 为立体声
BandwidthMode	带宽模式，0 为宽带编码，1 为超宽带编码
RateMode	位率模式，2 比特，包含 4 种编码位率模式
RtcSpeechMonoDec()	实时通信语音单声道解码函数
RtcSpeechStereoDec()	实时通信语音立体声解码函数
RtcSpeechMonoWbDec()	实时通信语音单声道宽带模式解码函数
RtcSpeechMonoSwbDec()	实时通信语音单声道超宽带模式解码函数
RtcSpeechStereoWbDec()	实时通信语音立体声宽带模式解码函数
RtcSpeechStereoSwbDec()	实时通信语音立体声超宽带模式解码函数
DecodeWbInforBits()	解析宽带特征信息函数
DecodeSwbInforBits()	解析超宽带特征信息函数
DecodeMcrInforBits()	解析立体声特征信息函数
DecodeSwbRInforBits()	解析超宽带特征信息函数（仅限立体声超宽带解码）
WbInvQuanCod()	宽带解码和逆量化函数
SwbInvQuanCod()	超宽带解码和逆量化函数
McrInvQuanCod()	立体声解码和逆量化函数
SwbRInvQuanCod()	超宽带解码和逆量化函数（仅限立体声超宽带解码）
WbNNetDec ()	宽带解码神经网络函数
SwbBweDec ()	带宽扩展解码函数
McrDec ()	MCR 立体声解码函数
SwbBweRDec ()	带宽扩展解码函数（仅限立体声超宽带解码）
QMFSynthesis()	QMF 合成滤波函数

8.2.3 通用信息

8.2.3.1 概述

AVS3P10解码器可以分为逆量化和解码、第一级解码卷积神经网络、多级解码块神经网络、后处理神经网络、第二级解码卷积神经网络、带宽扩展、MCR立体声解码、QMF合成滤波器（适用于超宽带

编码模式)等模块。根据输入信号声道数、带宽模式等编码模式参数,调用相关模块进行解码。解码过程涉及的解码神经网络结构描述性信息和网络权值参数,见本文件附录B。

8.2.3.2 单声道宽带解码

AVS3P10 单声道宽带解码流程图见图 5。单声道解码器通过熵解码和逆量化从位流中解析得到宽带特征向量量化值,经过包含多个神经网络的计算,生成单声道宽带重建信号。

a) 熵解码和逆量化

从位流中获取与宽带特征向量编码相关的位流信息,进行熵解码和逆量化处理以得到宽带特征向量量化值,作为深度神经网络计算的输入。

——输入: 位流

——输出: 宽带特征向量量化值

b) 第一级解码卷积神经网络

将宽带特征向量量化值,经过 1 维卷积运算后,生成第一级解码向量。宽带特征向量的第一维度是神经网络通道数,设置为 56;宽带特征向量的第二维度是样本数,设置为 1。第一级解码向量的第一维度是神经网络通道数,设置为 256;第一级解码向量的第二维度是样本数,设置为 1。

——输入: 宽带特征向量量化值

——输出: 第一级解码向量

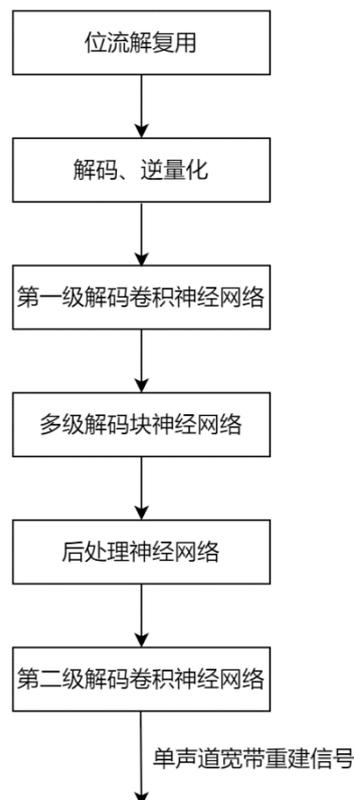


图5 AVS3P10 解码器单声道宽带解码流程图

c) 多级解码块神经网络

将第一级解码向量，经过 4 级解码块神经网络，生成第二级解码向量。每一级解码块神经网络包含多层神经网络计算。每一级解码块神经网络通过计算，将输入向量的通道数减少 1 倍，同时通过不同因子的复制，增加输出向量第二维度的数据量。第二级编码向量的第一维度是神经网络通道数，经过 4 级编码块神经网络，由 256 减少为 16；第二级编码向量的第二维度是样本数，设置为 160。多级解码块神经网络的模块细节见 8.4.3.3 至 8.4.3.6。

——输入：第一级解码向量

——输出：第二级解码向量

d) 后处理解码神经网络

后处理神经网络，将第二级解码向量进行一次一维因果卷积计算并激活后，生成第三级解码向量。第三级编码向量的第一维度是神经网络通道数，设置为 16；第三级编码向量的第二维度是样本数，设置为 320。

——输入：第二级解码向量

——输出：第三级解码向量

e) 第二级解码卷积神经网络

将第三级解码向量，进行一次一维因果卷积计算，生成单声道宽带重建信号。单声道宽带重建信号长度为 320。

——输入：第三级解码向量

——输出：单声道宽带重建信号

8.2.3.3 单声道超宽带解码

AVS3P10 单声道超宽带解码流程见图 6。单声道解码器通过熵解码和逆量化从位流中解析得到宽带特征向量量化值和超宽带特征向量量化值。对宽带特征向量量化值，经过包含多个神经网络的计算，生成低频子带信号。对低频子带信号和超宽带特征向量量化值，经过带宽扩展解码，生成高频子带信号。对低频子带重建信号和高频子带重建信号，经过 QMF 合成滤波器处理，生成单声道超宽带重建信号。

a) 熵解码和逆量化

从位流中获取与宽带特征向量编码和超宽带特征向量编码相关的位流信息，进行熵解码和逆量化处理以得到宽带特征向量量化值和超宽带特征向量量化值，作为深度神经网络计算和带宽扩展解码的输入。

——输入：位流

——输出：宽带特征向量量化值和超宽带特征向量量化值

b) 第一级解码卷积神经网络

将宽带特征向量量化值，经过 1 维卷积运算后，生成第一级解码向量。宽带特征向量量化值的第一维度是神经网络通道数，设置为 56；宽带特征向量量化值的第二维度是样本数，设置为 1。第一级解码向量的第一维度是神经网络通道数，设置为 256；第一级解码向量的第二维度是样本数，设置为 1。

——输入：低频特征向量量化值

——输出：第一级解码向量

c) 多级解码块神经网络

将第一级解码向量，经过 4 级解码块神经网络，生成第二级解码向量。每一级解码块神经网络包含多层神经网络计算。每一级解码块神经网络通过计算，将输入向量的通道数减少 1 倍，同时通过不同因子的复制，增加输出向量第二维度的数据量。第二级编码向量的第一维度是神经网络通道数，经过 4 级编码块神经网络，由 256 减少为 16；第二级编码向量的第二维度是样本数，设置为 160。多级解码块神经网络的模块细节见 8.4.3.3 至 8.4.3.6。

- 输入：第一级解码向量
- 输出：第二级解码向量

d) 后处理神经网络

后处理神经网络，将第二级解码向量进行一次一维因果卷积计算并激活后，生成第三级解码向量。第三级编码向量的第一维度是神经网络通道数，设置为 16；第三级编码向量的第二维度是样本数，设置为 320。

- 输入：第二级解码向量
- 输出：第三级解码向量

e) 第二级解码卷积神经网络

将第三级解码向量，进行一次一维因果卷积计算，生成低频子带信号。低频子带信号长度设置为 320。

- 输入：第三级解码向量
- 输出：低频子带重建信号

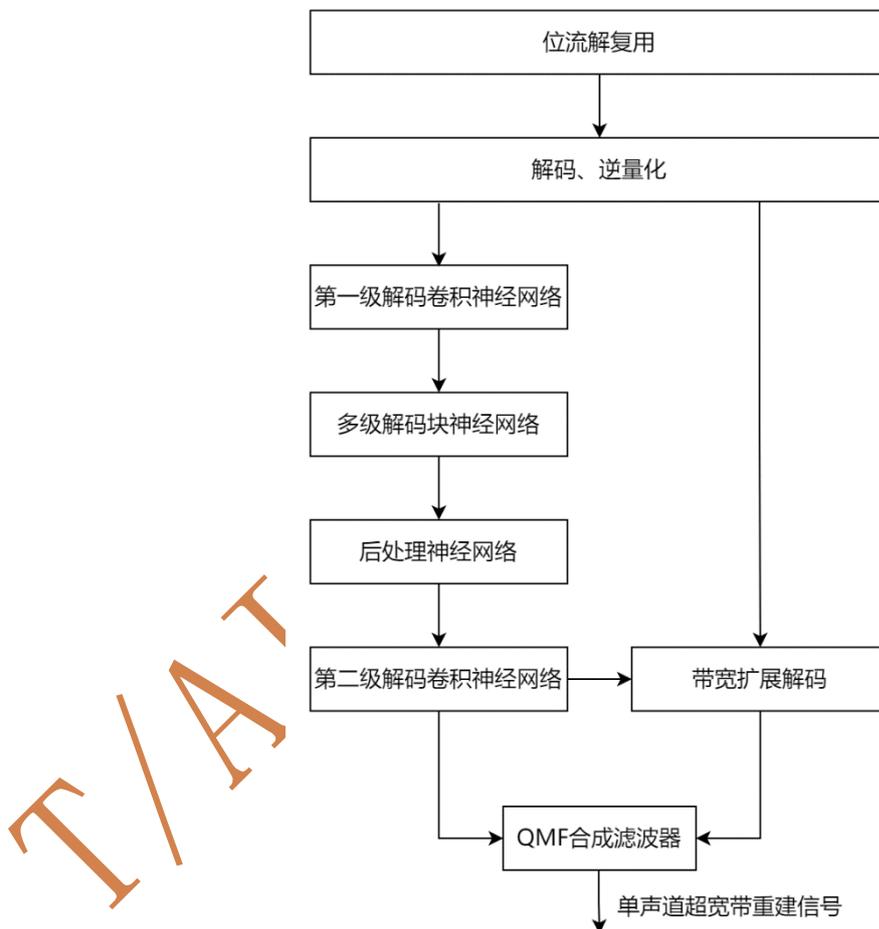


图6 AVS3P10 解码器单声道超宽带解码流程图

f) 带宽扩展解码

将低频子带重建信号和超宽带特征向量量化值，经过带宽扩展解码，获得高频子带重建信号。带宽扩展解码对低频子带重建信号进行 2 次长度 320 点 MDCT 变换，并通过频带复制，生成高频子带重建

信号对应的 MDCT 系数；然后，通过超宽带特征向量量化值，对高频 MDCT 系数进行增益控制；最后，通过 MDCT 逆变换，获得高频子带重建信号。高频子带信号的长度设置为 320。

——输入：低频子带重建信号和超宽带特征向量量化值

——输出：高频子带重建信号

g) QMF 合成滤波器

对于低频子带重建信号和高频子带重建信号，经过升 2 采样和合成滤波，生成单声道超宽带重建信号。单声道超宽带重建信号的长度设置为 640。

——输入：低频子带重建信号和高频子带重建信号

——输出：单声道超宽带重建信号

8.2.3.4 立体声宽带解码

AVS3P10 立体声宽带解码流程图见图 7。解码器通过熵解码和逆量化从位流中解析得到宽带特征向量量化值和 MCR 参数量化值。对宽带特征向量量化值，经过包含多个神经网络的计算，生成左声道宽带重建信号。然后，对宽带特征向量量化值和 MCR 参数量化值，经过 MCR 立体声解码，生成立体声宽带重建信号。

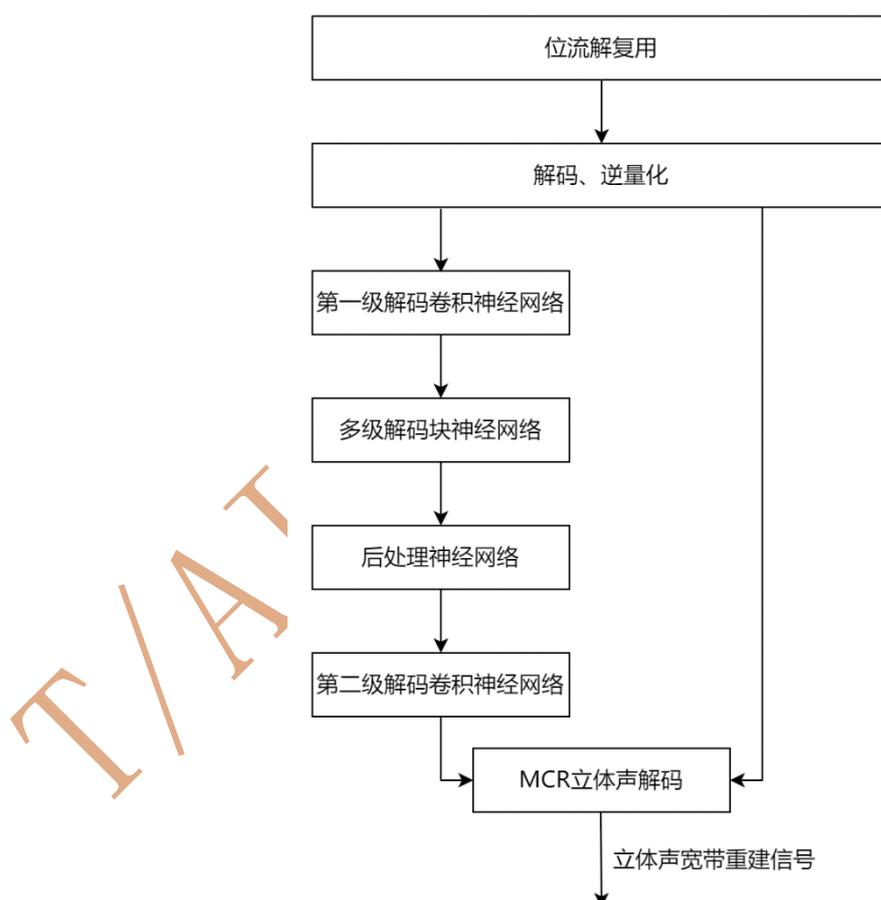


图7 AVS3P10 解码器立体声宽带解码流程图

a) 熵解码和逆量化

从位流中获取与宽带特征向量编码相关的位流信息，进行熵解码和逆量化处理以得到宽带特征向

量量化值，作为深度神经网络计算的输入。

- 输入：位流
- 输出：宽带特征向量量化值和 MCR 参数量化值

b) 第一级解码卷积神经网络

将宽带特征向量量化值，经过 1 维卷积运算后，生成第一级解码向量。宽带特征向量的第一维度是神经网络通道数，设置为 56；宽带特征向量的第二维度是样本数，设置为 1。第一级解码向量的第一维度是神经网络通道数，设置为 256；第一级解码向量的第二维度是样本数，设置为 1。

- 输入：宽带特征向量量化值
- 输出：第一级解码向量

c) 多级解码块神经网络

将第一级解码向量，经过 4 级解码块神经网络，生成第二级解码向量。每一级解码块神经网络包含多层神经网络计算。每一级解码块神经网络通过计算，将输入向量的通道数减少 1 倍，同时通过不同因子的复制，增加输出向量第二维度的数据量。第二级编码向量的第一维度是神经网络通道数，经过 4 级编码块神经网络，由 256 减少为 16；第二级编码向量的第二维度是样本数，设置为 160。多级解码块神经网络的模块细节见 8.4.3.3 至 8.4.3.6。

- 输入：第一级解码向量
- 输出：第二级解码向量

d) 后处理解码神经网络

后处理神经网络，将第二级解码向量进行一次一维因果卷积计算并激活后，生成第三级解码向量。第三级编码向量的第一维度是神经网络通道数，设置为 16；第三级编码向量的第二维度是样本数，设置为 320。

- 输入：第二级解码向量
- 输出：第三级解码向量

e) 第二级解码卷积神经网络

将第三级解码向量，进行一次一维因果卷积计算，生成单声道宽带重建信号。单声道宽带重建信号长度为 320。

- 输入：第三级解码向量
- 输出：左声道宽带重建信号

f) MCR 立体声解码

将左声道宽带重建信号和 MCR 参数量化值，进行 MCR 立体声解码，生成立体声宽带重建信号。立体声宽带重建信号每个声道的长度为 320。

- 输入：左声道宽带重建信号和 MCR 参数量化值
- 输出：立体声宽带重建信号

8.2.3.5 立体声超宽带解码

AVS3P10 立体声超宽带解码流程图见图 8。解码器通过熵解码和逆量化从位流中解析得到宽带特征向量量化值、MCR 参数量化值和超宽带特征向量量化值。对宽带特征向量量化值，经过包含多个神经网络的计算，生成左声道低频子带信号。对左声道低频子带信号和 MCR 参数量化值，经过 MCR 立体声解码，生成低频子带信号。对低频子带信号和超宽带特征向量量化值，经过带宽扩展解码，生成高频子带信号。对低频子带重建信号和高频子带重建信号，经过 QMF 合成滤波器处理，生成立体声超宽带重建信号。

a) 熵解码和逆量化

从位流中获取与宽带特征向量编码和超宽带特征向量编码相关的位流信息，进行熵解码和逆量化处理以得到宽带特征向量量化值和超宽带特征向量量化值，作为深度神经网络计算和带宽扩展解码的输入。

——输入：位流

——输出：宽带特征向量量化值和超宽带特征向量量化值

b) 第一级解码卷积神经网络

将宽带特征向量量化值，经过 1 维卷积运算后，生成第一级解码向量。宽带特征向量量化值的第一维度是神经网络通道数，设置为 56；宽带特征向量量化值的第二维度是样本数，设置为 1。第一级解码向量的第一维度是神经网络通道数，设置为 256；第一级解码向量的第二维度是样本数，设置为 1。

——输入：低频特征向量量化值

——输出：第一级解码向量

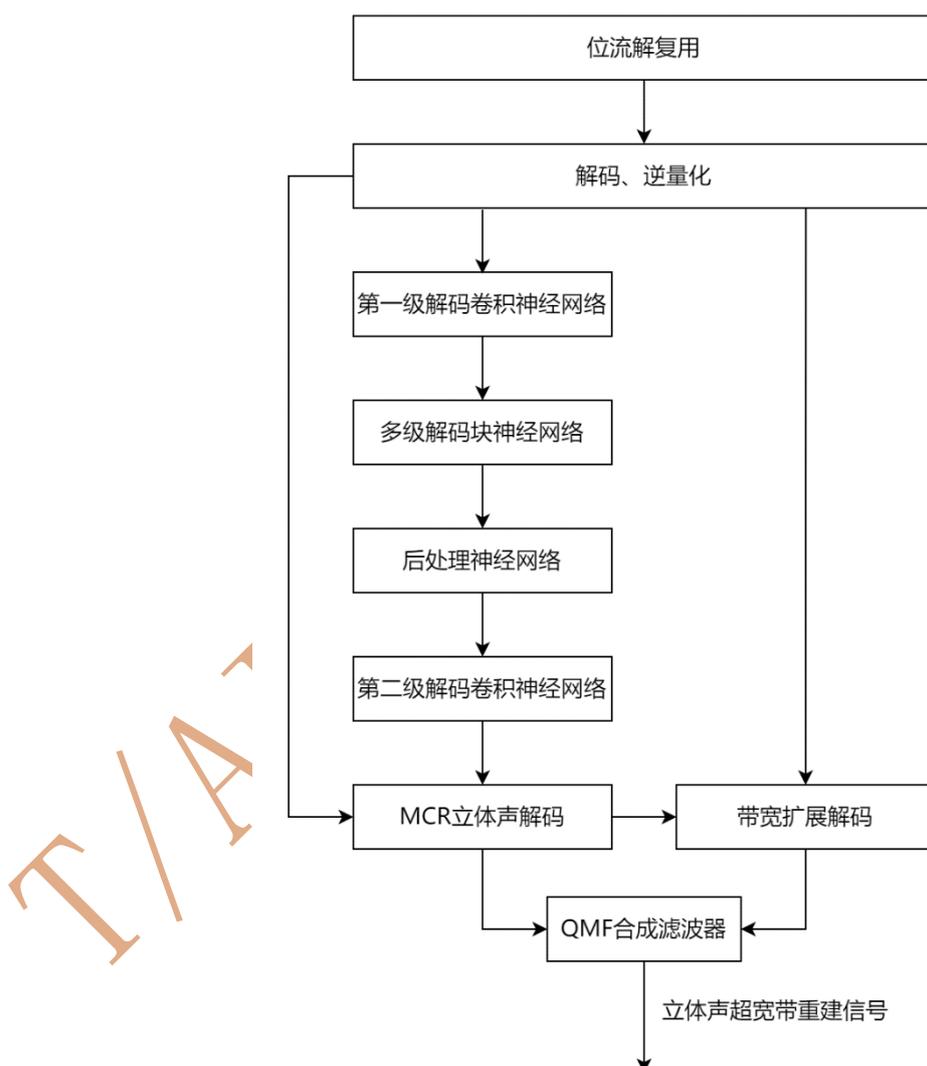


图8 AVS3P10 解码器立体声超宽带解码流程图

c) 多级解码块神经网络

将第一级解码向量，经过 4 级解码块神经网络，生成第二级解码向量。每一级解码块神经网络包含

多层神经网络计算。每一级解码块神经网络通过计算，将输入向量的通道数减少 1 倍，同时通过不同因子的复制，增加输出向量第二维度的数据量。第二级编码向量的第一维度是神经网络通道数，经过 4 级编码块神经网络，由 256 减少为 16；第二级编码向量的第二维度是样本数，设置为 160。多级解码块神经网络的模块细节见 8.4.3.3 至 8.4.3.6。

——输入：第一级解码向量

——输出：第二级解码向量

d) 后处理解码神经网络

后处理神经网络，将第二级解码向量进行一次一维因果卷积计算并激活后，生成第三级解码向量。第三级编码向量的第一维度是神经网络通道数，设置为 16；第三级编码向量的第二维度是样本数，设置为 320。

——输入：第二级解码向量

——输出：第三级解码向量

e) 第二级解码卷积神经网络

将第三级解码向量，进行一次一维因果卷积计算，生成低频子带信号。低频子带信号长度设置为 320。

——输入：第三级解码向量

——输出：左声道低频子带重建信号

f) MCR 立体声解码

将左声道宽带重建信号和 MCR 参数量化值，进行 MCR 立体声解码，生成立体声宽带重建信号。立体声宽带重建信号长度为 320。

——输入：左声道宽带重建信号和 MCR 参数量化值

——输出：低频子带重建信号

g) 带宽扩展解码

低频带子重建信号和超宽带特征向量量化值，分左右声道单独处理，经过带宽扩展解码，获得高频子带重建信号。带宽扩展解码对低频子带重建信号进行 2 次长度 320 点 MDCT 变换，并通过频带复制，生成高频子带重建信号对应的 MDCT 系数；然后，通过超宽带特征向量量化值，对高频 MDCT 系数进行增益控制；最后，通过 MDCT 逆变换，获得高频子带重建信号。高频子带信号的长度设置为 320。

——输入：低频子带重建信号和超宽带特征向量量化值

——输出：高频子带重建信号

h) QMF 合成滤波器

对于低频子带重建信号和高频子带重建信号，分左右声道单独处理，经过升 2 采样和合成滤波，生成立体声超宽带重建信号。立体声超宽带重建信号每个声道的长度设置为 640。

——输入：低频带子重建信号和高频子带重建信号

——输出：立体声超宽带重建信号

8.2.4 解码过程

实时通信语音编码原始数据块语法(RtcSpeechRawDataBlock)是构成ARSTF编码位流的基本单位，可获得一帧的音频样本。ARSTF格式见附录A。

解码器根据NumChans选择实时通信语音单声道解码(RtcSpeechMonoDec)或者实时通信语音立体声解码(RtcSpeechStereoDec)。单声道解码根据BandwidthMode选择单声道宽带解码(RtcSpeechMonoWbDec)或者单声道超宽带解码(RtcSpeechMonoSwbDec)。立体声解码根据BandwidthMode选择立体声宽带解码(RtcSpeechStereoWbDec)或者立体声超宽带解码(RtcSpeechStereoSwbDec)。

单声道宽带解码，首先解析宽带特征向量信息（DecodeWbInforBits）。然后进行宽带解码和逆量化（WbInvQuanCod）获得宽带特征向量量化值。最后，进行解码神经网络（WbNNetDec），得到单声道宽带重建信号。

单声道超宽带解码，首先解析宽带特征向量信息（DecodeWbInforBits）和解析超宽带特征向量信息（DecodeSwbInforBits）。然后进行宽带解码和逆量化（WbInvQuanCod）和进行超宽带解码和逆量化（SwbInvQuanCod）获得宽带特征向量量化值和超宽带特征向量量化值。进行解码神经网络（WbNNetDec）获得低频子带重建信号；进行带宽扩展解码（SwbBweDec）获得高频子带重建信号；最后，进行QMF合成滤波（QMFSynthesis），获得单声道超宽带重建信号。

立体声宽带解码，首先解析宽带特征向量信息（DecodeWbInforBits）和解析立体声特征信息（DecodeMcrInforBits）。然后，进行宽带解码和逆量化（WbInvQuanCod）获得宽带特征向量量化值；进行立体声解码和逆量化（McrInvQuanCod）获得MCR参数量化值。进行解码神经网络（WbNNetDec），得到左声道宽带重建信号；最后，进行MCR立体声解码（McrDec），得到立体声宽带重建信号。

立体声超宽带解码，首先解析宽带特征向量信息（DecodeWbInforBits）、解析超宽带特征向量信息（DecodeSwbInforBits）、解析立体声特征信息（DecodeMcrInforBits）和解析右声道相关的超宽带特征向量信息（DecodeSwbRInforBits）。然后进行宽带解码和逆量化（WbInvQuanCod）获得宽带特征向量量化值、进行立体声解码和逆量化（McrInvQuanCod）获得MCR参数量化值、进行超宽带解码和逆量化（SwbInvQuanCod）获得左声道相关超宽带特征向量量化值和进行超宽带解码和逆量化（SwbRInvQuanCod）获得右声道相关超宽带特征向量量化值。进行解码神经网络（WbNNetDec），得到左声道宽带重建信号；进行MCR立体声解码（McrDec），得到低频子带重建信号；进行带宽扩展解码（SwbBweDec和SwbBweRDec）获得高频子带重建信号；最后，进行QMF合成滤波（QMFSynthesis），获得立体声超宽带重建信号。

AVS3P10编码是一种基于子带编码的多模式编码方案。单声道宽带编码模式和单声道超宽带编码模式均支持4个位率，根据每种编码模式下每一种位率模式对应的运行模式见表8和表9。

表8 单声道宽带编码模式的位率模式信息

编码位率模式	编码位率（kbit/s）	宽带运行模式
0	5.9	WB_Rate1
1	7.1	WB_Rate2
2	8.0	WB_Rate3
3	9.2	WB_Rate4

表9 单声道超宽带编码模式的位率模式信息

编码位率模式	编码位率（kbit/s）	宽带运行模式	超宽带运行模式
0	7.9	WB_Rate1	SWB_Rate1
1	11.2	WB_Rate4	SWB_Rate2
2	12.6	WB_Rate4	SWB_Rate3
3	14.5	WB_Rate4	SWB_Rate4

立体声宽带编码模式和立体声超宽带编码模式，是在对应的单声道编码模式下增加必要的立体声信息和超宽带特征向量信息；其中，对左声道的重建与单声道模式一致。立体声宽带编码模式和立体声超宽带编码模式均支持2个位率。根据每种编码模式下每一种位率模式对应的运行模式见表10和表11。

表10 立体声宽带编码模式的位率模式信息

编码位率模式	编码位率 (kbit/s)	宽带运行模式
0	7.6	WB_Rate1
1	8.7	WB_Rate2

表11 立体声超宽带编码模式的位率模式信息

编码位率模式	编码位率 (kbit/s)	宽带运行模式	超宽带运行模式
0	11.4	WB_Rate1	SWB_Rate1
1	14.5	WB_Rate4	SWB_Rate2

8.3 宽带特征向量信息和宽带特征向量

8.3.1 语法

解码宽带特征向量信息（索引值）和宽带特征向量，见表12至13。

表12 解析宽带特征向量索引值语法

语法	位数	助记符
DecodeWbInforBits() {		
for (i = 0; i < 56; i++) {		
WbIndice[i]	2;3	uimsbf
}		
}		

表13 宽带解码和逆量化语法

语法	位数	助记符
WbInvQuanCod() {		
for (i = 0; i < 56; i++) {		
WbEmbedding[i]		
}		

8.3.2 语义

WbIndice

解析获得 56 维宽带特征向量索引值

WbEmbedding

通过每一维度的宽带特征向量索引值（WbIndice），获得第一维度的宽带特征向量量化值。

8.3.3 解码过程

从AVS3P10位流中解析宽带特征信息函数DecodeWbInforBits()（见表12），利用熵解码，解析出56维宽带特征向量索引值；通过宽带特征向量索引值（WbIndice）和表15至18，获得宽带特征向量量化值，用于解码神经网络。

对于表8至表11所示的任意一种宽带运行模式，熵解码具有唯一性。根据对应宽带运行模式下包含的码表中的标识数量和统计概率，通过熵解码可以获得对应的宽带特征向量索引值。每种宽带运行模式包含的标识数量和标识，见表14。

表14 宽带运行模式包含的标识数量和标识信息

宽带运行模式	标识数量	标识
WB_Rate1	5	-1.0, 0.5, 0, 0.5, 1.0
WB_Rate2	7	-1.0, -0.66667, -0.33333, 0, 0.33333, 0.66667, 1.0
WB_Rate3	9	-1.0, -0.75, -0.50, -0.25, 0, 0.25, 0.50, 0.75, 1.0
WB_Rate4	11	-1.0, -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1.0

每种宽带运行模式进行熵解码所需的统计概率表，见表15至18。标识概率统计值的最大值为255。根据标识概率统计值各分量，计算出对应标识的概率后，调用熵解码，获得宽带特征向量索引值。需要操作56次。

表15 宽带运行模式 WB_Rate1 标识统计表

宽带特征向量维度	标识概率统计值
1	224,113,36,4,0
2	245,162,61,5,0
3	245,174,63,3,0
4	245,179,74,10,0
5	252,198,88,6,0
6	247,199,105,17,0
7	250,177,71,6,0
8	238,180,88,9,0
9	250,186,76,5,0
10	247,169,62,2,0
11	237,179,88,18,0
12	243,182,81,5,0
13	239,147,49,4,0
14	240,157,57,6,0
15	228,133,43,3,0
16	250,195,89,3,0
17	250,177,80,8,0
18	245,174,66,5,0
19	247,195,100,10,0
20	240,146,62,13,0

表15 (续)

宽带特征向量维度	标识概率统计值
21	246,150,54,2,0
22	243,192,113,17,0
23	239,154,50,3,0
24	250,185,75,3,0
25	249,200,94,9,0
26	234,174,87,6,0
27	248,166,68,5,0
28	244,173,63,3,0
29	249,189,82,8,0
30	250,181,69,5,0
31	244,194,117,23,0
32	246,183,77,6,0
33	250,204,107,9,0
34	230,171,117,44,0
35	229,152,85,34,0
36	240,160,57,6,0
37	241,143,53,6,0
38	244,184,77,7,0
39	240,147,54,8,0
40	247,172,64,5,0
41	243,169,75,13,0
42	248,227,195,130,0
43	235,146,48,5,0
44	245,155,55,4,0
45	196,144,83,21,0
46	249,205,116,13,0
47	248,186,80,3,0
48	238,151,62,10,0
49	246,163,59,5,0
50	246,169,59,2,0
51	249,173,66,5,0
52	247,204,119,22,0
53	248,185,72,4,0
54	240,140,51,7,0
55	242,174,70,6,0

表15 (续)

宽带特征向量维度	标识概率统计值
56	251,189,81,9,0

表16 宽带运行模式 WB_Rate2 标识统计表

宽带特征向量维度	标识概率统计值
1	238,166,96,45,16,1,0
2	251,212,143,73,27,1,0
3	251,217,156,77,23,1,0
4	251,224,164,91,37,4,0
5	254,235,181,106,36,1,0
6	251,228,185,120,54,6,0
7	253,224,157,84,30,1,0
8	247,212,169,104,41,1,0
9	253,228,168,92,27,1,0
10	251,220,149,76,22,1,0
11	245,217,169,107,50,9,0
12	250,218,170,99,34,1,0
13	248,198,128,61,19,1,0
14	248,207,138,70,25,2,0
15	240,184,115,53,16,1,0
16	253,231,180,107,34,1,0
17	253,223,156,91,38,1,0
18	251,216,158,80,27,1,0
19	252,225,183,118,44,2,0
20	248,201,127,71,33,5,0
21	252,207,132,67,21,1,0
22	248,221,181,129,63,4,0
23	248,209,138,65,20,1,0
24	253,227,170,95,25,1,0
25	252,230,184,111,38,1,0
26	244,208,164,105,35,1,0
27	252,216,147,81,28,1,0
28	251,216,158,80,23,1,0
29	253,228,174,98,40,1,0
30	253,227,162,84,28,1,0
31	249,222,182,130,69,6,0

表16 (续)

宽带特征向量维度	标识概率统计值
32	252,222,169,95,33,1,0
33	253,234,190,124,48,1,0
34	240,201,162,126,89,19,0
35	244,188,135,90,58,20,0
36	248,208,142,70,25,1,0
37	248,195,121,62,22,1,0
38	250,220,168,90,34,1,0
39	249,203,129,65,24,2,0
40	252,219,154,77,27,1,0
41	249,213,152,88,37,5,0
42	251,240,225,205,173,110,0
43	245,201,129,61,21,1,0
44	251,210,134,66,23,1,0
45	211,175,140,98,54,11,0
46	252,233,194,135,59,1,0
47	252,226,172,100,29,1,0
48	249,196,132,72,29,4,0
49	251,214,143,71,25,1,0
50	252,217,151,73,22,1,0
51	252,224,153,80,27,1,0
52	251,230,193,135,65,9,0
53	253,224,170,90,28,1,0
54	249,194,121,61,23,3,0
55	250,213,159,84,30,1,0
56	254,230,171,94,35,1,0

表17 宽带运行模式 WB_Rate3 标识统计表

宽带特征向量维度	标识概率统计值
1	244,196,139,88,50,25,9,1,0
2	253,230,188,133,81,43,15,1,0
3	253,233,198,146,87,42,11,1,0
4	251,234,201,150,95,51,19,1,0
5	254,245,218,172,116,59,17,1,0
6	252,238,214,178,129,79,34,2,0
7	254,239,201,146,92,48,15,1,0

表17 (续)

宽带特征向量维度	标识概率统计值
8	249,223,194,159,111,61,22,1,0
9	254,241,208,158,102,49,13,1,0
10	252,236,195,138,84,39,9,1,0
11	246,226,197,158,112,66,32,4,0
12	252,231,200,160,106,55,15,1,0
13	250,220,172,119,68,32,9,1,0
14	249,225,181,128,77,38,13,1,0
15	243,205,156,105,58,25,6,1,0
16	254,243,216,173,118,61,13,1,0
17	254,239,201,148,100,59,22,1,0
18	253,231,196,148,89,45,14,1,0
19	253,237,210,176,126,70,23,1,0
20	251,223,172,120,78,46,22,1,0
21	253,230,176,121,73,35,8,1,0
22	250,232,206,174,136,87,38,1,0
23	250,224,179,123,69,30,7,1,0
24	254,241,208,161,105,49,10,1,0
25	253,241,217,178,124,66,22,1,0
26	247,220,189,154,111,57,15,1,0
27	253,236,189,137,87,44,12,1,0
28	252,230,194,148,89,40,10,1,0
29	254,240,210,165,106,59,21,1,0
30	254,241,206,152,94,47,14,1,0
31	251,233,207,175,136,92,46,3,0
32	253,235,203,159,104,52,16,1,0
33	254,244,221,184,135,77,25,1,0
34	244,214,184,155,129,102,70,9,0
35	248,209,172,132,97,70,47,15,0
36	250,224,184,131,78,39,14,1,0
37	251,221,169,115,70,36,13,1,0
38	252,233,203,158,99,54,18,1,0
39	251,223,174,118,71,36,14,1,0
40	253,234,196,142,85,44,14,1,0
41	251,229,191,143,96,53,23,1,0
42	251,242,231,218,202,180,149,85,0
43	246,216,170,114,65,31,10,1,0

表17 (续)

宽带特征向量维度	标识概率统计值
44	253,229,184,123,73,37,11,1,0
45	210,179,154,127,97,64,34,4,0
46	253,241,219,185,140,83,32,1,0
47	253,238,207,162,109,52,11,1,0
48	252,217,174,125,80,44,18,1,0
49	253,233,189,133,79,41,14,1,0
50	253,233,194,140,81,38,9,1,0
51	253,239,201,144,90,47,16,1,0
52	253,241,220,189,146,94,42,6,0
53	253,236,203,159,98,46,12,1,0
54	252,220,168,114,69,35,13,1,0
55	251,226,192,148,92,48,15,1,0
56	254,242,211,160,103,56,20,1,0

表18 宽带运行模式 WB_Rate4 标识统计表

宽带特征向量维度	标识概率统计值
1	247,214,168,124,84,52,30,14,3,1,0
2	253,239,211,172,126,85,52,26,6,1,0
3	254,241,217,184,139,92,53,24,5,1,0
4	253,242,221,189,145,101,63,33,11,1,0
5	254,249,234,205,165,119,73,34,7,1,0
6	253,244,228,205,173,133,93,54,21,1,0
7	254,246,222,185,140,96,58,28,6,1,0
8	252,233,211,187,157,118,77,42,14,1,0
9	254,247,227,194,152,107,64,28,6,1,0
10	253,243,219,179,132,88,50,20,2,1,0
11	249,234,214,186,152,114,76,45,20,1,0
12	253,239,217,189,155,112,69,33,7,1,0
13	252,232,198,156,112,72,41,19,5,1,0
14	251,235,206,168,124,82,48,23,6,1,0
15	246,220,183,143,102,64,35,15,3,1,0
16	254,247,230,202,164,120,74,33,5,1,0
17	254,246,222,186,142,104,70,39,12,1,0
18	254,240,215,183,142,95,57,28,7,1,0
19	254,244,225,201,172,133,87,44,13,1,0

表18 (续)

宽带特征向量维度	标识概率统计值
20	252,234,200,156,115,81,54,32,14,1,0
21	254,242,207,162,118,79,46,20,2,1,0
22	251,238,220,197,171,141,103,63,24,1,0
23	252,234,205,164,118,74,39,16,2,1,0
24	254,246,226,194,153,107,61,23,2,1,0
25	254,246,231,207,173,129,81,40,11,1,0
26	250,229,206,180,151,116,74,34,7,1,0
27	254,244,215,174,132,92,56,27,6,1,0
28	253,239,214,181,142,95,52,22,3,1,0
29	254,246,227,198,160,113,73,40,12,1,0
30	254,247,226,191,145,98,58,27,6,1,0
31	252,239,221,198,171,140,104,67,29,1,0
32	254,243,221,192,155,111,67,33,9,1,0
33	254,247,232,209,177,137,91,46,11,1,0
34	246,222,199,176,153,132,111,89,55,3,0
35	250,222,191,160,128,101,79,60,39,10,0
36	252,236,208,171,126,82,48,24,8,1,0
37	253,234,197,152,110,74,45,23,8,1,0
38	253,239,218,191,152,105,67,35,10,1,0
39	252,236,203,160,116,78,47,25,10,1,0
40	254,242,217,181,136,90,54,26,6,1,0
41	252,238,213,177,139,100,64,36,14,1,0
42	252,245,237,228,217,204,188,166,137,66,0
43	249,228,197,156,110,69,39,18,4,1,0
44	254,240,210,168,119,79,48,24,6,1,0
45	217,191,171,150,127,102,75,49,25,2,0
46	253,245,231,209,180,144,100,55,16,1,0
47	254,245,225,195,157,113,67,28,5,1,0
48	254,232,199,161,121,85,55,31,13,1,0
49	253,241,213,173,127,84,51,26,8,1,0
50	254,242,216,179,134,87,50,22,3,1,0
51	254,245,223,185,138,94,57,28,7,1,0
52	253,244,229,209,182,147,107,66,28,2,0
53	254,244,222,192,154,105,60,26,4,1,0
54	253,233,196,153,111,74,44,22,7,1,0
55	253,236,211,182,144,99,60,30,8,1,0

表18 (续)

宽带特征向量维度	标识概率统计值
56	254,247,228,196,153,107,67,34,9,1,0

8.4 解码神经网络

8.4.1 语法

AVS3P10解码神经网络信息，见附录B。解码神经网络语法WbNNNetDec()，见表19。

表19 解码神经网络语法

语法	位数	助记符
WbNNNetDec() {		
DecConvLayer1()		
DecBlocks1()		
DecBlocks2()		
DecBlocks3()		
DecBlocks4()		
DecPostProc()		
DecConvLayer2()		
}		

8.4.2 语义

DecConvLayer1()

第一级解码卷积神经网络，基于56维宽带特征向量量化值，获得第一级解码向量。

DecBlocks1()

解码块神经网络1，基于第一级解码向量，获得第一级解码块向量。

DecBlocks2()

解码块神经网络2，基于第一级解码块向量，获得第二级解码块向量。

DecBlocks3()

解码块神经网络3，基于第二级解码块向量，获得第三级解码块向量。

DecBlocks4()

解码块神经网络4，基于第三级解码块向量，获得第二级解码向量。

DecPostProc()

后处理解码神经网络，基于第二级解码向量，获得第三级解码向量。

DecConvLayer2()

第二级解码卷积神经网络，基于第三级解码向量，获得单声道宽带重建信号或者超宽带低频子带重建信号。

8.4.3 解码过程

8.4.3.1 概述

本节规定了宽带解码神经网络的解码过程，见图9，包含图5至8中的第一级解码卷积神经网络、多级解码块神经网络、后处理神经网络和第二级解码卷积神经网络。每个箭头右侧的[x, y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。宽带解码神经网络，将解码获得

的56维宽带特征向量量化值（WbEmbedding），转换成宽带重建信号或者超宽带低频子带重建信号。解码过程涉及的解码神经网络结构描述性信息和网络权值参数，见本文件附录B。

8.4.3.2 第一级解码卷积神经网络

将宽带特征向量量化值WbEmbedding，经过1维卷积运算DecConvLayer1 ()后，生成第一级解码向量DecVector1。宽带特征向量量化值的第一维度是神经网络通道数，设置为56；宽带特征向量量化值的第二维度是样本数，设置为1。第一级解码向量的第一维度是神经网络通道数，设置为256；第一级解码向量的第二维度是样本数，设置为1。第一级解码卷积神经网络包含因果卷积，对于特定的卷积核，需要前向填充对应长度的零值，即填充数（Padding）。



图9 宽带解码神经网络结构图

第一级解码卷积神经网络的卷积配置见表20。

表20 第一级解码卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	56

表20 (续)

网络参数	参数配置
输出神经网络通道数	256
卷积核大小	1x3
步长	1
填充数	2

8.4.3.3 解码块神经网络 1

多级解码块神经网络处理包括4个解码块神经网络。将第一级解码向量DecVector1，经过解码块神经网络1（DecBlocks1）运算后，生成第一级解码块向量DecBlockVector1。第一级解码向量的第一维度是神经网络通道数，设置为256；第一级解码向量的第二维度是样本数，设置为1。第一级解码块向量的第一维度是神经网络通道数，设置为128；第一级解码块向量的第二维度是样本数，设置为5。

解码块神经网络1的结构，见图10。每个箭头右侧的[x, y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。

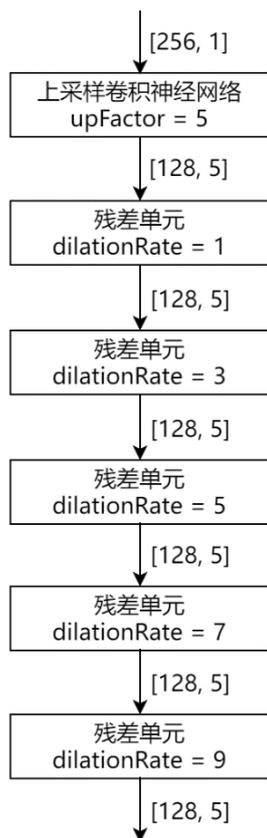


图10 解码块神经网络 1 结构图

解码块神经网络1的第一部分是上采样卷积神经网络。上采样卷积神经网络包括一维因果卷积，和上采样因子（UpFactor）为5的上采样模块。其中，上采样通过复制，增加向量第二维度的数据量；复制的倍数由上采样因子指定。

解码块神经网络1的上采样卷积神经网络的卷积配置见表21。

表21 解码块神经网络1的上采样卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	256
输出神经网络通道数	128
卷积核大小	1x3
步长	1
填充数	2
上采样因子	5

解码块神经网络1的上采样卷积神经网络后，级连5个不同空洞率（DilationRate）的残差单元。其中，任意一个残差单元，包含一个指定空洞率的空洞卷积和一个一维因果卷积再加一个跳连接。残差单元示意图，见图11。残差单元的输入（Input [x, y]）和输出（Output [x, y]）维度信息一致。

解码块神经网络1包含的5个残差单元，见图10，空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积，根据设置的分组数，将输入通道分成若干组，只在每个组内，输入和输出通道产生关联。该空洞卷积的配置见表22。

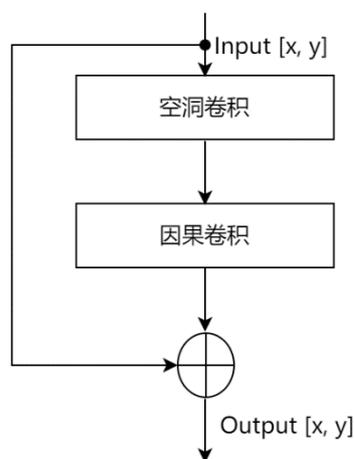


图11 残差单元示意图

表22 解码块神经网络1的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	128
输出神经网络通道数	128
卷积核大小	1x5
步长	1

表22 (续)

网络参数	参数配置
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	4
激活函数	PReLU

所有残差单元中的因果卷积配置见表23。

表23 解码块神经网络1的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	128
输出神经网络通道数	128
卷积核大小	1x1
步长	1
填充数	0
激活函数	PReLU

8.4.3.4 解码块神经网络2

将第一级解码块向量DecBlockVector1，经过解码块神经网络2（DecBlocks2）运算后，生成第二级解码块向量DecBlockVector2。第一级解码向量的第一维度是神经网络通道数，设置为128；第一级解码向量的第二维度是样本数，设置为5。第二级解码块向量的第一维度是神经网络通道数，设置为64；第二级解码块向量的第二维度是样本数，设置为20。

解码块神经网络2的结构，见图12。每个箭头右侧的[x, y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。

解码块神经网络2的第一部分是上采样卷积神经网络。上采样卷积神经网络包括一维因果卷积，和上采样因子上采样因子为4的上采样模块。其中，上采样通过复制，增加向量第二维度的数据量；复制的倍数由上采样因子指定。

解码块神经网络2的上采样卷积神经网络的卷积配置见表24。

表24 解码块神经网络2的上采样卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	128
输出神经网络通道数	64
卷积核大小	1x3
步长	1
填充数	2
上采样因子	4

解码块神经网络2包含的5个残差单元，见图12，空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积，根据设置的分组数，将输入通道分成若干组，只在每个组内，输入和输出通道产生关联。该空洞卷积的配置见表25。

表25 解码块神经网络 2 的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	64
输出神经网络通道数	64
卷积核大小	1x5
步长	1
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	4
激活函数	PReLU

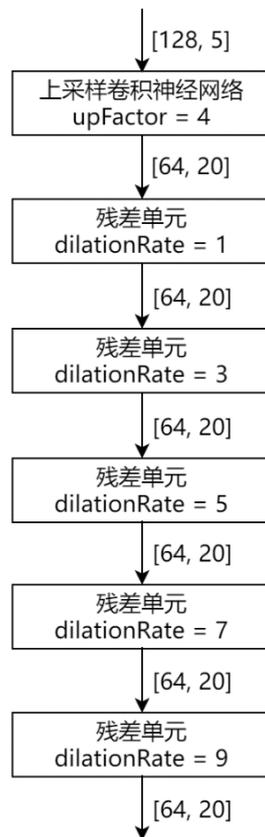


图12 解码块神经网络 2 结构图

所有残差单元中的因果卷积配置见表26。

表26 解码块神经网络 2 的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	64
输出神经网络通道数	64
卷积核大小	1x1
步长	1
填充数	0
激活函数	PReLU

8.4.3.5 解码块神经网络 3

将第二级解码块向量DecBlockVector2，经过解码块神经网络3（DecBlocks3）运算后，生成第三级解码块向量DecBlockVector3。第二级解码向量的第一维度是神经网络通道数，设置为64；第二级解码向量的第二维度是样本数，设置为20。第三级解码块向量的第一维度是神经网络通道数，设置为64；第三级解码块向量的第二维度是样本数，设置为80。

解码块神经网络3的结构，见图13。每个箭头右侧的[x, y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。

解码块神经网络3的第一部分是上采样卷积神经网络。上采样卷积神经网络包括一维因果卷积，和上采样因子为4的上采样模块。其中，上采样通过复制，增加向量第二维度的数据量；复制的倍数由上采样因子指定。

解码块神经网络3的上采样卷积神经网络的卷积配置见表27。

表27 解码块神经网络 3 的上采样卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	64
输出神经网络通道数	32
卷积核大小	1x3
步长	1
填充数	2
上采样因子	4

解码块神经网络3包含的5个残差单元，见图13，空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积，根据设置的分组数，将输入通道分成若干组，只在每个组内，输入和输出通道产生关联。该空洞卷积的配置见表28。

表28 解码块神经网络 3 的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	32
输出神经网络通道数	32
卷积核大小	1x5
步长	1

表28 (续)

网络参数	参数配置
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	2
激活函数	PReLU

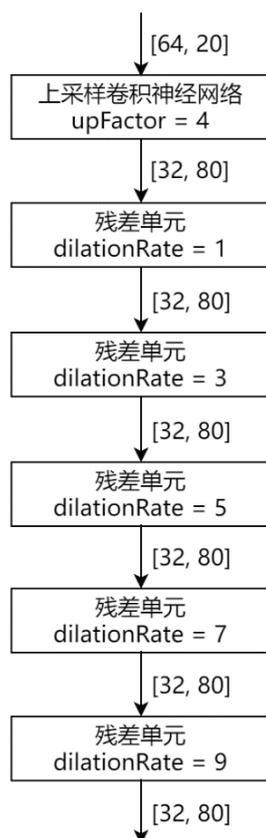


图13 解码块神经网络3 结构图

所有残差单元中的因果卷积配置如表29。

表29 解码块神经网络3 的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	32
输出神经网络通道数	32
卷积核大小	1x1
步长	1
填充数	0
上采样因子	2

8.4.3.6 解码块神经网络 4

将第三级解码块向量DecBlockVector3，经过解码块神经网络4（DecBlocks4）运算后，生成第二级解码向量DecVector2。第三级解码块向量的第一维度是神经网络通道数，设置为32；第三级解码块向量的第二维度是样本数，设置为80。第二级解码向量的第一维度是神经网络通道数，设置为16；第二级解码向量的第二维度是样本数，设置为160。

解码块神经网络4的结构，见图14。每个箭头右侧的[x, y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。

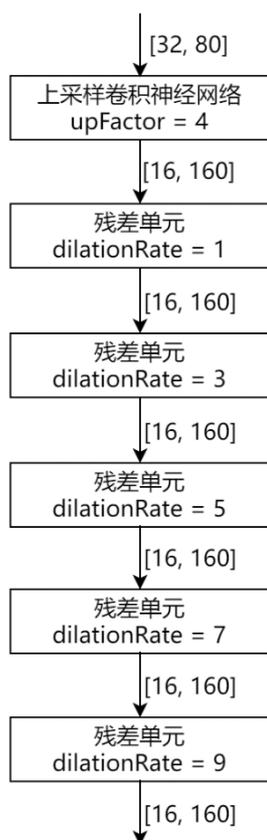


图14 解码块神经网络 4 结构图

解码块神经网络4的第一部分是上采样卷积神经网络。上采样卷积神经网络包括一维因果卷积，和上采样因子为2的上采样模块。其中，上采样通过复制，增加向量第二维度的数据量；复制的倍数由上采样因子指定。

解码块神经网络4的上采样卷积神经网络的卷积配置见表30。

表30 解码块神经网络 4 的上采样卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	32
输出神经网络通道数	16
卷积核大小	1x3

表30 (续)

网络参数	参数配置
步长	1
填充数	2
上采样因子	2

解码块神经网络4包含的5个残差单元，见图14，空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积，根据设置的分组数，将输入通道分成若干组，只在每个组内，输入和输出通道产生关联。该空洞卷积的配置见表31。

表31 解码块神经网络4的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	16
卷积核大小	1x5
步长	1
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	2
激活函数	PReLU

所有残差单元中的因果卷积配置见表32。

表32 解码块神经网络4的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	16
卷积核大小	1x1
步长	1
填充数	0
激活函数	PReLU

8.4.3.7 后处理神经网络

将第二级解码向量DecVector2，经过后处理解码神经网络(DecPostProc)运算后，生成第三级解码向量DecVector3。第二级解码向量的第一维度是神经网络通道数，设置为16；第二级解码向量的第二维度是样本数，设置为160。第三级解码向量的第一维度是神经网络通道数，设置为16；第三级解码向量的第二维度是样本数，设置为320。

后处理神经网络是一个上采样卷积神经网络。上采样卷积神经网络包括一维因果卷积，和上采样因子为2的上采样模块。其中，上采样通过复制，增加向量第二维度的数据量；复制的倍数由上采样因子指定。后处理神经网络的卷积配置见表33。

表33 后处理神经网络配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	16
卷积核大小	1x3
步长	1
填充数	2
上采样因子	2
激活函数	PReLU

8.4.3.8 第二级解码卷积神经网络

将第三级解码向量DecVector3，经过第二级解码卷积神经网络（DecConvLayer2）运算后，生成单声道宽带重建信号或者超宽带低频子带重建信号。第三级解码向量的第一维度是神经网络通道数，设置为16；第三级解码向量的第二维度是样本数，设置为320。单声道宽带重建信号或者超宽带低频子带重建信号的第一维度是神经网络通道数，设置为1；单声道宽带重建信号或者超宽带低频子带重建信号的第二维度是样本数，设置为320。

第二级解码卷积神经网络的卷积配置见表34。

表34 第二级解码卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	1
卷积核大小	1x3
步长	1
填充数	2

对于单声道宽带编码模式，单声道宽带重建信号即为解码器输出。对于单声道超宽带编码模式，还需要进行后续章节的解码流程，获得单声道超宽带重建信号。

对于立体声宽带编码模式或者立体声超宽带编码模式，单声道宽带重建信号即为左声道宽带重建信号。还需要进行后续章节的解码流程，获得立体声宽带重建信号或者立体声超宽带重建信号。

8.5 MCR 立体声解码

8.5.1 语法

解析立体声特征索引值，见表35。

表35 解析立体声特征索引值语法

语法	位数	助记符
DecodeMcrInforBits() {		
for (i = 0; i < 4; i++) {		
VqIdx[0][i]	4	uimsbf

表35 (续)

语法	位数	助记符
VqIdx[1][i]	4	uimsbf
}		
}		

立体声解码和逆量化，见表36。

表36 立体声解码和逆量化语法

语法	位数	助记符
McrInvQuanCod() {		
for (i = 0; i < 6; i++) {		
McrAngle0 [0][i]		
McrAngle0 [1][i]		
McrAngle1 [0][i]		
McrAngle1 [1][i]		
}		
}		

8.5.2 语义

VqIdx[0][i]	偶数 MDCT 频谱对应的 MCR 参数矢量中，第 i 个子矢量的矢量量化索引
VqIdx[1][i]	奇数 MDCT 频谱对应的 MCR 参数矢量中，第 i 个子矢量的矢量量化索引
McrAngle0[0][i]	第 1 子帧偶数 MDCT 频谱对应的第 i 个子带 MCR 参数量化值
McrAngle0[1][i]	第 1 子帧奇数 MDCT 频谱对应的第 i 个子带 MCR 参数量化值
McrAngle1[0][i]	第 2 子帧偶数 MDCT 频谱对应的第 i 个子带 MCR 参数量化值
McrAngle1[1][i]	第 2 子帧奇数 MDCT 频谱对应的第 i 个子带 MCR 参数量化值
McrDec()	MCR 立体声解码函数

8.5.3 解码过程

MCR立体声解码过程McrDec，参考GB/T 33475.3中7.4 couple_channel_element()的实现，但调整为320点MDCT变换配置，采用新的MCR子带边界配置和MCR参数码表。

每一帧分成两子帧，进行两次MCR立体声解码。每次MCR立体声解码，将左右声道的MDCT频谱分为奇频谱和偶频谱两部分，各分为6个子带，子带边界和子带频点数量定义要求见表37。

表37 MCR 子带边界配置

子带ID	起始索引idx_st	结束索引idx_ed
0	0	9
1	10	19
2	20	29

表37 (续)

子带ID	起始索引idx_st	结束索引idx_ed
3	30	44
4	45	59
5	60	79

立体声特征索引，需要通过熵解码，获得8个MCR参数索引值。熵解码所需的统计概率统计值见表38。

表38 MCR 参数概率统计值

标识概率统计值
240, 224, 208, 192, 176, 160, 144, 128, 112, 96, 80, 64, 48, 32, 16, 0

每个MCR参数索引值，可以获得相邻3个子带的MCR参数量化值，码表见表39。

表39 MCR 参数码表

索引值	量化值
0	-0.005, -0.005, -0.005
1	0.0015, 0.0015, 0.0015
2	0.035, 0.035, 0.035
3	0.035, 0.035, 0.07
4	-0.0015, -0.005, -0.005
5	-0.005, -0.005, -0.0015
6	0.035, 0.07, 0.035
7	-0.005, -0.0015, -0.005
8	0.07, 0.035, 0.035
9	0.035, 0.07, 0.07
10	0.015, 0.035, 0.035
11	-0.005, -0.005, -0.015
12	-0.015, -0.005, -0.005
13	0.07, 0.035, 0.07
14	0.07, 0.07, 0.07
15	-0.005, -0.015, -0.005

8.6 带宽扩展解码

8.6.1 语法

解析超宽带特征索引值，见表40。

表40 解析超宽带特征索引值语法

语法	位数	助记符
DecodeSwbInforBits () {		
for (i = 0; i < 8; i++) {		
SwbIndices [i]	1;2;3;4;5	uimsbf
}		
for (i = 0; i < 4; i++) {		
SwbIndices [8+i]	4	uimsbf
}		
if (RateMode >= 2) {		
for (i=0; i<16; i++) {		
SwbIndices [12+i]	1;2	uimsbf
}		
if (RateMode > 2) {		
for (i=0; i<16; i++) {		
SwbIndices [28+i]	1;2	uimsbf
}		
}		
}		

解析右声道相关的超宽带特征索引值，见表41。

表41 解析右声道相关的超宽带特征索引值语法

语法	位数	助记符
DecodeSwbRInforBits () {		
for (i = 0; i < 8; i++) {		
SwbRIndices [i]	1;2;3;4	uimsbf
}		
}		

超宽带解码和逆量化语法，见表42。

表42 超宽带解码和逆量化语法

语法	位数	助记符
SwbInvQuanCod() {		
for (i = 0; i < 8; i++) {		
BweEnvelope[i]		
}		
for (i = 0; i < 4; i++) {		
BweFlatness[i]		
}		
if (RateMode >= 2) {		

表42 (续)

语法	位数	助记符
for (i=0; i<16; i++) {		
BweEnvelopeR1[i]		
}		
if (RateMode > 2) {		
for (i=0; i<16; i++) {		
BweEnvelopeR2[i]		
}		
}		

右声道相关的超宽带解码和逆量化语法，见表43。

表43 右声道相关的超宽带解码和逆量化语法

语法	位数	助记符
SwbRInvQuanCod() {		
for (i = 0; i < 8; i++) {		
BweREnvelope[i]		
}		
}		

8.6.2 语义

SWBIndices	解析获得超宽带特征向量索引值，用于带宽扩展解码
SWBRIndices	解析获得右声道相关的超宽带特征向量索引值
BweEnvelope	高频 MDCT 子带谱包络
BweFlatenss	高频 MDCT 平坦度
BweEnvelopeR1	高频 MDCT 子带谱包络一级残差
BweEnvelopeR2	高频 MDCT 子带谱包络二级残差
BweREnvelopeDiff	右声道相关的高频 MDCT 子带谱包络残差
SwbBweDec()	带宽扩展解码函数
SwbBweRDec()	右声道相关的带宽扩展解码函数
QMFSynthesis()	QMF 合成滤波函数

8.6.3 解码过程

带宽扩展解码，仅限超宽带编码模式，重建的超宽带信号为32000Hz采样，对应的低频子带信号和高频子带信号为16000Hz采样。

对于单声道信号，低频子带信号，来自于8.4.3输出的超宽带低频子带重建信号。带宽扩展解码函数SwbBweDec()，调用DecodeSwbInforBits()从位流中，利用熵解码，解析出超宽带特征向量索引值SwbIndices用于后续解码模块解码。根据RateMode和表9定义的运行模式，超宽带特征向量信息长度，见表44。

表44 单声道超宽带编码模式的位率模式信息

编码位率模式	超宽带运行模式	SwbIndices 长度
0	SWB_Rate1	12
1	SWB_Rate2	12
2	SWB_Rate3	28
3	SWB_Rate4	44

熵解码具有唯一性。根据对应超宽带运行模式下包含的码表中的标识数量和统计概率，通过熵解码可以获得不同数量超宽带特征向量索引值。

当超宽带运行模式为SWB_Rate1和SWB_Rate2时，需要通过熵解码，获得8个高频MDCT子带谱包络索引值和4个高频MDCT平坦度索引值。熵解码所需的统计概率见表45和46。

表45 高频 MDCT 子带谱包络标识统计表

标识概率统计值
254, 253, 240, 206, 191, 168, 131, 116, 106, 97, 88, 79, 70, 61, 52, 43, 36, 29, 23, 18, 14, 11, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0

表46 高频 MDCT 平坦度标识统计表

标识概率统计值
127, 0

当超宽带运行模式为SWB_Rate3时，需要进一步通过熵解码，获得16个高频MDCT子带谱包络第一残差索引值。熵解码所需的统计概率见表47。

表47 高频 MDCT 子带谱包络第一残差和第二残差标识统计表

标识概率统计值
191, 127, 63, 0

当超宽带运行模式为SWB_Rate4时，需要进一步通过熵解码，获得16个高频MDCT子带谱包络第二残差索引值。熵解码所需的统计概率见表47。

超宽带解码和逆量化SwbInvQuanCod的输出为超宽带特征向量量化值。根据对应超宽带运行模式和超宽带特征向量索引值，通过对应的码表，获得对应的超宽带特征向量量化值。

当超宽带运行模式为SWB_Rate1和SWB_Rate2时，获得高频MDCT子带谱包络量化值和高频MDCT平坦度量化值，码表见表48和49。

表48 高频 MDCT 子带谱包络码表

量化值
-3.0, -2.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5

表49 高频 MDCT 平坦度码表

量化值
0, 1

当超宽带运行模式为SWB_Rate3时,进一步获得高频MDCT子带谱包络第一残差量化值,码表见表50。

表50 高频 MDCT 子带谱包络第一残差码表

量化值
-1.0, 0, 0.5, 1

当超宽带运行模式为SWB_Rate4时,进一步获得高频MDCT子带谱包络第二残差量化值,码表见表51。

表51 高频 MDCT 子带谱包络第二残差码表

量化值
-0.5, -0.25, 0, 0.25

根据表42中的语法,获得8个高频子带谱包络, $BweEnvelope$, 分别用于第一子帧和第二子帧的解码。带宽扩展解码基于320点MDCT变换。对于低频子带信号和高频子带信号,每帧包含320个样本点。每帧分成2个子帧,分别进行带宽扩展解码,完成高频子带信号的重建。

首先,计算低频子带信号的MDCT变换,见式(1)。

$$X_{LB}(k) = \sum_{n=0}^{319} x_{LB}(n) \times \cos\left(\frac{\pi}{160} \times (n + 0.5 + 80) \times (k + 0.5)\right) \quad (1)$$

式中:

x_{LB} ——低频子带信号;

X_{LB} ——对应的MDCT系数;

k ——MDCT系数索引值,范围是0至159。

此外,在MDCT之前,会对输入的时域信号进行加窗。窗函数见表52。

表52 MDCT 变换窗函数

MDCT 窗函数值				
0.0000185893186147,	0.0000486037914381,	0.0000927688434217,	0.0001545914301989,	0.0002380804864255,
0.0003477792867198,	0.0004887959888188,	0.0006668321962828,	0.0008882090673339,	0.0011598904899753,
0.0014895028542470,	0.0018853509650651,	0.0023564296541129,	0.0029124306679236,	0.0035637444323701,
0.0043214563216568,	0.0051973370927536,	0.0062038271840106,	0.0073540146193362,	0.0086616063066024,
0.0101408925705638,	0.0118067048161896,	0.0136743662774962,	0.0157596358692319,	0.0180786452235989,
0.0206478290609977,	0.0234838491119383,	0.0266035118761163,	0.0300236805735115,	0.0337611817105339,
0.0378327067509737,	0.0422547094461006,	0.0470432994399520,	0.0522141328239475,	0.0577823003687517,
0.0637622142101106,	0.0701674938085635,	0.0770108520398758,	0.0843039823032056,	0.0920574475569046,

表52 (续)

MDCT 窗函数值				
0.1002805722070290,	0.1089813377807450,	0.1181662833155482,	0.1278404113853892,	0.1380071006662601,
0.1486680259165248,	0.1598230862113012,	0.1714703422256777,	0.1836059633086797,	0.1962241850290194,
0.2093172778051547,	0.2228755271565480,	0.2368872260308000,	0.2513386795731948,	0.2662142226118255,
0.2814962500336617,	0.2971652601254847,	0.3131999108494487,	0.3295770889170314,	0.3462719914182706,
0.3632582196564186,	0.3805078847324590,	0.3979917243203169,	0.4156792299730271,	0.4335387842035685,
0.4515378064924636,	0.4696429072884857,	0.4878200489897543,	0.5060347128209521,	0.5242520704590991,
0.5424371592059419,	0.5605550594601741,	0.5785710732079099,	0.5964509022255281,	0.6141608246755578,
0.6316678687739247,	0.6489399822158286,	0.6659461960678338,	0.6826567818654332,	0.6990434006982884,
0.7150792431193708,	0.7307391587790614,	0.7459997747605440,	0.7608396016781199,	0.7752391266948591,
0.7891808927196848,	0.8026495631558977,	0.8156319716925321,	0.8281171567559753,	0.8400963803710784,
0.8515631313175482,	0.8625131126076963,	0.8729442134544646,	0.8828564660428210,	0.8922519875617956,
0.9011349080971681,	0.9095112851246163,	0.9173890054783417,	0.9247776757991240,	0.9316885025865936,
0.9381341630914126,	0.9441286683821094,	0.9496872200066174,	0.9548260617382377,	0.9595623279480006,
0.9639138901785685,	0.9678992035074742,	0.9715371542784315,	0.9748469107478829,	0.9778477781394401,
0.9805590595215044,	0.9829999238237467,	0.9851892821874428,	0.9871456737046944,	0.9888871614446633,
0.9904312394940361,	0.9917947515574048,	0.9929938214749322,	0.9940437958236645,	0.9949591985794346,
0.9957536976327569,	0.9964400827785989,	0.9970302546402955,	0.9975352238455668,	0.9979651196504936,
0.9983292071076030,	0.9986359117984107,	0.9988928510995576,	0.9991068709250405,	0.9992840868841527,
0.9994299288141757,	0.9995491876865389,	0.9996460639425623,	0.9997242163871434,	0.9997868108527277,
0.9998365679384202,	0.9998758092269605,	0.9999065014824681,	0.9999302984315374,	0.9999485798269191,
0.9999624875845040,	0.9999729588688779,	0.9999807560789712,	0.9999864937523628,	0.9999906624640352,
0.9999936498426482,	0.9999957588648086,	0.9999972236157885,	0.9999982227242898,	0.9999988906900084,
0.9999993273267994,	0.9999996055422485,	0.999999776673862,	0.9999998805392335,	0.9999999395247821,
0.9999999716588406,	0.9999999880507447,	0.9999999956969708,	0.999999988188357,	0.999999998272185,
0.999999998272185,	0.999999988188357,	0.9999999956969708,	0.9999999880507447,	0.9999999716588406,
0.9999999395247821,	0.9999998805392335,	0.999999776673862,	0.9999996055422485,	0.9999993273267994,
0.9999988906900084,	0.9999982227242898,	0.9999972236157885,	0.9999957588648086,	0.9999936498426482,
0.9999906624640352,	0.9999864937523628,	0.9999807560789712,	0.9999729588688779,	0.9999624875845040,
0.9999485798269191,	0.9999302984315374,	0.9999065014824681,	0.9998758092269605,	0.9998365679384202,
0.9997868108527277,	0.9997242163871434,	0.9996460639425623,	0.9995491876865389,	0.9994299288141757,
0.9992840868841527,	0.9991068709250405,	0.9988928510995576,	0.9986359117984107,	0.9983292071076030,
0.9979651196504936,	0.9975352238455668,	0.9970302546402955,	0.9964400827785989,	0.9957536976327569,
0.9949591985794346,	0.9940437958236645,	0.9929938214749322,	0.9917947515574048,	0.9904312394940361,
0.9888871614446633,	0.9871456737046944,	0.9851892821874428,	0.9829999238237467,	0.9805590595215044,
0.9778477781394401,	0.9748469107478829,	0.9715371542784315,	0.9678992035074742,	0.9639138901785685,
0.9595623279480006,	0.9548260617382377,	0.9496872200066174,	0.9441286683821094,	0.9381341630914126,
0.9316885025865936,	0.9247776757991240,	0.9173890054783417,	0.9095112851246163,	0.9011349080971681,
0.8922519875617956,	0.8828564660428210,	0.8729442134544646,	0.8625131126076963,	0.8515631313175482,
0.8400963803710784,	0.8281171567559753,	0.8156319716925321,	0.8026495631558977,	0.7891808927196848,
0.7752391266948591,	0.7608396016781199,	0.7459997747605440,	0.7307391587790614,	0.7150792431193708,

表52 (续)

MDCT 窗函数值				
0.6990434006982884,	0.6826567818654332,	0.6659461960678338,	0.6489399822158286,	0.6316678687739247,
0.6141608246755578,	0.5964509022255281,	0.5785710732079099,	0.5605550594601741,	0.5424371592059419,
0.5242520704590991,	0.5060347128209521,	0.4878200489897543,	0.4696429072884857,	0.4515378064924636,
0.4335387842035685,	0.4156792299730271,	0.3979917243203169,	0.3805078847324590,	0.3632582196564186,
0.3462719914182706,	0.3295770889170314,	0.3131999108494487,	0.2971652601254847,	0.2814962500336617,
0.2662142226118255,	0.2513386795731948,	0.2368872260308000,	0.2228755271565480,	0.2093172778051547,
0.1962241850290194,	0.1836059633086797,	0.1714703422256777,	0.1598230862113012,	0.1486680259165248,
0.1380071006662601,	0.1278404113853892,	0.1181662833155482,	0.1089813377807450,	0.1002805722070290,
0.0920574475569046,	0.0843039823032056,	0.0770108520398758,	0.0701674938085635,	0.0637622142101106,
0.0577823003687517,	0.0522141328239475,	0.0470432994399520,	0.0422547094461006,	0.0378327067509737,
0.0337611817105339,	0.0300236805735115,	0.0266035118761163,	0.0234838491119383,	0.0206478290609977,
0.0180786452235989,	0.0157596358692319,	0.0136743662774962,	0.0118067048161896,	0.0101408925705638,
0.0086616063066024,	0.0073540146193362,	0.0062038271840106,	0.0051973370927536,	0.0043214563216568,
0.0035637444323701,	0.0029124306679236,	0.0023564296541129,	0.0018853509650651,	0.0014895028542470,
0.0011598904899753,	0.0008882090673339,	0.0006668321962828,	0.0004887959888188,	0.0003477792867198,
0.0002380804864255,	0.0001545914301989,	0.0000927688434217,	0.0000486037914381,	0.0000185893186147

然后,选择MDCT系数中的第60到149点的系数,逆序后,复制两遍,生成高频子带信号相关的MDCT系数。伪码如下

```

for i = 0 To 79    // 第一遍复制
    XHB(159 - i) = XLB(60 + i)
end
for i = 0 To 79    // 第二遍复制
    XHB(79 - i) = XLB(60 + i)
end

```

其中, X_{HB} 为高频子带信号相关的MDCT系数。

进一步,对复制产生的高频子带信号相关的MDCT系数进行平滑处理。平滑处理,将高频子带的MDCT系数均分成2部分,每部分包含80个MDCT系数。对于每80个MDCT系数,当且仅当前文解码获得的高频MDCT平坦度(BweFlatenss)为真时,进行平滑处理;否则,保持不变。平滑处理基于滑动平均方式,伪码如下

```

for i = 0 To 79
    PowAve = 0
    for j = i-1 To i+1    //注: XHB(-1) = XLB(159)
        PowAve = PowAve + XHB(j)*XHB(j)
    end
    PowAve = √(PowAve/3)
    if XHB(i) ≠ 0
        XHB(i) = XHB(i)/PowAve
    end
end
end

```

```

for i = 80 To 159
    PowAve = 0
    for j = i-1 To i+1
        PowAve = PowAve + XHB(j)*XHB(j)
    end
    PowAve =  $\sqrt{(PowAve/3)}$ 
    if XHB(i) ≠ 0
        XHB(i) = XHB(i)/PowAve
    end
end
end

```

基于熵解码和逆量化，获得8个高频子带谱包络BweEnvelope，分别用于第一子帧和第二子帧的解码。对于高频子带信号的160点MDCT系数，分成4个子带，每个子带的索引信息见表53。

表53 高频子带信号 MDCT 系数分带信息

子带ID	起始索引idx_st	结束索引idx_ed
0	0	19
1	20	59
2	60	99
3	100	159

对于RateMode ≥ 2时，基于熵解码和逆量化，获得16个高频子带谱包络一级残差BweEnvelopeR1，分别用于第一子帧和第二子帧的解码。对于RateMode > 2时，基于熵解码和逆量化，获得16个高频子带谱包络二级残差BweEnvelopeR2，分别用于第一子帧和第二子帧的解码。

对于第一子帧或者第二子帧，获得对应高频子带谱包络的伪码如下

```

for i = 0 To 3
    env[2*i] = BweEnvelope[i]
    env[2*i+1] = BweEnvelope[i]
end
if RateMode ≥ 2
    for i = 0 To 7
        env[i] = env[i] + BweEnvelopeR1[i]
    end
end
if RateMode > 2
    for i = 0 To 7
        env[i] = env[i] + BweEnvelopeR2[i];
    end
end
for i = 0 To 7
    env[i] = 2env[i]
end
end

```

对通过复制和平滑处理生成的高频子带信号相关的MDCT系数,计算每个子带谱包络参考值, $env0$ 。子带谱包络参考值的计算过程,见C.7中的公式11。对于第一子帧或者第二子帧,当RateMode为0或者1时,高频子带谱包络的参考值数量为4;否则,数量为8。然后,根据高频子带谱包络和子带谱包络参考值,获得子带增益值。伪码如下

```

if RateMode < 2
  for i = 0 To 3
    g[2*i] = env[2*i] / env0[2*i]
    g[2*i+1] = g[2*i]
  end
else
  for i = 0 To 3
    g[2*i] = env[2*i] / env0[2*i]
    g[2*i+1] = env[2*i+1] / env0[2*i+1]
  end
end

```

对高频 MDCT 系数,使用对应子带谱包络进行增益控制。伪码如下

```

for i = 0 To 3
  for j = idx_st[i] To (idx_ed[i]/2)-1
    XHB(j) = XHB[j] * g[2*i]
  end
  for j = idx_ed[i]/2 To idx_ed[i]-1
    XHB(j) = XHB[j] * g[2*i+1]
  end
end

```

最后,进行MDCT逆变换,获得高频子带信号,见式(2)。

$$x_{HB}(n) = \frac{1}{160} \sum_{k=0}^{159} X_{HB}(k) \times \cos\left(\frac{\pi}{160} \times (n + 0.5 + 80) \times (k + 0.5)\right) \quad (2)$$

式中:

n ——信号样本点,范围是0至319。

此外,在MDCT逆变换之后,会对输出的时域信号进行加窗,获得高频子带重建信号。窗函数见表52。

对于立体声信号,低频子带信号,来自于8.5.3输出的超宽带低频子带重建信号。左声道相关的高频子带重建信号的生成过程,与单声道超宽带解码过程一致。

对于右声道相关的高频子带重建信号的生成,需调用SwbBweRDec()。首先,DecodeSwbRInforBits()从位流中,利用熵解码,解析出右声道超宽带特征向量索引值SwbRIndices用于后续解码模块解码。根据RateMode和表11定义的运行模式,右声道相关的超宽带特征向量信息长度,见表54。

表54 右声道相关的超宽带编码模式的位率模式信息

编码位率模式	超宽带运行模式	SwbIndices 长度
0	SWB_Rate1	8
1	SWB_Rate2	8

熵解码具有唯一性。见表54,根据对应超宽带运行模式下包含的码表中的标识数量和统计概率,通过熵解码可以获得不同长度右声道相关的超宽带特征向量索引值。

当超宽带运行模式为SWB_Rate1和SWB_Rate2时,需要通过熵解码,获得8个右声道相关的超宽带特征向量索引值。熵解码所需的统计概率表见表55。进一步,获得右声道相关的超宽带特征向量量化值,码表见表56。

表55 右声道相关的超宽带特征向量索引值统计表

标识概率统计值
238, 234, 226, 197, 157, 115, 71, 33,16, 8, 5, 4, 3, 2, 1, 0

表56 右声道相关的超宽带特征向量码表

量化值
-0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1

右声道相关的超宽带特征向量量化值,反映了左右声道对应的高频MDCT子带谱包络的残差。通过左声道相关的高频MDCT子带谱包络, $envL$, 和右声道相关的超宽带特征向量量化值,可以获得左声道相关的高频MDCT子带谱包络, $envR$, 见公式(3)。

$$envR(i) = envL(i) \times 2^{BweREnvelopDiff(i)}, i = 0,1,2,3 \quad (3)$$

基于右声道相关的低频子带信号和高频子带谱包络,可以获得右声道相关的高频子带重建信号。过程与单声道高频子带重建信号类似。因此,对于立体声超宽带编码模式,可获得左右声道对应的高频子带重建信号。

对于单声道超宽带编码模式或者立体声超宽带编码模式,每个声道的高频重建信号的进一步处理流程类似。特别地,对高频子带重建信号需要进行频谱翻折,见公式(4)。

$$x_{HB}(n) = (-1)^n \times x_{HB}(n) \quad (4)$$

对于低频子带重建信号和高频子带重建信号,分别进行升2采样后通过使用QMF合成滤波QMFsynthesis(),生成单声道超宽带重建信号或者立体声超宽带重建信号。

QMF(Quadrature Mirror Filter)是一个包含分析-合成的滤波器对。对于QMF分析滤波器,可以将输入的采样频率32000Hz的超宽带信号分解成两路子带信号后,分别进行降2采样后生成采样频率16000Hz的低频子带信号和高频子带信号。QMF分析滤波器的低通部分定义为 h_{Low} ,则QMF分析滤波器的高通部分可通过 h_{Low} 计算获得,见公式(5)。

$$h_{High}(n) = -1^k \times h_{Low}(n) \quad (5)$$

将QMF分析滤波器低通部分和高通部分的频响分别定义 $H_{Low}(z)$ 和 $H_{High}(z)$,则分别可以获得QMF合成滤波器的低通部分 $G_{Low}(z)$ 和高通部分 $G_{High}(z)$ 的频响,见公式(6)和(7)。

$$G_{Low}(z) = H_{Low}(z) \quad (6)$$

$$G_{High}(z) = (-1) \times H_{High}(z) \quad (7)$$

最后,对低频子带重建信号和高频子带重建信号进行QMF合成滤波,输出单声道超宽带重建信号或者立体声超宽带重建信号。QMF合成滤波需要原型滤波器参数,见表57。

表57 QMF 原型滤波器参数

滤波器参数			
0.457257900,	0.142022000,	-0.096089540,	-0.055443150,
0.054974440,	0.031331460,	-0.038449870,	-0.019830160,
0.029128720,	0.013062040,	-0.022925640,	-0.008630004,
0.018350770,	0.005559824,	-0.014772120,	-0.003382441,
0.011865500,	0.001816576,	-0.009445686,	-0.000708161,
0.007416155,	-0.000057608,	-0.005704768,	0.000551125,
0.004262906,	-0.000844748,	-0.003064472,	0.000983427,
0.002074592,	-0.001019671,	-0.001257778,	0.001079806

附录 A
(规范性)
ARSTF 语法及语义

A.1 ARSTF语法

A.1.1 概述

ARSTF包含同步字和解码必需的信息，是一种流传输格式。同步字使得解码器无需确定的起始点即可进行解码。

A.1.2 ARSTF语法

ARSTF包含同步字和解码必需的信息，其语法见表A.1。

表A.1 arstf_sequence() 语法

语法	位数	助记符
arstf_sequence()		
{		
arstf_frame()		
}		

ARSTF帧描述了音频传输帧的位流序列，该序列由相应的同步字、解码头信息、错误校验和原始数据块组成。arstf_frame() 语法见表A.2。

表A.2 arstf_frame() 语法

语法	位数	助记符
arstf_frame()		
{		
syncword	12	bslbf
arstf_frame_header()		
RtcSpeechRawDataBlock()		
arstf_error_check()		
}		

ARSTF解码头描述了解码头信息，由声道数、带宽模式、位率模式等组成。arstf_frame_header()语法见表A.3。

表A.3 arstf_frame_header() 语法

语法	位数	助记符
arstf_frame_header()		
{		
NumChans	1	Bslbf

表A.3 (续)

语法	位数	助记符
BandwidthMode	1	Uimsbf
BitRateIndex	2	Uimsbf
Reserved	4	Bslbf
}		

ARSTF错误校验描述了ARSTF的CRC数据。为在传输过程中快速查验位流正确性，减少延时，`arstf_error_check()`用于查验帧头的全部位以及所有原始数据块的前半部分，即其查验长度为 $\text{raw_frame_length}/2$ 。所有查验位根据其出现顺序进入CRC算法中。`arstf_error_check()`语法见表A.4。

表A.4 `arstf_error_check()` 语法

语法	位数	助记符
<code>arstf_error_check()</code>		
{		
crc_check	8	rpchof
}		

A.2 ARSTF语义

NumChans	1比特， ‘0’：单声道 ‘1’：立体声
BandwidthMode	1比特： ‘0’：16000 Hz 采样频率宽带信号 ‘1’：32000 Hz 采样频率超宽带信号
BitRateIndex	2比特 ‘0’：RateMode = 1 ‘1’：RateMode = 2 ‘2’：RateMode = 3 ‘3’：RateMode = 4
reserved	4比特，用于将来对本文件扩展的保留字段
<code>arstf_sequence()</code>	符合AVS_RTC_Speech_Transport_Format格式的序列，见表A.1
<code>arstf_frame()</code>	ARSTF帧，见表A.2
<code>arstf_frame_header()</code>	ARSTF的解码帧头，见表A.3
<code>raw_frame_length</code>	16比特，标识ARSTF帧的帧长

RtcSpeechRawDataBlock	实时通信语音编码原始数据块
arstf_error_check()	CRC 校验生成的数据。校验 ARSTF 帧的前一半数据位，并根据其出现的顺序进入 CRC-16 算法，见表 A.4
syncword	同步字，比特串 1111 1111 1111 (0xFFF)
crc_check	CRC 校验

T/AI 109.10-2024

附 录 B
(规范性)
AVS3P10 解码神经网络信息

本文件附录B规定了AVS3P10解码神经网络的结构描述性信息和网络权值参数，由于参数量大，为不影响文本阅读且方便使用，特提供名为AVS3P10_RM0_tnne_decoder.bin的数字文件，下载地址为：

http://standard.avscn.org.cn/avs3_download/index.asp

其中，解码神经网络数字文档的基本信息包括：

第1-3行，为文件头信息。包含推理库名称（tnne）、模型总层数（99）、训练平台（pytorch），如图B.1所示。

第4行，留空。

第5-134行，为解码神经网络结构的描述性信息，见8.4。

第135行，留空。

第136行至文件结束行，为解码神经网络的权值，以float32二进制格式保存。

```

1  tnne
2  num_layers: 99
3  training: pytorch
4
5  tnne layers
6  pad_1      input_0      56x1x1   56x1x3   share  share  padding:0,0,0,2,0,0,0,0  CacheMode
7  conv2d_2   pad_1      56x1x3   256x1x1  alloc  share  kernel:1x3
8
9  // DecoderBlock1
10 pad_3      conv2d_2   256x1x1  256x1x3  alloc  share  padding:0,0,0,2,0,0,0,0  CacheMode
11 conv2d_4   pad_3      256x1x3  128x1x1  alloc  share  kernel:1x3
12 repeat_5   conv2d_4   128x1x1  128x1x5  alloc  share  split      expand:1,1,5
13
14 pad_6      repeat_5   128x1x5  128x1x9  alloc  share  padding:0,0,0,4,0,0,0,0  CacheMode
15 conv2d_7   pad_6      128x1x9  128x1x5  alloc  share  actn_prelu kernel:1x5  groups:4  dilation:1x1
16 conv2d_8   conv2d_7   128x1x5  128x1x5  alloc  share  actn_prelu kernel:1x1
17 eltwise_9  conv2d_8|repeat_5  128x1x5  128x1x5  alloc  share  op_type:1
18

```

图B.1 文件 AVS3P10_RM0_tnne_decoder.bin 中的前几行示例

附录 C (资料性) AVS3P10 参考编码器

C.1 概述

本附录给出了AVS3P10编码器的示例。

编码器根据指定的NumChans选择实时通信语音单声道编码(RtcSpeechMonoEnc)或者实时通信语音立体声编码(RtcSpeechStereoEnc)。对于单声道编码,编码器根据指定的BandwidthMode选择单声道宽带编码(RtcSpeechMonoWbEnc)或者单声道超宽带编码(RtcSpeechMonoSwbEnc)。对于立体声编码,编码器根据指定的BandwidthMode选择立体声宽带编码(RtcSpeechStereoWbEnc)或者立体声超宽带编码(RtcSpeechStereoSwbEnc)。

单声道宽带编码,输入的单声道宽带信号采样频率为16000Hz,每帧包含320样本点。对输入信号经过宽带编码神经网络处理后,获得56维宽带特征向量。编码神经网络包含第一级编码卷积神经网络、预处理神经网络、多级编码块神经网络、第二级编码卷积神经网络等。对宽带特征向量进行标量量化和熵编码后,写入位流。

单声道超宽带编码,输入的单声道超宽带信号采样频率为32000Hz,每帧包含640样本点。对输入信号进行QMF分析滤波并进行降2采样,分别获得低频子带信号和高频子带信号。对低频子带信号进行类似单声道宽带编码过程,获得56维宽带特征向量。对低频子带信号和高频子带信号进行带宽扩展编码,获得超宽带特征向量。对宽带特征向量和超宽带特征向量进行标量量化和熵编码后,写入位流。

立体声宽带编码,输入的立体声宽带信号采样频率为16000Hz,每声道每帧包含320样本点。对输入左声道信号经过宽带编码神经网络处理后,获得56维宽带特征向量。编码神经网络包含第一级编码卷积神经网络、预处理神经网络、多级编码块神经网络、第二级编码卷积神经网络等。对左右声道信号进行MCR立体声编码,提取MCR参数。对宽带特征向量和MCR参数进行对应的标量量化或者矢量量化后,经熵编码后,写入位流。

立体声超宽带编码,输入的单声道超宽带信号采样频率为32000Hz,每声道每帧包含640样本点。对左右声道信号分别进行QMF分析滤波并进行降2采样,分别获得低频子带信号和高频子带信号。对左声道对应的低频子带信号进行类似单声道宽带编码过程,获得56维宽带特征向量。对左右声道信号进行MCR立体声编码,提取MCR参数。分别对左右声道对应的低频子带信号和高频子带信号进行带宽扩展编码,获得超宽带特征向量。对宽带特征向量、MCR参数和超宽带特征向量进行对应的标量量化或矢量量化,经熵编码后,写入位流。

AVS3P10编码器单声道模式流程图见图C.1。AVS3P10编码器单声道编码模式由QMF分析滤波器(仅适用于超宽带编码模式)、卷积神经网络、预处理神经网络、多线编码块神经网络、带宽扩展、量化和熵编码等构成核心编码器,将输入信号编码为位流。

AVS3P10编码器立体声模式流程图见图C.2。AVS3P10编码器立体声模式由QMF分析滤波器(仅适用于超宽带编码模式)、卷积神经网络、预处理神经网络、多线编码块神经网络、MCR立体声编码、带宽扩展、量化和熵编码等构成核心编码器,将输入信号编码为位流。

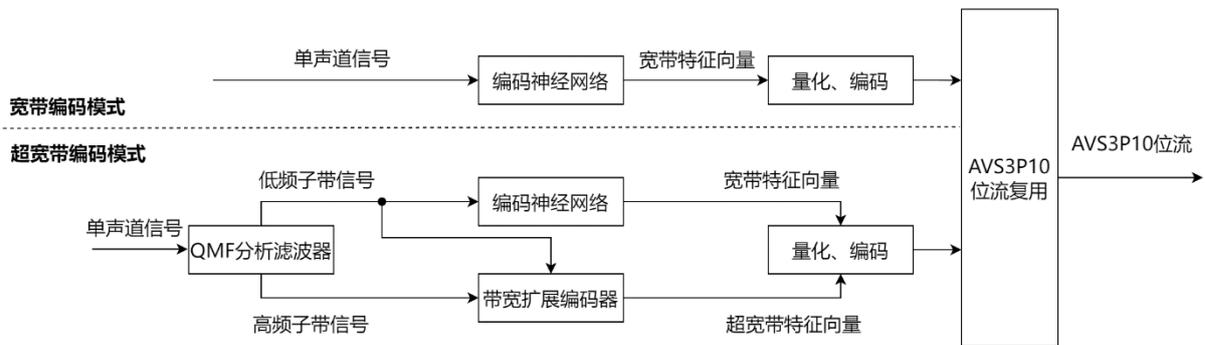
AVS3P10参考编码器所需编码神经网络的结构描述性信息和网络权值参数,由于参数量大,为不影响文本阅读且方便使用,特提供名为AVS3P10_RM0_tmne_encoder.bin的数字文件,下载地址为:

http://standard.avs.org.cn/avs3_download/index.asp

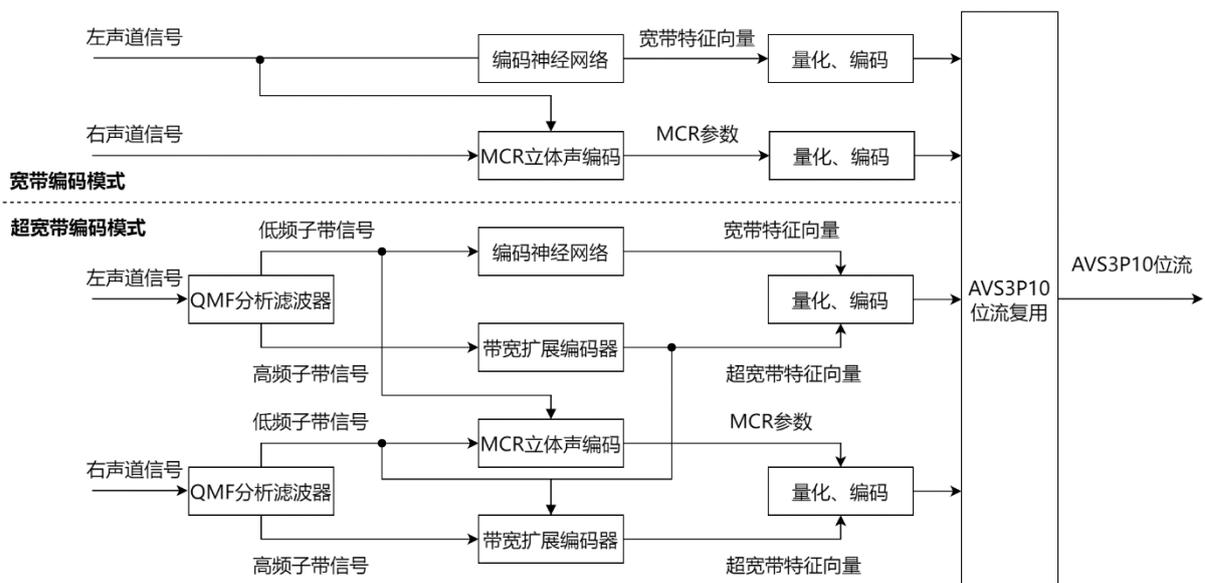
C.2 AVS3P10 编码通用信息

C.2.1 概述

AVS3P10编码器可以分为QMF分析滤波器（适用于超宽带编码模式）、第一级编码卷积神经网络、预处理神经网络、多级编码块神经网络、第二级编码卷积神经网络、MCR立体声编码、带宽扩展、量化和熵编码等模块。根据输入信号声道数、采样频率等编码模式参数，调用相关模块进行编码。以下对各个模式编解码模块进行介绍。



图C.1 AVS3P10 编码器单声道模式流程图



图C.2 AVS3P10 编码器立体声模式流程图

单声道宽带编码对于单声道宽带编码模式，单声道编码器将时域单声道信号，经过编码神经网络，获得宽带特征向量。最后经过量化和熵编码得到位流。其中，编码神经网络是分层的深度神经网络模块，包括卷积神经网络、预处理神经网络、多级编码块神经网络等子模块。

C.2.2 单声道宽带编码

a) 第一级编码卷积神经网络

第一级编码卷积神经网络将输入的宽带时域信号进行一次一维因果卷积计算，生成16通道神经网络

络向量。对于采样频率 16000Hz 的宽带信号，每帧长度为 20ms，包含 320 个样本点。卷积神经网络网络，将一维时域单声道信号，转换成二维的第一级编码向量。其中，第一级编码向量的第一维度是神经网络通道数，设置为 16；第一级编码向量的第二维度是样本数，设置为 320。

——输入：单声道宽带信号

——输出：第一级编码向量

b) 预处理神经网络

预处理神经网络，将第一级编码向量进行一次一维因果卷积计算并激活后，进行因子为 2 的池化操作，生成第二级编码向量。池化操作，可以减少特定维度上的数据量。第二级编码向量的第一维度是神经网络通道数，设置为 16；第二级编码向量的第二维度是样本数，设置为 160。

——输入：第一级编码向量

——输出：第二级编码向量

c) 多级编码块神经网络

将第二级编码向量，经过 4 级编码块神经网络，生成第三级编码向量。每一级编码块神经网络包含多层神经网络计算。每一级编码块神经网络通过计算，将输入向量的通道数增加 1 倍，同时通过不同因子的池化，降低输出向量第二维度的数据量。第三级编码向量的第一维度是神经网络通道数，经过 4 级编码块神经网络，由 16 增加为 256；第三级编码向量的第二维度是样本数，设置为 1。多级编码块神经网络的模块细节见 C.3.4 至 C.3.7 节。

——输入：第二级编码向量

——输出：第三级编码向量

d) 第二级编码卷积神经网络

将第三级编码向量，经过 1 维卷积运算并激活后，生成宽带特征向量。宽带特征向量的第一维度是神经网络通道数，设置为 56；宽带特征向量的第二维度是样本数，设置为 1。

——输入：第三级编码向量

——输出：宽带特征向量

e) 量化和熵编码

量化模块对宽带特征向量进行线性标量量化，编码模块对量化结果进行熵编码以得到位流。

——输入：宽带特征向量

——输出：位流

C.2.3 单声道超宽带编码

对于单声道超宽带编码模式，单声道编码器将单声道超宽带信号，经过 QMF 分析滤波器组，生成低频子带信号和高频子带信号。对于低频子带信号，经过编码神经网络，获得宽带特征向量。经过带宽扩展编码，对低频子带信号和高频子带信号进行分析，获得超宽带特征向量。最后经过量化和熵编码得到位流。其中，编码神经网络是分层的深度神经网络模块，包括卷积神经网络、预处理神经网络、多级编码块神经网络等子模块。

a) QMF 分析滤波器

QMF 分析滤波器，即 Quadrature Mirror Filters（正交镜像滤波器）。对于采样频率 32000Hz 的超宽带信号，每帧长度为 20ms，包含 640 个样本点，经过分析滤波和降 2 采样后，生成低频子带信号和高频子带信号。低频子带信号和高频子带信号的长度设置为 320。

——输入：超宽带单声道信号

——输出：低频子带信号和高频子带信号

b) 第一级编码卷积神经网络

第一级编码卷积神经网络将输入的低频子带信号进行一次一维因果卷积计算，生成 16 通道神经网络

络向量。对于采样频率 16000Hz 的宽带信号，每帧长度为 20ms，包含 320 个样本点。卷积神经网络网络，将一维低频子带信号，转换成二维的第一级编码向量。其中，第一级编码向量的第一维度是神经网络通道数，设置为 16；第一级编码向量的第二维度是样本数，设置为 320。

- 输入：低频子带信号
- 输出：第一级编码向量

c) 预处理神经网络

预处理神经网络，将第一级编码向量进行一次一维因果卷积计算并激活后，进行因子为2的池化操作，生成第二级编码向量。池化操作，可以减少特定维度上的数据量。第二级编码向量的第一维度是神经网络通道数，设置为16；第二级编码向量的第二维度是样本数，设置为160。

- 输入：第一级编码向量
- 输出：第二级编码向量

d) 多级编码块神经网络

将第二级编码向量，经过4级编码块神经网络，生成第三级编码向量。每一级编码块神经网络包含多层神经网络计算。每一级编码块神经网络通过计算，将输入向量的通道数增加1倍，同时通过不同因子的池化，降低输出向量第二维度的数据量。第三级编码向量的第一维度是神经网络通道数，经过4级编码块神经网络，由16增加为256；第三级编码向量的第二维度是样本数，设置为1。多级编码块神经网络的模块细节见C.3.4至C.3.7节。

- 输入：第二级编码向量
- 输出：第三级编码向量

e) 第二级编码卷积神经网络

将第三级编码向量，经过1维卷积运算并激活后，生成宽带特征向量。宽带特征向量的第一维度是神经网络通道数，设置为56；宽带特征向量的第二维度是样本数，设置为1。

- 输入：第三级编码向量
- 输出：宽带特征向量

f) 带宽扩展编码

将低频带子信号和高频子带信号，经过带宽扩展编码，获得超宽带特征向量。带宽扩展编码分别对低频子带信号和高频子带信号，进行2次长度320点MDCT变换，提取子带包络信息和频谱平坦度信息，合并为超宽带特征向量。

- 输入：低频带子信号和高频子带信号
- 输出：超宽带特征向量

g) 量化和熵编码

量化模块对宽带特征向量和超宽带特征向量进行线性标量量化，编码模块对量化结果进行熵编码以得到位流。

- 输入：宽带特征向量和超宽带特征向量
- 输出：位流

C.2.4 立体声宽带编码

对于立体声宽带编码模式，编码器将左声道信号，经过编码神经网络，获得宽带特征向量。将左右声道信号，经过MCR立体声编码，获得MCR参数。最后经过量化和熵编码得到位流。其中，编码神经网络是分层的深度神经网络模块，包括卷积神经网络、预处理神经网络、多级编码块神经网络等子模块。

a) 第一级编码卷积神经网络

第一级编码卷积神经网络将输入的宽带时域信号进行一次一维因果卷积计算，生成 16 通道神经网络

络向量。对于采样频率 16000Hz 的宽带信号，每帧长度为 20ms，包含 320 个样本点。卷积神经网络网络，将一维时域单声道信号，转换成二维的第一级编码向量。其中，第一级编码向量的第一维度是神经网络通道数，设置为 16；第一级编码向量的第二维度是样本数，设置为 320。

——输入：左声道宽带信号

——输出：第一级编码向量

b) 预处理神经网络

预处理神经网络，将第一级编码向量进行一次一维因果卷积计算并激活后，进行因子为 2 的池化操作，生成第二级编码向量。池化操作，可以减少特定维度上的数据量。第二级编码向量的第一维度是神经网络通道数，设置为 16；第二级编码向量的第二维度是样本数，设置为 160。

——输入：第一级编码向量

——输出：第二级编码向量

c) 多级编码块神经网络

将第二级编码向量，经过 4 级编码块神经网络，生成第三级编码向量。每一级编码块神经网络包含多层神经网络计算。每一级编码块神经网络通过计算，将输入向量的通道数增加 1 倍，同时通过不同因子的池化，降低输出向量第二维度的数据量。第三级编码向量的第一维度是神经网络通道数，经过 4 级编码块神经网络，由 16 增加为 256；第三级编码向量的第二维度是样本数，设置为 1。多级编码块神经网络的模块细节见 C.3.4 至 C.3.7 节。

——输入：第二级编码向量

——输出：第三级编码向量

d) 第二级编码卷积神经网络

将第三级编码向量，经过 1 维卷积运算并激活后，生成宽带特征向量。宽带特征向量的第一维度是神经网络通道数，设置为 56；宽带特征向量的第二维度是样本数，设置为 1。

——输入：第三级编码向量

——输出：宽带特征向量

e) MCR 立体声编码

将左右声道信号，经过 MCR 立体声编码，获得 MCR 参数。

——输入：左右声道宽带信号

——输出：MCR 参数

f) 量化和熵编码

量化模块对宽带特征向量、MCR 参数进行对应的标量量化或矢量量化，编码模块对量化结果进行熵编码以得到位流。

——输入：宽带特征向量和 MCR 参数

——输出：位流

C.2.5 立体声超宽带编码

对于立体声超宽带编码模式，编码器分别将左右声道超宽带信号，经过 QMF 分析滤波器组，生成对应的低频子带信号和高频子带信号。对于左声道相关的低频子带信号，经过编码神经网络，获得宽带特征向量。将左右声道信号，经过 MCR 立体声编码，获得 MCR 参数。经过带宽扩展编码，分别对左右声道对应的低频子带信号和高频子带信号进行分析，获得超宽带特征向量。最后经过量化和熵编码得到位流。其中，编码神经网络是分层的深度神经网络模块，包括卷积神经网络、预处理神经网络、多级编码块神经网络等子模块。

a) QMF 分析滤波器

QMF分析滤波器，即Quadrature Mirror Filters（正交镜像滤波器）。对于采样频率32000Hz的超宽带信号，每帧长度为20ms，包含640个样本点，经过分析滤波和降2采样后，生成低频子带信号和高频子带信号。低频子带信号和高频子带信号的长度设置为320。

——输入：超宽带单声道信号

——输出：低频带子信号和高频子带信号

b) 第一级编码卷积神经网络

第一级编码卷积神经网络将输入的低频子带信号进行一次一维因果卷积计算，生成16通道神经网络向量。对于采样频率16000Hz的宽带信号，每帧长度为20ms，包含320个样本点。卷积神经网络网络，将一维低频子带信号，转换成二维的第一级编码向量。其中，第一级编码向量的第一维度是神经网络通道数，设置为16；第一级编码向量的第二维度是样本数，设置为320。

——输入：低频子带信号

——输出：第一级编码向量

c) 预处理神经网络

预处理神经网络，将第一级编码向量进行一次一维因果卷积计算并激活后，进行因子为2的池化操作，生成第二级编码向量。池化操作，可以减少特定维度上的数据量。第二级编码向量的第一维度是神经网络通道数，设置为16；第二级编码向量的第二维度是样本数，设置为160。

——输入：第一级编码向量

——输出：第二级编码向量

d) 多级编码块神经网络

将第二级编码向量，经过4级编码块神经网络，生成第三级编码向量。每一级编码块神经网络包含多层神经网络计算。每一级编码块神经网络通过计算，将输入向量的通道数增加1倍，同时通过不同因子的池化，降低输出向量第二维度的数据量。第三级编码向量的第一维度是神经网络通道数，经过4级编码块神经网络，由16增加为256；第三级编码向量的第二维度是样本数，设置为1。多级编码块神经网络的模块细节见C.3.4至C.3.7节。

——输入：第二级编码向量

——输出：第三级编码向量

e) 第二级编码卷积神经网络

将第三级编码向量，经过1维卷积运算并激活后，生成宽带特征向量。宽带特征向量的第一维度是神经网络通道数，设置为56；宽带特征向量的第二维度是样本数，设置为1。

——输入：第三级编码向量

——输出：宽带特征向量

f) MCR 立体声编码

将左右声道信号，经过MCR立体声编码，获得MCR参数。

——输入：左右声道宽带信号

——输出：MCR参数

g) 带宽扩展编码

将左右声道对应的低频带子信号和高频子带信号，经过带宽扩展编码，获得超宽带特征向量。带宽扩展编码分别对低频子带信号和高频子带信号，进行2次长度320点MDCT变换，提取子带包络信息和频谱平坦度信息，合并为超宽带特征向量。

——输入：左右声道对应的低频带子信号和高频子带信号

——输出：超宽带特征向量

h) 量化和熵编码

量化模块对宽带特征向量、MCR参数和超宽带特征向量进行对应的标量量化或矢量量化，编码模块对量化结果进行熵编码以得到位流。

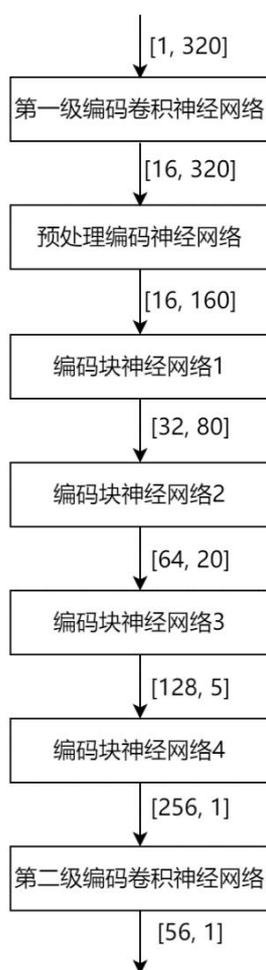
——输入：宽带特征向量、MCR参数和超宽带特征向量

——输出：位流

C.3 宽带编码神经网络

C.3.1 概述

宽带编码神经网络结构图，见图C.3。每个箭头右侧的 $[x, y]$ 符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。宽带编码神经网络，将单声道宽带信号或者超宽带低频子带信号，转换成56维宽带特征向量。图C.3显示了多级编码块神经网络由4个单独的编码块神经网络组成，每个箭头右侧的 $[x, y]$ 符号的第一维度神经网络通道数和第二维度样本数。



图C.3 宽带编码神经网络结构图

C.3.2 第一级编码卷积神经网络

将单声道宽带信号或者超宽带低频子带信号,经过1维卷积运算后,生成第一级编码向量EncVector1。单声道宽带信号或者超宽带低频子带信号的第一维度是神经网络通道数,设置为1;单声道宽带信号或者超宽带低频子带信号的第二维度是样本数,设置为320。第一级编码向量的第一维度是神经网络通道数,设置为16;第一级编码向量的第二维度是样本数,设置为320。第一级编码卷积神经网络包含因果卷积,对于特定的卷积核,需要前向填充对应长度的零值,即填充数(Padding)。

第一级编码卷积神经网络的卷积配置见表C.1。

表C.1 第一级编码卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	1
输出神经网络通道数	16
卷积核大小	1x3
步长	1
填充数	2

C.3.3 预处理神经网络

将第一级编码向量EncVector2,经过预理解码神经网络运算后,生成第二级编码向量EncVector2。第一级编码向量的第一维度是神经网络通道数,设置为16;第一级编码向量的第二维度是样本数,设置为320。第二级编码向量的第一维度是神经网络通道数,设置为16;第二级编码向量的第二维度是样本数,设置为160。

预处理神经网络是一个下采样卷积神经网络。下采样卷积神经网络包括一维因果卷积,和下采样因子(dnFactor)为2的池化(Pooling)模块。其中,池化增加向量第二维度的数据量;倍数由下采样因子指定。预处理神经网络的卷积配置见表C.2。

表C.2 预处理神经网络配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	16
卷积核大小	1x3
步长	1
填充数	2
池化因子	2
激活函数	PReLU

C.3.4 编码块神经网络1

多级编码块神经网络处理包括4个编码块神经网络。将第二级编码向量EncVector2,经过编码块神经网络1运算后,生成第一级编码块向量EncBlockVector1。第二级编码向量的第一维度是神经网络通道数,设置为16;第二级编码向量的第二维度是样本数,设置为160。第一级编码块向量的第一维度是神经网络通道数,设置为32;第一级编码块向量的第二维度是样本数,设置为80。

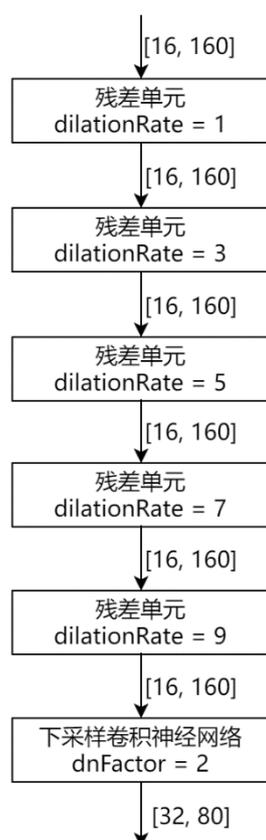
编码块神经网络1的结构，见图C.4。每个箭头右侧的[x, y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。

编码块神经网络1级连5个不同空洞率（DilationRate）的残差单元。其中，任意一个残差单元，包含一个指定空洞率的空洞卷积和一个一维因果卷积再加一个跳连接。残差单元示意图，见图11。残差单元的输入（Input [x, y]）和输出（Output [x, y]）维度信息一致。

编码块神经网络1包含的5个残差单元，见图C.4，空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积，根据设置的分组数，将输入通道分成若干组，只在每个组内，输入和输出通道产生关联。该空洞卷积的配置见表C.3。

表C.3 编码块神经网络1的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	16
卷积核大小	1x5
步长	1
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	2
激活函数	PReLU



图C.4 编码块神经网络1结构图

所有残差单元中的因果卷积配置见表C.4。

表C.4 编码块神经网络1的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	16
卷积核大小	1x1
步长	1
填充数	0
激活函数	PReLU

编码块神经网络1的5个残差单元处理后，是下采样卷积神经网络。下采样卷积神经网络包括一维因果卷积，和下采样因子（dnFactor）为2的池化（Pooling）模块。其中，池化增加向量第二维度的数据量；倍数由下采样因子指定。编码块神经网络1的下采样处理神经网络配置见表C.5。

表C.5 编码块神经网络1的下采样处理神经网络配置参数

网络参数	参数配置
输入神经网络通道数	16
输出神经网络通道数	32
卷积核大小	1x3
步长	1
填充数	2
池化因子	2
激活函数	PReLU

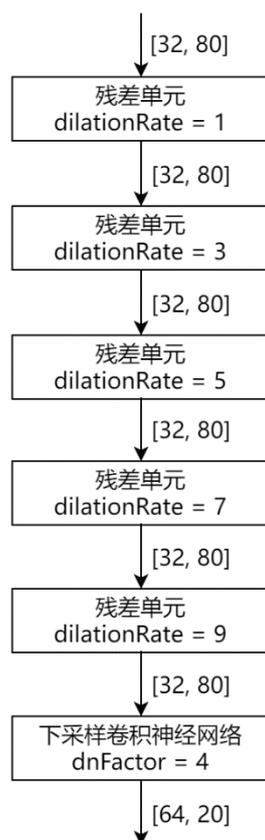
C.3.5 编码块神经网络2

将第一级编码块向量EncBlockVector1，经过编码块神经网络2运算后，生成第二级编码块向量EncBlockVector2。第一级编码块向量的第一维度是神经网络通道数，设置为32；第一级编码块向量的第二维度是样本数，设置为80。第二级编码块向量的第一维度是神经网络通道数，设置为64；第二级编码块向量的第二维度是样本数，设置为20。

编码块神经网络2的结构，见图C.5。每个箭头右侧的[x,y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。

编码块神经网络2级连5个不同空洞率（DilationRate）的残差单元。其中，任意一个残差单元，包含一个指定空洞率的空洞卷积和一个一维因果卷积再加一个跳连接。残差单元示意图，见图11。残差单元的输入（Input [x, y]）和输出（Output [x, y]）维度信息一致。

编码块神经网络2包含的5个残差单元，见图C.5，空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积，根据设置的分组数，将输入通道分成若干组，只在每个组内，输入和输出通道产生关联。该空洞卷积的配置见表C.6。



图C.5 编码块神经网络 2 结构图

表C.6 编码块神经网络 2 的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	32
输出神经网络通道数	32
卷积核大小	1x5
步长	1
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	4
激活函数	PReLU

所有残差单元中的因果卷积配置见表C.7。

表C.7 编码块神经网络 2 的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	32
输出神经网络通道数	32
卷积核大小	1x1

表C.7 (续)

网络参数	参数配置
步长	1
填充数	0
激活函数	PReLU

编码块神经网络2的5个残差单元处理后,是下采样卷积神经网络。下采样卷积神经网络包括一维因果卷积,和下采样因子(dnFactor)为4的池化(Pooling)模块。其中,池化增加向量第二维度的数据量;倍数由下采样因子指定。编码块神经网络2的下采样处理神经网络配置见表C.8。

表C.8 编码块神经网络2的下采样处理神经网络配置参数

网络参数	参数配置
输入神经网络通道数	32
输出神经网络通道数	64
卷积核大小	1x3
步长	1
填充数	2
池化因子	4
激活函数	PReLU

C.3.6 编码块神经网络3

将第二级编码块向量EncBlockVector2,经过编码块神经网络3运算后,生成第三级编码块向量EncBlockVector3。第二级编码块向量的第一维度是神经网络通道数,设置为64;第二级编码块向量的第二维度是样本数,设置为20。第三级编码块向量的第一维度是神经网络通道数,设置为128;第三级编码块向量的第二维度是样本数,设置为5。

编码块神经网络3的结构,见图C.6。每个箭头右侧的[x,y]符号表示当前向量的维度,其中,第一维度是神经网络通道数,第二维度是样本数。

编码块神经网络3级连5个不同空洞率(DilationRate)的残差单元。其中,任意一个残差单元,包含一个指定空洞率的空洞卷积和一个一维因果卷积再加一个跳连接。残差单元示意图,见图11。残差单元的输入(Input [x,y])和输出(Output [x,y])维度信息一致。

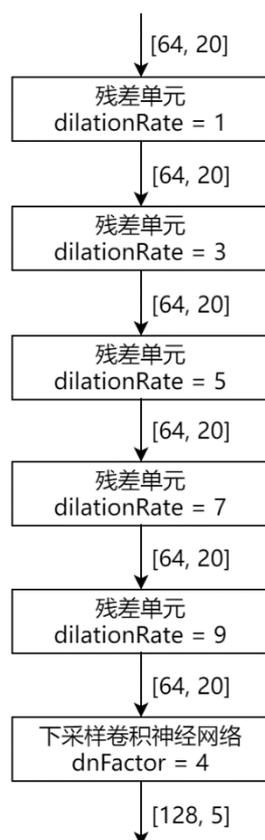
编码块神经网络3包含的5个残差单元,见图C.6,空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积,根据设置的分组数,将输入通道分成若干组,只在每个组内,输入和输出通道产生关联。该空洞卷积的配置见表C.9。

表C.9 编码块神经网络3的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	64
输出神经网络通道数	64
卷积核大小	1x5
步长	1

表C.9 (续)

网络参数	参数配置
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	4
激活函数	PReLU



图C.6 编码块神经网络3结构图

所有残差单元中的因果卷积配置见表C.10。

表C.10 编码块神经网络3的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	64
输出神经网络通道数	64
卷积核大小	1x1
步长	1
填充数	0
激活函数	PReLU

编码块神经网络3的5个残差单元处理后，是下采样卷积神经网络。下采样卷积神经网络包括一维因果卷积，和下采样因子（dnFactor）为4的池化（Pooling）模块。其中，池化增加向量第二维度的数据量；倍数由下采样因子指定。编码块神经网络3的下采样处理神经网络配置见表C.11。

表C.11 编码块神经网络3的下采样处理神经网络配置参数

网络参数	参数配置
输入神经网络通道数	64
输出神经网络通道数	128
卷积核大小	1x3
步长	1
填充数	2
池化因子	4
激活函数	PReLU

C.3.7 编码块神经网络4

将第三级编码块向量EncBlockVector3，经过编码块神经网络4运算后，生成第三级编码向量EncVector3。第三级编码块向量的第一维度是神经网络通道数，设置为128；第三级编码块向量的第二维度是样本数，设置为5。第三级编码向量的第一维度是神经网络通道数，设置为256；第三级编码向量的第二维度是样本数，设置为1。

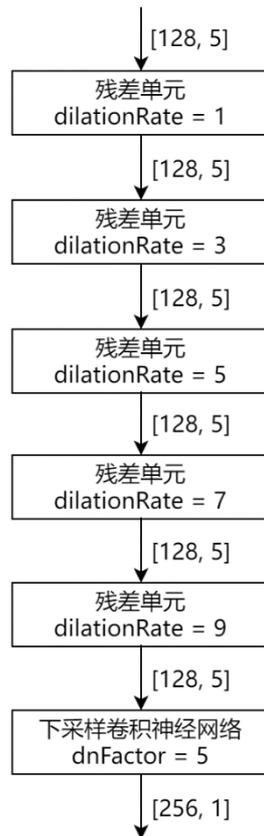
编码块神经网络4的结构，见图C.7。每个箭头右侧的[x, y]符号表示当前向量的维度，其中，第一维度是神经网络通道数，第二维度是样本数。

编码块神经网络3级连5个不同空洞率（DilationRate）的残差单元。其中，任意一个残差单元，包含一个指定空洞率的空洞卷积和一个一维因果卷积再加一个跳连接。残差单元示意图，见图11。残差单元的输入（Input [x, y]）和输出（Output [x, y]）维度信息一致。

编码块神经网络3包含的5个残差单元，见图C.7，空洞率分别是1、3、5、7、9。每个残差单元中包含的空洞卷积设置为分组卷积，根据设置的分组数，将输入通道分成若干组，只在每个组内，输入和输出通道产生关联。该空洞卷积的配置见表C.12。

表C.12 编码块神经网络4的残差单元中空洞卷积配置参数

网络参数	参数配置
输入神经网络通道数	128
输出神经网络通道数	128
卷积核大小	1x5
步长	1
填充数	5个残差单元的填充数分别是4、12、20、28、36
分组数	4
激活函数	PReLU



图C.7 编码块神经网络 4 结构图

所有残差单元中的因果卷积配置见表C.13。

表C.13 编码块神经网络 4 的残差单元中因果卷积配置参数

网络参数	参数配置
输入神经网络通道数	128
输出神经网络通道数	128
卷积核大小	1x1
步长	1
填充数	0
激活函数	PReLU

编码块神经网络4的5个残差单元处理后，是下采样卷积神经网络。下采样卷积神经网络包括一维因果卷积，和下采样因子（dnFactor）为5的池化（Pooling）模块。其中，池化增加向量第二维度的数据量；倍数由下采样因子指定。编码块神经网络4的下采样处理神经网络配置见表C.14。

表C.14 编码块神经网络 4 的下采样处理神经网络配置参数

网络参数	参数配置
输入神经网络通道数	128

表C.14 (续)

网络参数	参数配置
输出神经网络通道数	256
卷积核大小	1x3
步长	1
填充数	2
池化因子	5
激活函数	PReLU

C.3.8 第二级编码卷积神经网络

将第三级编码向量EncVector3，经过第二级编码卷积神经网络，生成宽带特征向量。第三级编码向量的第一维度是神经网络通道数，设置为256；第三级编码向量的第二维度是样本数，设置为1。宽带特征向量的第一维度是神经网络通道数，设置256；宽带特征向量的第二维度是样本数，设置为1。

第二级编码卷积神经网络的卷积配置见表C.15。

表C.15 第二级编码卷积神经网络配置参数

网络参数	参数配置
输入神经网络通道数	256
输出神经网络通道数	56
卷积核大小	1x5
步长	1
填充数	4
激活函数	Tanh

C.4 宽带量化编码和位流复用

根据系统指定的宽带编码位率模式，确认宽带运行模式，见表8至11，对56维宽带特征向量进行标量量化和熵编码。不同宽带运行模式的量化标识信息见表14；熵编码需使用的对应不同宽带运行模式的统计表见表15至18。然后，进行位流复用。

C.5 MCR立体声编码

立体声宽带编码，输入的立体声宽带信号采样频率为16000Hz，每帧包含320样本点。立体声超宽带编码，输入的立体声超宽带信号采样频率为32000Hz，每帧包含640样本点；对输入立体声超宽带信号左声道进行QMF分析滤波并降2采样，分别获得立体声低频子带信号和高频子带信号；QMF分析滤波的细节见8.6.3和表57。低频子带信号，每帧包含320样本点。MCR立体声编码，对立体声宽带信号或者立体声低频子带输入信号进行分析，提取MCR参数。每一帧分成2个子帧，分子帧完成MCR立体声编码。

首先，MCR立体声编码，对输入左右声道信号， x_{Left} 和 x_{Right} ，进行320点MDCT变换，获得MDCT系数， X_{Left} 和 X_{Right} 。MDCT变换计算过程见8.6.3。

将左右声道MDCT系数，分成奇偶频点两个部分， $\{X_{\text{Left}}^{(e)}, X_{\text{Left}}^{(o)}\}$ 和 $\{X_{\text{Right}}^{(e)}, X_{\text{Right}}^{(o)}\}$ 。对左右声道的奇偶频点MDCT系数，分成6个MCR子带，见表37。

对于每一个MCR子带对应的左右声道MDCT系数， X_{Left} 和 X_{Right} ，计算子带能量和互相关，见式(8)。

$$\begin{aligned} e_{\text{Left}}(i) &= \sum_{k=\text{idx_st}(i)}^{\text{idx_ed}(i)} X_{\text{Left}}(k)^2 \\ e_{\text{Right}}(i) &= \sum_{k=\text{idx_st}(i)}^{\text{idx_ed}(i)} X_{\text{Right}}(k)^2 \\ X_{\text{corr}}(i) &= \sum_{k=\text{idx_st}(i)}^{\text{idx_ed}(i)} X_{\text{Left}}(k) \times X_{\text{Right}}(k) \end{aligned} \quad (8)$$

式中：

i ——MCR子带数索引值，取值范围为0至5。

k ——MDCT系数索引值。

$\text{idx_st}(i)$ 和 $\text{idx_ed}(i)$ ——第 i 个MCR子带的起始和结束索引，见表37。

然后，获得左右声道MCR旋转角度，见公式(9)。

$$\theta(i) = 0.5 \times \text{atan} \left(\frac{0.5 \times (e_{\text{Right}}(i) - e_{\text{Left}}(i))}{X_{\text{corr}}(i)} \right) \quad (9)$$

最后，将相邻3个MCR子带的MCR旋转角度，合并成一个3维矢量，生成MCR参数。对MCR参数进行矢量量化，获得MCR参数量化值，码表见表39。

C.6 MCR参数量化编码和位流复用

对8个MCR参数量化索引值进行熵编码，标识统计表见表38。

最后，进行位流复用。

C.7 带宽扩展编码

单声道超宽带编码，输入的单声道超宽带信号采样频率为32000Hz，每帧包含640样本点。对输入信号进行QMF分析滤波分解成两路子带信号，分别进行降2采样后生成采样频率16000Hz低频子带信号和高频子带信号。低频子带信号和高频子带信号，每帧包含320样本点。QMF分析滤波的细节见8.6.3和表57。其中，高频子带信号需要进行频谱翻折，见公式(4)。

带宽扩展编码流程图，每一帧分成2个子帧，分子帧完成带宽扩展编码，见图C.8。

首先，对低频子带信号和高频子带信号， x_{LB} 和 x_{HB} ，进行320点MDCT变换，获得MDCT系数， X_{LB} 和 X_{HB} 。MDCT变换计算过程见8.6.3。

平坦度分析部分，分别计算低频MDCT系数和高频MDCT系数的平坦度。平坦度由MDCT功率谱的几何平均值和算术平均值的比值表征，见公式(10)。

$$Flatness_{LB} = \frac{\exp\left(\frac{1}{80} \sum_{k=0}^{79} \log_2(X_{LB}(k) \times X_{LB}(k))\right)}{\frac{1}{80} \times \sum_{k=0}^{79} X_{LB}(k) \times X_{LB}(k)}$$

$$Flatness_{HB} = \frac{\exp\left(\frac{1}{80} \sum_{k=0}^{79} \log_2(X_{HB}(k) \times X_{HB}(k))\right)}{\frac{1}{80} \times \sum_{k=0}^{79} X_{HB}(k) \times X_{HB}(k)}$$
(10)

式中：

k ——MDCT系数索引值。

基于低频MDCT系数和高频MDCT系数的平坦度，获得平坦度分析结果，伪码见下

$if (Flatness_{HB} < Flatness_{LB}) // (Flatness_{HB} < SF_THD)$

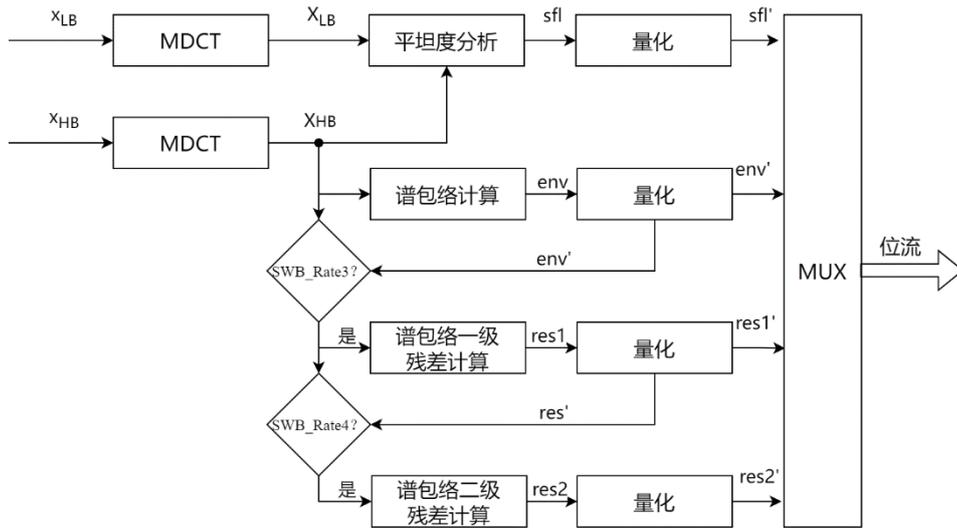
$Sfl = FALSE$

$else$

$Sfl = TRUE$

end

其中， sfl 为真，表示解码端需要进行相应的平坦化处理。SF_THD为预设阈值，0.50。每帧进行两次320点MDCT变换，因此，获得4个平坦化分析结果。



图C.8 带宽扩展编码流程图

谱包络计算部分，见表53，将高频MDCT系数分成4个子带；每个子带计算MDCT谱包络，见公式(11)。

$$env(i) = \frac{\log_2\left(\frac{\sum_{k=idx_st(i)}^{idx_ed(i)} X_{HB}(k) \times X_{HB}(k)}{idx_ed(i) - idx_st(i) + 1}\right)}{2}$$
(11)

式中：

i ——高频MDCT系数子带索引值，取值范围为0至3。

k ——MDCT系数索引值。

$idx_st(i)$ 和 $idx_ed(i)$ ——第 i 个高频MDCT系数子带的起始和结束索引，见表53。

每帧进行两次320点MDCT变换，因此，获得8个MDCT谱包络。对8个MDCT谱包络进行量化，获得对应MDCT谱包络量化值， env' 。码表见表48。

如果运行模式为SWB_Rate3,将表53中所示的每一个子带，内部进下分成两个子带，将160点MDCT系数分成8个子带。对8个子带中的任意一个子带，计算MDCT谱包络，并根据对应的MDCT谱包络量化值，获得MDCT谱包络一级残差， $res1$ 。对16个MDCT谱包络一级残差进行量化，获得对应MDCT谱包络一级残差量化值， $res1'$ 。码表见表50。

如果运行模式为SWB_Rate4,基于MDCT谱包络量化值和MDCT谱包络一级残差量化值，获得MDCT谱包络二级残差， $res2$ 。对16个MDCT谱包络二级残差进行量化，获得对应MDCT谱包络二级残差量化值， $res2'$ 。码表见表51。

对于立体声超宽带编码模式，左声道相关的带宽扩展编码与单声道超宽带编码一致；同时，获得左声道相关的谱包络量化值， $envL'$ 。右声道相关的谱包络， $envR$ ，计算过程见公式(11)。计算对应子带左右声道相关的谱包络的残差见公式(12)。

$$envD = envR - envL' \quad (12)$$

对8个右声道相关的谱包络残差进行量化，获得谱包络残差量化值， $envD'$ 。码表见表56。

C.8 超宽带量化编码和位流复用

对于单声道超宽带编码模式,根据系统指定的超宽带编码位率模式,确认超宽带运行模式,见表44,首先,对8个MDCT谱包络量化值进行熵编码,标识统计表见表45。然后,对4个平坦度分析结果进行熵编码,标识统计表见表46。

如果运行模式为SWB_Rate3,对16个MDCT谱包络一级残差量化值进行熵编码,标识统计表见表47。

如果运行模式为SWB_Rate4,对16个MDCT谱包络二级残差量化值进行熵编码,标识统计表见表47。

对于立体声超宽带编码模式,左声道相关的量化编码与单声道超宽带编码模式类似。右声道相关的谱包络残差量化值进行熵编码,标识统计表见表55。

最后,进行位流复用。

附录 D
(资料性)
丢包隐藏

丢包隐藏是一种基于已恢复语音帧预测丢失帧的解码方法。作为可选模块,仅在数据帧丢失时使用,且不涉及位流信息。

对于单声道宽带解码,不涉及丢包隐藏。当数据帧丢失时,将宽带特征向量量化值设置为零,进行单声道宽带解码。

对于立体声宽带解码或者立体声超宽带解码中MCR解码部分,当数据帧丢失时,将宽带解码获得的左声道信号对应的MDCT谱进行复制,获得右声道MDCT谱,并通过MDCT逆变换恢复右声道信号,完成丢包隐藏。

对于单声道超宽带解码或者立体声超宽带解码,当数据帧丢失时,基于上一帧的超宽带特征向量,完成丢包隐藏。具体地,复制上一帧高频MDCT平坦度作为当前帧高频MDCT平坦度。对上一帧高频子带谱包络和高频子带谱包络残差值,在对数域减去一个因子 de_factor ,抑制当前帧输出的高频子带重建信号能量。对于当前丢失帧中的第一子帧或者第二子帧,获得对应高频子带谱包络的伪码如下

```

for i = 0 To 3
    env[2*i] = BweEnvelope[i] - de_factor
    env[2*i+1] = BweEnvelope[i] - de_factor
end
if RateMode ≥ 2
    for i = 0 To 7
        env[i] = env[i] + BweEnvelopeR1[i] - de_factor
    end
end
if RateMode > 2
    for i = 0 To 7
        env[i] = env[i] + BweEnvelopeR2[i] - de_factor
    end
end
for i = 0 To 7
    env[i] = 2env[i]
end

```

其中,因子 de_factor 初始值为-0.32,并根据实际丢帧情况进一步负向成倍增加。例如,如果当前丢失帧为首次丢帧,则 de_factor 等于-0.32;如果当前丢失帧为连续第二次丢帧,则 de_factor 等于-0.64,以此类推。如果当前帧未丢失,则 de_factor 重置为-0.32。获得当前丢失帧的超宽带特征向量后,参考8.6.3中的带宽扩展流程,完成丢包隐藏。