

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日:

2005年10月13日(13.10.2005)

PCT

(10) 国际公布号:

WO 2005/096508 A1

(51) 国际分类号⁷: H03M 7/30 G10L 19/00

(21) 国际申请号: PCT/CN2004/001034

(22) 国际申请日: 2004年9月9日(09.09.2004)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
200410030947.3 2004年4月1日(01.04.2004) CN

(71) 申请人(对除美国以外的所有指定国): 北京宫羽数字技术有限责任公司(BEIJING MEDIA WORKS CO., LTD) [CN/CN]; 中国北京市海淀区三里河路1号西苑饭店5531室, Beijing 100044 (CN)。北京卓国数字技术有限公司(BEIJING E-WORLD TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市西城区西直门外大街甲143号凯旋大厦C座4层, Beijing 100044 (CN)。编码技术有限公司(CODING TECHNOLOGIES AB) [SE/SE]; 瑞典斯德哥尔摩比伦斯盖坦64号, 11352 Stockholm (SE)。

(72) 发明人;及

(75) 发明人/申请人(仅对美国): 潘兴德(PAN, Xingde) [CN/CN]; 中国北京市西城区西直门外大街甲143号凯旋大厦C座4层, Beijing 100044 (CN)。马丁·帝资(MARTIN, Dietz) [DE/DE]; 德国纽伦堡达赤和恩斯特拉贝15-19, 90429 Nürnberg (DE)。

(72) 发明人: 叶瑞特, 安德斯(EHRET, Andreas); 何瑞斯, 豪格(HÖRICH, Holger); 德国纽伦堡达赤和恩斯特拉贝15-19, 90429 Nürnberg (DE)。朱晓明(ZHU, Xiaoming); 中国北京市西城区西直门外大街甲143号凯旋大厦C座4层, Beijing 100044 (CN)。舒克, 麦可(SCHUG, Michael); 德国纽伦堡达赤和恩斯特拉贝15-19, 90429 Nürnberg (DE)。任为民(REN, Weimin);

王磊(WANG, Lei); 邓昊(DENG, Hao); 中国北京市西城区西直门外大街甲143号凯旋大厦C座4层, Beijing 100044 (CN)。海恩, 佛里德里克(HENN, Fredrik); 瑞典斯德哥尔摩比伦斯盖坦64号, 11352 Stockholm (SE)。

(74) 代理人: 北京同立钧成知识产权代理有限公司(LEADER PATENT & TRADEMARK FIRM) 中国北京市海淀区花园路13号道隆商务会馆, Beijing 100088 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚专利(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲专利(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

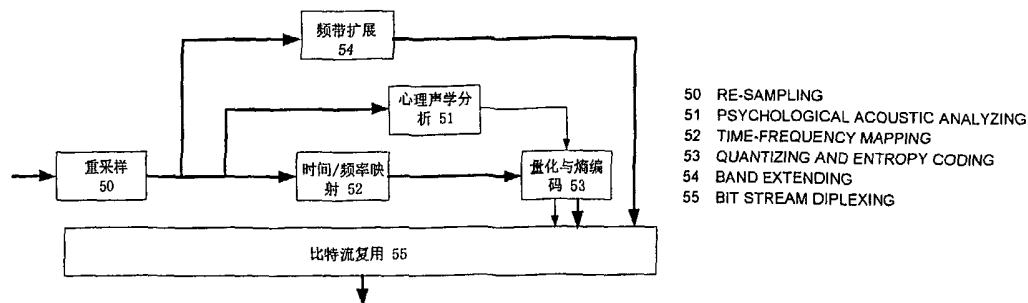
本国际公布:

— 包括国际检索报告。

所引用双字母代码和其它缩写符号, 请参考刊登在每期PCT公报期刊起始的“代码及缩写符号简要说明”。

(54) Title: ENHANCED AUDIO ENCODING AND DECODING EQUIPMENT, METHOD THEREOF

(54) 发明名称: 一种增强音频编解码装置及方法



(57) Abstract: The present invention relates to enhanced audio encoding equipment, including band extending module, re-sampling module, psychological acoustic analyzing module, time-frequency mapping module, quantizing module, entropy coding module and bit stream diplexing module; the band extending module analyzes the original inputted audio signals in whole band-width, extracts the spectral envelope of the high-frequency part and the parameter characterized of the dependency between the lower and higher parts of the frequency spectral, and then outputs them to the bit stream diplexing module; the re-sampling module re-samples the inputted audio signals, changes the sampling rate, and outputs them to the psychological acoustic analyzing module and the time-frequency mapping module; The present invention is applicable to hi-fi compress encoding of audio signals which have configurations about all sorts of sampling rates and sound channels, it can support audio signals whose sampling rate has a range of 8kHz - 192kHz; and it can support all the sound channel configurations which are available; it also can support audio encode/decode which has wide range object code rate.

[见续页]



(57) 摘要

增强音频编码装置，包括频带扩展模块、重采样模块、心理声学分析模块、时频映射模块、量化和熵编码模块以及比特流复用模块；所述频带扩展模块，用于将原始输入音频信号在整个频带上进行分析，提取高频部分的谱包络及表征低、高频谱之间相关性的有关参数，并输出到所述比特流复用模块；所述重采样模块，用于对输入音频信号进行重采样，改变音频信号的采样率，并输出到所述心理声学分析模块、所述时频映射模块；本发明适用于多种采样率、声道配置的音频信号的高保真压缩编码，可以支持采样率为 8kHz 到 192kHz 之间的音频信号；可支持所有可能的声道配置；并且支持范围很宽的目标码率的音频编/解码。

一种增强音频编解码装置及方法

技术领域

本发明涉及音频编解码技术领域，具体地说，涉及一种基于感知模型的增强音频编解码装置及方法。

背景技术

为得到高保真的数字音频信号，需对数字音频信号进行音频编码或音频压缩以便于存储和传输。对音频信号进行编码的目的是用尽可能少的比特数实现音频信号的透明表示，例如原始输入的音频信号与经编码后输出的音频信号之间几乎没有差别。

在二十世纪八十年代初，CD 的出现体现了用数字表示音频信号的诸多优点，例如高保真度、大动态范围和强鲁棒性。然而，这些优点都是以很高的数据速率为代价的。例如 CD 质量的立体声信号的数字化所要求的采样率为 44.1kHz，且每个采样值需用 15 比特进行均匀量化，这样，没有经过压缩的数据速率就达到了 1.41Mb/s，如此高的数据速率给数据的传输和存储带来极大的不便，特别是在多媒体应用和无线传输应用的场合下，更是受到带宽和成本的限制。为了保持高质量的音频信号，因此要求新的网络和无线多媒体数字音频系统必须降低数据的速率，且同时不损害音频的质量。针对上述问题，目前已提出了多种既能得到很高压缩比又能产生高保真的音频信号的音频压缩技术，典型的有国际标准化组织 ISO/IEC 的 MPEG-1/-2/-4 技术、杜比公司的 AC-2/AC-3 技术、索尼公司的 ATRAC/MiniDisc/SDDS 技术以及朗讯科技的 PAC/EPAC/MPAC 技术等。下面选择 MPEG-2 AAC 技术、杜比公司的 AC-3 技术进行具体的说明。

MPEG-1 技术和 MPEG-2 BC 技术是主要用于单声道及立体声音频信号的高音质编码技术，随着对在较低码率下达到较高编码质量的多声道音频编码的需求的日益增长，由于 MPEG-2 BC 编码技术强调与 MPEG-1 技术的后向兼容性，因此无法以低于 540kbps 的码率实现五声道的高音质编码。针对这一不足，提出了 MPEG-2 AAC 技术，该技术可采用 320kbps 的速率对五声道信号实现较高质量的编码。

图 1 给出了 MPEG-2 AAC 编码器的方框图，该编码器包括增益控制器 101、滤波器组 102、时域噪声整形模块 103、强度/耦合模块 104、心理声学模型、二阶后向自适应预测器 105、和差立体声模块 106、比特分配和量化编码模块 107 以及比特流复用模块 108，其中比特分配和量化编码模块 107 进一步包括压缩比/失真处理控制器、尺度因子模块、非均匀量化器和熵编码模块。

滤波器组 102 采用改进的离散余弦变换 (MDCT)，其分辨率是信号自适应的，即对于稳态信号采用 2048 点 MDCT 变换，而对于瞬态信号则采用 256 点 MDCT 变换；这样，对于 48kHz 采样的信号，其最大频率分辨率为 23Hz，最大时间分辨率为 2.6ms。同时在滤波器组 102 中可以使用正弦窗和 Kaiser-Bessel 窗，当输入信号的谐波间隔小于 140Hz 时使用正弦窗，当输入信号中很强的成分间隔大于 220Hz 时使用 Kaiser-Bessel 窗。

音频信号经过增益控制器 101 后进入滤波器组 102，根据不同的信号进行滤波，然后通过时域噪声整形模块 103 对滤波器组 102 输出的频谱系数进行处理，时域噪声整形技术是在频域上对频谱系数进行线性预测分析，然后依据上述分析控制量化噪声

在时域上的形状，以此达到控制预回声的目的。

强度/耦合模块 104 是用于对信号强度的立体声编码，由于对于高频段（大于 2kHz）的信号，听觉的方向感与有关信号强度的变化（信号包络）有关，而与信号的波形无关，即恒包络信号对听觉方向感无影响，因此可利用这一特点以及多声道间的相关信息，将若干声道合成一个共同声道进行编码，这就形成了强度/耦合技术。

二阶后向自适应预测器 105 用于消除稳态信号的冗余，提高编码效率。和差立体声（M/S）模块 106 是针对声道对进行操作，声道对是指诸如双声道信号或多声道信号中的左右声道或左右环绕声道的两个声道。M/S 模块 106 利用声道对中两个声道之间的相关性以达到减少码率和提高编码效率的效果。比特分配和量化编码模块 107 是通过一个嵌套循环过程实现的，其中非均匀量化器是进行有损编码，而熵编码模块是进行无损编码，这样可以去除冗余和减少相关。嵌套循环包括内层循环和外层循环，其中内层循环调整非均匀量化器的步长直到所提供的比特用完，外层循环则利用量化噪声与掩蔽阈值的比来估计信号的编码质量。最后经过编码的信号通过比特流复用模块 108 形成编码的音频流输出。

在采样率可伸缩的情况下，输入信号同时进行四频段多相位滤波器组（PQF）中产生四个等带宽的频带，每个频带利用 MDCT 产生 256 个频谱系数，总共有 1024 个。在每个频带内都使用增益控制器 101。而在解码器中可以忽略高频的 PQF 频带得到低采样率信号。

图 2 给出了对应的 MPEG-2 AAC 解码器的方框示意图。该解码器包括比特流解复用模块 201、无损解码模块 202、逆量化器 203、尺度因子模块 204、和差立体声（M/S）模块 205、预测模块 206、强度/耦合模块 207、时域噪声整形模块 208、滤波器组 209 和增益控制模块 210。编码的音频流经过比特流解复用模块 201 进行解复用，得到相应的数据流和控制流。上述信号通过无损解码模块 202 的解码后，得到尺度因子的整数表示和信号谱的量化值。逆量化器 203 是一组通过压扩函数实现的非均匀量化器组，用于将整数量化值转换为重建谱。由于编码器中的尺度因子模块是将当前尺度因子与前一尺度因子进行差分，然后将差分值采用 Huffman 编码，因此解码器中的尺度因子模块 204 进行 Huffman 解码可得到相应的差分值，再恢复出真实的尺度因子。M/S 模块 205 在边信息的控制下将和差声道转换成左右声道。由于在编码器中采用二阶后向自适应预测器 105 消除稳态信号的冗余并提高编码效率，因此在解码器中通过预测模块 206 进行预测解码。强度/耦合模块 207 在边信息的控制下进行强度/耦合解码，然后输出到时域噪声整形模块 208 中进行时域噪声整形解码，最后通过滤波器组 209 进行综合滤波，滤波器组 209 采用逆向改进离散余弦变换（IMDCT）技术。

对于采样频率可伸缩的情况，可通过增益控制模块 210 忽略高频的 PQF 频带，以得到低采样率信号。

MPEG-2 AAC 编解码技术适用于中高码率的音频信号，但对低码率或甚低码率的音频信号的编码质量较差；同时该编解码技术涉及的编解码模块较多，实现的复杂度较高，不利于实时实现。

图 3 给出了采用杜比 AC-3 技术的编码器的结构示意图，包括暂态信号检测模块 301、改进的离散余弦变换滤波器 MDCT 302、频谱包络/指数编码模块 303、尾数编码模块 304、前向-后向自适应感知模型 305、参数比特分配模块 306 和比特流复用模块 307。

音频信号通过暂态信号检测模块 301 判别是稳态信号还是瞬态信号，同时通过信号自适应 MDCT 滤波器组 302 将时域数据映射到频域数据，其中 512 点的长窗应用于稳态信号，一对短窗应用于瞬态信号。

频谱包络/指数编码模块 303 根据码率和频率分辨率的要求采用三种模式对信号的指数部分进行编码，分别是 D15、D25 和 D45 编码模式。AC-3 技术在频率上对频谱包络采取差分编码，因为最多需要 ± 2 增量，每个增量代表 6dB 的电平变化，对于第一个直流项采用绝对值编码，其余指数就采用差分编码。在 D15 频谱包络指数编码中，每个指数大约需要 2.33 比特，3 个差分组在一个 7 比特的字长中编码，D15 编码模式通过牺牲时间分辨率而提供精细的频率分辨率。由于只是对相对平稳的信号才需要精细的频率分辨率，而这样的信号在许多块上的频谱保持相对恒定，因此，对于稳态信号，D15 偶尔被传送，通常是每 6 个声音块（一个数据帧）的频谱包络被传送一次。当信号频谱不稳定时，需要常更新频谱估计值。估计值采用较小的频率分辨率编码，通常使用 D25 和 D45 编码模式。D25 编码模式提供了合适的频率分辨率和时间分辨率，每隔一个频率系数就进行差分编码，这样每个指数大约需要 1.15 比特。当频谱在 2 至 3 个块上都是稳定的，然后突然变化时，可以采用 D25 编码模式。D45 编码模式是每隔三个频率系数进行差分编码，这样每个指数大约需要 0.58 比特。D45 编码模式提供了很高的时间分辨率和较低的频率分辨率，所以一般应用在对瞬态信号的编码中。

前向-后向自适应感知模型 305 用于估计每帧信号的掩蔽阈值。其中前向自适应部分仅应用在编码器端，在码率的限制下，通过迭代循环估计一组最佳的感知模型参数，然后这些参数被传递到后向自适应部分以估计出每帧的掩蔽阈值。后向自适应部分同时应用在编码器端和解码器端。

参数比特分配模块 306 根据掩蔽准则分析音频信号的频谱包络，以确定给每个尾数分配的比特数。该模块 306 利用一个比特池对所有声道进行全局比特分配。在尾数编码模块 304 中进行编码时，从比特池中循环取出比特分配给所有的声道，根据可以获得的比特数来调整尾数的量化。为达到压缩编码的目的，AC-3 编码器还采用高频耦合的技术，将被耦合信号的高频部分按照人耳临界带宽划分成 18 个子频段，然后选择某些声道从某个子带开始进行耦合。最后通过比特流复用模块 307 形成 AC-3 音频流输出。

图 4 给出了采用杜比 AC-3 解码的流程示意图。首先输入经过 AC-3 编码器编码的比特流，对比特流进行数据帧同步和误码检测，如果检测到一个数据误码，则进行误码掩盖或弱音处理。然后对比特流进行解包，获得主信息和边信息，再进行指数解码。在进行指数解码时，需要有两个边信息：一是打包的指数数目；一个是所采用的指数策略，如 D15、D25 或 D45 模式。已经解码的指数和比特分配边信息再进行比特分配，指出每个打包的尾数所用的比特数，得到一组比特分配指针，每个比特分配指针对应一个编码的尾数。比特分配指针指出用于尾数的量化器以及在码流中每个尾数占用的比特数。对单个编码的尾数值进行解量化，将其转变成一个解量化的值，占用零比特的尾数被恢复成零，或者在抖动标志的控制下用一个随机抖动值代替。然后进行解耦合的操作，解耦合是从公共耦合声道和耦合因子中恢复出被耦合声道的高频部分，包括指数和尾数。如果在编码端采用 2/0 模式编码时，会对某子带采用矩阵处理，那么在解码端需通过矩阵恢复将该子带的和差声道值转换成左右声道值。在码流中包含有每个音频块的动态范围控制值，将该值进行动态范围压缩，以改变系数的幅度，包括

指数和尾数。将频域系数进行逆变换,转变成时域样本,然后对时域样本进行加窗处理,相邻的块进行重叠相加,重构出 PCM 音频信号。当解码输出的声道数小于编码比特流中的声道数时,还需要对音频信号进行下混处理,最后输出 PCM 流。

杜比 AC-3 编码技术主要针对高比特率多声道环绕声的信号,但是当 5.1 声道的编码比特率低于 384kbps 时,其编码效果较差;而且对于单声道和双声道立体声的编码效率也较低。

综上,现有的编解码技术无法全面解决从甚低码率、低码率到高码率音频信号以及单声道、双声道信号的编解码质量,实现较为复杂。

发明内容

本发明所要解决的技术问题在于提供一种增强音频编解码的装置及方法,以解决现有技术对于较低码率音频信号的编码效率低、质量差的问题。

本发明所述增强音频编码装置,包括频带扩展模块、重采样模块、心理声学分析模块、时频映射模块、量化和熵编码模块以及比特流复用模块;所述频带扩展模块,用于将原始输入音频信号在整个频带上进行分析,提取高频部分的谱包络及其与低频部分产生关系的特性,并输出到所述比特流复用模块;所述重采样模块,用于对输入音频信号进行重采样,改变音频信号的采样率,并将改变采样率的音频信号输出给所述心理声学分析模块和所述时频映射模块;所述心理声学分析模块用于计算输入音频信号的掩蔽阈值和信掩比,输出到所述量化和熵编码模块;所述时频映射模块用于将时域音频信号转变成频域系数;所述量化和熵编码模块用于在所述心理声学分析模块输出的信掩比的控制下对频域系数进行量化和熵编码,并输出到所述比特流复用模块;所述比特流复用模块用于将接收到的数据进行复用,形成音频编码的码流。

本发明所述增强音频解码装置,包括比特流解复用模块、熵解码模块、逆量化器组、频率-时间映射模块和频带扩展模块;所述比特流解复用模块用于对压缩音频数据流进行解复用,并向所述熵解码模块和所述频带扩展模块输出相应的数据信号和控制信号;所述熵解码模块用于对上述信号进行解码处理,恢复谱的量化值,输出到所述逆量化器组;所述逆量化器组用于重建逆量化谱,并输出到所述频率-时间映射模块;所述频率-时间映射模块用于对谱系数进行频率-时间映射,得到低频带的时域音频信号;所述频带扩展模块,用于接收所述比特流解复用模块和所述频率-时间映射模块输出的频带扩展控制信息和低频段的时域音频信号,重建高频信号部分,输出宽频带音频信号。

本发明适用于多种采样率、声道配置的音频信号的高保真压缩编码,可以支持采样率为 8kHz 到 192kHz 之间的音频信号;可支持所有可能的声道配置;并且支持范围很宽的目标码率的音频编/解码。

附图说明

- 图 1 是 MPEG-2 AAC 编码器的方框图;
- 图 2 是 MPEG-2 AAC 解码器的方框图;
- 图 3 是采用杜比 AC-3 技术的编码器的结构示意图;
- 图 4 是采用杜比 AC-3 技术的解码流程示意图;
- 图 5 是本发明音频编码装置的结构示意图;
- 图 6 是本发明音频解码装置的结构示意图;
- 图 7 是本发明编码装置的实施例一的结构示意图;

图 8 是本发明解码装置的实施例一的结构示意图;
图 9 是本发明编码装置的实施例二的结构示意图;
图 10 是本发明解码装置的实施例二的结构示意图;
图 11 是本发明编码装置的实施例三的结构示意图;
图 12 是本发明解码装置的实施例三的结构示意图;
图 13 是本发明编码装置的实施例四的结构示意图;
图 14 是采用 Harr 小波基小波变换的滤波结构示意图;
图 15 是采用 Harr 小波基小波变换得到的时频划分示意图;
图 16 是本发明解码装置的实施例四的结构示意图;
图 17 是本发明编码装置的实施例五的结构示意图;
图 18 是本发明解码装置的实施例五的结构示意图;
图 19 是本发明编码装置的实施例六的结构示意图;
图 20 是本发明解码装置的实施例六的结构示意图;
图 21 是本发明编码装置的实施例七的结构示意图;
图 22 是本发明解码装置的实施例七的结构示意图;
图 23 是本发明编码装置的实施例八的结构示意图;
图 24 是本发明编码装置的实施例九的结构示意图;
图 25 是本发明编码装置的实施例十的结构示意图;
图 26 是本发明编码装置的实施例十一的结构示意图;
图 27 是本发明编码装置的实施例十二的结构示意图;
图 28 是本发明编码装置的实施例十三的结构示意图;
图 29 是本发明编码装置的实施例十四的结构示意图;
图 30 是本发明解码装置的实施例八的结构示意图;
图 31 是本发明编码装置的实施例十五的结构示意图;
图 32 是本发明编码装置的实施例十六的结构示意图;
图 33 是本发明编码装置的实施例十七的结构示意图;
图 34 是本发明编码装置的实施例十八的结构示意图;
图 35 是本发明编码装置的实施例十九的结构示意图;
图 36 是本发明解码装置的实施例九的结构示意图;
图 37 是本发明解码装置的实施例十的结构示意图;
图 38 是本发明解码装置的实施例十一的结构示意图;
图 39 是本发明解码装置的实施例十二的结构示意图;
图 40 是本发明解码装置的实施例十三的结构示意图。

具体实施方式

图 1 至图 4 是现有技术的几种编码器的结构示意图,已在背景技术中进行了介绍,此处不再赘述。

需要说明的是:为方便、清楚地说明本发明,下述编解码装置的具体实施例是采用对应的方式说明的,但并不表明编码装置与解码装置必须是一一对应的。

如图 5 所示,本发明提供的音频编码装置包括重采样模块 50、心理声学分析模块 51、时频映射模块 52、量化和熵编码模块 53、频带扩展模块 54 以及比特流复用模块 55;其中重采样模块 50 用于对输入的音频信号进行重采样;心理声学分析模块 51 用

于计算重采样后音频信号的掩蔽阈值和信掩比, 分析信号类型; 时频映射模块 52 用于将时域音频信号转变成频域系数; 量化和熵编码模块 53 用于在心理声学分析模块 51 输出的信掩比的控制下对频域系数进行量化和熵编码, 并输出到比特流复用模块 55; 频带扩展模块 54 用于将输入的音频信号在整个频带上进行分析, 提取高频部分的谱包络及其与低频部分产生关系的特性, 并输出到比特流复用模块 55; 比特流复用模块 55 用于将重采样模块 50、量化和熵编码模块 53 以及频带扩展模块 54 输出的数据进行复用, 形成音频编码码流。

数字音频信号在重采样模块 50 中进行重采样, 改变了采样率, 重采样后的信号分别输入心理声学分析模块 51 和时频映射模块 52, 一方面在心理声学分析模块 51 中计算该帧音频信号的掩蔽阈值和信掩比, 然后将信掩比作为控制信号传送给量化和熵编码模块 53; 另一方面时域的音频信号通过时频映射模块 52 转变成频域系数。上述频域系数在心理声学分析模块 51 输出的信掩比的控制下, 在量化和熵编码模块 53 中进行量化和熵编码。另一方面, 原始数字音频信号经过频带扩展模块 54 的分析, 获得高频部分的谱包络及谱特性参数, 输出到比特流复用模块 55。经过编码后的数据和控制信号在比特流复用模块 55 进行复用, 形成增强音频编码的码流。

下面对上述音频编码装置的各个组成模块进行具体详细地说明。

重采样模块 50 用于对输入音频信号进行重采样, 重采样包括上采样和下采样两种, 下面以下采样为例说明重采样。在本实施例中, 重采样模块 50 包括低通滤波器和下采样器, 其中低通滤波器用于限制音频信号的频带, 消除下采样可能引起的混叠。输入的音频信号经过低通滤波后, 进行下采样。假设输入的音频信号为 $s(n)$, 经过脉冲响应为 $h(n)$ 的低通滤波器滤波后的输出为 $v(n)$, 则有
$$v(n) = \sum_{k=-\infty}^{\infty} h(k)s(n-k);$$
 对 $v(n)$

进行 M 倍的下采样后的序列为 $x(n)$, 则
$$x(m) = v(Mm) = \sum_{k=-\infty}^{\infty} h(k)s(Mm-k)。$$
 这样, 重采样后的音频信号 $x(n)$ 的采样率就比原始输入的音频信号 $s(n)$ 的采样率降低了 M 倍。

心理声学分析模块 51 主要用于计算输入音频信号的掩蔽阈值、信掩比和感知熵, 分析信号类型。根据心理声学分析模块 51 计算出的感知熵可动态地分析当前帧信号进行透明编码所需的比特数, 从而调整帧间的比特分配。心理声学分析模块 51 输出各个子带的信掩比到量化和熵编码模块 53, 对其进行控制。

时频映射模块 52 用于实现音频信号从时域信号到频域系数的变换, 由滤波器组构成, 具体可以是离散傅立叶变换 (DFT) 滤波器组、离散余弦变换 (DCT) 滤波器组、修正离散余弦变换 (MDCT) 滤波器组、余弦调制滤波器组、小波变换滤波器组等。

通过时频映射得到的频域系数被输出到量化和熵编码模块 53 中, 进行量化和编码处理。

量化和熵编码模块 53 进一步包括了非线性量化器组和编码器, 其中量化器可以是标量量化器或矢量量化器。矢量量化器进一步分为无记忆矢量量化器和有记忆矢量量化器两大类。对于无记忆矢量量化器, 每个输入矢量是独立进行量化的, 与以前的各矢量无关; 有记忆矢量量化器是在量化一个矢量时考虑以前的矢量, 即利用了矢量之间的相关性。主要的无记忆矢量量化器包括全搜索矢量量化器、树搜索矢量量化器、多级矢量量化器、增益/波形矢量量化器和分离均值矢量量化器; 主要的有记忆矢量量化器包括预测矢量量化器和有限状态矢量量化器。

如果采用标量量化器,则非线性量化器组进一步包括M个子带量化器。在每个子带量化器中主要利用尺度因子进行量化,具体是:对M个尺度因子带中所有的频域系数进行非线性压缩,再利用尺度因子对该子带的频域系数进行量化,得到整数表示的量化谱输出到编码器,将每帧信号中的第一个尺度因子作为公共尺度因子输出到比特流复用模块55,其它尺度因子与其前一个尺度因子进行差分处理后输出到编码器。

上述步骤中的尺度因子是不断变化的值,按照比特分配策略来调整。本发明提供了一种全局感知失真最小的比特分配策略,具体如下:

首先,初始化每个子带量化器,所有子带中的谱系数的量化值均为零。此时每个子带的量化噪声等于每个子带的能量值,每个子带的噪声掩蔽比NMR等于它的信掩比SMR,量化所消耗的比特数为0,剩余比特数 B_i 等于目标比特数 B 。

其次,查找噪声掩蔽比NMR最大的子带,若最大噪声掩蔽比NMR小于等于1,则尺度因子不变,输出分配结果,比特分配过程结束;否则,将对应的子带量化器的尺度因子减小一个单位,然后计算该子带所需增加的比特数 $\Delta B_i(Q_i)$ 。若该子带的剩余

比特数 $B_i \geq \Delta B_i(Q_i)$,则确认此次尺度因子的修改,并将剩余比特数 B_i 减去 $\Delta B_i(Q_i)$,

重新计算该子带的噪声掩蔽比NMR,然后继续查找噪声掩蔽比NMR最大的子带,重复执行后续步骤。如果该子带的剩余比特数 $B_i < \Delta B_i(Q_i)$,则取消此次修改,保留上一次的尺度因子以及剩余比特数,最后输出分配结果,比特分配过程结束。

如果采用矢量量化器,则频域系数组成多个M维矢量输入到非线性量化器组中,对于每个M维矢量都根据平整因子进行谱平整,即缩小谱的动态范围,然后由矢量量化器根据主观感知距离测度准则在码书中找到与待量化矢量距离最小的码字,将对应的码字索引传递给编码器。平整因子是根据矢量量化的比特分配策略调整的,而矢量量化的比特分配则根据不同子带间感知重要度来控制。

在经过上述量化处理后,利用熵编码技术进一步去除量化后的系数以及边信息的统计冗余。熵编码是一种信源编码技术,其基本思想是:对出现概率较大的符号给予较短长度的码字,而对出现概率小的符号给予较长的码字,这样平均码字的长度最短。根据Shannon的无噪声编码定理,如果传输的N个源消息的符号是独立的,那么使用

适当的变长度编码,码字的平均长度 $\bar{n}$ 将满足 $\left\lceil \frac{H(x)}{\log_2(D)} \right\rceil \leq \bar{n} < \left\lceil \frac{H(x)}{\log_2(D)} + \frac{1}{N} \right\rceil$,其中 $H(x)$

表示信源的熵, x 表示符号变量。由于熵 $H(x)$ 是平均码字长度的最短极限,上述公式表明此时码字的平均长度很接近于它的下限熵 $H(x)$,因此这种变长度编码技术又成为“熵编码”。熵编码主要有Huffman编码、算术编码或游程编码等方法,本发明中的熵编码均可采用上述编码方法的任一种。

经过标量量化器量化后输出的量化谱和差分处理后的尺度因子在编码器中进行熵编码,得到码书序号、尺度因子编码值和无损编码量化谱,再对码书序号进行熵编码,得到码书序号编码值,然后将尺度因子编码值、码书序号编码值和无损编码量化谱输出到比特流复用模块55中。

经过矢量量化器量化后得到的码字索引在编码器中进行一维或多维熵编码,得到

码字索引的编码值,然后将码字索引的编码值输出到比特流复用模块 55 中。

原始输入音频信号输入到频带扩展模块 54 后,在整个频带上进行分析,提取出高频部分的谱包络及其与低频部分产生关系的特性,作为频带扩展控制信息输出到比特流复用模块 55。

频带扩展的基本原理是:对于大多数音频信号,其高频部分的特性与低频部分的特性存在很强的相关性,因此音频信号的高频部分可以通过其低频部分有效地重构出来,这样,音频信号的高频部分可以不传输。为确保高频部分能够正确的重构,在压缩音频码流中仅传输少量的频带扩展控制信号就可以了。

频带扩展模块 54 包括参数提取模块和谱包络提取模块,输入信号进入参数提取模块中,提取在不同时频区域表示输入信号谱特性的参数,然后在谱包络提取模块中,以一定的时频分辨率估计信号高频部分的谱包络。为了确保时频分辨率最适合于当前输入信号的特性,谱包络的时频分辨率可以自由选择。输入信号谱特性的参数和高频部分的谱包络作为频带扩展的输出送到比特流复用模块 55 中复用。

比特流复用模块 55 收到量化和熵编码模块 53 输出的包括公共尺度因子、尺度因子编码值、码书序号编码值和无损编码量化谱的码流或者是码字索引的编码值以及频带扩展模块 54 输出的有关信息后,对其进行复用,得到压缩音频数据流。

基于上述编码器的编码方法,具体包括:在整个频带上分析输入音频信号,提取其高频谱包络和信号谱特性参数作为频带扩展控制信号;对输入音频信号进行重采样处理;计算重采样后信号的信掩比;对重采样后的信号进行时频映射,获得音频信号的频域系数;对频域系数进行量化和熵编码;将频带扩展控制信号和编码后的音频码流进行复用,得到压缩音频码流。

重采样处理包括两个步骤:限制音频信号的频带;对限制频带的音频信号进行多倍的下采样。

对时域音频信号进行时频变换的方法有很多,如离散傅立叶变换(DFT)、离散余弦变换(DCT)、修正离散余弦变换(MDCT)、余弦调制滤波器组、小波变换等。下面以修正离散余弦变换 MDCT 和余弦调制滤波为例说明时频映射的过程。

对于采用修正离散余弦变换 MDCT 进行时频变换的情况,首先选取前一帧 M 个样本和当前帧 M 个样本的时域信号,再对这两帧共 $2M$ 个样本的时域信号进行加窗操作,然后对经过加窗后的信号进行 MDCT 变换,从而获得 M 个频域系数。

MDCT 分析滤波器的脉冲响应为:

$$h_k(n) = w(n) \sqrt{\frac{2}{M}} \cos \left[\frac{(2n+M+1)(2k+1)\pi}{4M} \right],$$

则 MDCT 变换为: $X(k) = \sum_{n=0}^{2M-1} x(n) h_k(n)$ $0 \leq k \leq M-1$, 其中: $w(n)$ 为窗函数; $x(n)$ 为 MDCT 变换的输入时域信号; $X(k)$ 为 MDCT 变换的输出频域信号。

为满足信号完全重构的条件, MDCT 变换的窗函数 $w(n)$ 必须满足以下两个条件:

$$w(2M-1-n) = w(n) \quad \text{且} \quad w^2(n) + w^2(n+M) = 1.$$

在实际中,可选用 Sine 窗作为窗函数。当然,也可以通过使用双正交变换,用特定的分析滤波器和合成滤波器修改上述对窗函数的限制。

对于采用余弦调制滤波进行时频变换的情况,则首先选取前一帧 M 个样本和当前帧 M 个样本的时域信号,再对这两帧共 $2M$ 个样本的时域信号进行加窗操作,然后对经过加窗后的信号进行余弦调制变换,从而获得 M 个频域系数。

传统的余弦调制滤波技术的冲击响应为

$$\begin{aligned} h_k(n) &= 2p_a(n) \cos\left(\frac{\pi}{M}(k+0.5)(n-\frac{D}{2}) + \theta_k\right), \\ n &= 0, 1, \dots, N_h - 1 \\ f_k(n) &= 2p_s(n) \cos\left(\frac{\pi}{M}(k+0.5)(n-\frac{D}{2}) - \theta_k\right), \\ n &= 0, 1, \dots, N_f - 1 \end{aligned}$$

其中 $0 \leq k < M-1$, $0 \leq n < 2KM-1$, K 为大于零的整数, $\theta_k = (-1)^k \frac{\pi}{4}$ 。

假设 M 子带余弦调制滤波器组的分析窗(分析原型滤波器) $p_a(n)$ 的冲击响应长度为 N_a , 综合窗(综合原型滤波器) $p_s(n)$ 的冲击响应长度为 N_s 。当分析窗和综合窗相等时,即 $p_a(n) = p_s(n)$, 且 $N_a = N_s$, 由上面两式所表示的余弦调制滤波器组为正交滤波器组,此时矩阵 H 和 F ($[H]_{n,k} = h_k(n)$, $[F]_{n,k} = f_k(n)$) 为正交变换矩阵。为获得线性相位滤波器组,进一步规定对称窗满足 $p_a(2KM-1-n) = p_a(n)$ 。为保证正交和双正交系统的完全重构性,窗函数还需满足一定的条件,详见文献“Multirate Systems and Filter Banks”, P.P. Vaidynathan, Prentice Hall, Englewood Cliffs, NJ, 1993。

计算重采样后信号的掩蔽阈值和信掩比包括以下步骤:

第一步、将信号进行时域到频域的映射。可采用快速傅立叶变换和汉宁窗(hanning window)技术将时域数据转换成频域系数 $X[k]$, $X[k]$ 用幅度 $r[k]$ 和相位 $\phi[k]$ 表示为 $X[k] = r[k]e^{j\phi[k]}$, 那么每个子带的能量 $e[b]$ 是该子带内所有谱线能量的和,

即 $e[b] = \sum_{k=k_l}^{k=k_h} r^2[k]$, 其中 k_l 和 k_h 分别表示子带 b 的上下边界。

第二步、确定信号中的音调和非音调成分。信号的音调性是通过每个谱线进行帧间预测来估计的,每个谱线的预测值和真实值的欧氏距离被映射为不可预测测度,高预测性的谱成分被认为是音调性很强的,而低预测性的谱成分被认为是类噪声的。

预测值的幅度 r_{pred} 和相位 ϕ_{pred} 可用以下公式来表示:

$$\begin{aligned} r_{pred}[k] &= r_{t-1}[k] + (r_{t-1}[k] - r_{t-2}[k]) \\ \phi_{pred}[k] &= \phi_{t-1}[k] + (\phi_{t-1}[k] - \phi_{t-2}[k]) \end{aligned}$$

其中, t 表示当前帧的系数; $t-1$ 表示前一帧的系数; $t-2$ 表示前两帧的系数。

那么, 不可预测度 $c[k]$ 的计算公式为:

$$c[k] = \frac{\text{dist}(X[k], X_{\text{pred}}[k])}{r[k] + |r_{\text{pred}}[k]|}$$

其中, 欧氏距离 $\text{dist}(X[k], X_{\text{pred}}[k])$ 采用下式计算:

$$\begin{aligned} \text{dist}(X[k], X_{\text{pred}}[k]) &= |X[k] - X_{\text{pred}}[k]| \\ &= \left((r[k] \cos(\phi[k]) - r_{\text{pred}}[k] \cos(\phi_{\text{pred}}[k]))^2 + (r[k] \sin(\phi[k]) - r_{\text{pred}}[k] \sin(\phi_{\text{pred}}[k]))^2 \right)^{1/2} \end{aligned}$$

因此, 每个子带的不可预测度 $c[b]$ 是该子带内所有谱线的能量对其不可预测度的

加权和, 即 $c[b] = \sum_{k=k_l}^{k=k_h} c[k] r^2[k]$ 。子带能量 $e[b]$ 和不可预测度 $c[b]$ 分别与扩展函数进

行卷积运算, 得到子带能量扩展 $e_s[b]$ 和子带不可预测度扩展 $c_s[b]$, 掩模 i 对于子带 b 的扩展函数表示为 $s[i, b]$ 。为了消除扩展函数对能量变换的影响, 需要对子带不可预测度扩展 $c_s[b]$ 做归一化处理, 其归一化的结果用 $\tilde{c}_s[b]$ 表示为 $\tilde{c}_s[b] = \frac{c_s[b]}{e_s[b]}$ 。同样, 为

消除扩展函数对子带能量的影响, 定义归一化能量扩展 $\tilde{e}_s[b]$ 为 $\tilde{e}_s[b] = \frac{e_s[b]}{n[b]}$, 其中归一

化因子 $n[b]$ 为: $n[b] = \sum_{i=1}^{b_{\max}} s[i, b]$, $b_{\max}$ 为该帧信号所分的子带数。

根据归一化不可预测度扩展 $\tilde{c}_s[b]$, 可计算子带的音调性 $t[b]$:

$t[b] = -0.299 - 0.43 \log_e(\tilde{c}_s[b])$, 且 $0 \leq t[b] \leq 1$ 。当 $t[b] = 1$ 时, 表示该子带信号为纯音调;

当 $t[b] = 0$ 时, 表示该子带信号为白噪声。

第三步、计算每个子带所需的信噪比 (Signal-to-Noise Ratio, 简称 SNR)。将所有子带的噪声掩蔽音调 (Noise-Masking-Tone, 简称 NMT) 的值设为 5dB, 音调掩蔽噪声 (Tone-Masking-Noise, 简称 TMN) 的值设为 18dB, 若要使噪声不被感知, 则每个子带所需的信噪比 $\text{SNR}[b]$ 是 $\text{SNR}[b] = 18t[b] + 6(1 - t[b])$ 。

第四步、计算每个子带的掩蔽阈值以及信号的感知熵, 并进行信号类型分析。根据前述步骤得到的每个子带的归一化信号能量和所需的信噪比 SNR, 计算每个子带的噪声能量阈值 $n[b]$ 为 $n[b] = \tilde{e}_s[b] 10^{-\text{SNR}[b]/10}$ 。

为了避免预回声的影响, 将当前帧的噪声能量阈值 $n[b]$ 与前一帧的噪声能量阈值 $n_{\text{prev}}[b]$ 进行比较, 得到信号的掩蔽阈值为 $n[b] = \min(n[b], 2n_{\text{prev}}[b])$, 这样可以确保掩蔽阈值不会因为在分析窗的近末端有高能量的冲击产生而出现偏差。

进一步地, 考虑静止掩蔽阈值 $qsthr[b]$ 的影响, 选择最终的信号的掩蔽阈值为静

止掩蔽阈值与上述计算的掩蔽阈值两者中的数值大者, 即 $n[b] = \max(n[b], qsthr[b])$ 。然后采用如下公式计算感知熵, 即 $pe = - \sum_{b=0}^{b_{\max}} (cbwidth_b \times \log_{10}(n[b]/(e[b]+1)))$, 其中 $cbwidth_b$ 表示各子带所包含的谱线个数。

判断某一帧信号的感知熵是否超过指定的门限 PE-SWITCH, 如果超过, 则该帧信号为快变类型, 否则为缓变类型。

第五步: 计算每个子带信号的信掩比 (Signal-to-Mask Ratio, 简称 SMR)。每个子带的信掩比 SMR [b] 为 $SMR[b] = 10 \log_{10} \left(\frac{e[b]}{n[b]} \right)$ 。

在获得了子带信号的信掩比后, 根据信掩比对频域系数进行量化和熵编码, 其中量化可以是标量量化或矢量量化。

标量量化包括以下步骤: 对所有尺度因子带中的频域系数进行非线性压缩; 再利用每个子带的尺度因子对该子带的频域系数进行量化, 得到整数表示的量化谱; 选择每帧信号中的第一个尺度因子作为公共尺度因子; 其它尺度因子与其前一个尺度因子进行差分处理。

矢量量化包括以下步骤: 将频域系数构成多个多维矢量信号; 对于每个 M 维矢量都根据平整因子进行谱平整; 根据主观感知距离测度准则在码书中查找与待量化矢量距离最小的码字, 获得其码字索引。

熵编码步骤包括: 对量化谱和差分处理后的尺度因子进行熵编码, 得到码书序号、尺度因子编码值和无损编码量化谱; 对码书序号进行熵编码, 得到码书序号编码值。

或者是: 对码字索引进行一维或多维熵编码, 得到码字索引的编码值。

上述的熵编码方法可以采用现有的 Huffman 编码、算术编码或游程编码等方法中的任一种。

经过量化和熵编码处理后, 得到编码后的音频码流, 将该码流与公共尺度因子、频带扩展控制信号一起进行复用, 得到压缩音频码流。

图 6 是本发明音频解码装置的结构示意图。音频解码装置包括比特流解复用模块 601、熵解码模块 602、逆量化器组 603、频率-时间映射模块 604 和频带扩展模块 605。压缩音频码流经过比特流解复用模块 601 的解复用后, 得到相应的数据信号和控制信号, 输出到熵解码模块 602 和频带扩展模块 605。数据信号和控制信号在熵解码模块 602 中进行解码处理, 恢复出谱的量化值。上述量化值在逆量化器组 603 中重建, 得到逆量化后的谱, 逆量化谱输出到频率-时间映射模块 604 中, 经过频率-时间映射得到时域的音频信号, 并在频带扩展模块 605 中, 重建高频信号部分, 得到宽频带的时域音频信号。

比特流解复用模块 601 对压缩音频码流进行分解, 得到相应的数据信号和控制信号, 为其他模块提供相应的解码信息。压缩音频数据流经过解复用后, 输出到熵解码模块 602 的信号包括公共尺度因子、尺度因子编码值、码书序号编码值和无损编码量化谱, 或者是码字索引的编码值; 输出频带扩展控制信息到频带扩展模块 605。

在编码装置中, 如果量化和熵编码模块 53 中采用标量量化器, 则在解码装置中, 熵解码模块 602 收到的是比特流解复用模块 601 输出的公共尺度因子、尺度因子编码值、码书序号编码值和无损编码量化谱, 然后对其进行码书序号解码、谱系数解码和

尺度因子解码,重建出量化谱,并向逆量化器组 603 输出尺度因子的整数表示和谱的量化值。熵解码模块 602 采用的解码方法与编码装置中熵编码的编码方法相对应,如 Huffman 解码、算术解码或游程解码等。

逆量化器组 603 接收到谱的量化值和尺度因子的整数表示后,将谱的量化值逆量化为无缩放的重建谱(逆量化谱),并向频率-时间映射模块 604 输出逆量化谱。逆量化器组 603 可以是均匀量化器组,也可以是通过压扩函数实现的非均匀量化器组。在编码装置中,量化器组采用的是标量量化器,则在解码装置中逆量化器组 603 也采用标量逆量化器。在标量逆量化器中,首先对谱的量化值进行非线性扩张,然后利用每个尺度因子得到对应尺度因子带中所有的谱系数(逆量化谱)。

如果量化和熵编码模块 53 中采用矢量量化器,则在解码装置中,熵解码模块 602 收到比特流解复用模块 601 输出的码字索引的编码值,将码字索引的编码值采用与编码时的熵编码方法对应的熵解码方法进行解码,得到对应的码字索引。

码字索引输出到逆量化器组 603 中,通过查询码书,得到量化值(逆量化谱),输出到频率-时间映射模块 604。逆量化器组 603 采用逆矢量量化器。

逆量化谱通过频率-时间映射模块 604 的映射处理,得到低频段的时域音频信号。频率-时间映射模块 604 可以是逆离散余弦变换(IDCT)滤波器组、逆离散傅立叶变换(IDFT)滤波器组、逆修正离散余弦变换(IMDCT)滤波器组、逆小波变换滤波器组以及余弦调制滤波器组等。

频带扩展模块 605 接收比特流解复用模块 601 输出的频带扩展信息和频率-时间映射模块 604 输出的低频段的时域音频信号,通过频谱搬移和高频调整重建高频信号部分,输出宽频带音频信号。

基于上述解码器的解码方法包括:对压缩音频码流进行解复用,得到数据信息和控制信息;对上述信息进行熵解码,得到谱的量化值;对谱的量化值进行逆量化处理,得到逆量化谱;将逆量化谱进行频率-时间映射,得到低频段的时域音频信号;根据频带扩展控制信息,重建时域音频信号的高频部分,得到宽频带音频信号。

如果解复用后的信息中包括码书序号编码值、公共尺度因子、尺度因子编码值和无损编码量化谱,则表明在编码装置中谱系数是采用标量量化技术进行量化,则熵解码的步骤包括:对码书序号编码值进行解码,获得所有尺度因子带的码书序号;根据码书序号对应的码书,解码所有尺度因子带的量化系数;解码所有尺度因子带的尺度因子,重建量化谱。上述过程所采用的熵解码方法对应编码方法中的熵编码方法,如游程解码方法、Huffman 解码方法、算术解码方法等。

下面以采用游程解码方法解码码书序号、采用 Huffman 解码方法解码量化系数和采用 Huffman 解码方法解码尺度因子为例,说明熵解码的过程。

首先通过游程解码方法获得所有尺度因子带的码书序号,解码后的码书序号为某一区间的整数,如假设该区间为 $[0, 11]$,那么只有位于该有效范围内的,即 0 至 11 之间的码书序号才与对应的谱系数 Huffman 码书相对应。对于全零子带,可选择某一码书序号对应,典型的可选 0 序号。

当解码得到各尺度因子带的码书号后,使用与该码书号对应的谱系数 Huffman 码书,对所有尺度因子带的量化系数进行解码。如果一个尺度因子带的码书号在有效范围内,本实施例如在 1 至 11 之间,那么该码书号对应一个谱系数码书,则使用该码书从量化谱中解码得到尺度因子带的量化系数的码字索引,然后从码字索引中解包得

到量化系数。如果尺度因子带的码书号不在 1 至 11 之间,那么该码书号不对应任何谱系数码书,该尺度因子带的量化系数也就不需要解码,直接将该子带的量化系数全部置为零。

尺度因子用于在逆量化谱系数基础上重构谱值,如果尺度因子带的码书号处于有效范围内,则每一个码书号都对应一个尺度因子。在对上述尺度因子进行解码时,首先读取第一个尺度因子所占用的码流,然后对其它尺度因子进行 Huffman 解码,依次得到各尺度因子与前一尺度因子之间的差值,将该差值与前一尺度因子值相加,得到各尺度因子。如果当前子带的量化系数全部为零,那么该子带的尺度因子不需要解码。

经过上述熵解码过程后,得到谱的量化值和尺度因子的整数表示,然后对谱的量化值进行逆量化处理,获得逆量化谱。逆量化处理包括:对谱的量化值进行非线性扩张;根据每个尺度因子得到对应尺度因子带中的所有谱系数(逆量化谱)。

如果解复用后的信息中包括码字索引的编码值,则表明编码装置中采用矢量量化技术对谱系数进行量化,则熵解码的步骤包括:采用与编码装置中熵编码方法对应的熵解码方法对码字索引的编码值进行解码,得到码字索引。然后对码字索引进行逆量化处理,获得逆量化谱。

对逆量化谱进行频率-时间映射处理的方法与编码方法中的时-频映射处理方法相对应,可以采用逆离散余弦变换(IDCT)、逆离散傅立叶变换(IDFT)、逆修正离散余弦变换(IMDCT)、逆小波变换等方法完成。

下面以逆修正离散余弦变换 IMDCT 为例说明频率-时间映射过程。频率-时间映射过程包括三个步骤:IMDCT 变换、时域加窗处理和时域叠加运算。

首先对预测前的谱或逆量化谱进行 IMDCT 变换,得到变换后的时域信号 $x_{i,n}$ 。IMDCT

变换的表达式为: $x_{i,n} = \frac{2}{N} \sum_{k=0}^{\frac{N}{2}-1} spec[i][k] \cos\left(\frac{2\pi}{N}(n+n_0)(k+\frac{1}{2})\right)$, 其中, n 表示样本序号,

且 $0 \leq n < N$, N 表示时域样本数,取值为 2048, $n_0 = (N/2+1)/2$; i 表示帧序号; k 表示谱序号。

其次,对 IMDCT 变换获得的时域信号在时域进行加窗处理。为满足完全重构条件,窗函数 $w(n)$ 必须满足以下两个条件: $w(2M-1-n) = w(n)$ 且 $w^2(n) + w^2(n+M) = 1$ 。

典型的窗函数有 Sine 窗、Kaiser-Bessel 窗等。本发明采用一种固定的窗函数,其窗函数为: $w(N+k) = \cos(\pi/2 * ((k+0.5)/N - 0.94 * \sin(2 * \pi / N * (k+0.5)) / (2 * \pi)))$, 其中 $k = 0 \dots N-1$; $w(k)$ 表示窗函数的第 k 个系数,有 $w(k) = w(2 * N - 1 - k)$; N 表示编码帧的样本数,取值为 $N=1024$ 。另外可以利用双正交变换,采用特定的分析滤波器和合成滤波器修改上述对窗函数的限制。

最后,对上述加窗时域信号进行叠加处理,得到时域音频信号。具体是:将加窗操作后获得的信号的前 $N/2$ 个样本和前一帧信号的后 $N/2$ 个样本重叠相加,获得 $N/2$ 个输出的时域音频样本,即 $timeSam_{i,n} = preSam_{i,n} + preSam_{i-1,n+N/2}$, 其中 i 表示帧序号,

n 表示样本序号, 有 $0 \leq n \leq \frac{N}{2}$, 且 N 的取值为 2048。

在获得时域音频信号后, 根据频带扩展控制信息和时域音频信号, 重构音频信号的高频部分, 得到宽频带音频信号。

图 7 是本发明编码装置的第一个实施例的示意图。该实施例在图 5 的基础上, 增加了和差立体声 (M/S) 编码模块 56, 该模块位于时频映射模块 52 的输出与量化和熵编码模块 53 的输入之间, 心理声学分析模块 51 向其输出信号类型分析结果。对于多声道信号, 心理声学分析模块 51 除了计算音频信号单声道的掩蔽阈值, 还要计算和差声道的掩蔽阈值, 输出到量化和熵编码模块 53。和差立体声编码模块 56 还可以位于量化和熵编码模块 53 中的量化器组与编码器之间。

和差立体声编码模块 56 是利用声道对中两个声道之间的相关性, 将左右声道的频域系数/残差序列转换为和差声道的频域系数/残差序列, 以达到提高编码效率和立体声声像效果的目的, 因此只适用于信号类型一致的声道对信号。如果是单声道信号或者信号类型不一致的声道对信号, 则不进行和差立体声编码处理。

基于图 7 所示编码装置的编码方法与基于图 5 所示编码装置的编码方法基本相同, 区别在于增加了下述步骤: 在对频域系数进行量化和熵编码处理之前, 判断音频信号是否为多声道信号, 如果是多声道信号, 则判断左、右声道信号的信号类型是否一致, 如果信号类型一致, 则判断两声道对应的尺度因子带之间是否满足和差立体声编码条件, 如果满足, 则对其进行和差立体声编码, 得到和差声道的频域系数; 如果不满足, 则不进行和差立体声编码; 如果是单声道信号或信号类型不一致的多声道信号, 则对频域系数不进行处理。

和差立体声编码除了可以应用在量化处理之前, 还可以应用在量化之后、熵编码之前, 即: 在对频域系数量化后, 判断音频信号是否为多声道信号, 如果是多声道信号, 则判断左、右声道信号的信号类型是否一致, 如果信号类型一致, 则判断两声道对应的尺度因子带之间是否满足和差立体声编码条件, 如果满足, 则对其进行和差立体声编码; 如果不满足, 则不进行和差立体声编码处理; 如果是单声道信号或信号类型不一致的多声道信号, 则对频域系数不进行和差立体声编码处理。

判断尺度因子带是否可进行和差立体声编码的方法很多, 本发明采用的判断方法是: 通过 K-L 变换。具体判断过程如下:

假如左声道尺度因子带的谱系数为 $l(k)$, 右声道相对应的尺度因子带的谱系数为

$r(k)$, 其相关矩阵 C 为 $C = \begin{pmatrix} C_{ll} & C_{lr} \\ C_{lr} & C_{rr} \end{pmatrix}$, 其中, $C_{ll} = \frac{1}{N} \sum_{k=0}^{N-1} l(k) * l(k)$; $C_{lr} = \frac{1}{N} \sum_{k=0}^{N-1} l(k) * r(k)$;

$C_{rr} = \frac{1}{N} \sum_{k=0}^{N-1} r(k) * r(k)$; N 是尺度因子带的谱线数目。

对相关矩阵 C 进行 K-L 变换, 得到

$$RCR^T = \Lambda = \begin{pmatrix} \lambda_{ll} & 0 \\ 0 & \lambda_{rr} \end{pmatrix}, \text{ 其中, } R = \begin{pmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{pmatrix} \quad a \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]。$$

旋转角度 a 满足 $\tan(2a) = \frac{2C_{lr}}{C_{ll} - C_{rr}}$, 当 $a = \pm \pi/4$ 时, 就是和差立体声编码模式。

因此当旋转角度 a 的绝对值偏离 $\pi/4$ 较小时 (如 $3\pi/16 < |a| < 5\pi/16$) , 两声道对应的尺度因子带之间可以进行和差立体声编码。

如果和差立体声编码应用在量化处理之前, 则将左右声道对应尺度因子带内的频域系数通过线性变换用和差声道的频域系数代替:

$$\begin{bmatrix} M \\ S \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix},$$

其中, M 表示和声道频域系数; S 表示差声道频域系数; L 表示左声道频域系数; R 表示右声道频域系数。

如果和差立体声编码应用在量化之后, 则左右声道在尺度因子带的量化后的频域系数通过线性变换用和差声道的频域系数代替:

$$\begin{bmatrix} \hat{M} \\ \hat{S} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{L} \\ \hat{R} \end{bmatrix},$$

其中: $\hat{M}$ 表示量化后的和声道频域系数; $\hat{S}$ 表示量化后的差声道频域系数; $\hat{L}$ 表示量化后的左声道频域系数; $\hat{R}$ 表示量化后的右声道频域系数。

将和差立体声编码放在量化处理之后, 不仅可以有效的去除左右声道的相关, 而且由于在量化后进行, 因此可以达到无损编码。

图 8 是解码装置的实施例一的示意图。该解码装置在图 6 所示的解码装置的基础上, 增加了和差立体声解码模块 606, 位于逆量化器组 603 的输出与频率-时间映射模块 604 的输入之间, 接收比特流解复用模块 601 输出的信号类型分析结果与和差立体声控制信号, 用于根据上述控制信息将和差声道的逆量化谱转换成左右声道的逆量化谱。

在和差立体声控制信号中, 有一个标志位用于表明当前声道对是否需要和差立体声解码, 若需要, 则在每个尺度因子带上也有一个标志位表明对应尺度因子带是否需要和差立体声解码, 和差立体声解码模块 606 根据尺度因子带的标志位, 确定是否需要某些尺度因子带中的逆量化谱/谱的量化值进行和差立体声解码。如果在编码装置中进行了和差立体声编码, 则在解码装置中必须对逆量化谱进行和差立体声解码。

和差立体声解码模块 606 还可以位于熵解码模块 602 的输出与逆量化器组 603 的输入之间, 接收比特流解复用模块 601 输出的和差立体声控制信号和信号类型分析结果。

基于图 8 所示解码装置的解码方法基本与基于图 6 所示解码装置的解码方法相同, 区别在于增加了下述步骤: 在得到逆量化谱后, 如果信号类型分析结果表明信号类型一致, 则根据和差立体声控制信号判断是否需要逆量化谱进行和差立体声解码; 如果需要, 则根据每个尺度因子带上的标志位判断该尺度因子带是否需要和差立体声解码, 如果需要, 则将该尺度因子带中的和差声道的逆量化谱转换成左右声道的逆量化谱, 再进行后续处理; 如果信号类型不一致或者不需要进行和差立体声解码, 则对逆量化谱不进行处理, 直接进行后续处理。

和差立体声解码还可以在熵解码处理之后、逆量化处理之前进行, 即: 当得到谱

的量化值后,如果信号类型分析结果表明信号类型一致,则根据和差立体声控制信号判断是否需要对谱的量化值进行和差立体声解码;如果需要,则根据每个尺度因子带上的标志位判断该尺度因子带是否需要和差立体声解码,如果需要,则将该尺度因子带中的和差声道的谱的量化值转换成左右声道的谱的量化值,再进行后续处理;如果信号类型不一致或者不需要进行和差立体声解码,则对谱的量化值不进行处理,直接进行后续处理。

如果和差立体声解码在熵解码之后、逆量化之前,则左右声道在尺度因子带的频域系数采用下列运算通过和差声道的频域系数得到:
$$\begin{bmatrix} \hat{l} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{s} \end{bmatrix}$$
, 其中: $\hat{m}$ 表示

量化后的和声道频域系数; $\hat{s}$ 表示量化后的差声道频域系数; $\hat{l}$ 表示量化后的左声道频域系数; $\hat{r}$ 表示量化后的右声道频域系数。

如果和差立体声解码在逆量化之后,则左右声道在子带的逆量化后的频域系数根据下面的矩阵运算通过和差声道的频域系数得到:
$$\begin{bmatrix} l \\ r \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} m \\ s \end{bmatrix}$$
, 其中: m 表示和声道频域系数; s 表示差声道频域系数; l 表示左声道频域系数; r 表示右声道频域系数。

图 9 给出了本发明编码装置的第二个实施例的结构示意图。该实施例在图 5 的基础上增加了频域线性预测及矢量量化模块 57, 其中频域线性预测及矢量量化模块 57 位于时频映射模块 52 的输出与量化和熵编码模块 53 的输入之间, 并输出残差序列到量化和熵编码模块 53, 同时将量化得到的码字索引作为边信息输出到比特流复用模块 55。

时频映射模块 52 输出的频域系数传送到频域线性预测及矢量量化模块 57 中, 如果频域系数的预测增益值满足给定的条件, 则对频域系数进行线性预测滤波, 获得的预测系数转换成线谱对频率系数 LSF (Line Spectrum Frequency), 再采用最佳的失真度量准则搜索计算出各级码本的码字索引, 并将码字索引作为边信息传送到比特流复用模块 55, 而经过预测分析得到的残差序列则输出到量化和熵编码模块 53。

频域线性预测及矢量量化模块 57 由线性预测分析器、线性预测滤波器、转换器和矢量量化器构成。频域系数输入到线性预测分析器中进行预测分析, 得到预测增益和预测系数, 如预测增益的值满足一定条件, 则将频域系数输出到线性预测滤波器中进行线性预测误差滤波, 得到频域系数的预测残差序列; 残差序列直接输出到量化和熵编码模块 53 中, 而预测系数则通过转换器转换成线谱对频率系数 LSF, 再将 LSF 参数送入矢量量化器中进行多级矢量量化, 对应的码字索引被传送到比特流复用模块 55 中。

对音频信号进行频域线性预测处理能够有效地抑制预回声并获得较大的编码增益。假设实信号 $x(t)$, 其平方 Hilbert 包络 $e(t)$ 表示为:
$$e(t) = F^{-1} \left\{ C(\xi) \cdot C^*(\xi - f) d\xi \right\},$$

其中 $C(f)$ 为对应于信号 $x(t)$ 正频率成分的单边谱, 即信号的 Hilbert 包络是与该信号谱的自相关函数有关的。而信号的功率谱密度函数与其时域波形的自相关函数的关系为:
$$PSD(f) = F \left\{ \int x(\tau) \cdot x^*(\tau - t) d\tau \right\},$$
 因此信号在时域的平方 Hilbert 包络与信号在

频域的功率谱密度函数是互为对偶关系的。由上可知, 每个一定频率范围内的部分带通信号, 如果它的 Hilbert 包络保持恒定, 那么相邻谱值的自相关也将保持恒定, 这

就意味着谱系数序列相对于频率而言是稳态序列，从而可以用预测编码技术来对谱值进行处理，用公用的一组预测系数来有效地表示该信号。

基于图 9 所示编码装置的编码方法与基于图 5 所示编码装置的编码方法基本相同，区别在于增加了下述步骤：对频域系数进行标准的线性预测分析，得到预测增益和预测系数；判断预测增益是否超过设定的阈值，如果超过，则根据预测系数对频域系数进行频域线性预测误差滤波，得到频域系数的预测残差序列；将预测系数转化成线谱对频率系数，并对线谱对频率系数进行多级矢量量化处理，得到边信息；对残差序列进行量化和熵编码；如果预测增益未超过设定的阈值，则对频域系数进行量化和熵编码。

当获得了频域系数后，首先对频域系数进行标准的线性预测分析，包括计算自相关矩阵、递推执行 Levinson-Durbin 算法获得预测增益和预测系数。然后判断计算的预测增益是否超过预先设定的阈值，如果超过，则根据预测系数对频域系数进行线性预测误差滤波；否则对频域系数不作处理，执行下一步骤，对频域系数进行量化和熵编码。

线性预测可分为前向预测和后向预测两种，前向预测是指利用某一时刻之前的值预测当前值，而后向预测是指利用某一时刻之后的值预测当前值。下面以前向预测为例说明线性预测误差滤波，线性预测误差滤波器的传递函数为 $A(z) = 1 - \sum_{i=1}^p a_i z^{-i}$ ，其中

a_i 表示预测系数， p 为预测阶数。经过时间-频率变换后的频域系数 $X(k)$ 经过滤波后，得到预测误差 $E(k)$ ，也称残差序列，两者之间满足关系 $E(k) = X(k) \cdot A(z) = X(k) - \sum_{i=1}^p a_i X(k-i)$ 。

这样，经过线性预测误差滤波，时间-频率变换模块输出的频域系数 $X(k)$ 就可以用残差序列 $E(k)$ 和一组预测系数 a_i 表示。然后将这组预测系数 a_i 转换成线谱对频率系数 LSF，并对其进行多级矢量量化。矢量量化选择最佳的失真度量准则（如最近邻准则），依次分别在各级码本中搜索得到与待量化的 LSF 参数矢量（残差矢量）最匹配的码字，将对应的码字索引作为边信息输出。同时，对残差序列 $E(k)$ 进行量化和熵编码。由线性预测分析编码原理可知，谱系数的残差序列的动态范围小于原始谱系数的动态范围，因此在量化时可以分配较少的比特数，或者对于相同比特数的条件，可以获得改进的编码增益。

图 10 是解码装置的实施例二的示意图，该解码装置在图 6 所示解码装置的基础上，增加了逆频域线性预测及矢量量化模块 607，位于逆量化器组 603 的输出与频率-时间映射模块 604 的输入之间，并且比特流解复用模块 601 向其输出逆频域线性预测矢量量化控制信息，用于对逆量化谱（残差谱）进行逆量化处理和逆线性预测滤波，得到预测前的谱，并输出到频率-时间映射模块 604 中。

在编码器中，采用频域线性预测矢量量化技术来抑制预回声，并获得较大的编码增益。因此在解码器中，逆量化谱和比特流解复用模块 601 输出的逆频域线性预测矢量量化控制信息输入到逆频域线性预测及矢量量化模块 607 中恢复出线性预测前的谱。

逆频域线性预测及矢量量化模块 607 包括逆矢量量化器、逆转换器和逆线性预测滤波器, 其中逆矢量量化器用于对码字索引进行逆量化得到线谱对频率系数 LSF; 逆转换器则用于将线谱对频率系数 LSF 逆转换为预测系数; 逆线性预测滤波器用于根据预测系数对逆量化谱进行逆滤波, 得到预测前的谱, 并输出到频率-时间映射模块 604。

基于图 10 所示解码装置的解码方法与基于图 6 所示解码装置的解码方法基本相同, 区别在于增加了下述步骤: 在得到逆量化谱后, 判断控制信息中是否包含逆量化谱需要经过逆频域线性预测矢量量化的信息, 如果含有, 则进行逆矢量量化处理, 得到预测系数, 并根据预测系数对逆量化谱进行线性预测合成, 得到预测前的谱; 将预测前的谱进行频率-时间映射。

在获得逆量化谱后, 根据控制信息判断该帧信号是否经过频域线性预测矢量量化, 如果是, 则从控制信息中获取预测系数矢量量化后的码字索引; 再根据码字索引得到量化的线谱对频率系数 LSF, 并以此计算出预测系数; 然后将逆量化谱进行线性预测合成, 得到预测前的谱。

线性预测误差滤波处理所采用的传递函数 $A(z)$ 为: $A(z) = 1 - \sum_{i=1}^p a_i z^{-i}$, 其中: a_i 是预测系数; p 为预测阶数。因此残差序列 $E(k)$ 与预测前的谱 $X(k)$ 满足:

$$X(k) = E(k) \cdot \frac{1}{A(z)} = E(k) + \sum_{i=1}^p a_i X(k-i)。$$

这样, 残差序列 $E(k)$ 和计算出的预测系数 a_i 经过频域线性预测合成, 就可得到预测前的谱 $X(k)$, 将预测前的谱 $X(k)$ 进行频率-时间映射处理。

如果控制信息表明该帧信号没有经过频域线性预测矢量量化, 则不进行逆频域线性预测矢量量化处理, 将逆量化谱直接进行频率-时间映射处理。

图 11 给出了本发明编码装置的第三个实施例的结构示意图。该实施例在图 9 的基础上, 增加了和差立体声编码模块 56, 位于频域线性预测及矢量量化模块 57 的输出与量化和熵编码模块 53 的输入之间, 心理声学分析模块 51 向其输出信号类型分析结果, 并将和差声道的掩蔽阈值输出到量化和熵编码模块 53。

和差立体声编码模块 56 也可以位于量化和熵编码模块 53 中的量化器组与编码器之间, 接收心理声学分析模块 51 输出的信号类型分析结果。

在本实施例中, 和差立体声编码模块 56 的功能及工作原理与其在图 7 中的相同, 此处不再赘述。

基于图 11 所示编码装置的编码方法与基于图 9 所示编码装置的编码方法基本相同, 区别在于增加了下述步骤: 在对频域系数进行量化和熵编码处理之前, 判断音频信号是否为多声道信号, 如果是多声道信号, 则判断左、右声道信号的信号类型是否一致, 如果信号类型一致, 则判断尺度因子带是否满足编码条件, 如果满足, 则对该尺度因子带进行和差立体声编码; 如果不满足, 则不进行和差立体声编码处理; 如果是单声道信号或信号类型不一致的多声道信号, 则不进行和差立体声编码处理。

和差立体声编码除了可以应用在量化处理之前, 还可以应用在量化之后、熵编码之前, 即: 在对频域系数量化后, 判断音频信号是否为多声道信号, 如果是多声道信号, 则判断左、右声道信号的信号类型是否一致, 如果信号类型一致, 则判断尺度因子带是否满足编码条件, 如果满足, 则对该尺度因子带进行和差立体声编码; 如果不

满足, 则不进行和差立体声编码处理; 如果是单声道信号或信号类型不一致的多声道信号, 则不进行和差立体声编码处理。

图 12 是解码装置的实施例三的结构图。该解码装置在图 10 所示解码装置的基础上, 增加了和差立体声解码模块 606, 位于逆量化器组 603 的输出与逆频域线性预测及矢量量化模块 607 的输入之间, 比特流解复用模块 601 向其输出和差立体声控制信号。

和差立体声解码模块 606 也可以位于熵解码模块 602 的输出与逆量化器组 603 的输入之间, 接收比特流解复用模块 601 输出的和差立体声控制信号。

在本实施例中, 和差立体声解码模块 606 的功能及工作原理与其在图 8 中的相同, 此处不再赘述。

基于图 12 所示解码装置的解码方法与基于图 10 所示解码装置的解码方法基本相同, 区别在于增加了下述步骤: 在得到逆量化谱后, 如果信号类型分析结果表明信号类型一致, 则根据和差立体声控制信号判断是否需要逆量化谱进行和差立体声解码; 如果需要, 则根据每个尺度因子带上的标志位判断该尺度因子带是否需要和差立体声解码, 如果需要, 则将该尺度因子带中的和差声道的逆量化谱转换成左右声道的逆量化谱, 再进行后续处理; 如果信号类型不一致或者不需要进行和差立体声解码, 则对逆量化谱不进行处理, 直接进行后续处理。

和差立体声解码还可以在逆量化处理之前进行, 即: 当得到谱的量化值后, 如果信号类型分析结果表明信号类型一致, 则根据和差立体声控制信号判断是否需要谱的量化值进行和差立体声解码; 如果需要, 则根据每个尺度因子带上的标志位判断该尺度因子带是否需要和差立体声解码, 如果需要, 则将该尺度因子带中的和差声道的谱的量化值转换成左右声道的谱的量化值, 再进行后续处理; 如果信号类型不一致或者不需要进行和差立体声解码, 则对谱的量化值不进行处理, 直接进行后续处理。

图 13 给出了本发明编码装置的第四个实施例的结构示意图。该实施例在图 5 的基础上, 增加了多分辨率分析模块 59, 其中多分辨率分析模块 59 位于时频映射模块 52 的输出与量化和熵编码模块 53 的输入之间, 心理声学分析模块 51 向其输出信号类型分析结果。

对于快变类型信号, 为有效克服编码过程中产生的预回声现象, 提高编码质量, 本发明编码装置通过多分辨率分析模块 59 来提高快变信号的频域系数的时间分辨率。时频映射模块 52 输出的频域系数输入到多分辨率分析模块 59 中, 如果是快变类型信号, 则进行频域小波变换或频域修正离散余弦变换 (MDCT), 获得对频域系数的多分辨率表示, 输出到量化和熵编码模块 53 中。如果是缓变类型信号, 则对频域系数不进行处理, 直接输出到量化和熵编码模块 53。

多分辨率分析模块 59 对输入的频域数据进行时-频域的重新组织, 以频率精度的降低为代价提高频域数据的时间分辨率, 从而自动地适应快变类型信号的时频特性, 达到抑制预回声的效果。此时频映射模块 52 中滤波器组的形式可以无需随时调整。多分辨率分析模块 59 包括频域系数变换模块和重组模块, 其中频域系数变换模块用于将频域系数变换为时频平面系数; 重组模块用于将时频平面系数按照一定的规则进行重组。频域系数变换模块可采用频域小波变换滤波器组、频域 MDCT 变换滤波器组等。

下面以频域小波变换和频域 MDCT 变换为例, 说明多分辨率分析模块 59 的工作过

程。

1) 频域小波变换

假设时序序列 $x(i), i = 0, 1, \dots, 2M-1$, 经过时频映射后得到的频域系数为 $X(k), k = 0, 1, \dots, M-1$ 。频域小波或小波包变换的小波基可以是固定的, 也可以是自适应的。

下面以最简单的 Harr 小波基小波变换为例, 说明对频域系数进行多分辨率分析的过程。

Harr 小波基的尺度系数为 $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$, 小波系数为 $[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$, 图 14 示出了采用

Harr 小波基进行小波变换的滤波结构示意图, 其中 H_0 表示低通滤波 (滤波系数为 $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$), H_1 表示高通滤波 (滤波系数为 $[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$), “ $\downarrow 2$ ” 表示 2 倍的下采样操作。对于频域系数的中低频部分 $X_1(k), k = 0, \dots, k_1$ 不进行小波变换, 对频域系数的高频部分进行 Harr 小波变换, 得到不同的时间-频率区间的系数 $X_2(k), X_3(k), X_4(k),$

$X_5(k), X_6(k)$ 和 $X_7(k)$, 对应的时间-频率平面划分如图 15 所示。选择不同的小波基,

可选用不同的小波变换结构进行处理, 得到其他类似的时间-频率平面划分。因此可以根据需要, 任意调整信号分析时的时频平面划分, 满足不同的时间和频率分辨率的分析要求。

上述时频平面系数在重组模块中按照一定的规则进行重组, 例如: 可先将时频平面系数在频率方向组织, 每个频带中的系数在时间方向组织, 然后将组织好的系数按照子窗、尺度因子带的顺序排列。

2) 频域 MDCT 变换

设输入频域 MDCT 变换滤波器组的频域数据为 $X(k), k = 0, 1, \dots, N-1$, 依次对这 N 点频域数据进行 M 点的 MDCT 变换, 使得时频域数据的频率精度有所下降, 而时间精度则相应地提高了。在不同的频域范围内使用不同长度的频域 MDCT 变换, 可以获得不同的时-频平面划分即不同的时、频精度。重组模块对频域 MDCT 变换滤波器组输出的时-频域数据进行重组, 一种重组方法是先将时频平面系数在频率方向组织, 同时每个频带中的系数在时间方向组织, 然后将组织好的系数按照子窗、尺度因子带的顺序排列。

基于图 13 所示编码装置的编码方法, 基本流程与基于图 5 所示编码装置的编码方法相同, 区别在于增加了下述步骤: 如果是快变类型信号, 则对频域系数进行多分辨率分析, 然后对频域系数的多分辨率表示进行量化和熵编码; 如果不是快变类型信号, 则直接将频域系数进行量化和熵编码。

多分辨率分析可采用频域小波变换法或频域 MDCT 变换法。频域小波分析法包括: 对频域系数进行小波变换, 得到时频平面系数; 将上述时频平面系数按照一定的规则重组。而 MDCT 变换法则包括: 对频域系数进行 n 次 MDCT 变换, 得到时频平面系数; 将上述时频平面系数按照一定的规则重组。重组的方法可以包括: 先将时频平面系数在频率方向组织, 每个频带中的系数在时间方向组织, 然后将组织好的系数按照子窗、尺度因子带的顺序排列。

图 16 是解码装置的实施例四的结构示意图。该解码装置在图 6 所示解码装置的基础上, 增加多分辨率综合模块 609。多分辨率综合模块 609 位于逆量化器组 603 的输出与频率-时间映射模块 604 的输入之间, 用于对逆量化谱进行多分辨率综合。

在编码器中, 对快变类型信号采用了多分辨率滤波技术以提高编码快变类型信号的时间分辨率。相应地, 在解码器中, 需采用多分辨率综合模块 609 对快变类型信号恢复多分辨率分析前的频域系数。多分辨率综合模块 609 包括: 系数重组模块和系数变换模块, 其中系数变换模块可以采用频域逆小波变换滤波器组或频域 IMDCT 变换滤波器组。

基于如图 16 所示的解码装置的解码方法, 基本流程与基于图 6 所示解码装置的解码方法相同, 区别在于增加了下述步骤: 在获得了逆量化谱后, 对逆量化谱进行多分辨率综合, 再将得到的预测系数进行频率-时间映射。

下面以频域 IMDCT 变换为例说明多分辨率综合的方法, 具体包括: 对逆量化谱系数进行重组; 再对每个系数进行多个 IMDCT 变换, 得到多分辨率分析前的逆量化谱。下面以 128 个 IMDCT 变换 (8 个输入, 16 个输出) 详细说明该过程。首先, 将逆量化谱系数按照子窗、尺度因子带的顺序排列; 再按照频序进行重组, 这样每个子窗的 128 个系数按频序被组织在一起。然后, 将按子窗排列的系数每 8 个一组按频率方向组织, 每组 8 个系数按时序排列, 这样在频率方向共有 128 组系数。将每组系数进行 16 点 IMDCT 变换, 每组 IMDCT 变换后输出的 16 个系数重叠相加, 获得 8 个频域数据。依次由低频向高频方向进行 128 次类似的操作, 获得 1024 个频域系数。

图 17 给出了本发明编码装置的第五个实施例的结构示意图。该实施例是在图 13 所示编码装置的基础上, 增加了频域线性预测及矢量量化模块 57, 位于多分辨率分析模块 59 的输出与量化和熵编码模块 53 的输入之间, 心理声学分析模块 51 向其输出信号类型分析结果; 频域线性预测及矢量量化模块 57 用于对经过多分辨率分析的频域系数进行线性预测和多级矢量量化, 输出残差序列到量化和熵编码模块 53, 同时将量化得到的码字索引输出到比特流复用模块 55。

由于频域系数在经过多分辨率分析后得到的是具有特定时频平面划分的时频系数, 因此频域线性预测及矢量量化模块 57 需对每个时间段上的频域系数进行线性预测和多级矢量量化。

频域线性预测及矢量量化模块 57 也可位于时频映射模块 52 的输出与多分辨率分析模块 59 的输入之间, 对频域系数进行线性预测和多级矢量量化, 输出残差序列到多分辨率分析模块 59, 同时将量化得到的码字索引输出到比特流复用模块 55。

基于图 17 所示编码装置的编码方法, 与基于图 13 所示编码装置的编码方法基本相同, 区别在于增加了下述步骤: 在对频域系数进行多分辨率分析后, 对每个时间段上的频域系数进行标准的线性预测分析; 判断预测增益是否超过设定的阈值, 如果超过, 则对频域系数进行频域线性预测误差滤波, 得到预测系数与频域系数的残差序列; 将预测系数转化为线谱对频率系数, 并对线谱对频率系数进行多级矢量量化处理, 得到边信息; 对残差序列进行量化和熵编码; 如果预测增益未超过设定的阈值, 则对频域系数进行量化和熵编码。或者在进行多分辨率分析前, 先对频域系数进行线性预测和多级矢量量化处理, 再将残差序列进行多分辨率分析。

图 18 是解码装置的实施例五的结构示意图。该解码装置在图 16 所示解码装置的基础上, 增加逆频域线性预测及矢量量化模块 607, 位于逆量化器组 603 的输出与多

分辨率综合模块 609 的输入之间, 并且比特流解复用模块 601 向其输出逆频域线性预测矢量量化控制信息, 用于对逆量化谱进行逆量化处理和线性预测合成, 得到预测前的谱, 并输出到多分辨率综合模块 609 中。

逆频域线性预测及矢量量化模块 607 也可位于多分辨率综合模块 609 的输出与频率-时间映射模块 604 的输入之间, 用于对多分辨率综合后的逆量化谱进行线性预测合成。

基于图 18 所示解码装置的解码方法, 与基于图 16 所示解码装置的解码方法基本相同, 区别在于增加了下述步骤: 在得到逆量化谱后, 判断控制信息中是否包含有逆量化谱需要经过逆频域线性预测矢量量化的信息, 如果含有, 则进行逆矢量量化处理, 得到预测系数, 并对逆量化谱进行线性预测合成, 得到预测前的谱; 将预测前的谱进行多分辨率综合。也可以是: 在得到逆量化谱后, 对逆量化谱进行多分辨率综合; 判断控制信息中是否包含有逆量化谱需要经过逆频域线性预测矢量量化的信息, 如果含有, 则进行逆矢量量化处理, 得到预测系数, 并对逆量化谱进行线性预测合成, 得到预测前的谱; 将预测前的谱进行频率-时间映射。

图 19 给出了本发明编码装置的第六个实施例的结构示意图。该实施例是在图 17 所示编码装置的基础上, 增加了和差立体声编码模块 56, 位于频域线性预测及矢量量化模块 57 的输出与量化和熵编码模块 53 的输入之间, 接收心理声学分析模块 51 输出的信号类型分析结果。和差立体声编码模块 56 也可位于量化和熵编码模块 53 中的量化器组与编码器之间。在本实施例中, 和差立体声编码模块 56 的功能及工作原理与图 11 中的相同, 此处不再赘述。

基于图 19 所示编码装置的编码方法与基于图 17 所示编码装置的编码方法基本相同, 区别在于增加了下述步骤: 在获得残差序列后, 根据音频信号是否为多声道信号且为信号类型一致的信号且满足编码条件, 确定是否对其进行和差立体声编码; 然后进行后续处理。具体的流程在上面已经介绍, 此处不再赘述。

图 20 给出了解码装置的实施例六的结构示意图。该解码装置是在图 18 所示解码装置的基础上, 增加了和差立体声解码模块 606, 位于逆量化器组 603 的输出与逆频域线性检测及矢量量化模块 607 的输入之间。和差立体声解码模块 606 也可位于熵解码模块 602 的输出与逆量化器组 603 的输入之间。本实施例中, 和差立体声解码模块 606 的功能及工作原理与其在图 12, 此处不再赘述。

基于图 20 所示解码装置的解码方法, 与基于图 18 所示解码装置的解码方法基本相同, 区别在于增加了下述步骤: 在得到逆量化谱后, 根据信号类型分析结果与和差立体声控制信息判断是否需要逆量化谱进行和差立体声解码; 然后进行后续处理。具体的流程在上面已经介绍, 此处不再赘述。

图 21 给出了本发明编码装置的第七个实施例的示意图, 该实施例是在图 13 的基础上, 增加了和差立体声编码模块 56, 位于多分辨率分析模块 59 的输出与量化和熵解码模块 53 的输入之间。和差立体声编码模块 56 也可位于量化和熵编码模块 53 中的量化器组与编码器之间。和差立体声编码模块 56 已在前面详述过, 此处不再赘述。

基于图 21 所示编码装置的编码方法与基于图 13 所示编码装置的编码方法基本相同, 区别在于增加了下述步骤: 在频域系数经过多分辨率分析后, 根据音频信号是否为多声道信号且为信号类型一致的信号且满足编码条件, 确定是否对其进行和差立体声编码; 然后进行后续处理。具体的流程在上面已经介绍, 此处不再赘述。

图 22 是解码装置的实施例七的示意图。该解码装置在图 16 所示的解码装置的基础上,增加了和差立体声解码模块 606,位于逆量化器组 603 的输出与多分辨率综合模块 609 的输入之间。和差立体声解码模块 606 也可位于熵解码模块 602 的输出与逆量化器组 603 的输入之间。和差立体声解码模块 606 已在前面详述过,此处不再赘述。

基于图 22 所示解码装置的解码方法与基于图 16 所示解码装置的解码方法基本相同,区别在于增加了下述步骤:在得到逆量化谱后,根据信号类型分析结果与和差立体声控制信息判断是否需要逆量化谱进行和差立体声解码;然后进行后续处理。具体的流程在上面已经介绍,此处不再赘述。

图 23 给出了本发明编码装置的第八个实施例的示意图,该实施例是在图 13 所示编码装置的基础上,增加了信号性质分析模块 510,用于对重采样模块 50 输出的信号进行信号类型分析,并将重采样后的信号输出到时频映射模块 52 和心理声学分析模块 51,同时将信号类型分析结果输出比特流复用模块 55。

信号性质分析模块 510 基于自适应阈值和波形预测进行前、后向掩蔽效应分析来确定信号的类型为缓变信号还是快变信号,若是快变类型信号,则继续计算突变成分的相关参数信息,如突变信号发生的位置以及突变信号的强度等。

基于图 23 所示编码装置的编码方法与基于图 13 所示编码装置的编码方法基本相同,区别在于增加了下述步骤:分析重采样后的信号的类型,作为信号复用的一部分。

信号类型是基于自适应阈值和波形预测进行前、后向掩蔽效应分析来确定的,具体步骤是:将输入的音频数据分解成帧;把输入帧分解成多个子帧,并查找各个子帧上 PCM 数据绝对值的局部最大点;在各子帧的局部最大点中选出子帧峰值;对某个子帧峰值,利用该子帧前面的多个(典型的可取 3 个)子帧峰值预测相对该子帧前向延迟的多个(典型的可取 4 个)子帧的典型样本值;计算该子帧峰值与所预测出的典型样本值的差值和比值;如果预测差值和比值都大于设定的阈值,则判断该子帧存在突跃信号,确认该子帧具备后向掩蔽预回声能力的局部最大峰点,如果在该子帧前端与掩蔽峰点前 2.5ms 处之间存在一个峰值足够小的子帧,则判断该帧信号属于快变类型信号;如果预测差值和比值不大于设定的阈值,则重复上述步骤直到判断出该帧信号是快变类型信号或者到达最后一个子帧,如果到达最后一个子帧仍未判断出该帧信号是快变类型信号,则该帧信号属于缓变类型信号。

图 24 至图 28 给出了本发明编码装置的第九个实施例至第十三个实施例的示意图。上述实施例分别是在图 17、图 19、图 21、图 9 和图 11 所示编码装置的基础上,增加了信号性质分析模块 510,用于对重采样模块 50 输出的信号进行信号类型分析,并将重采样后的信号输出到时频映射模块 52 和心理声学分析模块 51,同时将信号类型分析结果输出比特流复用模块 55。

基于图 24 至图 28 所示编码装置的编码方法,基本与基于图 17、图 19、图 21 和图 11 所示编码装置的编码方法相同,区别在于增加了下述步骤:分析重采样后的信号的类型,作为信号复用的一部分。

图 29 给出了本发明编码装置的第十四个实施例的示意图。本实施例是在图 23 所示编码装置的基础上,增加了增益控制模块 511,接收信号性质分析模块 510 输出的音频信号,控制快变类型信号的动态范围,消除音频处理中的预回声,其输出连接到时频映射模块 52 和心理声学分析模块 51。

增益控制模块 511 只对快变类型信号进行控制,而对缓变类型信号,则不进行处

理, 直接输出。对于快变类型信号, 增益控制模块 511 调整信号的时域能量包络, 提升快变点前信号的增益值, 使得快变点前、后的时域信号幅度较为接近; 然后将调整了时域能量包络的时域信号输出到时-频映射模块 52, 同时将增益调整量输出到比特流复用模块 55。

基于图 29 所示编码装置的编码方法, 基本与基于图 23 所示编码装置的编码方法相同, 区别在于增加了下述步骤: 对经过信号类型分析的信号进行增益控制。

图 30 是解码装置的实施例八的结构示意图, 该实施例是在图 16 所示解码装置的基础上增加了逆增益控制模块 610, 位于频率-时间映射模块 604 的输出与频带扩展模块 605 的输入之间, 接收比特流解复用模块 601 输出的信号类型分析结果和增益调整量信息, 用于调整时域信号的增益, 控制预回声。逆增益控制模块 610 接收到频率-时间映射模块 604 输出的重建时域信号后, 对快变类型信号进行控制, 而对缓变类型信号不进行处理, 直接输出到频带扩展模块 605。对于快变类型信号, 逆增益控制模块 610 根据增益调整量信息调整重建时域信号的能量包络, 减小快变点前信号的幅度值, 将能量包络调回原先的前低后高的状态, 这样快变点前的量化噪声的幅度值会和信号的幅度值一起相应地减小, 从而控制了预回声。

基于图 30 所示解码装置的解码方法, 基本与基于图 16 所示解码装置的解码方法相同, 区别在于增加了下述步骤: 在对重建时域信号进行频带扩展之前, 对重建时域信号进行逆增益控制。

图 31 至图 35 给出了本发明编码装置的第十五个至第十九个实施例的示意图。这五个实施例是分别在图 24 至图 28 所示编码装置的基础上, 增加了增益控制模块 511, 用于控制经过信号类型分析的音频信号的动态范围, 消除音频处理中的预回声, 其输出连接到时频映射模块 52 和心理声学分析模块 51。

基于上述五个编码装置的编码方法, 基本与基于图 24 至图 28 所示编码装置的编码方法相同, 区别在于增加了下述步骤: 对经过信号类型分析的信号进行增益控制。

图 36 至图 40 给出了解码装置的实施例九至实施例十三的结构示意图, 上述 5 个解码装置分别在图 18、图 20、图 22、图 10 和图 12 所示解码装置的基础上, 增加了逆增益控制模块 610, 位于频率-时间映射模块 604 的输出与频带扩展模块 605 的输入之间, 接收比特流解复用模块 601 输出的信号类型分析结果, 用于调整时域信号的增益, 控制预回声。

基于上述五个解码装置的解码方法, 也分别与基于图 18、图 20、图 22、图 10 和图 12 所示解码装置的解码方法基本相同, 区别在于增加了下述步骤: 在对重建时域信号进行频带扩展之前, 对重建时域信号进行逆增益控制。

最后所应说明的是, 以上实施例仅用以说明本发明的技术方案而非限制, 尽管参照较佳实施例对本发明进行了详细说明, 本领域的普通技术人员应当理解, 可以对本发明的技术方案进行修改或者等同替换, 而不脱离本发明技术方案的精神和范围, 其均应涵盖在本发明的权利要求范围当中。

权利要求

1、一种增强音频编码装置，包括心理声学分析模块、时频映射模块、量化和熵编码模块以及比特流复用模块，其特征在于，还包括频带扩展模块和重采样模块；其中

所述频带扩展模块，用于将原始输入音频信号在整个频带上进行分析，提取高频部分的谱包络及表征低、高频谱之间相关性的有关参数，并输出到所述比特流复用模块；

所述重采样模块，用于对输入音频信号进行重采样，改变音频信号的采样率，并将改变采样率后的音频信号输出到所述心理声学分析模块和所述时频映射模块；

所述心理声学分析模块用于计算输入音频信号的掩蔽阈值和信掩比，输出到所述量化和熵编码模块；

所述时频映射模块用于将时域音频信号变换为频域系数；

所述量化和熵编码模块用于在所述心理声学分析模块输出的信掩比的控制下对频域系数进行量化和熵编码，并输出到所述比特流复用模块；

所述比特流复用模块用于将接收到的数据进行复用，形成音频编码的码流。

2、根据权利要求1所述的增强音频编码装置，其特征在于，所述重采样模块包括低通滤波器和下采样器；其中所述低通滤波器用于限制音频信号的频带，所述下采样器用于对限制频带的音频信号进行下采样，降低信号的采样率。

3、根据权利要求1所述的增强音频编码装置，其特征在于，所述频带扩展模块包括参数提取模块和谱包络提取模块；所述参数提取模块用于提取输入信号在不同时频区域表示输入信号谱特性的参数；所述谱包络提取模块用于以一定的时频分辨率估计信号高频部分的谱包络，然后将输入信号的谱特性参数和高频部分的谱包络输出到所述比特流复用模块。

4、根据权利要求1所述的增强音频编码装置，其特征在于，还包括频域线性预测及矢量量化模块，位于所述时频映射模块的输出与所述量化和熵编码模块的输入之间，具体由线性预测分析器、线性预测滤波器、转换器和矢量量化器构成；

所述线性预测分析器，用于对频域系数进行预测分析，得到预测增益和预测系数，并将满足一定条件的频域系数输出到所述线性预测滤波器；对于不满足条件的频域系数直接输出到所述量化和熵编码模块；

所述线性预测滤波器，用于对频域系数进行线性预测误差滤波，得到残差序列，并将残差序列输出到所述量化和熵编码模块，将预测系数输出到转换器；

所述转换器，用于将预测系数转换成线谱对频率系数；

所述矢量量化器，用于对线谱对频率系数进行多级矢量量化，有关边信息被传送到所述比特流复用模块。

5、根据权利要求1所述的增强音频编码装置，其特征在于，还包括多分辨率分析模块，位于所述时频映射模块的输出与所述量化和熵编码模块的输入之间，接收所述心理声学分析模块输出的信号类型分析结果，用于对快变类型信号进行多分辨率分析；具体包括频域系数变换模块和重组模块，其中所述频域系数变换模块是频域小波变换滤波器组或频域 MDCT 变换滤波器组，用于将频域系数变换为时频平面系数；所述重组模块用于将时频平面系数按照一定的规则进行重组。

6、根据权利要求5所述的增强音频编码装置，其特征在于，还包括频域线性预

测及矢量量化模块,位于所述时频映射模块的输出与所述多分辨率分析模块的输入之间,用于对频域系数进行线性预测和多级矢量量化,输出残差序列到所述多分辨率分析模块,同时将边信息输出到所述比特流复用模块。

7、根据权利要求6所述的增强音频编码装置,其特征在于,还包括和差立体声编码模块,位于所述多分辨率分析模块的输出与所述量化和熵编码模块的输入之间或者位于所述量化和熵编码模块中的量化器组和编码器之间,接收所述心理声学分析模块输出的信号类型分析结果。

8、一种增强音频编码方法,其特征在于,包括以下步骤:

步骤一、在整个频带上分析输入音频信号,提取其高频谱包络和信号谱特性参数;

步骤二、对输入音频信号进行重采样处理;

步骤三、计算重采样后信号的信掩比;

步骤四、对重采样后的信号进行时频映射,获得音频信号的频域系数;

步骤五、对频域系数进行量化和熵编码;

步骤六、将高频谱包络、信号谱特性参数和编码后的音频信号进行复用,得到压缩音频码流。

9、根据权利要求8所述增强音频编码方法,其特征在于,所述步骤五的量化是标量量化,具体包括:对所有尺度因子带中的频域系数进行非线性压缩;再利用每个子带的尺度因子对该子带的频域系数进行量化,得到整数表示的量化谱;选择每帧信号中的第一个尺度因子作为公共尺度因子;其它尺度因子与其前一个尺度因子进行差分处理;

所述熵编码具体包括:对量化谱和差分处理后的尺度因子进行熵编码,得到码书序号、尺度因子编码值和量化谱的无损编码值;对码书序号进行熵编码,得到码书序号的编码值。

10、根据权利要求8所述增强音频编码方法,其特征在于,在所述步骤四与步骤五之间,还包括:对频域系数进行标准的线性预测分析,得到预测增益和预测系数;判断预测增益是否超过设定的阈值,如果超过,则根据预测系数对频域系数进行频域线性预测误差滤波,得到频域系数的预测残差序列;将预测系数转换成线谱对频率系数,并对线谱对频率系数进行多级矢量量化处理,得到边信息;对残差序列进行量化和熵编码;如果预测增益未超过设定的阈值,则对频域系数进行量化和熵编码。

11、根据权利要求8所述增强音频编码方法,其特征在于,在所述步骤四与步骤五之间还包括:如果是快变类型信号,则对频域系数进行多分辨率分析;如果不是快变类型信号,则直接将频域系数进行量化和熵编码;其中

所述多分辨率分析步骤包括:对频域系数进行 MDCT 变换,得到时频平面系数;将上述时频平面系数按照一定的规则重组;

所述重组包括:先将时频平面系数按频率方向组织,每个频带中的系数按时间方向组织,然后将组织好的系数按照子窗、尺度因子带的顺序排列。

12、根据权利要求11所述增强音频编码方法,其特征在于,在执行所述多分辨率分析步骤之前,还包括:对频域系数进行标准的线性预测分析,得到预测增益和预测系数;判断预测增益是否超过设定的阈值,如果超过,则根据预测系数对频域系数进行频域线性预测误差滤波,得到频域系数的预测残差序列;将预测系数转换成线谱对频率系数,并对线谱对频率系数进行多级矢量量化处理,得到边信息;对残差序列进

行多分辨率分析;如果预测增益未超过设定的阈值,则对频域系数进行多分辨率分析。

13、根据权利要求 12 所述增强音频编码方法,其特征在于,所述步骤五进一步包括:对残差序列进行量化;判断音频信号是否为多声道信号,如果是多声道信号,则判断左、右声道信号的信号类型是否一致,如果信号类型一致,则判断两声道对应的尺度因子带之间是否满足和差立体声编码条件,如果满足,则对该尺度因子带中的谱系数进行和差立体声编码,得到和差声道的残差序列;如果不满足,则该尺度因子带中的谱系数不进行和差立体声编码;如果是单声道信号或信号类型不一致的多声道信号,则对残差序列不进行处理;对残差序列进行熵编码;其中

所述判断尺度因子带是否满足编码条件的方法是:K-L 变换,具体是:计算左右声道尺度因子带的谱系数的相关矩阵;对相关矩阵进行 K-L 变换;如果旋转角度 α 的绝对值偏离 $\pi/4$ 较小时(如 $3\pi/16 < |\alpha| < 5\pi/16$),则两声道信号对应的尺度因子带可以进行和差立体声编码;

所述和差立体声编码为:
$$\begin{bmatrix} \hat{M} \\ \hat{S} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{L} \\ \hat{R} \end{bmatrix}$$
, 其中: $\hat{M}$ 表示量化后的和声道频域

系数; $\hat{S}$ 表示量化后的差声道频域系数; $\hat{L}$ 表示量化后的左声道频域系数; $\hat{R}$ 表示量化后的右声道频域系数。

14、一种增强音频解码装置,包括:比特流解复用模块、熵解码模块、逆量化器组、频率-时间映射模块,其特征在于,还包括频带扩展模块;

所述比特流解复用模块用于对压缩音频数据流进行解复用,并向所述熵解码模块和所述频带扩展模块输出相应的数据信号和控制信号;

所述熵解码模块用于对上述信号进行解码处理,恢复谱的量化值,输出到所述逆量化器组;

所述逆量化器组用于重建逆量化谱,并输出到所述频率-时间映射模块;

所述频率-时间映射模块用于对谱系数进行频率-时间映射,得到低频带的时域音频信号;

所述频带扩展模块,用于接收所述比特流解复用模块和所述频率-时间映射模块输出的频带扩展控制信息和低频段的时域音频信号,重建高频信号部分,输出宽频带音频信号。

15、根据权利要求 14 所述的增强音频解码装置,其特征在于,还包括和差立体声解码模块,位于所述逆量化器组的输出与所述频率-时间映射模块的输入之间或者位于所述熵解码模块的输出与所述逆量化器组的输入之间,接收所述比特流解复用模块输出的和差立体声控制信号,用于根据和差立体声控制信息将和差声道的逆量化谱/谱的量化值转换成左右声道的逆量化谱/谱的量化值。

16、根据权利要求 14 所述的增强音频解码装置,其特征在于,还包括逆频域线性预测及矢量量化模块,位于所述逆量化器组的输出与所述频率-时间映射模块的输入之间,接收所述比特流解复用模块输出的逆频域线性预测矢量量化控制信息,对逆量化谱进行线性预测合成过程,得到预测前的谱,并输出到所述频率-时间映射模块中;具体包括逆矢量量化器、逆转换器和逆线性预测滤波器;所述逆矢量量化器用于对码字索引进行逆量化,得到线谱对频率系数;所述逆转换器则用于将线谱对频率系

数逆转换为预测系数；所述逆线性预测滤波器用于根据预测系数将逆量化谱进行线性预测合成过程，得到预测前的谱。

17、根据权利要求 16 所述的增强音频解码装置，其特征在于，还包括和差立体声解码模块，位于所述逆量化器组的输出与所述逆频域线性预测及矢量量化模块的输入之间或者位于所述熵解码模块的输出与所述逆量化器组的输入之间，接收所述比特流解复用模块输出的和差立体声控制信号，用于根据和差立体声控制信息将和差声道的逆量化谱/谱的量化值转换成左右声道的逆量化谱/谱的量化值。

18、根据权利要求 14 所述的增强音频解码装置，其特征在于，还包括多分辨率综合模块，位于所述逆量化器组的输出与所述频率-时间映射模块的输入之间，接收所述比特流解复用模块输出的信号类型分析结果，用于对逆量化谱进行多分辨率综合；所述多分辨率综合模块具体包括：系数重组模块和系数变换模块；所述系数变换模块是频域逆小波变换滤波器组或频域逆修正离散余弦变换滤波器组。

19、根据权利要求 18 所述的增强音频解码装置，其特征在于，还包括逆频域线性预测及矢量量化模块，位于所述逆量化器组的输出与所述多分辨率综合模块的输入之间或者位于所述多分辨率综合模块的输出与所述频率-时间映射模块的输入之间，接收所述比特流解复用模块输出的逆频域线性预测矢量量化控制信息，用于对逆量化谱/多分辨率综合后的逆量化谱进行逆量化处理和线性预测合成过程，得到预测前的谱。

20、根据权利要求 19 所述的增强音频解码装置，其特征在于，还包括和差立体声解码模块，位于所述逆量化器组的输出与所述逆频域线性预测及矢量量化模块的输入或所述多分辨率综合模块之间，或者位于所述熵解码模块的输出与所述逆量化器组的输入之间，接收所述比特流解复用模块输出的和差立体声控制信号，用于根据和差立体声控制信息将和差声道的逆量化谱/谱的量化值转换成左右声道的逆量化谱/谱的量化值。

21、一种增强音频解码方法，其特征在于，包括以下步骤：

步骤一、对压缩音频数据流进行解复用，得到数据信息和控制信息；

步骤二、对上述信息进行熵解码，得到谱的量化值；

步骤三、对谱的量化值进行逆量化处理，得到逆量化谱；

步骤四、将逆量化谱进行频率-时间映射，得到时域音频信号；

步骤五、根据频带扩展控制信息，重建时域音频信号的高频部分，得到宽频带音频信号。

22、根据权利要求 21 所述的增强音频解码方法，其特征在于，所述步骤四进一步包括：对逆量化谱进行逆修正离散余弦变换，得到变换后的时域信号；对变换后的时域信号在时域进行加窗处理；对上述加窗时域信号进行叠加处理，得到时域音频信号；其中所述加窗处理中的窗函数为：

$w(N+k) = \cos(\pi/2 * ((k+0.5)/N - 0.94 * \sin(2 * \pi/N * (k+0.5)) / (2 * \pi)))$ ，其中 π 为圆周率， $k = 0 \dots N-1$ ； $w(k)$ 表示窗函数的第 k 个系数，有 $w(k) = w(2 * N - 1 - k)$ ； N 表示编码帧的样本数。

23、根据权利要求 21 所述的增强音频解码方法，其特征在于，在所述步骤二与步骤三之间，还包括：如果信号类型分析结果表明信号类型一致，则根据和差立体声控制信号判断是否需要逆量化谱进行和差立体声解码；如果需要，则根据每个尺度

因子带上的标志位判断该尺度因子带是否需要和差立体声解码,如果需要,则将该尺度因子带中的和差声道的逆量化谱转换成左右声道的逆量化谱,转至步骤三;如果信号类型不一致或者不需要进行和差立体声解码,则对逆量化谱不进行处理,转至步骤

三;在上述步骤中,所述和差立体声解码是:
$$\begin{bmatrix} \hat{l} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{s} \end{bmatrix}$$
, 其中: $\hat{m}$ 表示逆量化

后的和声道频域系数; $\hat{s}$ 表示逆量化后的差声道频域系数; $\hat{l}$ 表示逆量化后的左声道频域系数; $\hat{r}$ 表示逆量化后的右声道频域系数。

24、根据权利要求 21 所述的增强音频解码方法,其特征在于,在所述步骤三与步骤四之间,还包括:判断控制信息中是否包含有逆量化谱需要经过逆频域线性预测矢量量化的信息,如果含有,则进行逆矢量量化处理,得到预测系数,并对逆量化谱进行线性预测合成,得到预测前的谱;将预测前的谱进行频率-时间映射;其中所述逆矢量量化处理进一步包括:从控制信息中获得预测系数矢量量化后的码字索引;再根据码字索引得到量化的线谱对频率系数,并由线谱对频率系数计算出预测系数。

25、根据权利要求 24 所述的增强音频解码方法,其特征在于,在所述步骤二与步骤三之间,还包括:如果信号类型分析结果表明信号类型一致,则根据和差立体声控制信号判断是否需要谱的量化值进行和差立体声解码;如果需要,则根据每个尺度因子带上的标志位判断该尺度因子带是否需要和差立体声解码,如果需要,则将该尺度因子带中的和差声道的谱的量化值转换成左右声道的谱的量化值,转至步骤三;如果信号类型不一致或者不需要进行和差立体声解码,则对谱的量化值不进行处理,转至步骤三。

26、根据权利要求 21 所述的增强音频解码方法,其特征在于,在所述步骤三与步骤四之间,还包括:对逆量化谱进行多分辨率综合的步骤;具体是:对逆量化谱系数按照子窗、尺度因子带的顺序排列,再按照频序进行重组,然后对重组的系数进行多个逆修正离散余弦变换,得到多分辨率分析前的逆量化谱。

27、根据权利要求 26 所述的增强音频解码方法,其特征在于,在执行多分辨率综合步骤之前,还包括:判断控制信息中是否包含有逆量化谱需要经过逆频域线性预测矢量量化的信息,如果含有,则进行逆矢量量化处理,得到预测系数,并对残差谱进行线性预测合成,得到预测前的谱;将预测前的谱进行多分辨率综合。

28、根据权利要求 26 所述的增强音频解码方法,其特征在于,在执行所述步骤四之前,还包括:判断控制信息中是否包含有逆量化谱需要经过逆频域线性预测矢量量化的信息,如果含有,则进行逆矢量量化处理,得到预测系数,并对残差谱进行线性预测合成,得到预测前的谱;将预测前的谱进行频率-时间映射。

29、根据权利要求 26 至 28 任一所述的增强音频解码方法,其特征在于,在所述步骤二与步骤三之间,还包括:如果信号类型分析结果表明信号类型一致,则根据和差立体声控制信号判断是否需要谱的量化值进行和差立体声解码;如果需要,则根据每个尺度因子带上的标志位判断该尺度因子带是否需要和差立体声解码,如果需要,则将该尺度因子带中的和差声道的谱的量化值转换成左右声道的谱的量化值,转至步骤三;如果信号类型不一致或者不需要进行和差立体声解码,则对谱的量化值不进行处理,转至步骤三。

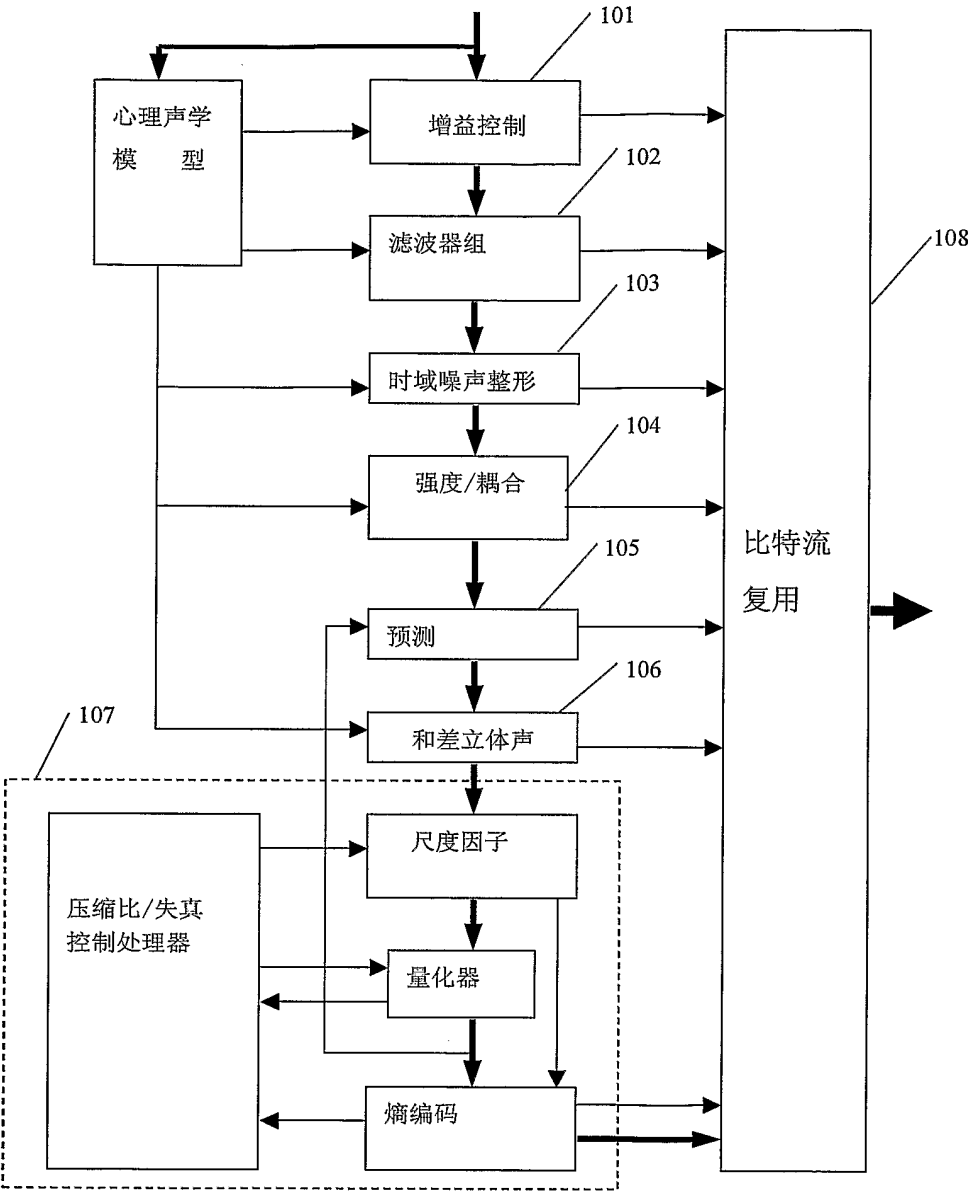


图 1

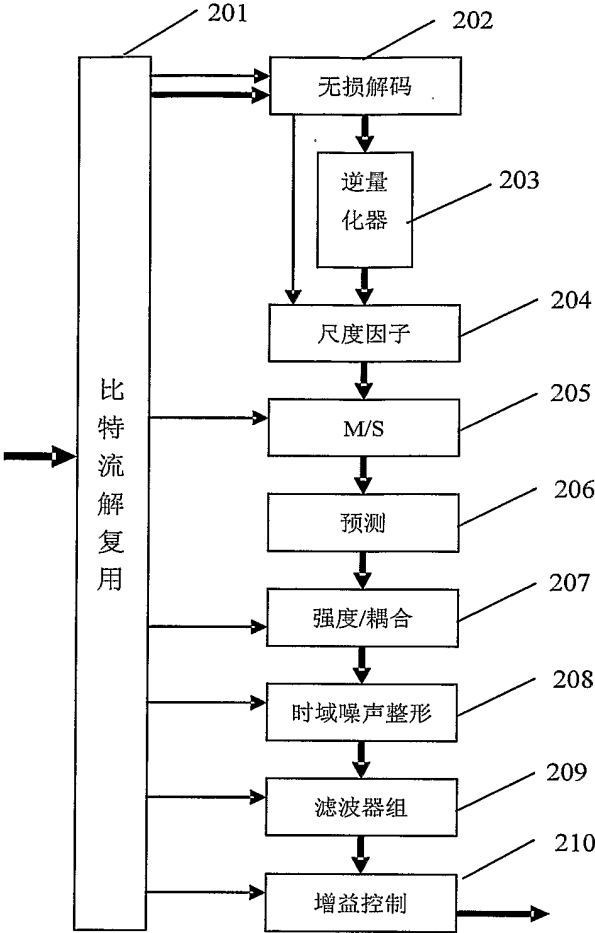


图 2

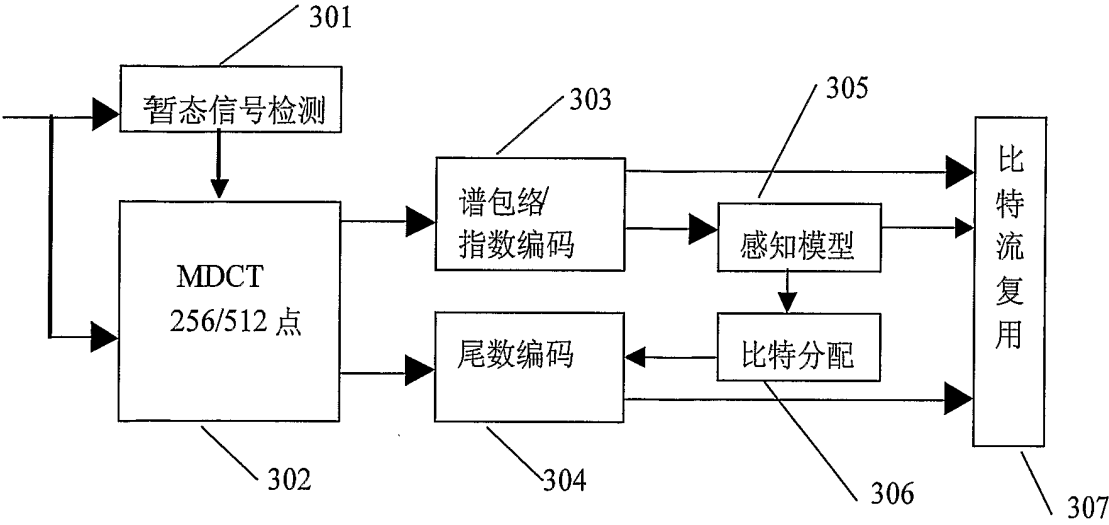


图 3

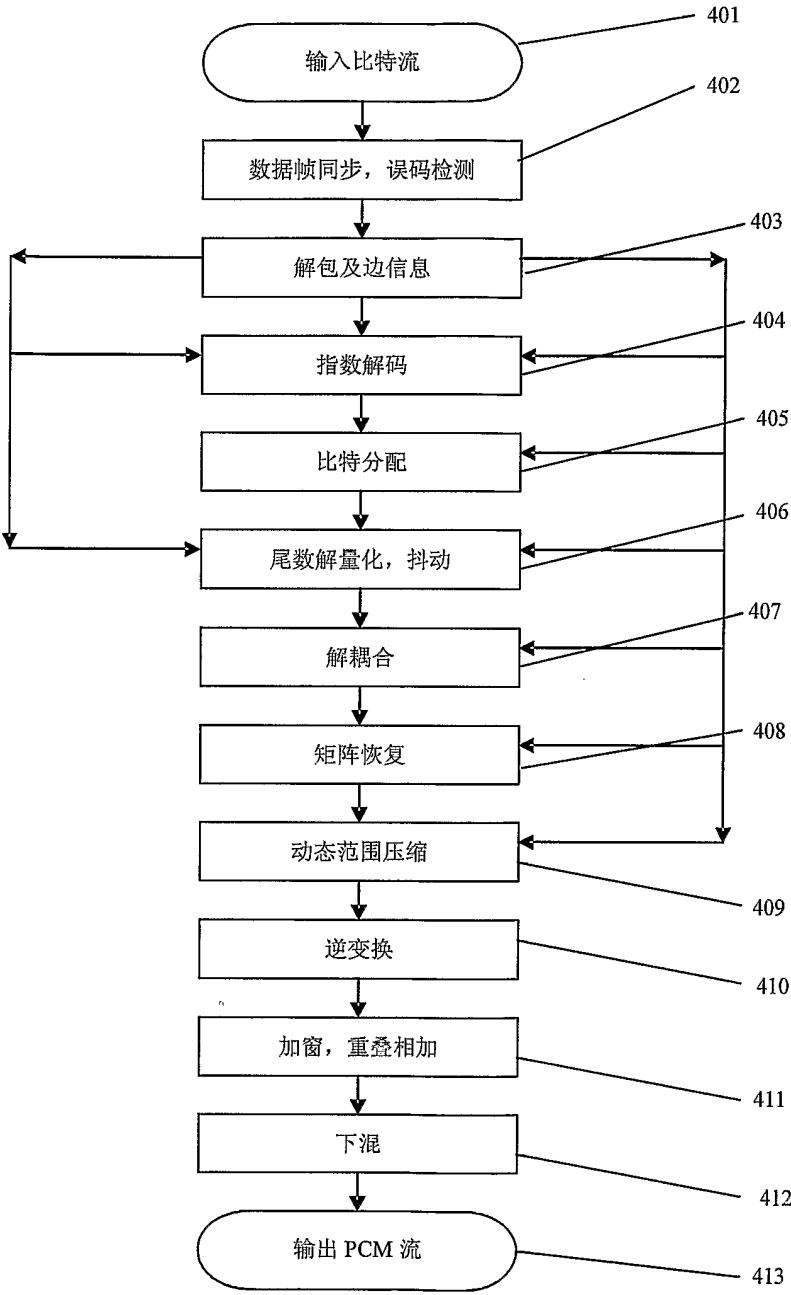


图 4

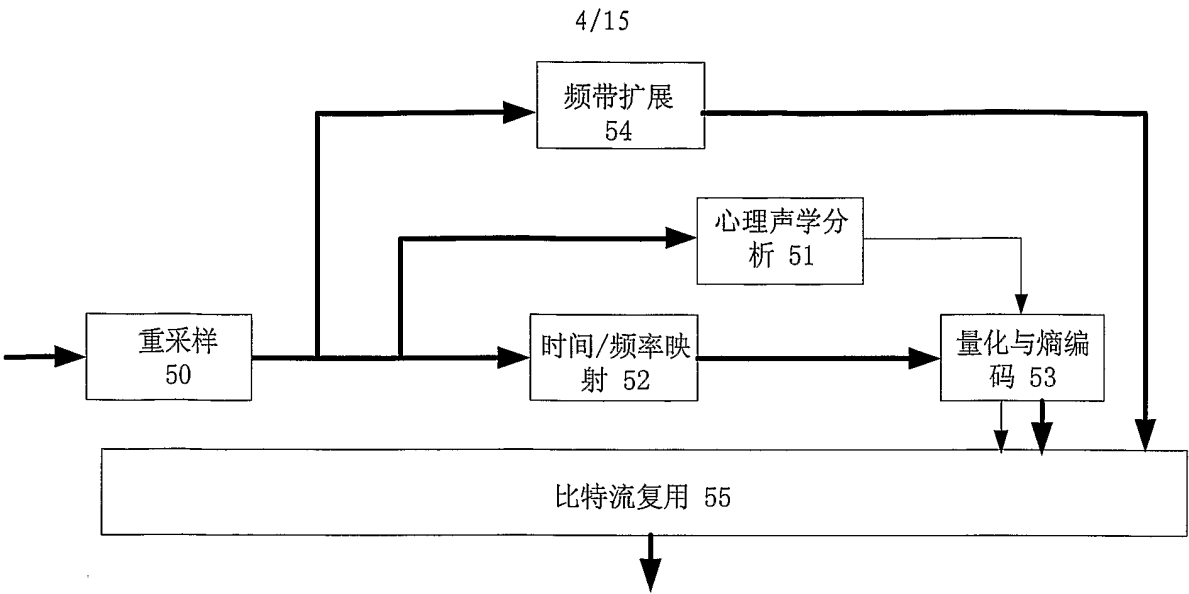


图 5

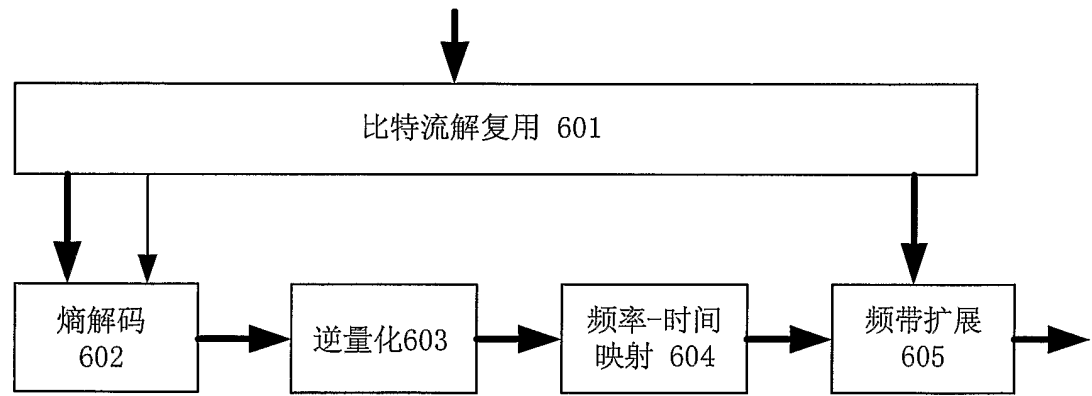


图 6

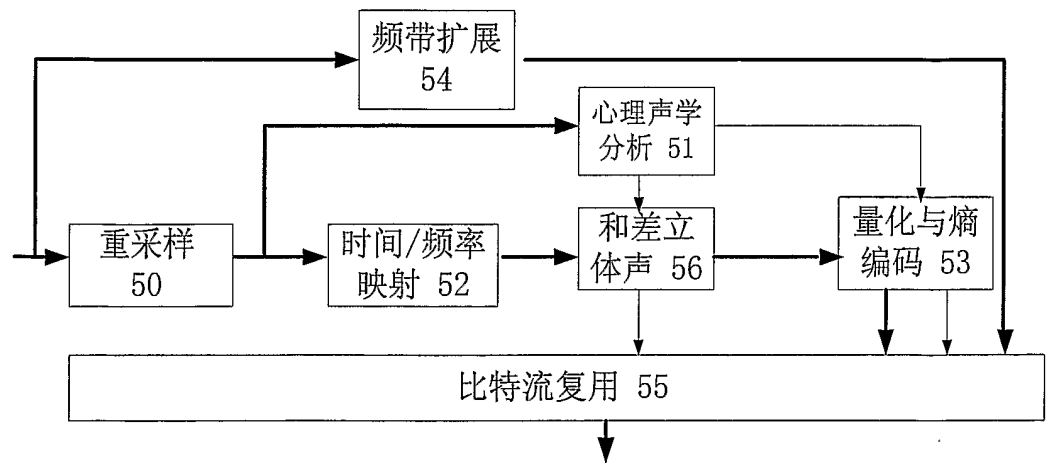


图 7

5/15

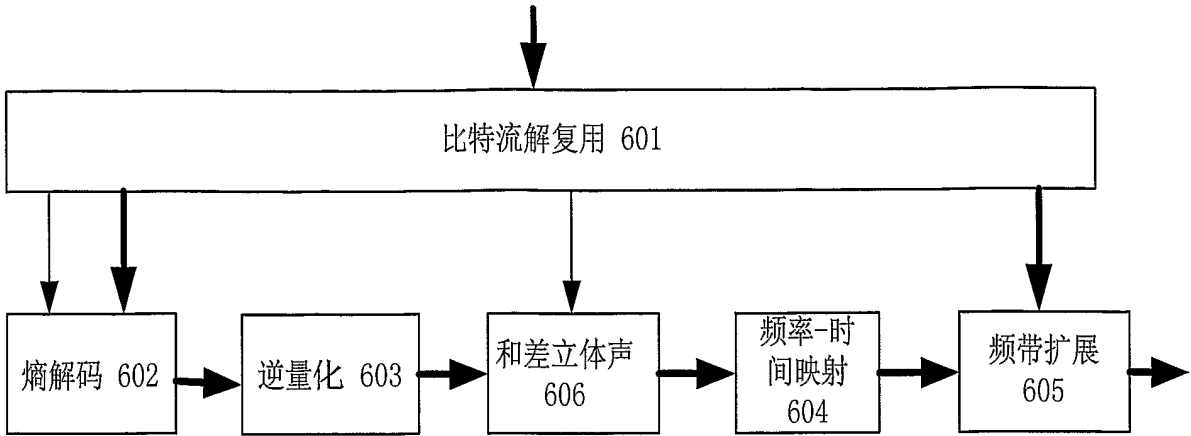


图 8

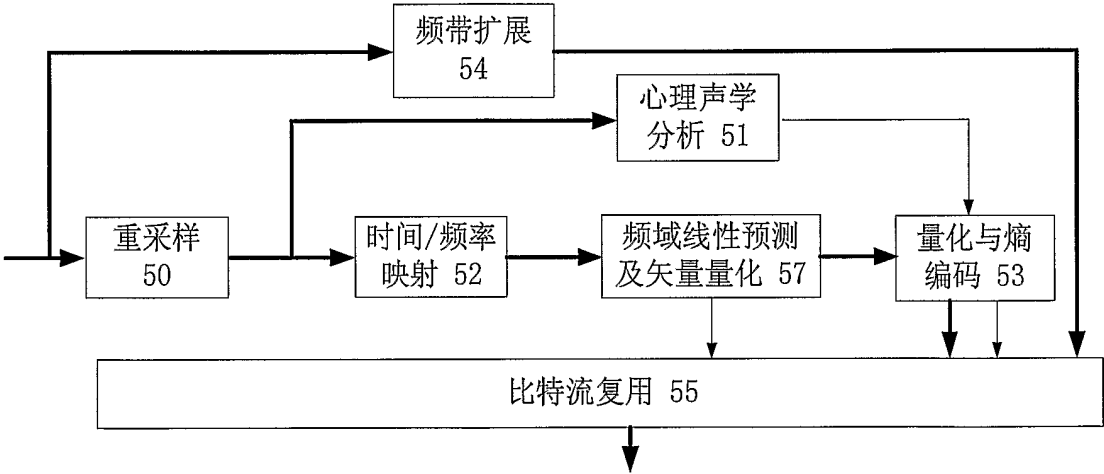


图 9

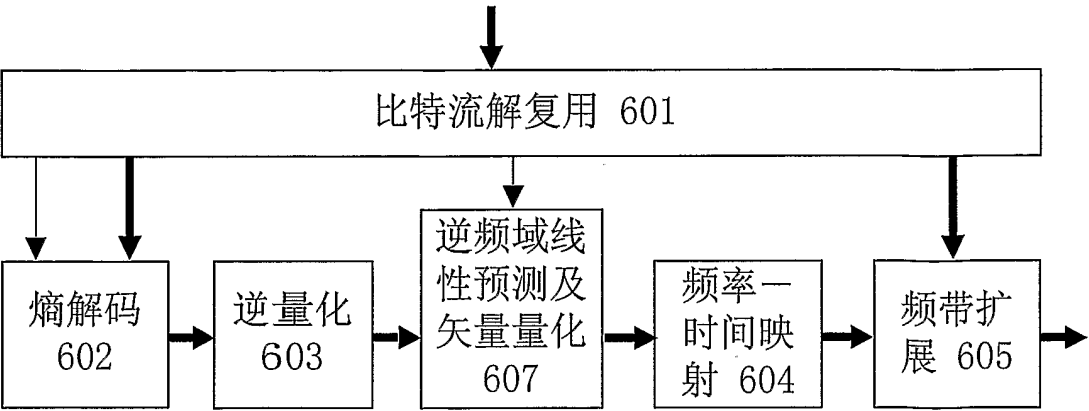


图 10

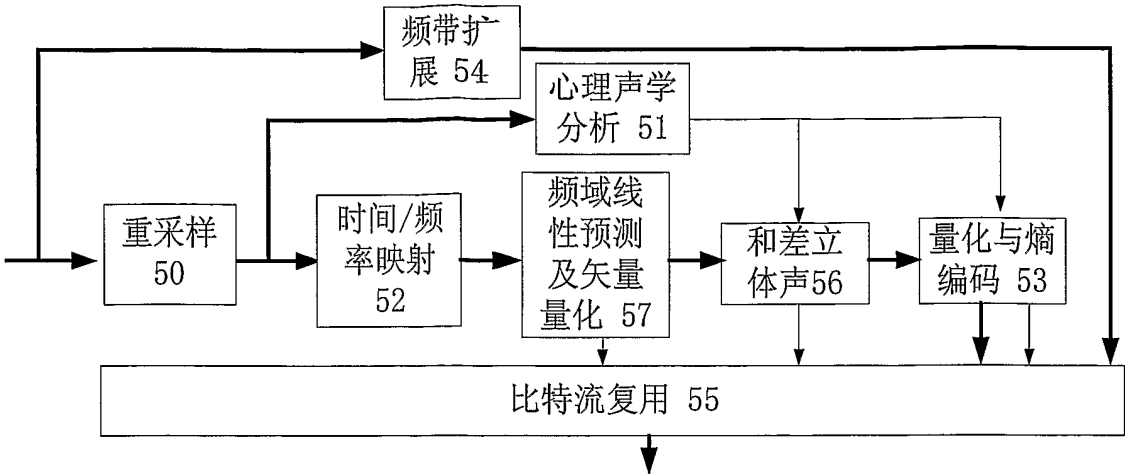


图 11

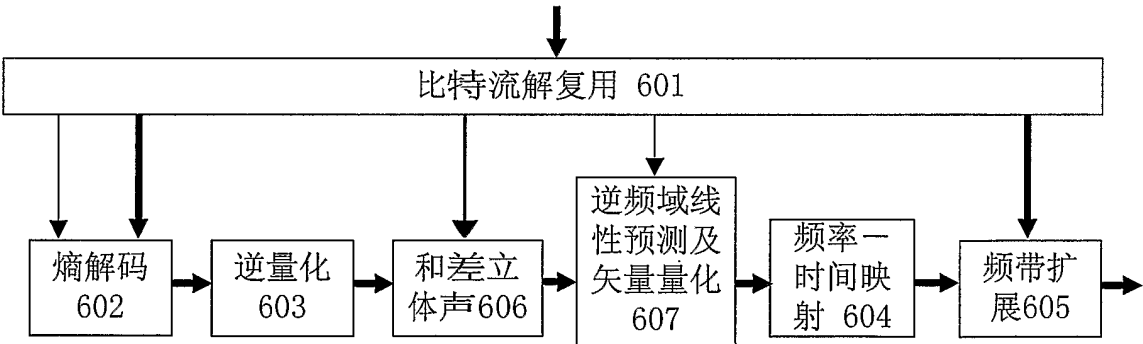


图 12

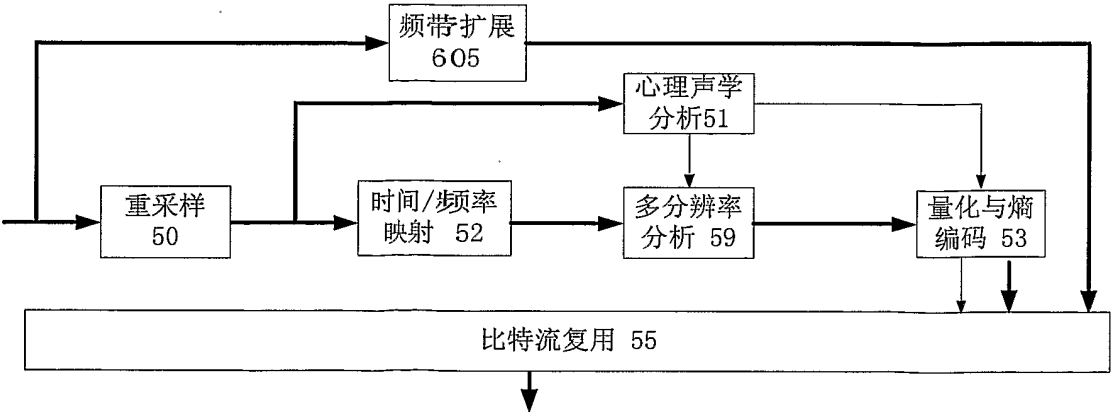


图 13

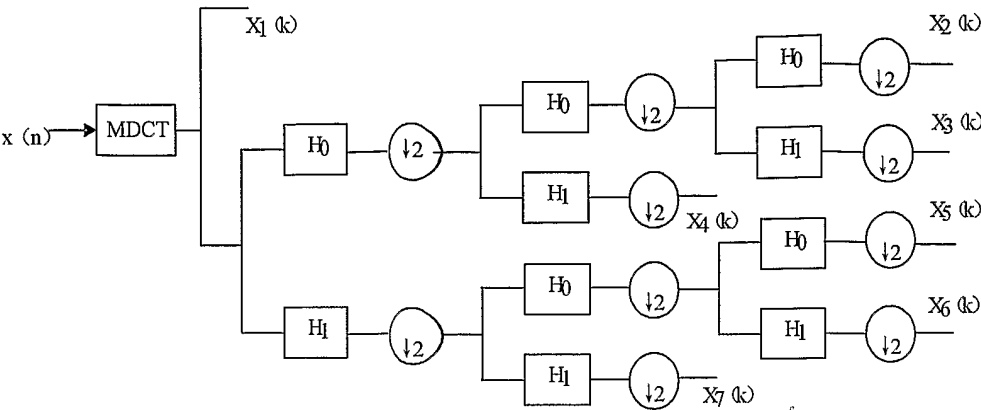


图 14

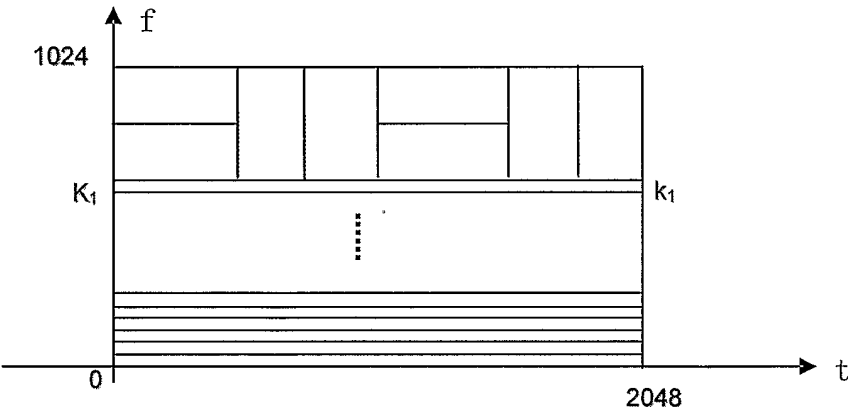


图 15

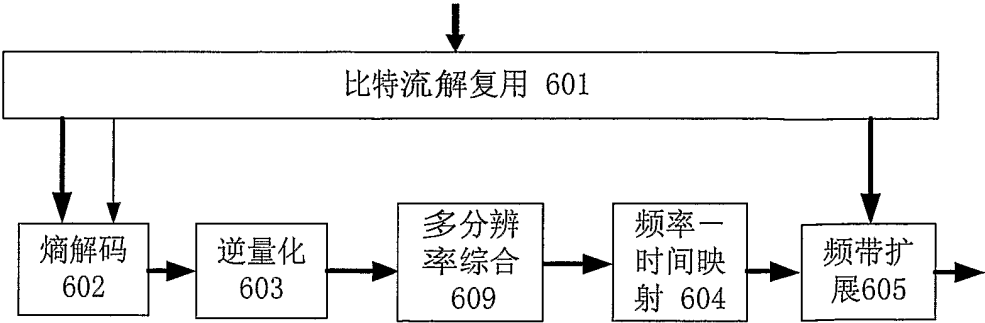


图 16

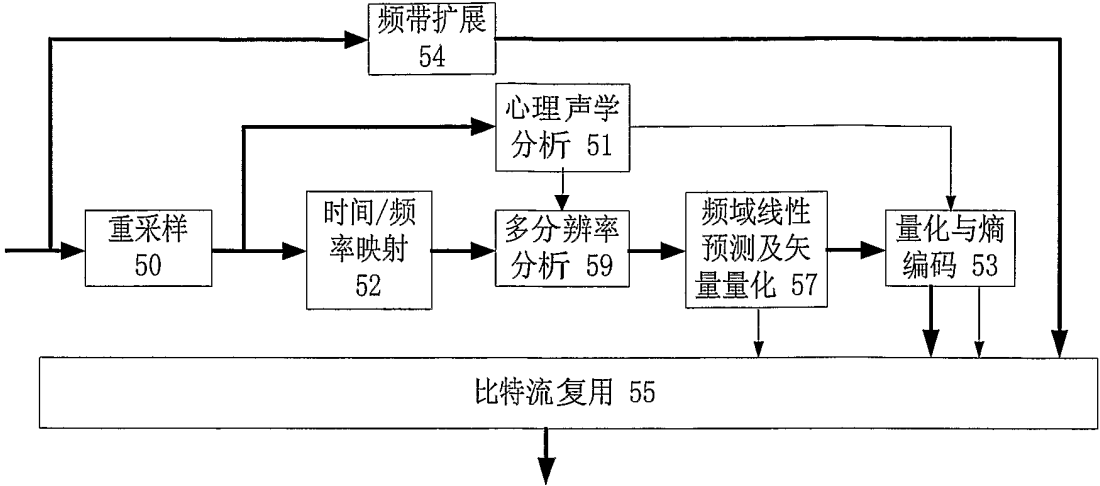


图 17

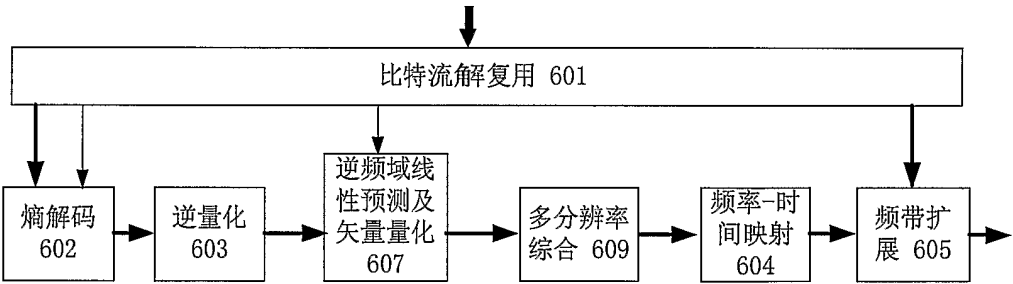


图 18

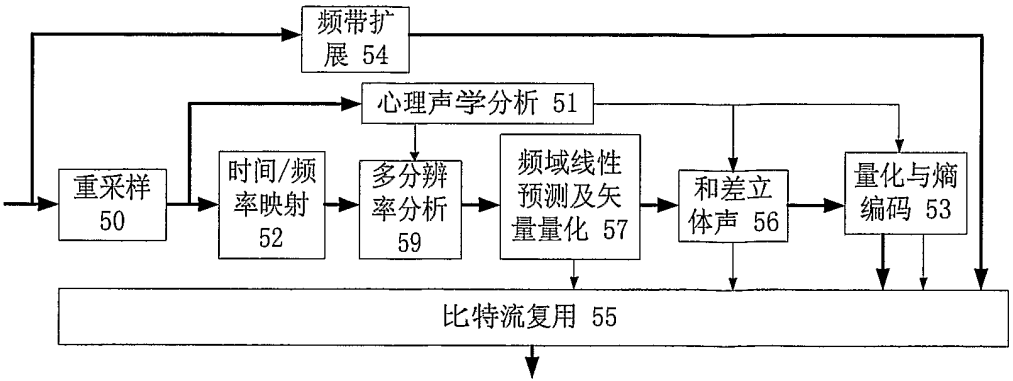


图 19

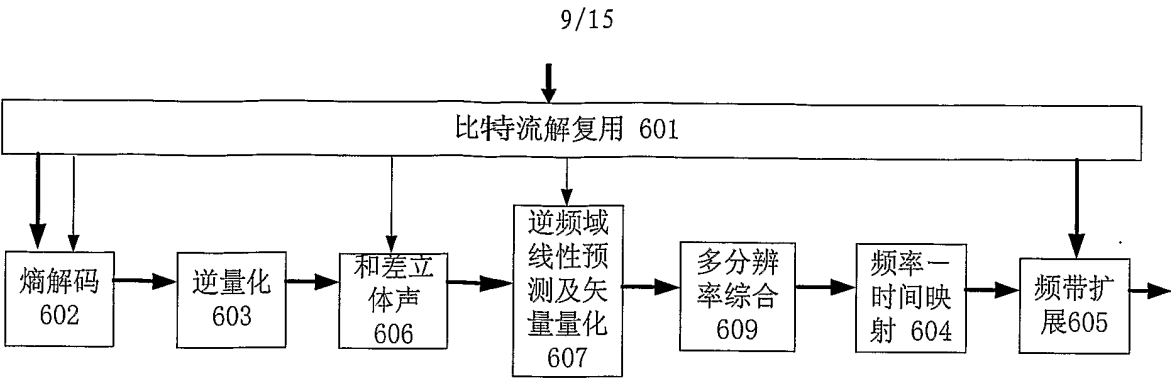


图 20

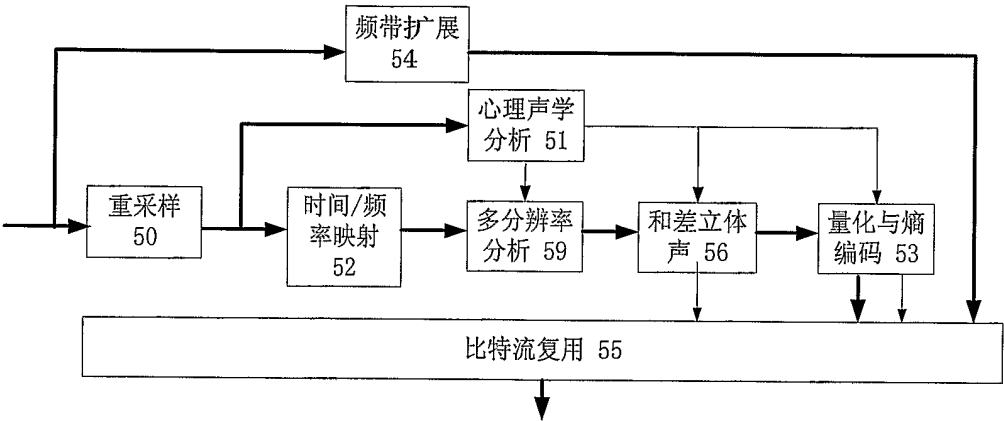


图 21

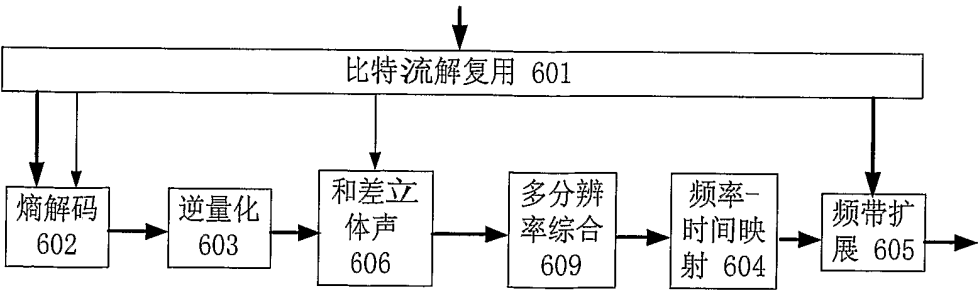


图 22

10/15

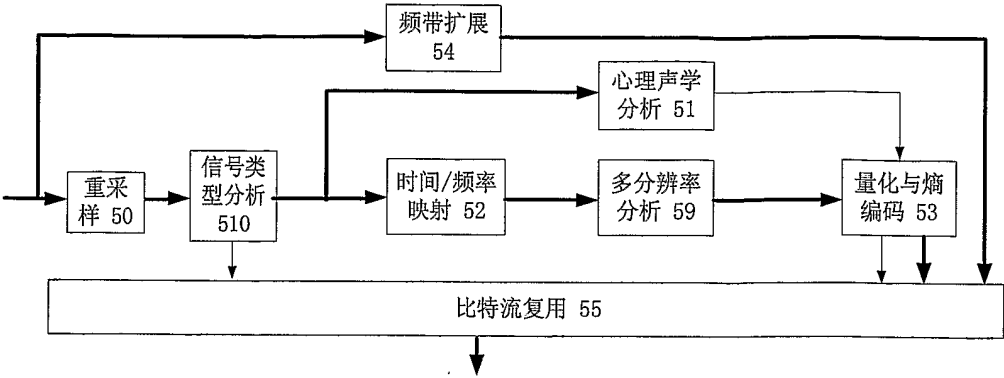


图 23

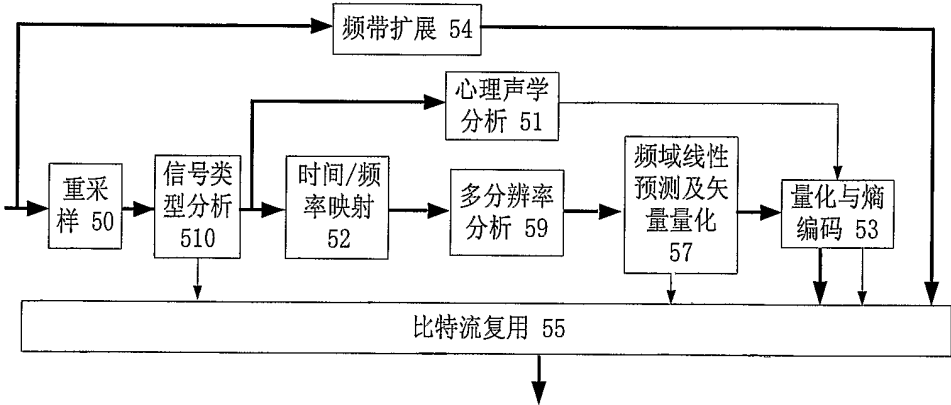


图 24

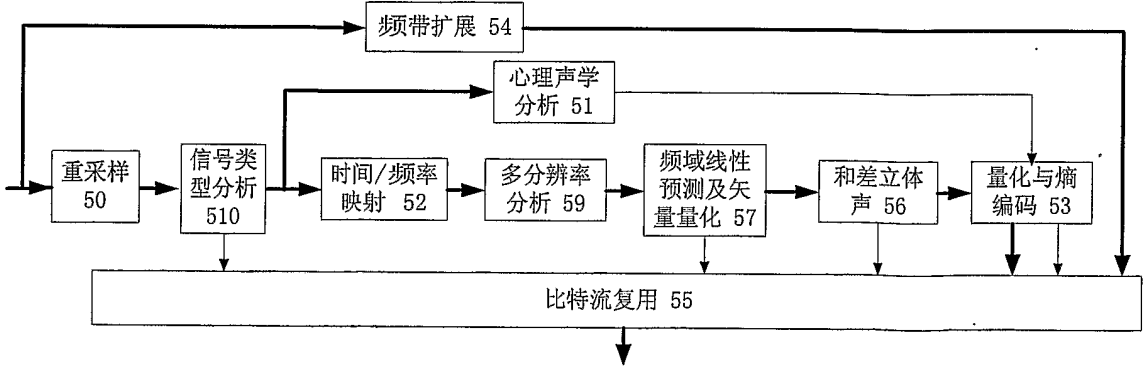


图 25

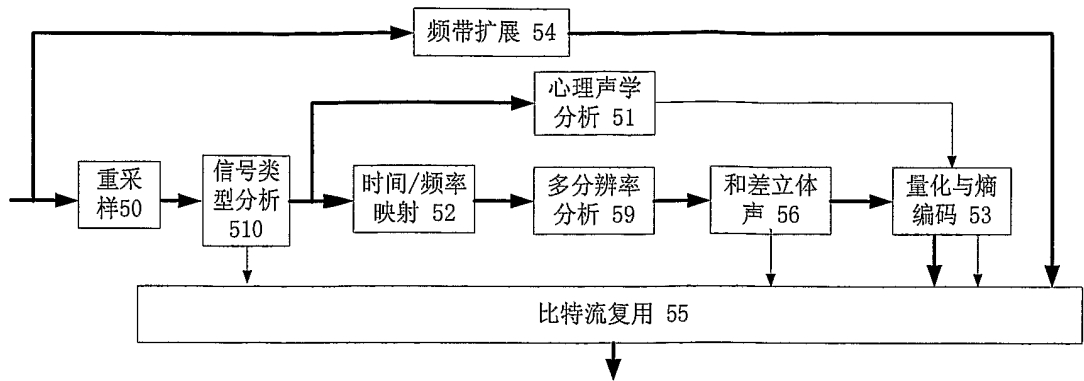


图 26

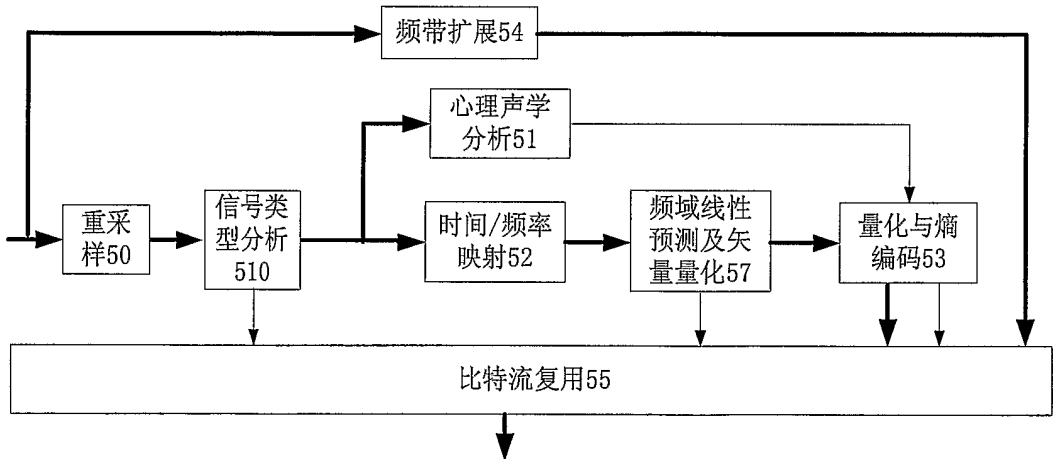


图 27

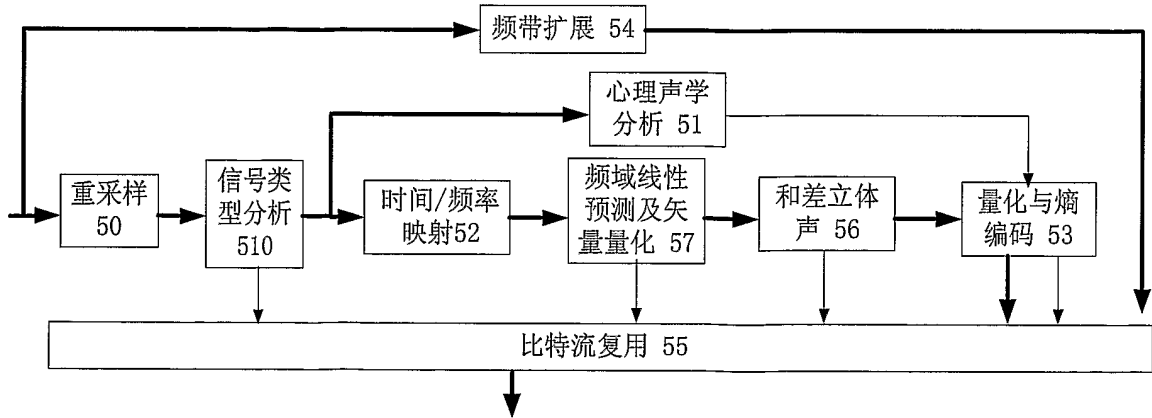


图 28

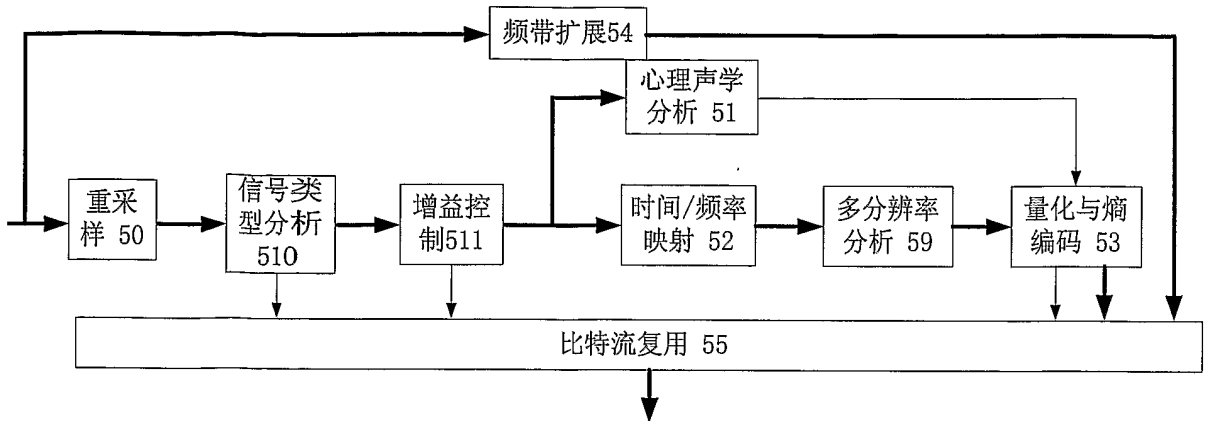


图 29

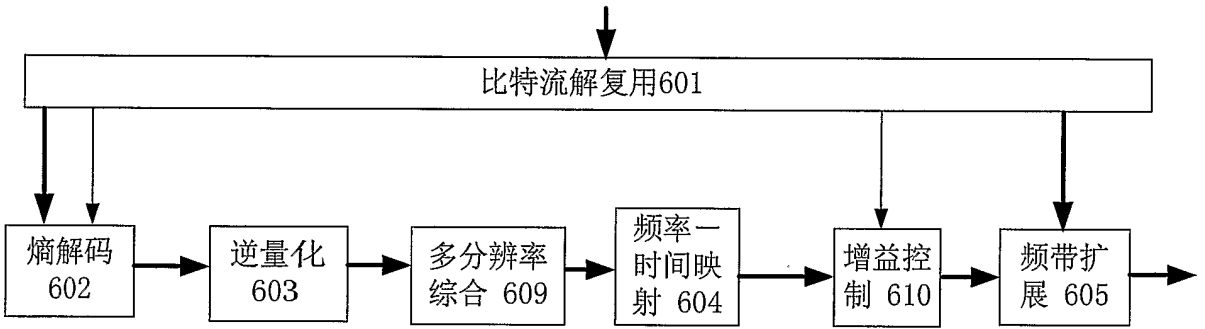


图 30

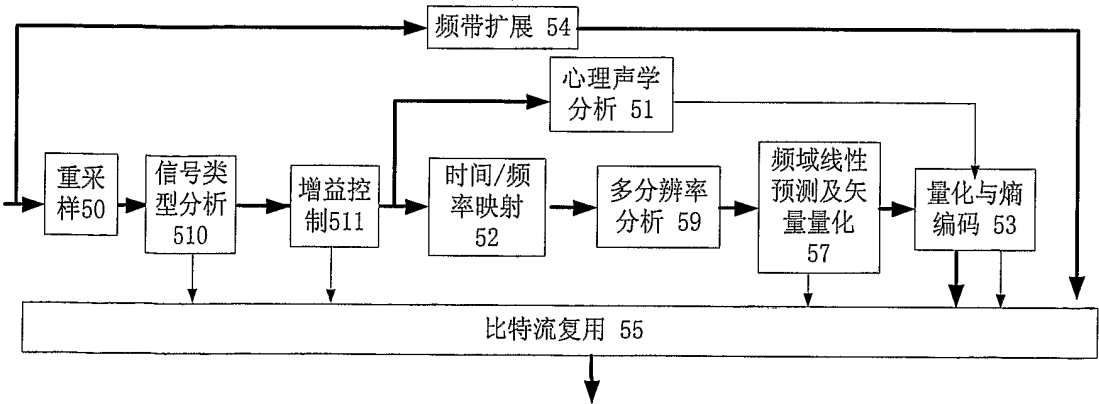


图 31

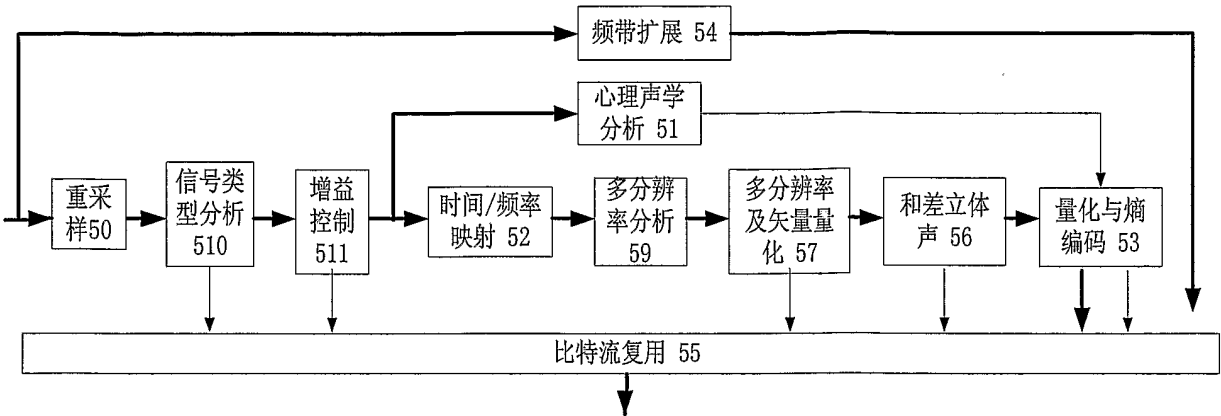


图 32

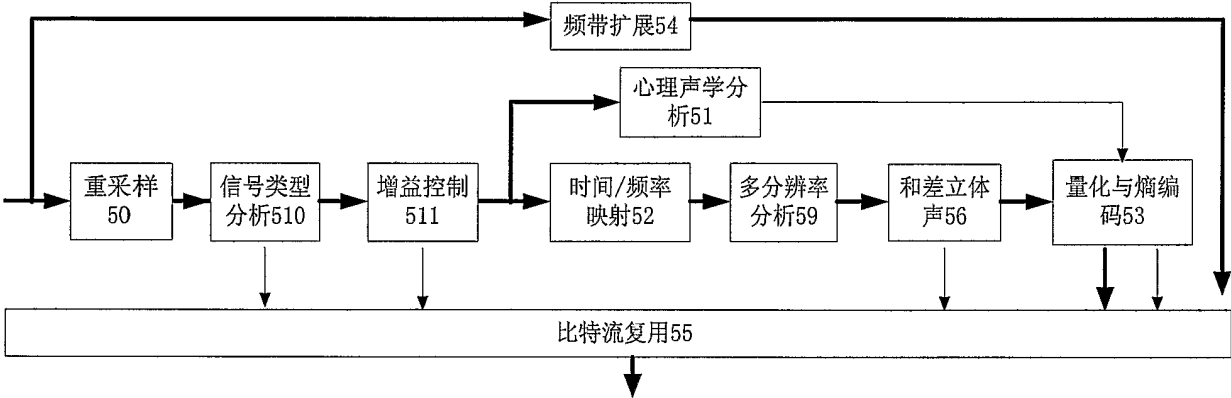


图 33

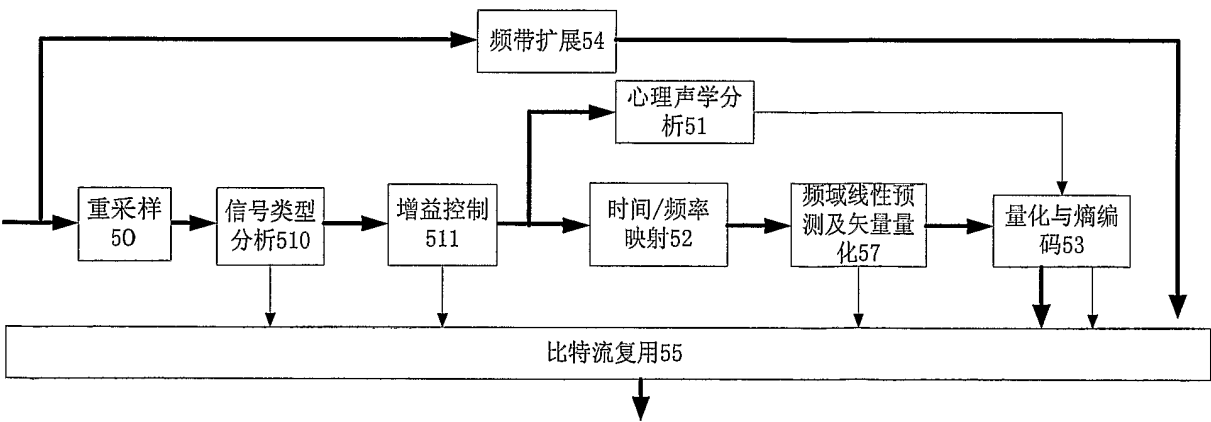


图 34

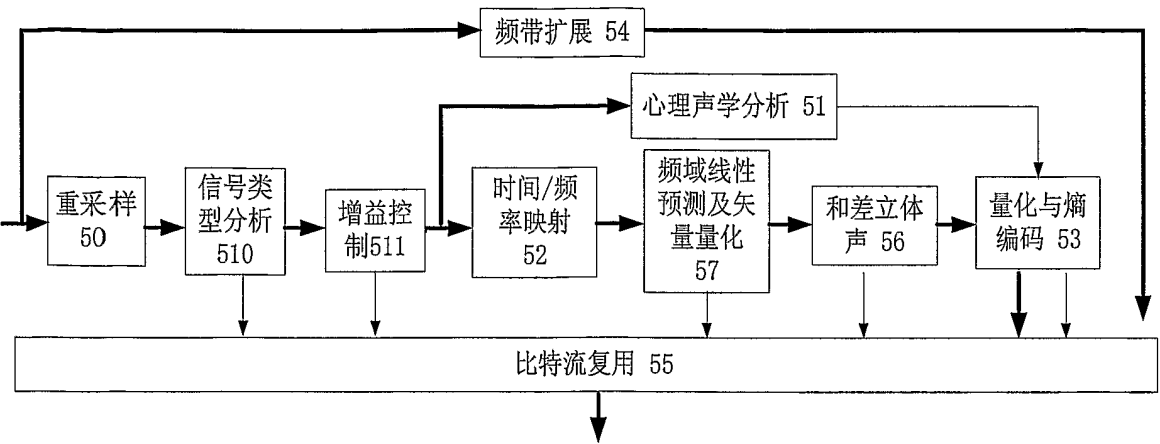


图 35

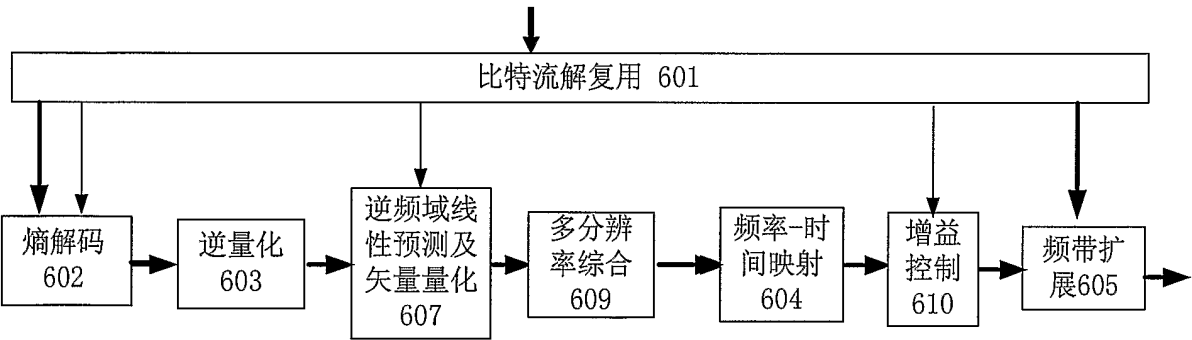


图 36

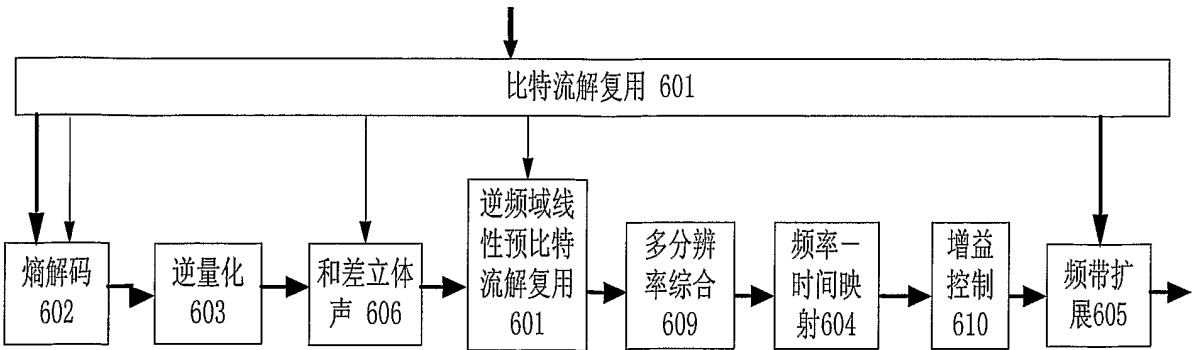


图 37

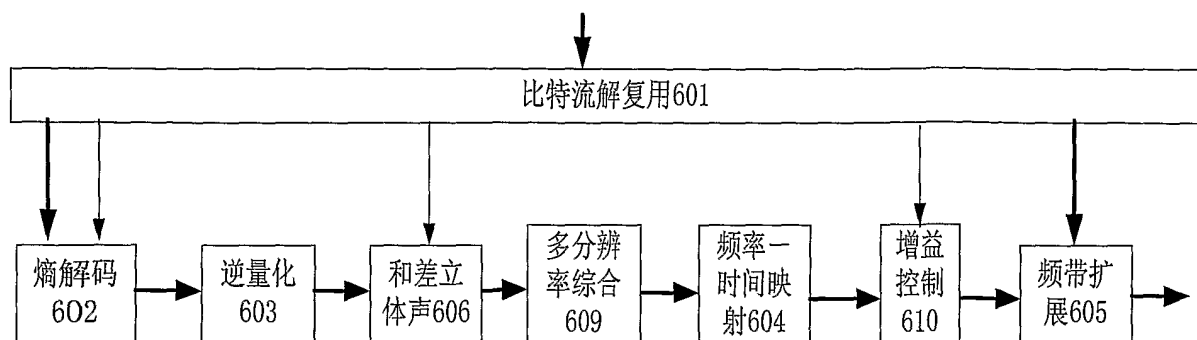


图 38

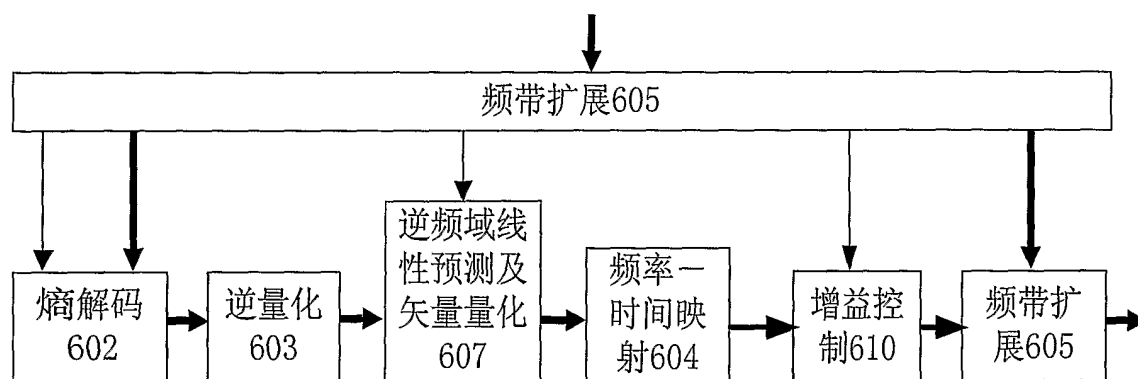


图 39

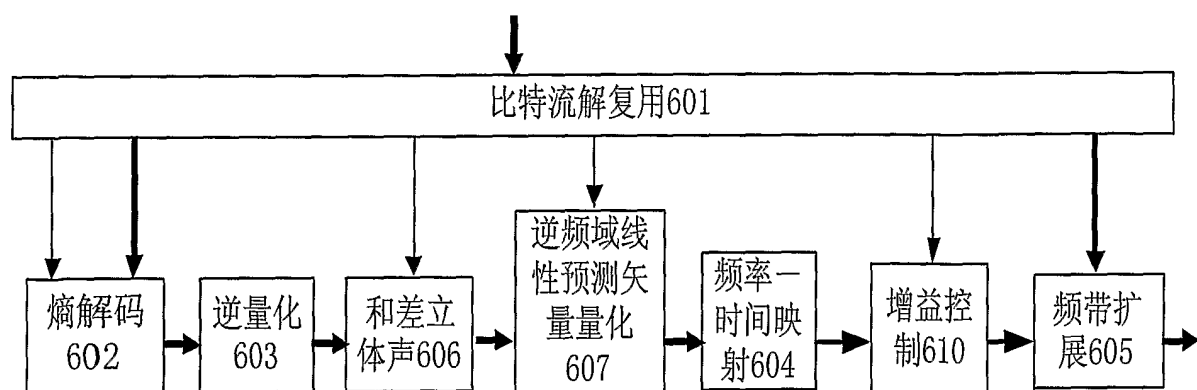


图 40

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2004/001034

A. CLASSIFICATION OF SUBJECT MATTER

H03M7/30 G10L19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H03M7/30 G10L+

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT WPI EPODOC PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP1351218A2 (TOSHIBA KK) 8.Oct.2003 (08/10/2003)	1-29
A	CN1213935A (MATSUSHITA ELECTRIC IND CO LTD) 14.Apr. 1999 (14/04/1999)	1-29

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
04.Jan.2005(04/01/2005)

Date of mailing of the international search report
20 · JAN 2005 (20 · 01 · 2005)

Name and mailing address of the ISA/
6 Xitucheng Rd., Jimen Bridge, Haidian District,
100088 Beijing, China
Facsimile No. (86-10)62019451

Authorized officer

wangqilin

Telephone No. 86-010-62085792

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2004/001034

Patent document Sited in search report	Publication date	Patent family Member(s)	Publication date
EP1351218A2	8.Oct.2003 (08/10/2003)	JP2003255997A	10.Sep.2003(10/09/2003)
		US2003171916A	11.Sep.2003(11/09/2002)
		CN1452155A	29.Oct.2003(29/10/2003)
CN1213935A	14.Apr.1999 (14/04/1999)	EP0901289A	10.Mar.1999(10/03/1999)
		JP11088876A	30.Mar.1999(30/03/1999)
		JP11098478A	09.Apr.1999(09/04/1999)
		JP11112998A	23.Apr.1999(23/04/1999)
		JP11136683A	21.May.1999(21/05/1999)
		US2001048719A	06.Dec.2001(06/12/2001)
		US6477202B	05.Nov.2002(05/11/2002)
		US6801574B	05.Oct.2004(05/10/2004)

国际检索报告

国际申请号
PCT/CN2004/001034

A. 主题的分类

IPC7 H03M7/30 G10L19/00

按照国际专利分类表(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC7 H03M7/30 G10L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT WPI EPODOC PAJ

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	EP 1351218A2 (TOSHIBA KK) 8.10 月.2003 (08.10.2003)	1-29
A	CN1213935A (MATSUSHITA ELECTRIC IND CO LTD) 14.4 月 1999 (14.04.1999)	1-29

☐ 其余文件在 C 栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

04.1 月 2005 (04.01.2005)

国际检索报告邮寄日期

20.1月 2005 (20.01.2005)

中华人民共和国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

授权官员

王琦琳

电话号码: (86-10)62085792

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2004/001034

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
EP1351218A2	8.10 月.2003 (08.10.2003)	JP2003255997A	10.09 月.2003(10.09.2003)
		US2003171916A	11.09 月.2003(11.09.2003)
		CN1452155A	29.10 月.2003(29.10.2003)
CN1213935A	14.4 月 1999 (14.04.1999)	EP0901289A	10.03 月.1999(10.03.1999)
		JP11088876A	30.03 月.1999(30.03.1999)
		JP11098478A	09.04 月.1999(09.04.1999)
		JP11112998A	23.04 月.1999(23.04.1999)
		JP11136683A	21.05 月.1999(21.05.1999)
		US2001048719A	06.12 月.2001(06.12.2001)
		US6477202B	05.11 月.2002(05.11.2002)
		US6801574B	05.10 月.2004(05.10.2004)