MIDI and Audio via ISDN

Ole Nielsen
Tele Danmark, Jysk Telefoni, Svejvej 30, DK-8300 Trondheim
Tel. +45 89 45 21 79, Fax +45 86 15 67, E-mail: oo@kbh.jt.dk

Abstract

This paper gives an introduction to the existing ISDN system and the future B-ISDN based on ATM. An application of the ISDN network for transmission of MIDI signals and high-quality audio signals using a bit rate reduction system like MPEG-1 is treated.

1 Introduction

The implementation of the common protocol Euro-ISDN (1) in December 1993 has improved the growth of ISDN (Integrated Services Digital Network) and availability of ISDN equipment. The ISDN network can be used for higher quality telephony, faster data transfer like LAN-LAN interconnections and transmission of high quality audio signals for exchange of music material and real-time playing of music with other remotely placed musicians. In order to keep the transmission cost low, a bit rate reduction system like MPEG-1 can be used for the audio transmission and still maintain a high audio quality. As an example the ISDN network can be used for real-time data transfer of MIDI signals for controlling remotely placed synthesizers, and returning the synthesizer sound as an MPEG-1 encoded signal. Finally an introduction to the fast progressing future wide bandwidth network, B-ISDN, is given.

2 ISDN

ISDN is designed to be the digital upgrade to the existing public switched telephone network and is based upon circuit switching of synchronous 64 kbit/s channels, so-called B-channels. The specification of ISDN has been time consuming and differences in practice have sidetracked the fast growth of ISDN. Since January 1992 Tele Danmark has offered commercial operation based on the agreed Euro-ISDN protocol and is one of the operators committed to 100% basic access availability in the whole service area.

For ISDN two types of access are defined:

Basic Access (BA) at 144 kbit/s comprising two B-channels at 64 kbit/s and one D-channel at 16 kbit/s for signalling.

Primary Access (PA) at 2 Mbit/s comprising 30 B-channels and one D-channel at 64 kbit/s in Europe. (In the US the figures are 1.5 Mbit/s and 23 B-channels).

Nearly all telephone line connections can be used for RA, eventual with repeaters for the few very long lines. For PA two pairs of lines must be used to accommodate the higher bit rate, either by standard 2 Mbit/s transmission equipment, where repeaters are needed beyond line lengths of 1.5 km, or of the new HDLU (High bit rate Digital Subscriber Line) technology which is able to reach 4.5 km line lengths without repeaters using two or three pairs of line.

The architecture for the ISDN user access is shown in fig.1. The telephone line (U-interface) is terminated by a network termination unit (NT) in the so-called S-interface to which dedicated ISDN equipment can be connected. It is possible to connect up to 8 ISDN terminals to the same S-interface through an S-bus. The terminals can be special ISDN terminals (TE) like an ISDN telephone or terminal adaptors (TA) used for connecting existing non-ISDN equipment to the S-interface through common interfaces like RS232, X.21, X.21 etc.

For Euro-ISDN is defined two bearer services:

* 3.1 kbit/s audio for telephony.
* 64 kbit/s unrestricted for applications like data and general audio transmission.

In today systems the B-channels are treated as individual 64 kbit/s channels, but these are developed methods for handling signals at higher bit rates like 256 kbit/s. By inverse multiplexing this signal is split into 4 separate 64 kbit/s B-channels at the transmitting side of the ISDN network and recombined into the original 256 kbit/s stream at the receiving side, as indicated in fig.2.

There is no mechanism for ensuring that the same routing is used through the ISDN network for all.
the 8 channels, and random differential time delays will occur. A standardized method of resynchronizing the individual channels into a single high-bit-rate channel at the receiving side is needed. One widely used method in connection with audio signals is the BONDING method (Band-with ON Demand Interoperability Group), [2] providing for inverse multiplexing of up to 63 bi-channels with a maximum relative delay of up to 1 s.

![Figure 2: Inverse multiplexing where a 256 kbit/s signal is split in 4 separate 64 kbit/s streams and reconnected.](image)

The inverse multiplexing process can be part of an intelligent Terminal Adaptor or included in the associated terminal equipment.

Basic access ISDN has for the last year been used by Danmarks Radio for audio transmission, either 7 kHz speech or MPEGe-1 encoded audio, for contribution purposes, e.g. for transmission from church services and local music festivals, with the benefit of much lower transmission costs in relation the previous solution through lines at 2 Mbit/s.

Primary access ISDN can be used for transmission of audio signals as studio quality in accordance with the AES/EBU digital audio interface by near instantaneous companding from 20 to 15 bit/sample [3]. Inverse multiplexing must be used to ensure correct recombination of the audio signal.

3 Audio via ISDN

Digitizing of high quality audio leads to a high bit rate e.g. digital audio at CD-quality gives a bit rate of 705 kbit/s for a single channel. With the purpose of saving storage space or frequency spectrum for broadcast of digital audio the overall bit rate must be reduced either by companding methods or a sample basis or by the more efficient perceptual coding methods. The implementation of standards for perceptual coding have been driven by a joint ISO/IEC work group, Moving Picture Experts Group, which have specified the so-called MPEGe-1 standard. One part of this standard [4] covers the audio only and three layers of increasing encoder complexity and performance is defined.

In perceptual coding the emphasis is on the removal of the data which are irrelevant to the auditory system, i.e. to the human ear. The main question in perceptual coding is: What amount of noise can be introduced to the signal without being audible? Answers to the question are derived from psychoacoustics, and the basic mechanism is masking both in the frequency and in the time domain.

The sensitivity of the ear is very frequency dependent (most sensitive around 3 kHz) and a loud signal masks weak signals close in frequency to the loud signal. When a loud signal is switched on and off it produces also a pre- and post masking effect. There are two competing methods of analysis, subband coding and transform coding for forming of spectral components from the time domain input signal. Subband coding with 32 bands are used in Layer I and II of the MPEGe-1 standard, while an increased frequency resolution is introduced in Layer III by a transform coding.

By defining a psychoacoustic model to calculate a just noticeable noise-level for each band in the filterbank, followed by adaptive bit allocation and quantization, using a block companding, it is possible to obtain bit rate reductions in the order of 10 times and still maintain a high quality signal. The digital signal can have a sample frequency of 48, 44.1 or 32 kHz and can be encoded as a single channel (mono), two channels, stereo or in a joint stereo mode signal, exploiting the stereophonic irrelevance or stereophonic redundancy, and thereby increase the quality of the encoding for a given fixed bitrate.

The encoded subband information is transferred in frames of 384 samples for Layer I which introduces delay, with the following theoretical limits:

- Layer I: 19 ms
- Layer II: 35 ms
- Layer III: 59 ms

The actual delay is dependent on the practical implementation of the encoder and decoder, leading to actual delays of up to three times these figures. Delay is mainly undesired for real-time bi-directional transmissions of audio.

A wide area of bit rates is defined to each layer:

- Layer I: 32-448 kbit/s
- Layer II: 32-334 kbit/s
- Layer III: 32-320 kbit/s

The three layers have found specific applications. Layer I at 384 kbit/s for consumer systems like the DCC-system, and Layer II at 256 kbit/s has been

*Audio Hardware, Networking* 452

ICMC Proceedings 1994
chosen for the proposed digital audio broadcast system (DAB) which is intended to replace the existing FM system and improve especially mobile radio reception. Layer III, which has been designed for best performance at “low” bit rates around 64 kbit/s per audio channel, is the candidate for ISDN. Both layer II and layer III is today in use in various telecommunications networks, like ISDN and satellite links for exchange of audio material, contribution applications from local music events and retrieval from audio databases (e.g. EU-project JUKEBOX).

A competitor to the MPEG-1 standard is AC-2 from Dolby, which is based on adaptive transform coding. AC-2 has derived into AC-3 implemented as a 5.1 channel arrangement for surround sound (5 full bandwidth channels and one low-frequency subwoofer channel) at bit rates as low as 320 kbit/s by taking advantage of redundancies that may exist across the channels. The AC-3 system is chosen for the audio part for the American digital HD-TV project, Grand Alliance. Both AC-3 and a multichannel version of MPEG-1 will be included in the audio part of the new MPEG-2 standard for higher bit rates video and audio coding.

The use of perceptual coders might give rise to new problems in the mixing process as the encode is not just reducing the bit rates but also actually throws away a part of the signal, that part the ear cannot hear. Special care must be taken in systems using decoding (i.e. multiple encoding and decoding processes).

MPEG-1 does not include any specific protection to error originating from the transport system, but CRC error protection of the most vital information in the header is included in the standard.

4 MIDI via ISDN

The Musical Instrument Digital Interface (MIDI) protocol was developed to allow musical instruments to be connected together. It operates at 31.25 kbps and synchronously with one start bit, 8 data bits and one stop bit. This data rate does not fit any of the common rates for data transmission, 19.2, 24, 28.2 or 38.4 kbps. For true real-time transmission of MIDI signals e.g. for controlling a remotely placed synthesizer, the instantaneous bit rate may vary from 0 to the maximum value of 31.25 kbps. Any type of bit rate reduction to a lower rate than 31.25 kbps can lead to bit corruption or undelayed delay in buffers necessary for the rate conversion. Using one B-channel at ISDN basic access the telephone line gives a constant synchronous capacity of 64 kbit/s for the MIDI signal and this rate sets no restrictions for the instantaneous bit rate of the MIDI signals. The inevitable conversion from 31.25 kbps asynchronous to 64 kbit/s synchronous can be done by using UARTs, a cheap microprocessor and parallel-serial shiftregisters synchronized to the clock from the ISDN-network. An ISDN connection is bi-directional giving the same capacity in both directions. The return channel can be used to transmit MPEG-1 encoded audio from the remotely placed synthesizer back to the person who has generated the MIDI signals. In fig.3 is shown a set-up for this application. The second B-channel of an ISDN connection can be used for a normal 3.1 kHz speech channel for communicating with a person at the remotely placed synthesizer, as shown in fig.3, or both B-channels in the return direction can be combined for encoded audio at 128 kbit/s using inverse multiplexing.

Figure 3: Set-up for ISDN transmission of MIDI signals to a remotely placed synthesizer and MPEG-1 encoded audio in the opposite direction through one B-channel. The second B-channel is used for telephony.
5 B-ISDN

In contrary to the circuit switching technology used for ISDN a dynamic bandwidth allocation technology has been chosen by ITU-T for the future broadband integrated services digital network (B-ISDN). The transfer technology is called asynchronous transfer mode (ATM) (5) and it is a fast packet technology used for multiplexing and transporting voice, video and data information of different bit rates over the same transport medium at high throughput rates up to 155 Mbit/s.

The information is encoded in packets called cells with a fixed length of 53 bytes, consisting of a 48-byte payload and a 5-byte header, which contains information to control the transport of the payload, payload type information (voice, data, video) and error checking and congestion control information.

The cell size is a compromise between transporting real-time voice (which requires short, infrequent packets) and bursty data traffic which requires long, infrequent packets, and is not optimized for either. The choice of a fixed cell size makes switching at a high bit rate fast and cost-effective, as the switching is handled in hardware. Despite the fact that ATM is a connection-oriented technology, it can handle both connection-oriented and connectionless traffic. Speed lies behind this sort of double-duty.

Typically an ATM switch will add a variable amount of delay to a signal. Internal buffering systems create this delay so that the switches can handle a variety of signals - including delay-sensitive video and voice (like telephony) and bursty, non-delay sensitive data. There is defined a number of ATM adaptation layers (AAL1-5) providing the means for information of different types to be adapted for transport in the 48-byte payloads.

Customer premises equipment is connected to the ATM network via a user-network interface (UNI). A number of physical layers which may be used for ATM transport have been defined:

- 155 Mbit/s, SDH/SONET STM-1 (for synchronous transportation).
- 100 Mbit/s FDDI (Fast Distributed Data Interface).
- 45 Mbit/s (DS3).

In addition the ATM Forum is working on standards for lower bit rate low-cost unshielded twisted-pair copper connections running on voice-grade telephone-grade cables at a limited distance of about 100 m and even on standards at 2 Mbit/s. ATM is a connection-oriented transport technology; that is to say, during the call set-up, the routing information for the cells is defined as both UNIs and all intermediate network-node interfaces (NNIs) within the network. Connections are virtual by virtue of the fact that they don’t use bandwidth unless they are in active service carrying traffic. This dynamic bandwidth allocation will lead to a future ATM billing structure which is very different from what is known for telephony.

The ATM Forum was created in 1991 to accelerate the deployment of ATM products and services and comprises more than 450 members from all segments of the industry, all over the world. It is not a standards body, but is the driving force behind ATM by promoting industry cooperation and encouraging the convergence of interoperability specifications. The existing specifications only define permanent virtual channel service support and future releases will incorporate specifications for the signalling requirements.

In Europe an ATM pilot project between 17 operators has been established to gain experience from early broadband applications. Potential ATM pilot users will be connected in the autumn of 1994. The ATM transfer mode has already found application in switches for Local Area Networks.

6 Summary

The existing ISDN network is based on synchronous 64 kbit/s channels and can be used for real-time playing of music between musicians using MIDI signals and exchange of real-time music material coded as MPEG-1 or at 2 Mbit/s. In the future the ISDN network will be supplemented by a Broadband-ISDN network providing dynamic bandwidth allocation.

References

[2] Interoperability Requirements for N56/64 kbit/s Calls, Bandwidth On Demand Interoperability Group, BSDP2CS, 1992