(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 September 2011 (29.09.2011)

PCT

(10) International Publication Number
WO 2011/119111 A1

(51) International Patent Classification:
H04N 7/26 (2006.01) *G10L 19/00* (2006.01)

(21) International Application Number:
PCT/SG2011/000112

(22) International Filing Date:
22 March 2011 (22.03.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201002108-7 26 March 2010 (26.03.2010) SG

(71) Applicant *(for all designated States except US)*: **AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH** [SG/SG]; 1 Fusionopolis Way #20-10, Connexis, Singapore 138632 (SG).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YU, Rongshan** [SG/SG]; PKM, Institute for Infocomm Research, 1 Fusionopolis Way #21-01, Connexis South Tower, Singapore 138632 (SG). **LI, Te** [CN/SG]; PKM, Institute for Infocomm Research, 1 Fusionopolis Way #21-01, Connexis South Tower, Singapore 138632 (SG). **SHU, Haiyan** [CN/SG]; PKM, Institute for Infocomm Research, 1 Fusionopolis Way #21-01, Connexis South Tower, Singapore 138632 (SG). **RAHARDJA, Susanto** [ID/SG]; PKM, Institute for Infocomm Research, 1 Fusionopolis Way #21-01, Connexis South Tower, Singapore 138632 (SG).

(74) Agent: **VIERING, JENTSCHURA & PARTNER LLP**; P.O. Box 1088, Rochor Post Office, Rochor Road, Singapore 911833 (SG).

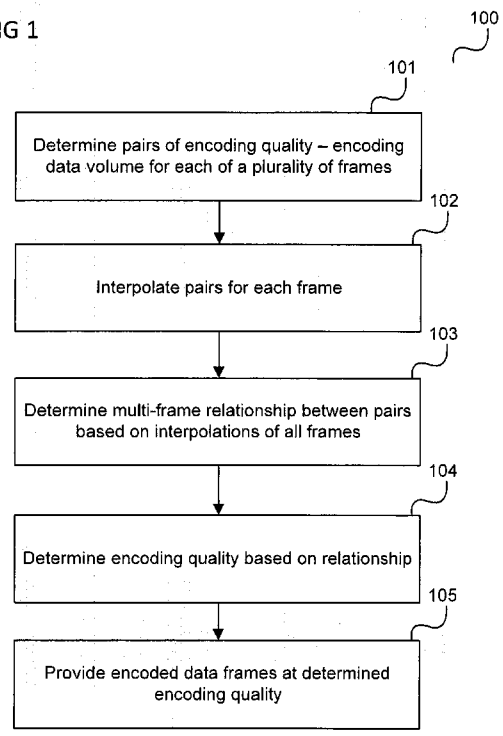
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

[Continued on next page]

(54) Title: METHODS AND DEVICES FOR PROVIDING AN ENCODED DIGITAL SIGNAL

FIG 1



(57) Abstract: In one embodiment, a method for providing an encoded digital signal is described comprising determining, for each data frame of a plurality of data frames of a digital signal, a plurality of pairs of an encoding data volume and an encoding quality, wherein each pair of an encoding data volume and an encoding quality specifies the encoding data volume required for achieving the encoding quality; determining for each data frame at least one or more interpolations between the plurality of determined pairs; determining a multi-frame relationship between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality based on a combination of the at least one or more interpolations for the plurality of data frames; determining an encoding quality for the plurality of data frames based on the relationship; and providing at least one data frame of the plurality of data frames encoded at the determined encoding quality.

WO 2011/119111 A1

GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

METHODS AND DEVICES FOR PROVIDING AN ENCODED DIGITAL SIGNAL

Field of the invention

Embodiments of the invention generally relate to methods and devices for providing an encoded digital signal.

Background of the invention

Audio streaming typically refers to constantly distributing audio content over a communication network from a streaming provider to an end-user. Usually, the audio content is compressed to a lower data rate (compared to the data rate of the original audio content) prior to streaming by using an audio coding technology so that the communication network bandwidth can be used efficiently.

Typically, in an audio encoder, audio content is segmented into a sequence of audio frames of constant time duration (referred to as frame length), and the audio frames are further processed so that redundancies and/or irrelevant information are removed from the audio frames, resulting in a compressed audio bit-stream with reduced data rate compared to the data rate of the original audio content.

Traditional audio codecs such as mp3 or MPEG-4 AAC produce a Constant Bit-Rate (CBR) bit stream that consists of compressed audio frames of equal size throughout the audio content. Due to the non-stationary or unstationary nature of audio signals, a CBR audio bit-stream typically exhibits quality fluctuation at multi time scales. As a result, streaming of CBR audio may result in unstable quality which is perceptually annoying to the end user and poor perceptual

quality at critical frames of audio signal, i.e., audio frames requiring more transmission bits to achieve the same quality compared with other frames of the audio signal.

This may be addressed using a Variable Bit-Rate (VBR) audio codec which generates variable bit-rate, but constant quality bit-streams. However, although VBR coding can be used to avoid quality fluctuation, VBR audio is in general not communication network friendly as the bit rate fluctuation of VBR encoded audio signals is typically content dependent and fixed after the encoding process. Therefore, it can conflict with actual available resource of the communication network during streaming.

The introduction of Fine Granular Scalable (FGS) audio coding such as MPEG-4 Scalable to Lossless (SLS) coding may allow solving the above issues.

Unlike other audio codecs the compressed audio frames produced by an FGS encoder can be further truncated to lower data rates at little or no additional computational cost. This feature allows an audio streaming system to adapt the streaming quality/rate in real-time depending on both the available bandwidth for streaming and the criticalness of the audio frames being streamed so that both constant quality streaming and network friendliness may be achieved.

Efficient methods for controlling FGS encoding with regard to the achieved audio quality and available bandwidth usage are desirable.

Documents [1] and [2] describe rate-quality models based on pre-measured data points and linear interpolation for rate

control of video coding and adaptive FGS video streaming, respectively. The method of [2] relies on iterative bisectional search, which has relatively high computational complexity.

In document [3] an idea on constant quality adaptive streaming has been proposed for video streaming wherein the target quality selection is over the entire media file. The rate-quality model, which is based on parameterized non-linear functions, is customized for naïve MSE quality measure for video/image in general.

Summary of the invention

In one embodiment, a method for providing an encoded digital signal is provided including determining, for each data frame of a plurality of data frames of a digital signal, a plurality of pairs of an encoding data volume and an encoding quality, wherein each pair of an encoding data volume and an encoding quality specifies the encoding data volume required for achieving the encoding quality; determining for each data frame at least one or more interpolations between the plurality of determined pairs; determining a multi-frame relationship between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality based on a combination of the at least one or more interpolations for the plurality of data frames; determining an encoding quality for the plurality of data frames based on the relationship; and providing at least one data frame of the plurality of data frames encoded at the determined encoding quality.

Short description of the figures

Illustrative embodiments of the invention are explained below with reference to the drawings.

Fig. 1 shows a flow diagram according to an embodiment.

Fig. 2 shows a device for providing an encoded digital signal according to an embodiment.

Fig. 3 shows a communication arrangement according to an embodiment.

Fig. 4 shows frame structures according to an embodiment.

Fig. 5 shows a flow diagram according to an embodiment.

Fig. 6 shows a quality-bit rate diagram according to an embodiment.

Fig. 7 shows an encoding data volume-encoding quality diagram according to an embodiment.

Fig. 8 shows a data rate-time diagram.

Fig. 9 shows a communication arrangement according to an embodiment.

Fig. 10 shows a flow diagram according to an embodiment.

Fig. 11 shows a device for providing an encoded digital signal.

Fig. 12 shows a communication arrangement according to an embodiment.

Detailed description

According to one embodiment, an adaptive streaming system (specifically an encoder, e.g. being part of a transmitter and an encoding method) for FGS audio is provided that maintains a constant quality streaming as much as possible while at the same time fully utilizing the bandwidth available for the streaming.

To this end, according to an embodiment, a target quality is first selected, and the sizes of the audio frames to be streamed are truncated accordingly so that this target quality is achieved.

To ensure best possible quality of the streamed audio while at the same time not to over-utilize the network resource, according to one embodiment a target encoding quality is selected such that the rate of the truncated bit-stream, on average, is within the constraint of available network bandwidth for the streaming. In order to effectively determine the target quality and the sizes of the truncated audio frames the adaptive streaming server (i.e. the transmitter or the encoder) is according to one embodiment made aware of the rate-quality relationship (i.e. the relationship between the encoding rate and the encoding quality achieved with the encoding rate) of the audio to be streamed at the audio frame level. This rate-quality relationship may be highly non-uniform and highly dynamic in general. As a result, it may not be easy to convey this information to the streaming server.

According to one embodiment, the streaming server (specifically a data rate (or encoding data volume) controller) is provided with the rate-quality relationship of the audio to be streamed by using a rate-quality model based on pre-measured data points and linear interpolation. This rate-quality model allows highly effective adaptive streaming at low complexity.

According to one embodiment, a sliding window is introduced so that the target quality selection can be seen to be "localized" to audio frames from a window of limited duration (e.g. in terms of a certain number of frames). The introduction of the sliding window can be seen to localize bit-rate fluctuation of the streamed audio so that it is more accommodating with available network bandwidth estimated during streaming.

Further, according to one embodiment, a pre-measured rate-quality table based model is used which is suitable for FGS audio and leads to an easy solution for the problem of selecting the target encoding quality/data rate for streaming.

According to one embodiment, a rate-quality model is used based on piece-wise linear functions and a closed-form low-complexity solution for selecting the target quality/rates for streaming is used. This allows lower computational complexity than for example using a Newton search algorithm.

A method for providing an encoded digital signal according to an embodiment is illustrated in figure 1.

Fig. 1 shows a flow diagram 100 according to an embodiment.

The flow diagram 100 illustrates a method for providing an encoded digital signal.

In 101, for each data frame of a plurality of data frames of a digital signal, a plurality of pairs of an encoding data volume and an encoding quality are determined, wherein each pair of an encoding data volume and an encoding quality specifies the encoding data volume required for achieving the encoding quality.

In 102, for each data frame at least one or more interpolations between the plurality of determined pairs are determined.

In 103, a multi-frame relationship between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality is determined based on a combination of the at least one or more interpolations for the plurality of data frames.

In 104, an encoding quality for the plurality of data frames is determined based on the relationship.

In 105 at least one data frame of the plurality of data frames is provided encoded at the determined encoding quality.

According to one embodiment, in other words, approximations for the dependence between encoding data volume and encoding quality for each of a plurality of frames are determined by interpolation of pre-determined (e.g. measured) pairs of

encoding data volume and encoding quality. These approximations are combined to have a multi-frame dependence between encoding data volume and encoding quality, i.e. a dependence between encoding data volume and encoding quality for the whole plurality of data frames. This overall dependence is then used to determine an encoding quality to be used for the frames (or at least a part of the frames until the encoding quality to be used is re-determined, e.g. on a periodic basis).

The digital signal is for example a media data signal, such as an audio or a video signal.

According to one embodiment the relationship specifies for each encoding quality of a plurality of encoding qualities a corresponding encoding data volume required to encode the plurality of data frames at the encoding quality.

According to one embodiment the encoding quality for the plurality of data frames is determined such that the encoding data volume corresponding to the determined encoding quality according to the relationship fulfils a predetermined criterion.

According to one embodiment the criterion is that the encoding data volume is below a pre-determined threshold.

According to one embodiment the threshold is based on a maximum data rate.

According to one embodiment the multi-frame relationship is determined based on a combination of the at least one or more

interpolations for at least two different data frames of the plurality of data frames.

According to one embodiment the at least one interpolation of a data frame of the plurality of data frames is an interpolation of the plurality of encoding data volume and encoding quality pairs of the data frame.

According to one embodiment the at least one interpolation of a data frame of the plurality of data frames is a linear interpolation of the plurality of encoding data volume and encoding quality pairs of the data frame.

According to one embodiment the plurality of data frames is a plurality of successive data frames.

According to one embodiment the at least one data frame of the plurality of data frames provided encoded at the determined encoding quality includes the first data frame of the plurality of successive data frames encoded at the determined encoding quality.

The method may further include determining a further encoding quality to be used for a further plurality of successive data frames including the plurality of data frames without the at least one data frame provided encoded at the determined encoding quality.

According to one embodiment each interpolation of the at least one or more interpolations between the plurality of determined pairs for a data frame is an interpolated pair of an encoding data volume and an encoding quality specifying

the encoding data volume required for achieving the encoding quality for the data frame.

According to one embodiment the multi-frame relationship is determined based on a summing of the encoding data volumes required for achieving an encoding quality for different data frames for the same encoding quality.

According to one embodiment the result of the summing is specified by the relationship for an encoding quality as a corresponding encoding data volume required to encode the plurality of data frames at the encoding quality.

According to one embodiment the multi-frame relationship is a piecewise linear correspondence between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality.

According to one embodiment the plurality of pairs of an encoding data volume and an encoding quality for each data frame are generated by measuring, for each of a plurality of encoding data volumes, the encoding quality achieved when encoding the data frame using the encoding data volume.

According to one embodiment the digital signal is an audio signal.

It should be noted that, as in the example described below, providing an encoded frame at a quality may include having a frame encoded at a higher quality (e.g. stored in a memory) and reducing the quality of the frame encoded at the higher quality e.g. by truncating the frame encoded at the higher quality.

The method illustrated in figure 1 is for example carried out by a device as illustrated in figure 2.

Fig. 2 shows a device for providing an encoded digital signal 200 according to an embodiment.

The device 200 includes a first determining circuit 201 configured to determine, for each data frame of a plurality of data frames of a digital signal, a plurality of pairs of an encoding data volume and an encoding quality, wherein each pair of an encoding data volume and an encoding quality specifies the encoding data volume required for achieving the encoding quality.

Further, the device 200 includes an interpolator 202 configured to determine for each data frame at least one or more interpolations between the plurality of determined pairs and a combiner 203 configured to determine a multi-frame relationship between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality based on a combination of the at least one or more interpolations for the plurality of data frames.

The device 200 further includes a second determining circuit 204 configured to determine an encoding quality for the plurality of data frames based on the relationship and an output circuit 205 providing at least one data frame of the plurality of data frames encoded at the determined encoding quality.

The device 200 is for example part of a server computer (e.g. a streaming server (computer)) providing encoded data, e.g.

encoded media data such as encoded audio data or encoded video data.

In an embodiment, a "circuit" may be understood as any kind of a logic implementing entity, which may be special purpose circuitry or a processor executing software stored in a memory, firmware, or any combination thereof. Thus, in an embodiment, a "circuit" may be a hard-wired logic circuit or a programmable logic circuit such as a programmable processor, e.g. a microprocessor (e.g. a Complex Instruction Set Computer (CISC) processor or a Reduced Instruction Set Computer (RISC) processor). A "circuit" may also be a processor executing software, e.g. any kind of computer program, e.g. a computer program using a virtual machine code such as e.g. Java. Any other kind of implementation of the respective functions which will be described in more detail below may also be understood as a "circuit" in accordance with an alternative embodiment. Further, it should be noted that different circuits may be implemented by the same circuitry, e.g. by only one processor.

An adaptive streaming system according to an embodiment, for example including a device as shown in figure 1 on the transmitter side is described in the following with reference to figure 3.

Fig. 3 shows a communication arrangement 300 according to an embodiment.

The communication arrangement 300 includes a transmitter 301 and a receiver 302. The transmitter 301 includes a scalable audio encoder 303 providing a scalable audio file 304 and a rate-quality table 305. The transmitter 301 further includes

a frame truncator 306 receiving the scalable audio file 304 as input and a rate controller 307 receiving the rate-quality table 305 as input. The transmitter 301 further includes a network bandwidth estimator 308 and a transmitting module 309. The receiver 302 includes a receiving module 310 and a streaming client 311. The streaming client 311 may for example be a software application running on the receiver 302 for playing audio to the user of the receiver 302.

The transmitter 301 streams encoded audio content at a certain encoding quality to the receiver 302 over a communication network 312, e.g. via a computer network such as the Internet or via a radio communication network such as a cellular mobile communication network, to the receiver 302.

The audio content is transmitted in a plurality of encoded audio frames, wherein each audio frame is encoded at a certain encoding quality.

The rate controller 307 selects the target encoding quality of the audio frames based on information from both the rate-quality table 305 and the available network bandwidth of the communication network 312 estimated by the network bandwidth estimator 308. Once the target quality is selected, the scalable audio file 304 is truncated accordingly, and sent to via the communication network 312 for streaming to the receiver (and ultimately to the streaming client 311).

The scalable audio file 304 may be provided by the scalable audio encoder 303, e.g. from audio content supplied to the transmitter 301. However, it should be noted that the scalable audio file 304 may also be pre-stored in the

transmitter 312, i.e. the scalable audio encoder 303 does not need to be part of the transmitter.

Examples for the detailed implementation of components of the transmitter 301 and the receiver 302 are described in more detail in the following.

The scalable audio file 304 may include the audio content to be streamed at high (or even lossless) quality. According to one embodiment, the scalable audio file 304 (including the audio content to be streamed, e.g. at high quality) is encoded according to MPEG-4 scalable lossless (SLS) coding. MPEG-4 scalable lossless (SLS) coding was released as a standard audio coding tool in June 2006. It allows the scaling up of a perceptually coded representation such as MPEG-4 AAC to a lossless representation with a wide range of intermediate bit rate representations.

One of the major merits of a FGS audio codec like MPEG-4 SLS can be seen in that the bit-stream generated by the encoding can be further truncated to lower data rates.

This is illustrated in figure 4.

Fig. 4 shows a first frame structure 401 and a second frame structure 402.

The first frame structure 401 for example corresponds to the scalable audio file 304 (e.g. is contained in the audio file 304) and second frame structure 402 for example corresponds to the output of the truncator 306.

The first frame structure 401 includes data for a plurality of losslessly encoded frames 403 and the second frame structure 401 includes data for a plurality of lossy encoded frames 404 (as an example three frames numbered from n-1 to n+1 are illustrated in this example).

Data sections 405 may be removed from the data of the losslessly encoded frames 403 to generate the data of the lossy encoded frames 404. The data section 405 of the data for a losslessly encoded frame 403 is for example an end section of the data (which are for example in the form of a bit-stream) for the losslessly encoded frame 403 such that the data for the losslessly encoded frame 403 may be simply truncated (e.g. by frame truncator 306) to generate the data for the lossy encoded frame 404.

The truncation can be done at any stage between the provider of the lossless bit-stream (e.g. included in first frame structure 401) and the streaming client (e.g. at a server or at a communication network gateway) and requires little computational resources. This merit may be particularly relevant for a streaming server or gateway that needs to handle large numbers of simultaneous streaming sessions.

For example, the first frame structure 401 includes a Lossless SLS bit-stream with frame size rn where n is the frame index and the second frame structure 402 includes the truncated SLS bit-stream with reduced bit-rate r'n. The truncation operation of SLS is done by simply dropping the end of each SLS frame of certain length from the SLS bit-stream of higher bit-rates (i.e. the data sections 405) according to the desired quality/rate of the truncated SLS bit-stream.

According to one embodiment, this possibility of truncation in FGS audio is used whereby the full-fidelity FGS audio (i.e. the losslessly or high quality encoded audio content as included in the scalable audio file 304) is truncated to lower data rates according to available bandwidth and quality demands before it is sent via the communication network 312 for streaming. It should be noted that MPEG-4 SLS is used as an example and embodiments are not limited to MPEG-4 SLS as scalable encoding process used for generating the scalable audio file 304.

According to one embodiment, the rate controller 307 determines the data rate of the encoded audio stream sent by the transmitter 301. Specifically, according to one embodiment, the rate controller 307 determines the sizes of the streamed FGS (encoded) audio frames based on a rate-quality relationship of the audio frames as well as the available network bandwidth. For this, according to one embodiment, the rate-quality table 305 is used.

The rate-quality relationship of the audio frames for example gives for each audio frame and each encoding quality of the audio frame the required encoding data rate (or, equivalently in case of a fixed frame rate, the encoding data volume) to achieve this encoding quality.

The detailed process for generating the rate-quality table according to one embodiment is illustrated in Fig. 5.

Fig. 5 shows a flow diagram 500 according to an embodiment.

The flow illustrates a process of constructing the rate-quality table 305 according to an embodiment.

As can be seen from the flow diagram 500, the process of constructing the rate-quality table 305 can be integrated with the encoding process of FGS audio, i.e. with the generation of the scalable audio file 304 generated by the scalable audio encoder 303. Accordingly, according to one embodiment (and as illustrated in figure 3) the scalable audio encoder 303 generates the scalable audio file 304.

The process is started for a frame in 501. In 502, a set of predetermined encoding data volumes r_i, $i = 1,\ldots,n$ (which can be seen to correspond to a data rate for the frame for a certain frame rate) are input.

In 502, a counter indicated by counter variable j is set to 1.

In 504, the frame is encoded such that the encoded frame has the data volume r_j.

In 505, the quality of the encoded audio frame is determined.

In 506, the pair of the data volume r_j and the determined quality is output as entry into the rate-quality table 305.

In 507, it is checked whether $j<J$ (i.e. whether the last encoding data volume has not already been reached in the process). If $j<J$, j is increased by one and the process continues with 504. If $j=J$, the process is ended (for this frame) in 509.

The process illustrated in figure 5 can be seen to include, during the encoding process, monitoring the size of the compressed (i.e. encoded) audio frame encoded so far and once the size is matches a certain pre-determined criterion, e.g., a pre-determined data rate r_j, computing the quality of the partially encoded audio frame, i.e., the quality of the resulting audio frame after decoding the encoded audio frame if the audio frame is encoded using the pre-determined data rate r_j (e.g. is truncated from the losslessly encoded audio frame to the size corresponding to r_j), and storing the computed quality together with the pre-determined data rate (or size) into the rate-quality table 305.

According to one embodiment, the process as described above with reference to figure 5 is performed for every audio frame during the encoding process. The resulting rate-quality table 305 may then be stored together with the scalable audio file 304, and may be used by the transmitter 301 (e.g. an audio streaming server) for the truncation process carried out by the frame truncator 306.

According to one embodiment, the data stored in the rate-quality table 305 resides only on the server side and is not sent to the receiver 302. Thus, these data do not increase the burden on the communication network 312 for the streaming process.

The encoding quality of an encoded audio frame is for example calculated as the minimum value of the Masking-to-Noise Ratios (MNRs) of all scale factor bands (sfb) for which the audio frame includes data. Other quality metrics (or measures) may be used.

Since the rate-quality table 305 generated according to the process explained above with reference to figure 5 only records a limited number of rate-quality points (i.e. pairs of encoding data rate (or encoding data volume) and encoding quality), the rate-quality points not recorded in the rate-quality table 305 are according to one embodiment determined by linear interpolation. This is for example done by the audio streaming server, e.g. by the rate controller 307 of the transmitter 301. This is illustrated in Fig. 6.

Fig. 6 shows a quality-bit rate diagram 600 according to an embodiment.

The bit rate (as example for data rate) is given by a first axis 601 in kbps (kilobits per second) and the quality is given in dB (decibel) as the masking to noise ratio.

Circles 603 indicate points (i.e. quality-data rate pairs) that have been determined for a frame, for example in the process illustrated in figure 5. A line 604 indicates the approximation of points determined by linear interpolation of the determined points. In other words, the line 604 indicates an interpolated piecewise linear quality-rate (or rate-quality) function for the frame generated from the determined quality-data rate pairs.

Crosses 605 indicate actual quality-data rate pairs for the frame.

As can be seen from the diagram 600 the linear interpolation is only an approximation of the actual rate-quality function and it introduces approximation error for "real" points

(which are marked by the crosses 605) in-between the interpolation points (marked by the circles 604).

In practical application, the approximation error is usually tolerable if the density of the data points for interpolation is carefully chosen. Further, as shown below, the linearly interpolated rate-quality function can be used to simplify the determination of a (target) encoding quality to be used for a rate-quality optimized audio streaming solution, namely to solving linear equations.

In the following it is explained how the rate controller 307 may derive the target encoding quality based on the rate-quality table 305 and the available bandwidth estimated by the bandwidth estimator 308. Assuming a rate quality table 305 of n different encoding data volumes (or, equivalently for a certain frame rate, encoding rate) r_i, $i = 1, \ldots, n$, where r_i is the audio frame size. The quality of frame j at encoding rate r_i is denoted as $q_{i,j}$. Let $\bar{r}_j(q)$ be the interpolated rate-quality function of frame j generated from the points $(r_i, q_{i,j})$ as explained with reference to figure 6. The goal of the rate controller 307 is to find a target encoding quality q_T for the streaming to follow in at least a period of time (e.g. to use for a certain number of frames), for example until the network situation is changed, e.g. the bandwidth constraint given by the communication network 312 for the streaming changes. To this end, according to one embodiment, a sliding look-ahead window is used and a constant quality streaming is kept within this look-ahead window under the available bandwidth constraint. In the following, it is assumed that the available streaming bit budget for a look-ahead window $[j_0, j_0 + L)$ is R_N, where j_0 is

the index of the current frame and L is the length of the look-ahead window. In other words, R_N bits are available for transmitting the L frames of the sliding window (e.g. according to the bandwidth constraint imposed by the current capacity of the communication network 312).

The aggregated R-D (rate distortion) function is defined as

$$R(q) = \sum_{j=j_0}^{j_0+L-1} \bar{r}_j(q). \quad (1)$$

The aggregated R-D function can be seen as a multi-frame relationship between the encoding quality and encoding data rate (or encoding data volume) for a plurality of frames (namely the L frames of the sliding window) determined based on a combination of the rate-quality functions for the frames of the sliding window (specifically, in this example, a sum of the rate-quality functions for the frames of the sliding window).

According to one embodiment, the target quality is determined by the rate controller 307 according to the following equation:

$$R(q_T) = R_N. \quad (2)$$

Since $\bar{r}_j(q)$ are piece-wise linear functions as a result of the linear interpolation, $R(q)$ is a piece-wise linear function as well. As a result, equation (2) is a linear equation and its solution is straightforwardly given by:

$$q_T = \frac{R_N - R_L}{R_H - R_L} q_L + \frac{R_H - R_N}{R_H - R_L} q_H \qquad (3)$$

where R_L and R_H are, respectively, lower and upper ends of the linear segment of $R(q)$ in which R_N is located, and q_L and q_H the corresponding qualities. Once the target quality is obtained the size of each streamed audio frame (i.e. the encoding data volume for each audio frame of the sliding window) is selected from the interpolated rate-quality function as $\bar{r}_j(q_T)$. The frame truncator 306 truncates the data for the audio frames of the sliding window included in the scalable audio file 304 according to this encoding data size.

The calculation according to equation (3) is illustrated in Fig 7.

Fig. 7 shows an encoding data volume-encoding quality diagram 700 according to an embodiment.

The quality increases along a first axis 701 and is given as a value for parameter q. This may for example be a measure of the mask-to-noise ratio or the value of a quantization parameter (e.g. an accuracy of the quantization which is done when truncating the encoding data or encoding bit-stream of a frame). The encoding data volume increases along a second axis 602 and is for example given in bits.

In this example, it is assumed that the sliding window has only two audio frames (i.e. L=2). As shown, a first (interpolated) rate-quality function 703 for a first frame ($j=j_0$) and a second (interpolated) rate-quality function 704

for a second frame ($j=j_0+1$) are piece-wise linear functions in-between adjacent points (adjacent in terms of encoding quality) included in the quality-rate table. The aggregated quality-rate function $R(q)$ 705 (given by equation (1)) is also piece-wise linear and the target quality q_T is thus obtained by the intersection of $R(q)$ and the total available transmission bits R_N. Once the target quality is determined the encoding data volume (or encoding data rate) for each audio frame is given by the quality-rate functions 703, 704, i.e., $r_0(q_T)$ and $r_1(q_T)$ which are indicated on the second axis 702 in figure 7.

According to one embodiment, the rate controller 307 performs the target quality selection periodically during the streaming process in order to cater for the potential bandwidth fluctuation of the communication channel offered by the communication network 312 for the streaming. This is illustrated in figure 8 with an example.

Fig. 8 shows a data rate-time diagram 800.

Time increases along a first axis 801 and rate increases along a second axis 802.

The required encoding data volume (in other words the bit consumption) for streaming at a first quality q_1 at a certain time is indicated by a first graph 803 and the required encoding data volume (in other words the bit consumption) for streaming at a second quality q_1 at a certain time is indicated by a second graph 804.

In this example, at time t_1 the target quality is selected as q_1 such that the total bits consumption for the streaming of the frames in the sliding window starting at t_1 (indicated by dashed lines 805) is under the constraint of a current measured available bandwidth $R(t_1)$. The target quality is updated at time t_2 again. Since it is assumed that the available bandwidth is increased to $R(t_2)$ at time t_2 the target quality is adjusted to q_2 accordingly such that the total bits consumption for the streaming of the frames in the sliding window starting at t_2 (indicated by solid lines 806) is under the constraint $R(t_2)$.

The effectiveness of the embodiment described above may be verified by simulation.

For example, for a simulation, MPEG-4 SLS (with an AAC core running at 32kbps/channel) is used as the FGS audio codec and the rate-quality table 305 is generated at a step size of 32kbps from the AAC core rate up to 256kbps/channel. The qualities of the audio frames are measured in minimum MNR. The available bandwidth is set to 96kbps. For example, the quality of streamed audio of three different cases are simulated: CBR streaming at 96kbps, streaming according to the embodiment as described above with sliding window length 20, and streaming according to an embodiment as described above with a sliding window length of 200. In the simulation, the target quality is updated for every audio frame in the streaming according to the embodiment as described above. From the result it can be seen that the embodiment as described above leads to much smoother streamed audio quality, and the qualities of critical frames are dramatically improved. It can also be seen from simulation that a longer sliding window leads to smoother streamed audio

quality. However, in practical application, care should be taken to avoid using a sliding window that is too long as smoothing streamed audio quality within an over-lengthy sliding window may not only increase the complexity of the target quality calculation, but also introduce bit-rate fluctuation over a large time-scale which may plague the buffer control of the streaming system.

The bandwidth estimator 308 may be seen to play an important role in the embodiment for a streaming system as described above. The accuracy of the bandwidth estimator decides, to a large degree, the degree of match between data rate of the streamed audio and available bandwidth of the communication network 312. Any mismatch between these two may either result in under-utilization of communication network resources which is inefficient, or in over-utilization which increases the chance of packet delivery failure and eventually deteriorate the streaming quality.

Other than this accuracy requirement, it is also desirable that the output of the bandwidth estimator 308 should be smooth enough to avoid quality fluctuation in the streamed audio, and meanwhile respond fast enough when the communication network conditions change so that the streaming server (i.e. the transmitter 301) always utilizes the communication network resources safely and efficiently. The selection of the bandwidth estimator 308 also may also depend on the actual communication network used for the streaming service whereby elements to consider include the rate/congestion control protocols employed in the streaming server, network gateway designs, and network QoS (Quality of Service) parameters, etc.

According to one embodiment, the streaming service is provided using TCP/IP (Transport Control Protocol/IP Protocol) for communicating via the communication network 312 and there is no network parameter feedback from intermediate nodes of the communication network 312 so that the only information available for bandwidth estimation is from both ends of the communication network 312, i.e. the transmitter 301 and the receiver 302. This may be typical setup for a general purpose communication network such as the Internet. In this situation, the available bandwidth for streaming follows the TCP throughput function given by

$$T = \frac{s}{R\sqrt{\frac{2p}{3}} + t_{RTO}\left(3\sqrt{\frac{3p}{8}}\right)p\left(1 + 32p^2\right)}$$, where s is the packet size, R is the round-trip time, p is the steady-state loss event rate, and t_{RTO} is the TCP retransmit timeout value.

This can for example be used by the bandwidth estimator 308 to estimate the available streaming bandwidth. However, it should be noted that this choice of the type of bandwidth estimator is only an example.

The adaptive audio streaming in accordance with the various embodiments maintains constant audio quality as much as possible during a streaming session to minimize the audio quality variance. It reserves available streaming bits during non-critical audio frames and uses them in streaming of critical audio frames, resulting in improved quality of the critical audio frames. Furthermore, it adapts the rate/quality of the streamed audio based on the available network bandwidth to avoid under-utilizing or over-utilizing the network resource.

In accordance with the various embodiments, the quality adaptation is done based on information from a rate-quality table generated from audio encoder, and real-time network condition during the streaming session.

The quality adaptation problem according to various embodiments can be seen to be based on simple linear interpolation that can be implemented with very low computational costs.

The adaptive streaming system according to an embodiment improves the audio streaming quality by reducing the quality variation during streaming, and boosting the quality of critical audio frames. This further leads to smoother audio playback during streaming since the demanded bandwidth is adapted to the available bandwidth in real-time during streaming.

The adaptive streaming system according to an embodiment further enables the service provider to use only one copy of FGS audio file to cater for users with different service preferences and network conditions. This reduces both implementation and running cost compared with conventional methods based on multiple copies of different quality/rate for the same contents.

The quality adaptation according to various embodiments is therefore suitable and applicable for multimedia streaming service over Internet (such as Internet audio) and over wired or wireless (including Mobile) networks.

According to one embodiment, the buffer level of the receiver 302 is considered. This may be done to avoid receiver buffer

level staggering to a randomly low level and underflow during bursts of critical frames that have higher-than-average frame sizes. Embodiments taking into account the buffer level of the receiver 302 are described in the following.

In an adaptive streaming system, since the streamed audio bit-streams are of variable bit-rate in nature and hence their bit-rate may not necessarily match the available network bandwidth at all time, FIFO (first-in-first-out) buffers may be utilized in both the transmitter (i.e. the streaming server) 301 and the receiver (including the streaming client) 302 to absorb the mismatch between the audio bit-rate and the actual communication network throughput in order to ensure smooth playback. Since such buffers have only limited length, a buffer control is used according to one embodiment to maintain appropriate buffer levels for these buffers to avoid overflow (i.e. the case that data is supplied to a full buffer) which may cause data loss or buffer underflow (i.e. the case that an empty buffer is to provide data) which may cause discontinuity in audio playback. In case that only the available streaming bandwidth is considered as a constraint in determining the streaming bit-rate (i.e. the encoding data volume of the frames) buffer constraints may be violated during a streaming session. To avoid this, a buffer control is introduced according to an embodiment. This is illustrated in figure 9.

Fig. 9 shows a communication arrangement 900 according to an embodiment.

The communication arrangement 900 includes, similarly to the communication arrangement 300 described above with reference to figure 3, a transmitter 901 and a receiver 902 connected

via a communication network 912. The transmitter 901 includes a scalable audio encoder 903 providing a scalable audio file 904 and a rate-quality table 905. The transmitter 901 further includes a frame truncator 906 receiving the scalable audio file 904 as input and a rate controller 907 receiving the rate-quality table 905 as input. The transmitter 901 further includes a network bandwidth estimator 908 and a transmitting module 909. The receiver 902 includes a receiving module 910 and a streaming client 911.

In addition, the transmitter 901 includes a buffer controller 913 connected to the output of the network estimator 908, and both the output and an input of the rate controller 907.

The rate controller 913 selects the target quality of the streamed audio based on information from both the rate-quality model 905 and the available network bandwidth estimated by the bandwidth estimator 908. Meanwhile, the selection to meets the conditions as set by the rate control. Once the target quality is selected, the data of the scalable audio file 904 is truncated accordingly and the resulting data are sent via the communication network 912 for streaming to the streaming client 914.

According to one embodiment, a method for providing an encoded digital signal is carried out as illustrated in figure 10.

Fig. 10 shows a flow diagram 1000 according to an embodiment.

The flow diagram 1000 illustrates a method for providing an encoded digital signal.

In 1001, a data transmission capacity available for transmitting the encoded digital signal from a transmitter to a receiver is determined.

In 1002, a transmission buffer filling level of the transmitter is determined.

In 1003, a decreased transmission capacity is calculated by decreasing the transmission capacity based on the transmission buffer filling level.

In 1004, a data volume for the encoded digital signal is determined based on the decreased transmission capacity.

In 1005, the encoded digital signal is provided at an encoding quality such that the encoded digital signal has the determined data volume.

According to one embodiment, in other words, the transmitter buffer level is taken into account when determining the encoding data volume to be used for a digital signal (e.g. for a plurality of data frames). According to one embodiment, the encoding quality at which the encoded digital signal is provided is determined with the method described above with reference to figure 1. In other words, according to one embodiment, the encoding quality is determined based on the multi-frame relationship determined as described above with reference to figure 1. For example, the encoding quality is determined as the encoding quality corresponding to the determined data volume (as encoding data volume) in accordance with the multi-frame relationship. In other words, the method described with reference to figure 1 and the

method described with reference to figure 10 may be combined. The same holds for corresponding devices.

According to one embodiment, decreasing the transmission capacity includes decreasing the transmission capacity by the transmission buffer filling level scaled with a pre-determined scaling factor.

According to one embodiment, determining the available data transmission capacity for transmitting the encoded digital signal includes estimating the available bandwidth of a communication channel between the transmitter and the receiver.

The method illustrated in figure 10 is for example carried out by a device as illustrated in figure 11.

Fig. 11 shows a device for providing an encoded digital signal 1100.

The device 1100 includes a capacity determining circuit 1101 configured to determine a data transmission capacity available for transmitting the encoded digital signal from a transmitter to a receiver and a filling level determining circuit 1102 configured to determine a transmission buffer filling level of the transmitter.

The device 1100 further includes a calculating circuit 1103 configured to calculate a decreased transmission capacity by decreasing the transmission capacity based on the transmission buffer filling level and a data volume determining circuit 1104 configured to determine a data

volume for the encoded digital signal based on the decreased transmission capacity.

Additionally, the device 1100 includes an output circuit 1105 configured to provide the encoded digital signal at an encoding quality such that the encoded digital signal has the determined data volume.

It should be noted that embodiments described in the context of one of the methods for providing an encoded digital signal are analogously valid for the other method for providing an encoded digital signal and for the devices for providing an encoded digital signal and vice versa.

According to one embodiment, FIFO buffers are used in both the transmitter (streaming server) 901 and the receiver (receiver buffer) 902 to absorb discrepancies between the rate of the VBR audio bit-stream and the actual network throughput. This is illustrated in figure 12.

Fig. 12 shows a communication arrangement 1200 according to an embodiment.

The communication arrangement 1200 includes a transmitter 1201 for example corresponding to transmitter 901 and a receiver 1202 for example corresponding to the receiver 1202 connected via a communication network 1207. The transmitter includes a frame truncator 1203 for example corresponding to frame truncator 906 and the receiver includes an audio decoder 1204 (which is for example part of the streaming client 914).

The transmitter 1201 includes a transmit buffer 1205 and the receiver includes a receiver buffer 1206. The transmitter 1201 sends data to the communication network 1207 via the transmitter buffer 1205 and the transmitter 1202 receives data from the communication network via the receiver buffer 1206.

The transmitter buffer 1205 and the receiver buffer 1206 are FIFO (first in - first out) buffers.

Figure 12 can be seen to illustrate a network model of the adaptive streaming system as illustrated in figures 3 and 9. As can be seen from figure 12, the task of buffer control is to properly control the data rates that audio data enter and leave the buffers 1205, 1206 so that the buffers 1205, 1206 do not get underflowed (i.e. data is to leave an empty buffer) or overflowed (i.e. data is to enter a full buffer).

In the case of file-based streaming (unconstrained streaming) audio data to be streamed is pre-encoded and stored on a disk and hence there is no constraint on the rate that audio data enter the transmitter buffer 1205. In this case the buffer control in the transmitter buffer 1205 is not an issue and there is only a need to consider the receiver buffer 1206.

For a situation of live streaming (constrained streaming) the audio data is generated in real-time during streaming and as a result they have to enter the transmitted buffer 1205 in a constrained rate. In this case the buffer control needs to be considered at both buffers 1205, 1206. However, receiver side buffer 1206 underflow is only considered because receiver/transmitter buffer overflow can be easily avoided if sufficient memory is available, and transmission side buffer

underflow can be solved by either reducing the transmission rate or using stiff bits.

Regarding the buffer level calculation of the receiver buffer 1206, the audio data being streamed is assumed to have a constant frame rate F in frame/sec, and each frame has a frame size of r_i bits, $i = 0, 1, \ldots$. Meanwhile, it is assumed that at each frame interval i the communication network 1207 transmits in total C_i bits of data from the transmitter 1201 to the receiver 1202. To simplify the problem, it is assumed that there is no transmission delay and transmission error so that the bits being moved out from the transmitter buffer 1205 reach the receiver buffer 1206 immediately. Furthermore, an initial receiver side delay of Δ frames is assumed, i.e., the receiver 1202 waits for Δ frames before removing the first frame from the receiver buffer 1206 after it is received, and there is no other delay present in the streaming system. Under these assumptions, the transmitter buffer level $B^T(i)$ and receiver buffer level $B^R(i)$ at frame interval i are given respectively as:

$$B^T(i) = \sum_{j=1}^{i} r_j - \sum_{j=1}^{i} C_j,$$

$$B^R(i) = \begin{cases} \sum_{j=1}^{i} C_j - \sum_{j=1}^{i-\Delta} r_j, & i \geq \Delta; \\ \sum_{j=1}^{i} C_j, & \text{otherwise.} \end{cases} \quad (4)$$

That is, the transmitter buffer level is simply the total number of bits being generated from the encoder minus the total bits being transmitted, and the receiver buffer

contains all the received bits minus those of the decoded frames. It should be noted that due to the initial receiver side delay at time i only $(i - \Delta)$ frames have been decoded. Here it is assumed that there is no transmitter buffer underflow to preserve the linearity of transmitter side buffer level calculation.

Combining the transmitter buffer level at time i and the receiver buffer level at time $i + \Delta$ gives

$$B^R(i+\Delta) = \sum_{j=1}^{i+\Delta} c_j - \sum_{j=1}^{i} r_j = \sum_{j=i+1}^{i+\Delta} c_j - \left(\sum_{j=1}^{i} r_j - \sum_{j=1}^{i} c_j \right) = \sum_{j=i+1}^{i+\Delta} c_j - B^T(i). \quad (5)$$

To prevent the receiver buffer 1206 from underflowing the right-hand side of equation (5) should be kept always greater than zero, i.e., the transmitter buffer size should not exceed $\sum_{j=i+1}^{i+\Delta} c_j$. It should be noted that given that there is sufficient memory available at the transmitter 1201 and receiver, this constraint is actually imposed by the initial delay Δ and the network condition c_j rather than by memory considerations. Therefore the amount of $\sum_{j=i+1}^{i+\Delta} c_j$ may also be referred to as effective buffer size to reflect this fact.

Since the prevention of receiver buffer underflow is equivalent to prevention of encoder buffer level from exceeding the effective buffer size, according to one embodiment, the transmitter buffer level is incorporated in the rate control equation in an appropriate manner to prevent

it from going too high. This can be implemented by modifying equation (1) as follows so that the overall bit-budget for each sliding window is further constrained by the transmission buffer level:

$$\sum_{j=i}^{i+L} r_j(q_T) = LFR_i - \alpha \cdot B^T(i) \qquad (6)$$

where $0 < \alpha$ is a predefined constant and R_i is the available bit budget for the transmission of the ith frame (assumed to be constant for all frames of the sliding window).

In other words, the transmission capacity provided by the communication network 1207 as for example estimated by bandwidth estimator 908, is decreased based on the transmission buffer filling level for purposes of encoding quality determination.

It can be seen that with equation (6)

$$\begin{aligned} B^T(i+L) &= B^T(i) + \sum_{j=i}^{i+L} r_j(q_T) - \sum_{j=i}^{i+L} C_j \\ &= B^T(i) + LFR_i - \alpha B^T(i) - \sum_{j=i}^{i+L} C_j, \\ &\approx (1-\alpha) B^T(i), \end{aligned} \qquad (7)$$

if target quality q_T is used for the whole sliding window and the bandwidth estimation made at frame index i is sufficiently close to actual amount of date being transferred within the sliding window. As a result, the transmitter buffer level will be pulled towards zero at the end of each

sliding window and the larger the value of the constant α, the transmitter buffer level will be pulled towards zero more aggressively. Therefore, the value of α plays an important role in determining the aggressiveness of the buffer control algorithm. As a rule of thumb, care should be taken to avoid using an overlarge α as it will discourage buffer usage and may lead to suboptimal quality at critical audio frames; on the other hand, α should be large enough so that the transmitter buffer level never exceeds the effective buffer level to avoid decoder buffer underflow. The minimum value of α can be determined from the network characters as well as other streaming parameters such as the amount of the initial buffer size and the length of the sliding window.

Mathematically, it can be shown that the transmitter buffer level is bounded by:

$$B^T(i) < \frac{LFR_{max}}{\alpha}, \tag{8}$$

where $R_{max} = \max(R_i)$ is the maximum possible available bandwidth for streaming. Therefore, receiver buffer underflow can be completely avoided if it can be guaranteed that the effective buffer size is larger than this upper bound for the transmitter buffer, i.e., that

$$\sum_{j=i+1}^{i+\Delta} C_j \geq \frac{LFR_{max}}{\alpha} > B^T(i). \tag{9}$$

Unfortunately, this condition may not be very helpful in practice where the actual amount of data C_j being transmitted from frame index i to $i + \Delta$ is, in general,

unknown a priori, in particular for a channel with variable bit rate. However, if it is assumed that the variable bit rate channel is characterized with a minimum bandwidth R_{min}, then $\sum_{j=i}^{i+\Delta} C_j \geq \Delta FR_{min}$, and equation (7) is satisfied as long as

$$\Delta FR_{min} \geq \frac{1}{\alpha} LFR_{max}, \tag{10}$$

or

$$\alpha \geq \frac{L}{\Delta} \cdot \frac{R_{max}}{R_{min}}. \tag{11}$$

Therefore, inequality (9) can be used as design guideline for selecting α once other design parameters such as the initial delay Δ and the sliding window length L are fixed, and the range of the bandwidth variation of the streaming network is known. In a simpler case if the channel has constant bit rate (CBR), resulting in $R_{min} = R_{max}$, equation (11) simplifies to

$$\alpha \geq \frac{L}{\Delta}. \tag{12}$$

It should be noted that the above bound for α (according to equation (11)) is a bit pessimistic and in practical application it may be possible to use a smaller value of α without leading to receiver buffer underflow.

The effectiveness of the buffer control as described can be verified by simulation. The buffer control algorithm may be integrated with the adaptive streaming system according to

the embodiment described above where MPEG-4 SLS (with an AAC core at 32 kbps/channel) is used as the FGS audio codec and the rate-quality table is generated at a step size of 32 kbps from the AAC core rate up to 256 kbps/channel. The qualities of the audio frames are measured in minimum MNR. CBR channel is assumed in this simulation where the available bandwidth is set at 96 kbps. The sliding window size for the adaptive streaming system is set to 10 frames, i.e., $L = 10$ and the target quality update is performed for each frame during streaming. The size of the receiver buffer is set to 20 kilobits and the receiver 902 starts to decode the first audio frame as long as the receiver buffer 1206 is full at beginning. Given the transmission data rate of 96 kbps, this is approximated to 20 kilobits/96 kbps $\approx$ 208.3 ms of delay or roughly 10 SLS frames, i.e., $\Delta \approx 10$.

As can be seen from a comparison between $\alpha = 0$ (no buffer control) and $\alpha = 1$ for a testing sequence buffer underflow may start at a certain frame and exaggerates with the progress of the streaming session when there is no buffer control. However, the buffer underflow problem may be solved with the introduction of the buffer control. In addition, from the quality data it can be seen that the buffer control only introduces negligible impact to the streaming quality.

According to an embodiment, as described above, a method and system for streaming scalable audio, in particularly, adaptively streaming fine grain scalable audio in a network with varying bandwidth is provided wherein quality of each audio frame in the audio stream being streamed is determined based on a function of two or more Rate-Quality data measured for each audio frame from a given window in which said frame

being streamed resides. A method of buffer control is also introduced to manage the receiver underflow problem.

While the invention has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

The following documents are cited in the above description

[1] J. Lin and A. Ortega, "Bit-rate Control using piecewise approximation rate-distortion characteristics," IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 4, pp. 446-459, Aug. 1998.

[2] L. Zhao, J. W. Kim, and C.-C. Kuo, "MPEG-4 FGS video streaming with constant-quality rate control and differentiated forwarding," in SPIE VCIP, Jan. 2002, pp. 230-241.

[3] M. Dai et al, "Rate-Distortion Analysis and Quality Control in Scalable Internet Streaming", IEEE Transactions on Multimedia, Vol. 8, No. 6, Dec. 2006.

Claims

1. A method for providing an encoded digital signal comprising
determining, for each data frame of a plurality of data frames of a digital signal, a plurality of pairs of an encoding data volume and an encoding quality, wherein each pair of an encoding data volume and an encoding quality specifies the encoding data volume required for achieving the encoding quality;
determining for each data frame at least one or more interpolations between the plurality of determined pairs;
determining a multi-frame relationship between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality based on a combination of the at least one or more interpolations for the plurality of data frames;
determining an encoding quality for the plurality of data frames based on the relationship; and
providing at least one data frame of the plurality of data frames encoded at the determined encoding quality.

2. Method according to claim 1, wherein the relationship specifies for each encoding quality of a plurality of encoding qualities a corresponding encoding data volume required to encode the plurality of data frames at the encoding quality.

3. Method according to claim 2, wherein the encoding quality for the plurality of data frames is determined such that the encoding data volume corresponding to the

determined encoding quality according to the relationship fulfils a predetermined criterion.

4. Method according to claim 3, wherein the criterion is that the encoding data volume is below a pre-determined threshold.

5. Method according to claim 4, wherein the threshold is based on a maximum data rate.

6. Method according to any one of claims 1 to 5, wherein the multi-frame relationship is determined based on a combination of the at least one or more interpolations for at least two different data frames of the plurality of data frames.

7. Method according to any one of claims 1 to 6, wherein the at least one interpolation of a data frame of the plurality of data frames is an interpolation of the plurality of encoding data volume and encoding quality pairs of the data frame.

8. Method according to any one of claims 1 to 7, wherein the at least one interpolation of a data frame of the plurality of data frames is a linear interpolation of the plurality of encoding data volume and encoding quality pairs of the data frame.

9. Method according to any one of claims 1 to 8, wherein the plurality of data frames are a plurality of successive data frames.

10. Method according to claim 9, wherein the at least one data frame of the plurality of data frames provided encoded at the determined encoding quality comprises the first data frame of the plurality of successive data frames encoded at the determined encoding quality.

11. Method according to claim 9 or 10, further comprising determining a further encoding quality to be used for a further plurality of successive data frames comprising the plurality of data frames without the at least one data frame provided encoded at the determined encoding quality.

12. Method according to any one of claims 1 to 11, wherein each interpolation of the at least one or more interpolations between the plurality of determined pairs for a data frame is an interpolated pair of an encoding data volume and an encoding quality specifying the encoding data volume required for achieving the encoding quality for the data frame.

13. Method according to any one of claims 1 to 12, wherein the multi-frame relationship is determined based on a summing of the encoding data volumes required for achieving an encoding quality for different data frames for the same encoding quality.

14. Method according to claim 13, wherein the result of the summing is specified by the relationship for an encoding quality as a corresponding encoding data volume required to encode the plurality of data frames at the encoding quality.

15. Method according to any one of claims 1 to 14, wherein the multi-frame relationship is a piecewise linear correspondence between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality.

16. Method according to any one of claims 1 to 15, wherein the plurality of pairs of an encoding data volume and an encoding quality for each data frame are generated by measuring, for each of a plurality of encoding data volumes, the encoding quality achieved when encoding the data frame using the encoding data volume.

17. Method according to any one of claims 1 to 16, wherein the digital signal is an audio signal.

18. A device for providing an encoded digital signal comprising
a first determining circuit configured to determine, for each data frame of a plurality of data frames of a digital signal, a plurality of pairs of an encoding data volume and an encoding quality, wherein each pair of an encoding data volume and an encoding quality specifies the encoding data volume required for achieving the encoding quality;
an interpolator configured to determine for each data frame at least one or more interpolations between the plurality of determined pairs;
a combiner configured to determine a multi-frame relationship between encoding quality and encoding data volume required to encode the plurality of data frames at the encoding quality based on a combination of the at

least one or more interpolations for the plurality of data frames;

a second determining circuit configured to determine an encoding quality for the plurality of data frames based on the relationship; and

an output circuit providing at least one data frame of the plurality of data frames encoded at the determined encoding quality.

19. A method for providing an encoded digital signal comprising

determining a data transmission capacity available for transmitting the encoded digital signal from a transmitter to a receiver;

determining a transmission buffer filling level of the transmitter;

calculating a decreased transmission capacity by decreasing the transmission capacity based on the transmission buffer filling level;

determining a data volume for the encoded digital signal based on the decreased transmission capacity;

providing the encoded digital signal at an encoding quality such that the encoded digital signal has the determined data volume.

20. The method according to claim 19, wherein decreasing the transmission capacity comprises decreasing the transmission capacity by the transmission buffer filling level scaled with a pre-determined scaling factor.

21. The method according to claim 19 or 20, wherein determining the available data transmission capacity for transmitting the encoded digital signal comprises

estimating the available bandwidth of a communication channel between the transmitter and the receiver.

22. A device for providing an encoded digital signal comprising

a capacity determining circuit configured to determine a data transmission capacity available for transmitting the encoded digital signal from a transmitter to a receiver;

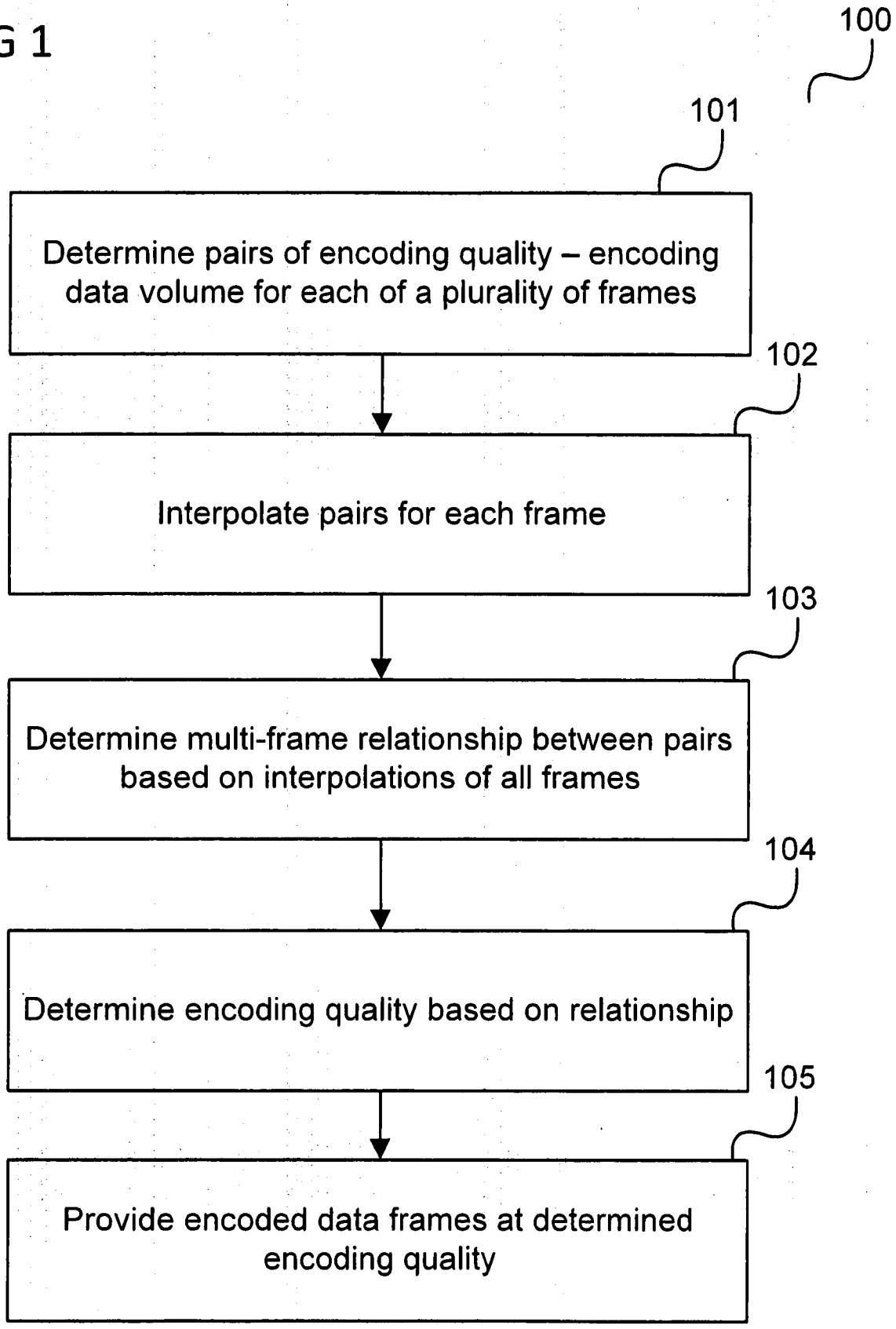
a filling level determining circuit configured to determine a transmission buffer filling level of the transmitter;

a calculating circuit configured to calculate a decreased transmission capacity by decreasing the transmission capacity based on the transmission buffer filling level;

a data volume determining circuit configured to determine a data volume for the encoded digital signal based on the decreased transmission capacity; and

an output circuit configured to provide the encoded digital signal at an encoding quality such that the encoded digital signal has the determined data volume.

FIG 1


100
101
Determine pairs of encoding quality – encoding data volume for each of a plurality of frames
102
Interpolate pairs for each frame
103
Determine multi-frame relationship between pairs based on interpolations of all frames
104
Determine encoding quality based on relationship
105
Provide encoded data frames at determined encoding quality

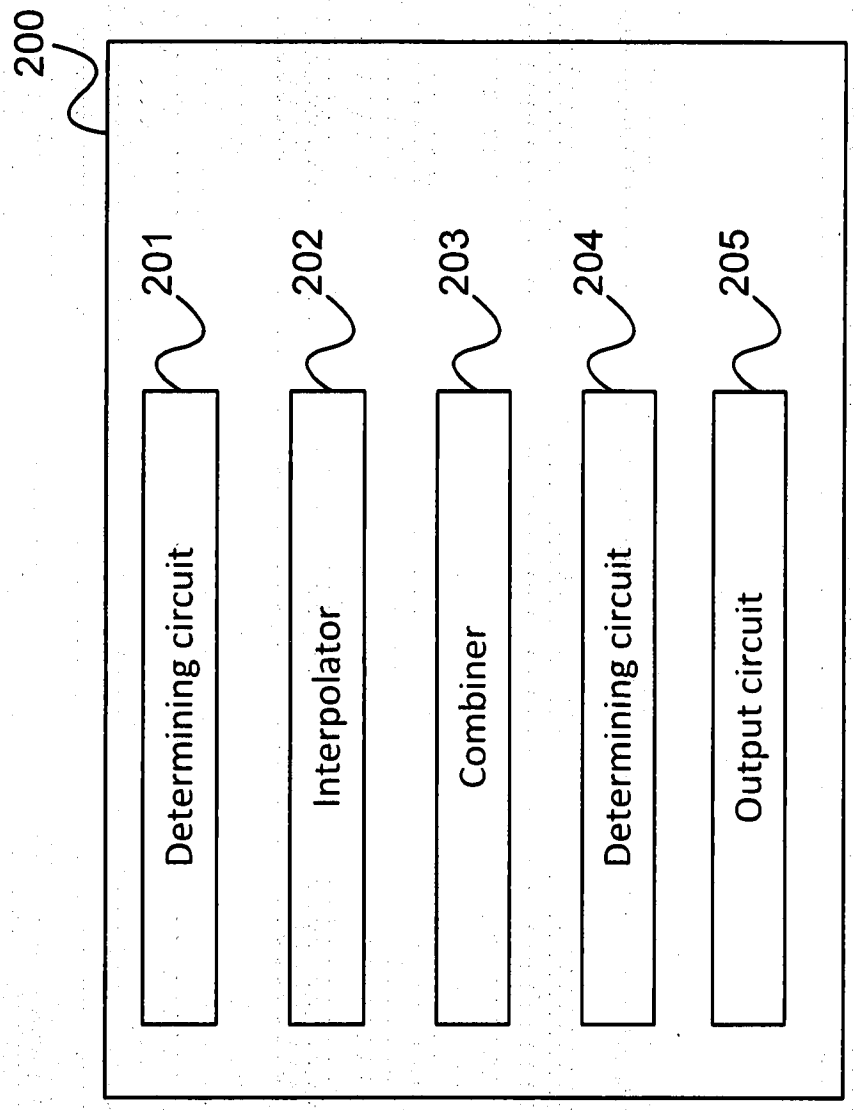

200
201
Determining circuit
202
Interpolator
203
Combiner
204
Determining circuit
205
Output circuit

FIG 2

FIG 3

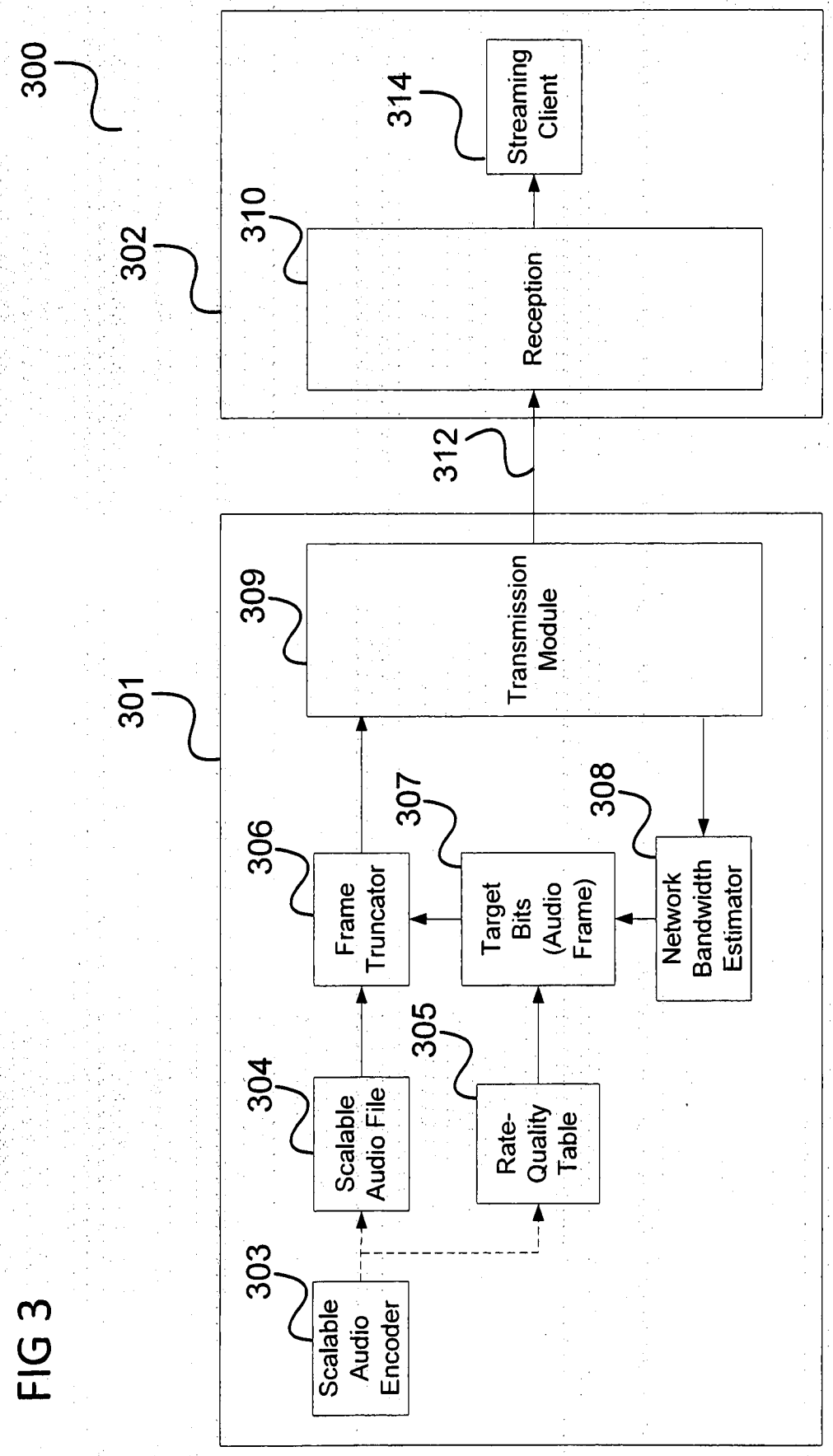

300
302
310
Reception
314
Streaming Client
312
301
309
Transmission Module
306
Frame Truncator
307
Target Bits (Audio Frame)
308
Network Bandwidth Estimator
304
Scalable Audio File
305
Rate-Quality Table
303
Scalable Audio Encoder

FIG 4

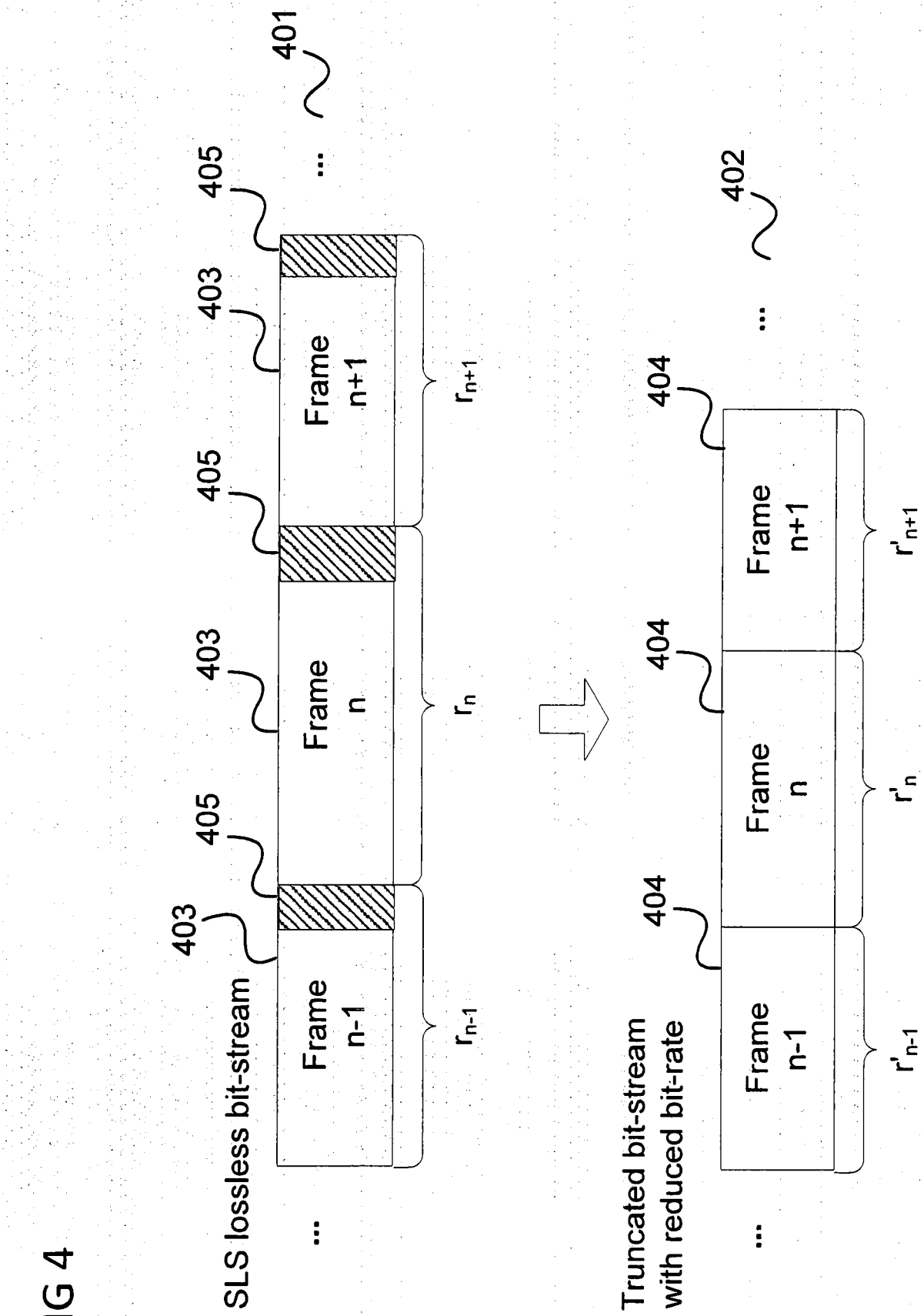
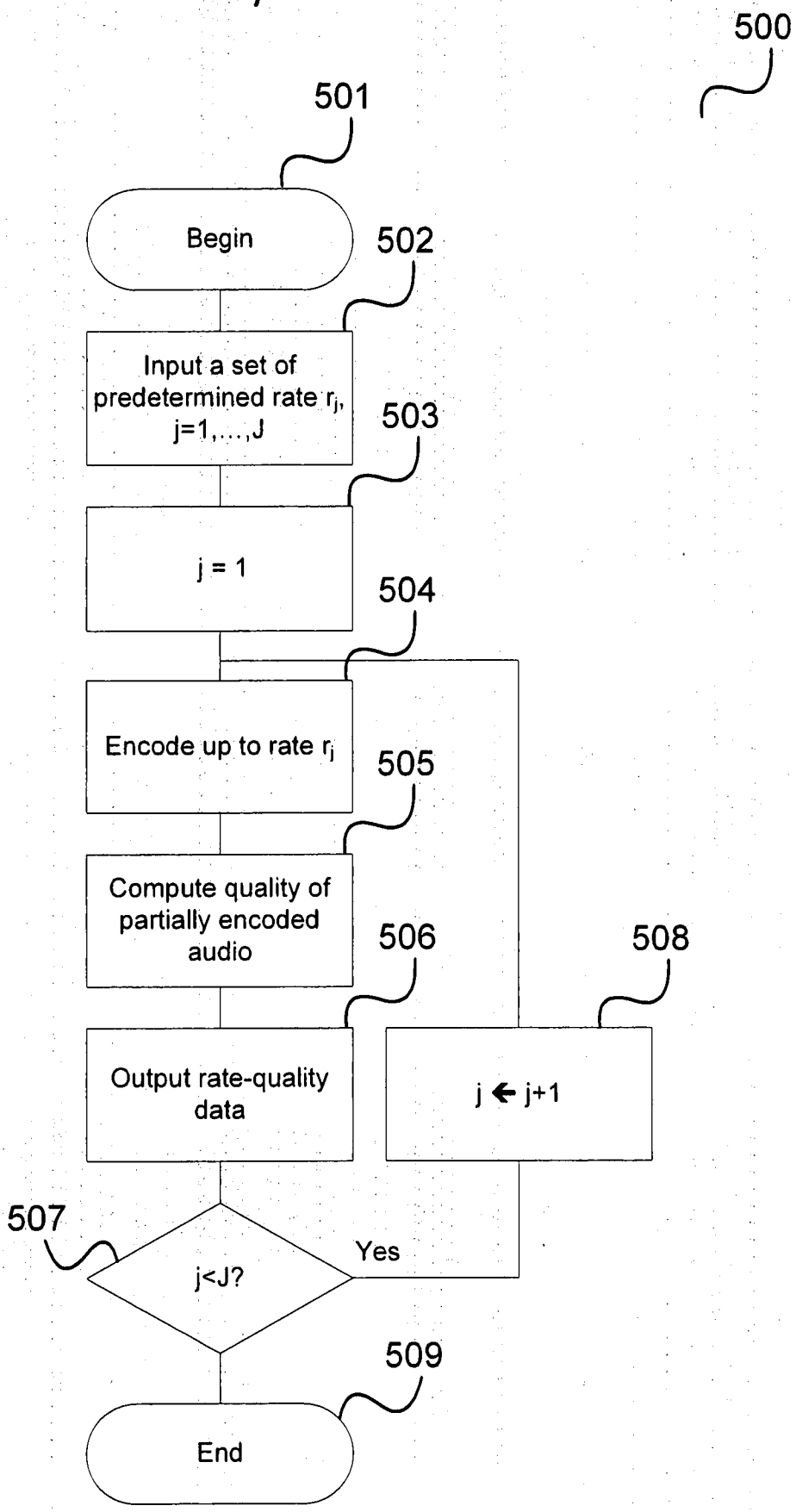

401
402
403
404
405
SLS lossless bit-stream
Truncated bit-stream with reduced bit-rate
Frame n-1
Frame n
Frame n+1
r_{n-1}
r_n
r_{n+1}
r'_{n-1}
r'_n
r'_{n+1}

FIG 5


500
501
Begin
502
Input a set of predetermined rate r_j, j=1,...,J
503
j = 1
504
Encode up to rate r_j
505
Compute quality of partially encoded audio
506
Output rate-quality data
508
j ← j+1
507
j<J?
Yes
509
End

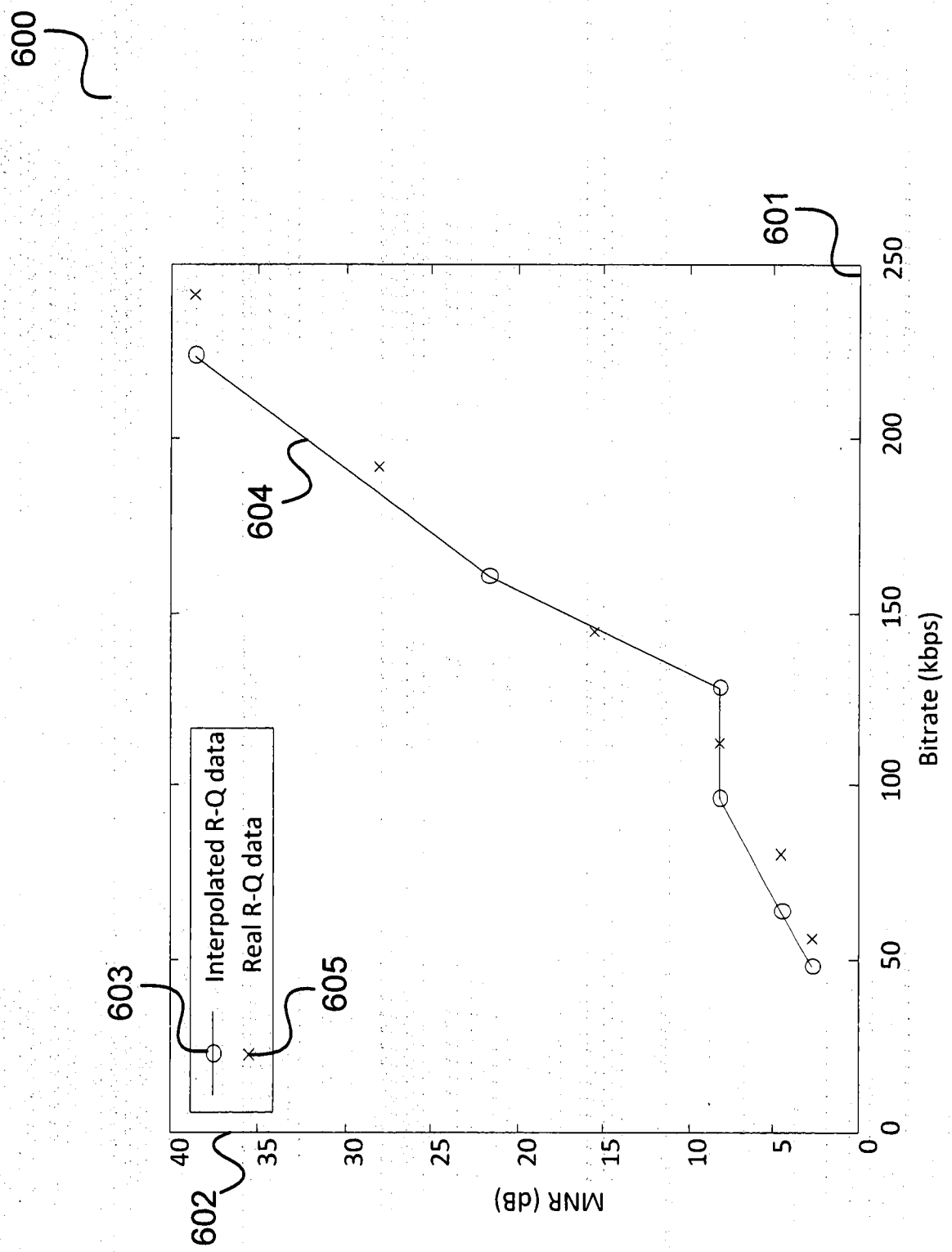

600
601
602
603
604
605
MNR (dB)
Bitrate (kbps)
Interpolated R-Q data
Real R-Q data
0
5
10
15
20
25
30
35
40
0
50
100
150
200
250

FIG 6

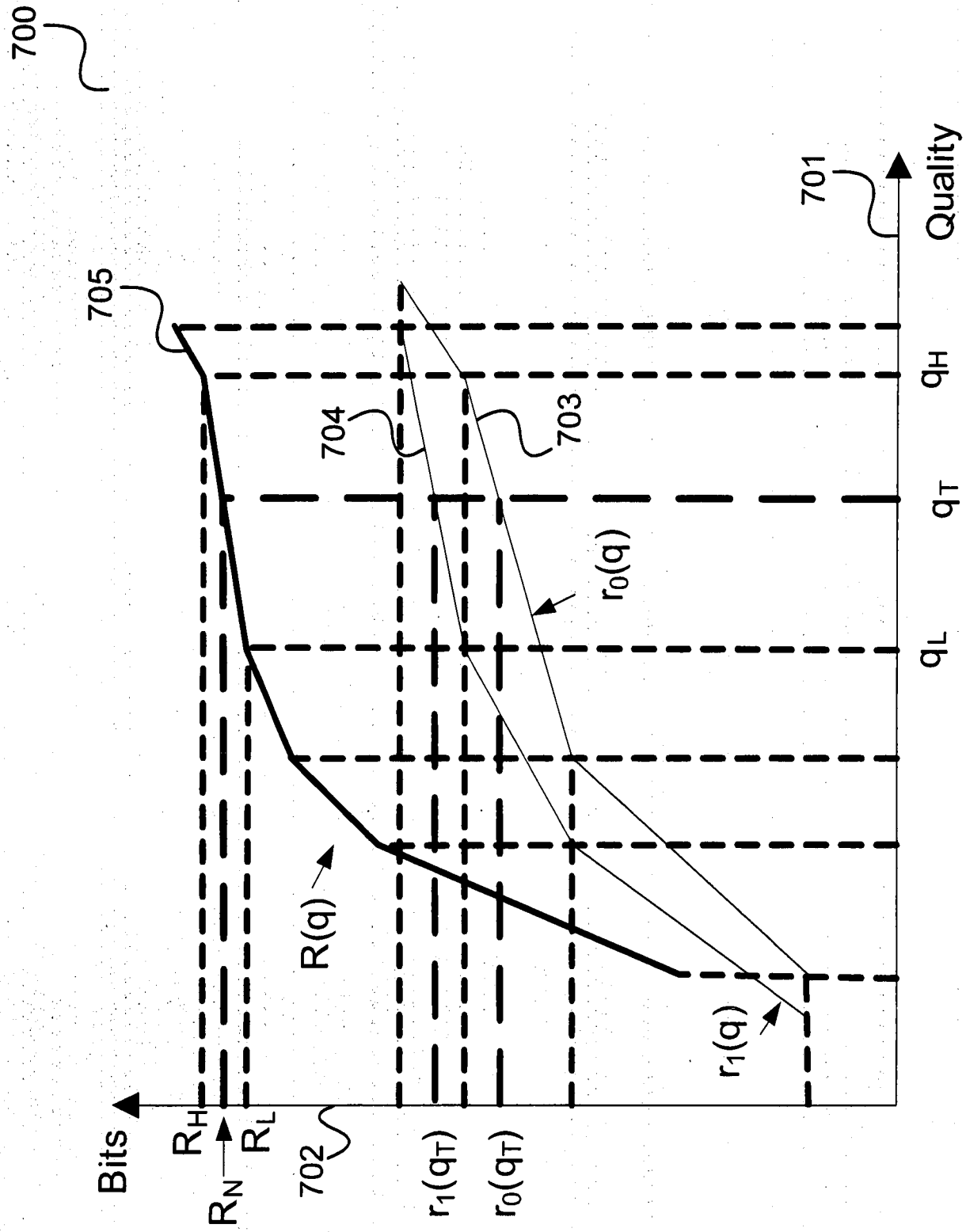

700
701
Quality
705
704
703
q_H
q_T
q_L
$r_0(q)$
R(q)
$r_1(q)$
Bits
R_H
R_N
R_L
702
$r_1(q_T)$
$r_0(q_T)$

FIG 7

FIG 8

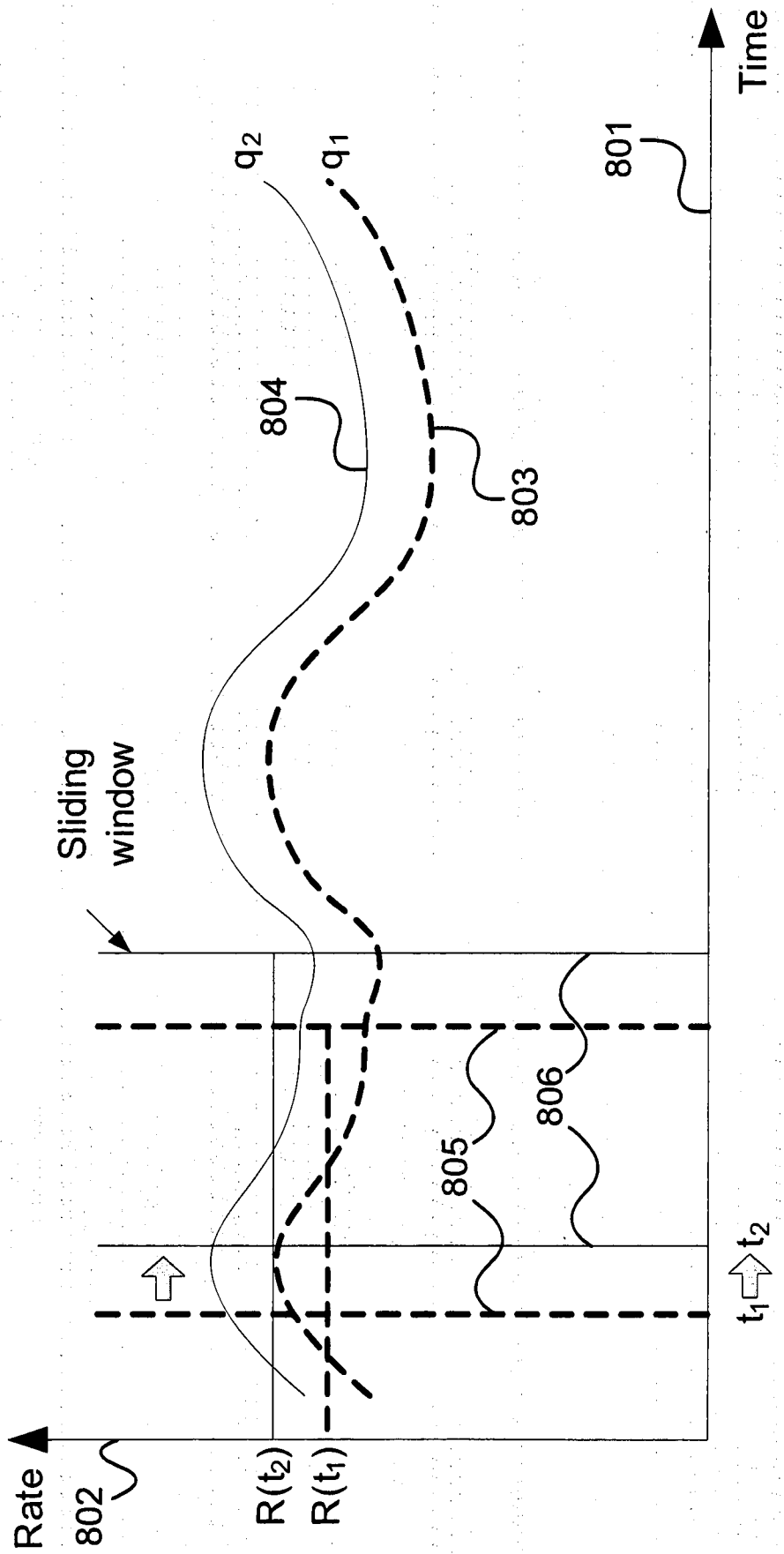

800
Time
801
q_2
q_1
804
803
Sliding window
805
806
t_1 t_2
Rate
802
$R(t_2)$
$R(t_1)$

FIG 9

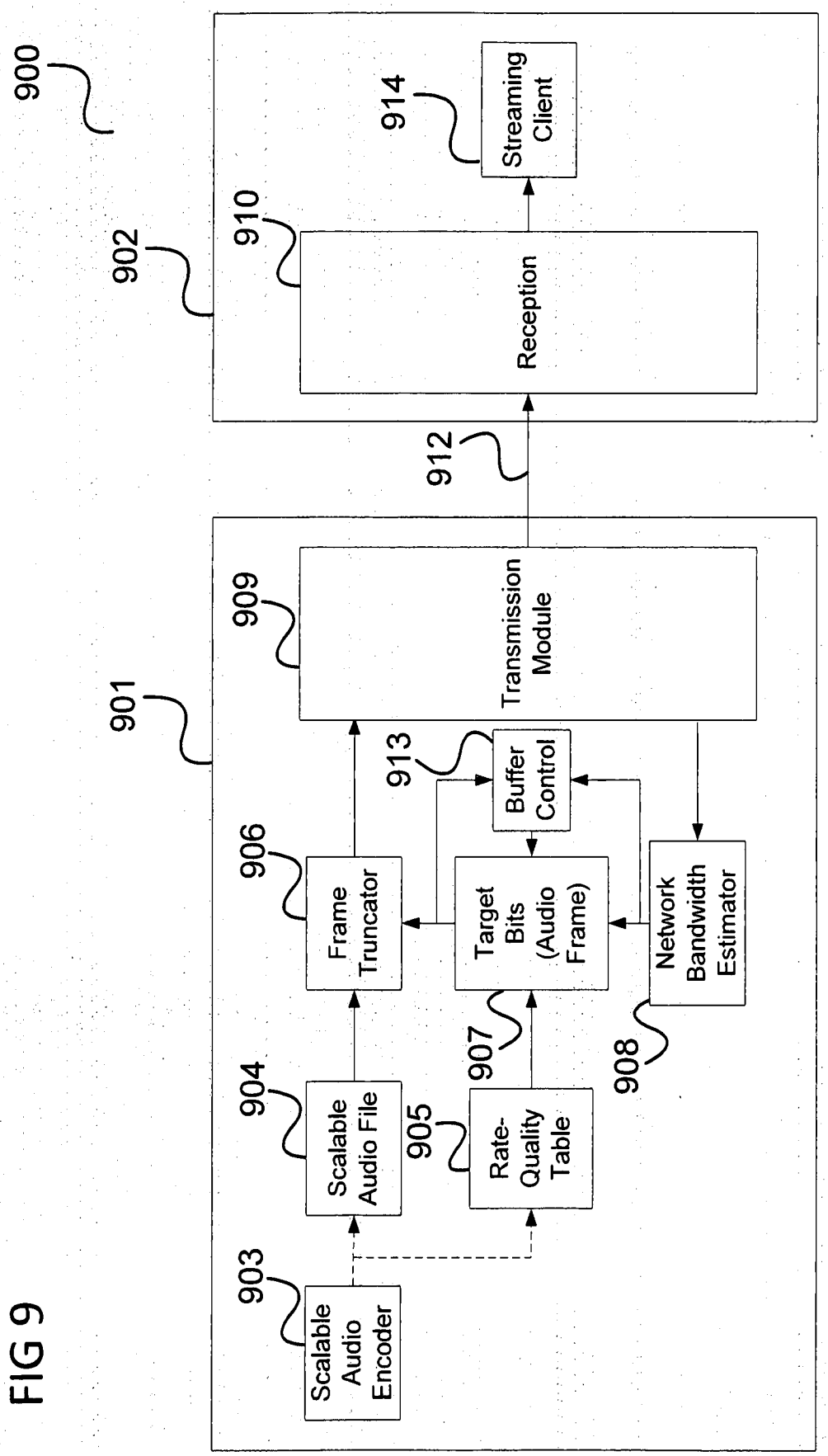
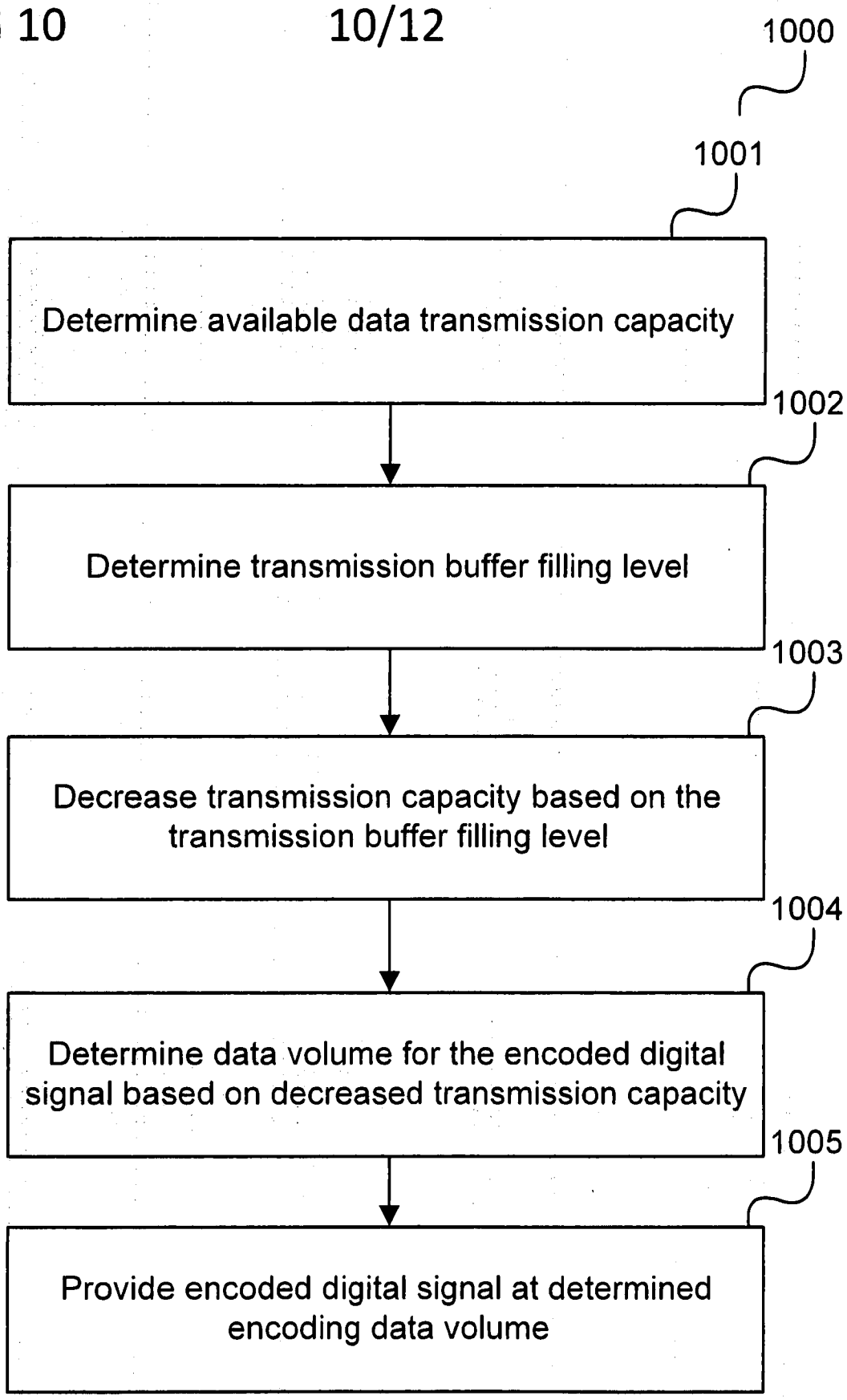

900
902
910
Reception
914
Streaming Client
912
901
909
Transmission Module
913
Buffer Control
906
Frame Truncator
Target Bits (Audio Frame)
Network Bandwidth Estimator
904
Scalable Audio File
905
Rate-Quality Table
907
908
903
Scalable Audio Encoder

FIG 10


1000
1001
Determine available data transmission capacity
1002
Determine transmission buffer filling level
1003
Decrease transmission capacity based on the transmission buffer filling level
1004
Determine data volume for the encoded digital signal based on decreased transmission capacity
1005
Provide encoded digital signal at determined encoding data volume

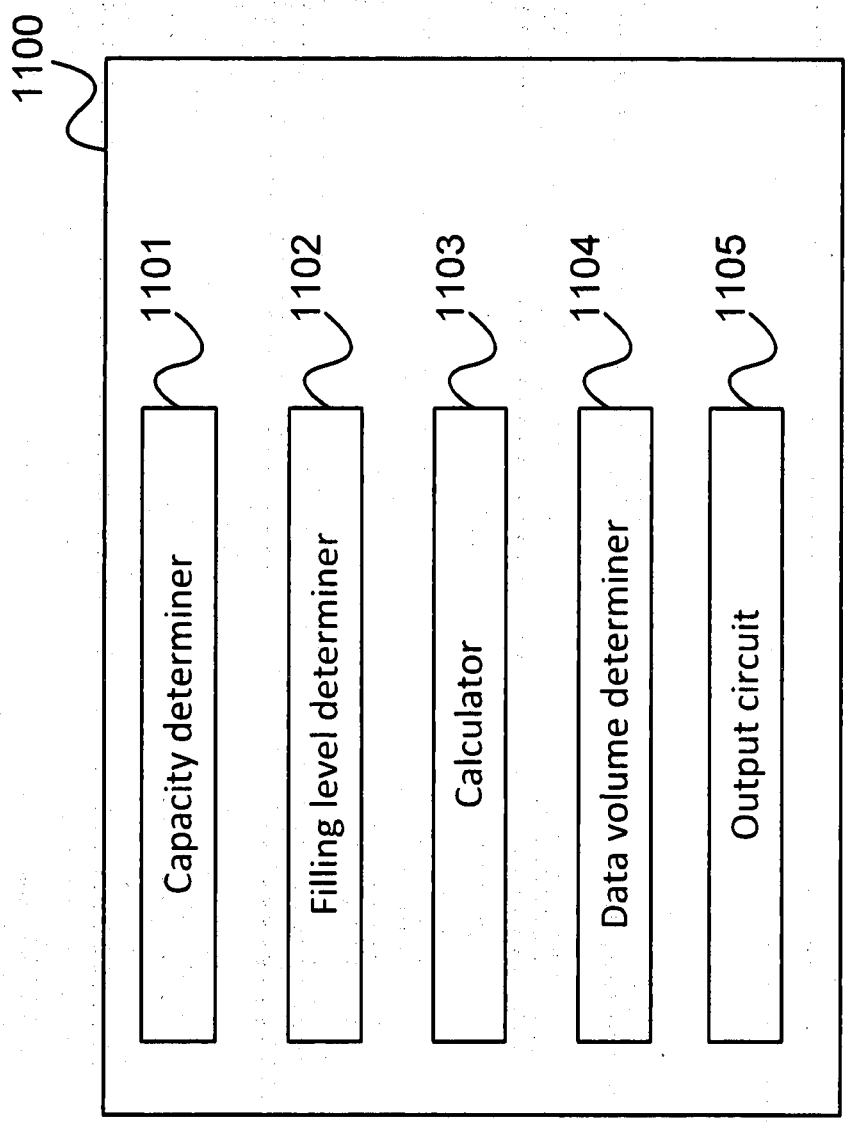
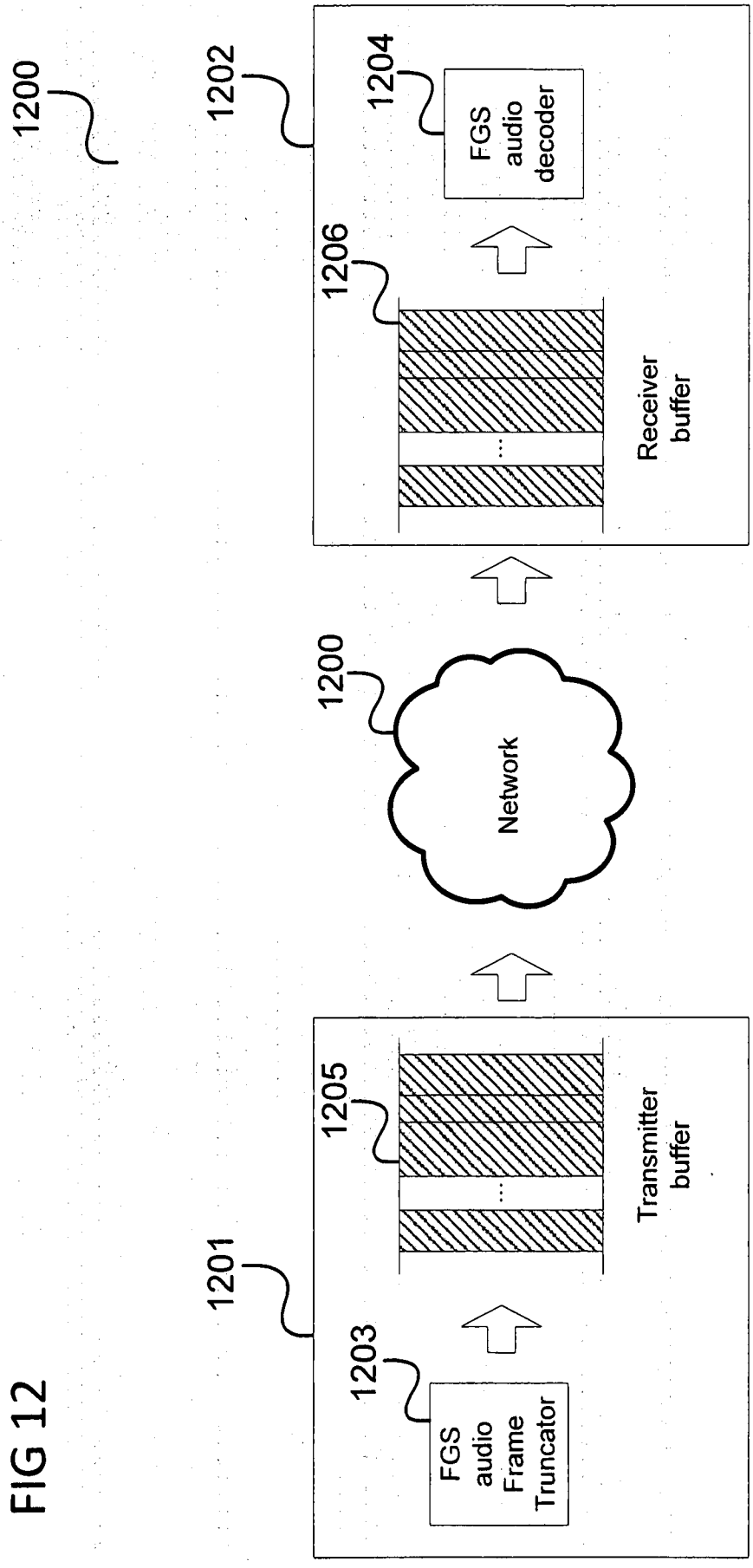

1100
1101
Capacity determiner
1102
Filling level determiner
1103
Calculator
1104
Data volume determiner
1105
Output circuit

FIG 11

FIG 12


1200
1202
1204
FGS audio decoder
1206
Receiver buffer
1200
Network
1201
1205
Transmitter buffer
1203
FGS audio Frame Truncator

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2011/000112

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

H04N 7/26 (2006.01) *G10L 19/00* (2006.01)

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)

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)

TXTE and INTERNET: : MPEG-4, AAC, base layer, enhancement layer, digital signal, encoding, quality, data, frame, audio, network, bandwidth, estimation, rate, control, bit-rate, rate, distortion and other terms and phrases.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Huang, C-M et. al; "A multilayered Audiovisual Streaming System Using the Network Bandwidth Adaptation and the Two Phase Synchronization"; IEEE Transactions on Multimedia, vol. 11, no.5, August 2009. section III and IV, pages 799-801, Table I & II , Figures 2 and 3	1-18
Y	US 2002/0010938 A1 (ZHANG et al.) 24 January 2002 Figure 2, [0047] – [0130]	19-22
Y	US 2004/0049379 A1 (THUMPUDI et al.) 11 March 2004 Abstract	19-22

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

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search
20 May 2011

Date of mailing of the international search report 31 MAY 2011

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaustralia.gov.au
Facsimile No. +61 2 6283 7999

Authorized officer
AMBARISH NATU
AUSTRALIAN PATENT OFFICE
(ISO 9001 Quality Certified Service)
Telephone No : +61 2 6283 2414

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2011/000112

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
See Supplemental Box I

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Supplemental Box I
(To be used when the space in any of Boxes I to IV is not sufficient)

Continuation of Box No: III

This international application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

In assessing whether there is more than one invention claimed, I have given consideration to those features which can be considered to potentially distinguish the claimed combination of features from the prior art.

The international searching authority has found that there are different inventions as follows:

1. Claims 1-18 are directed to a method of encoding a digital signal. It is considered that a method for providing an encoded digital signal comprising determining for each data frame, pairs of encoding data volume and its encoding quality and further, determining a multi-frame relationship between encoding quality and encoding data volume based on an interpolation function comprises as the first distinguishing feature.
2. Claims 19-22 are directed to a method of encoding a digital signal. It is considered that a method for providing an encoded digital signal comprising determining a data transmission capacity available for transmitting the encoded signal from a transmitter to a receiver, determining a transmission buffer filling level, calculating transmission capacity and then determining a data volume based on transmission capacity comprises as the second distinguishing feature.

PCT Rule 13.2, first sentence, states that unity of invention is only fulfilled when there is a technical relationship among the claimed inventions involving one or more of the same or corresponding special technical features. PCT Rule 13.2, second sentence, defines a special technical feature as a feature which makes a contribution over the prior art.

Each of the abovementioned groups of claims has a different distinguishing feature and they do not share any feature which could satisfy the requirement for being a special technical feature. Because there is no common special technical feature it follows that there is no technical relationship between the identified inventions. Therefore the claims do not satisfy the requirement of unity of invention *a priori*.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/SG2011/000112

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report	Patent Family Member					
US 2002010938	US	7260826	US	2005155080	US	7450571
	US	2006156201	US	7559078	US	2006156200
	US	7574726				
US 2004049379	EP	1396842	EP	1396843	EP	1396844
	EP	1400954	EP	1400955	EP	1403854
	EP	1734511	EP	2006840	EP	2023340
	EP	2028648	EP	2261897	EP	2267698
	EP	2270777	EP	2282310	JP	2004264814
	JP	2004258603	JP	2004264810	JP	2004264811
	JP	2004264812	JP	2004264813	JP	2010160517
	JP	2010160518	JP	2010176151	JP	2010217900
	US	2004044527	US	7299190	US	2004044534
	US	7328150	US	2004044521	US	7424434
	US	2005015249	US	7433824	US	7502743
	US	2004044520	US	7536305	US	2008021704
	US	7801735	US	2008262855	US	7822601
	US	2008228476	US	7840403	US	2008221908
	US	7860720	US	2009228290	US	2010318368
	US	2011035225	US	2011054916	US	2011060597

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX