Digital audio circuits: computer simulations and listening tests

L. D. J. Eggermont and P. J. Berkhout

Digital audio technology has arrived — we now have digital recording on tape and on the Compact Disc [*]. *We should now be able to enjoy a great improvement in quality. Representing the sound signal by a series of digits and recording it in this form make it possible to choose the level of accuracy required. In practice, of course, it is necessary to compromise between the desired accuracy on the one hand and the size, speed and dissipation of the circuits on the other. At Philips Research Laboratories in Eindhoven a computer system has been installed and a listening room built for the specific purpose of simulating digital signal processing on a computer and subjectively appraising the audible result. In this way it is the ear that ultimately decides how many bits are needed in each signal processing step in order to reach the high quality that is possible with digital audio technology.*

Introduction

In audio engineering wide use is now made of digital techniques both for storing audio signals and for processing them. This enables a higher audio quality to be maintained than is possible with traditional analog methods. A frequency-response curve flat from 20 Hz to 20 kHz, a signal-to-noise ratio of more than 90 dB and nonlinear distortion of no more than a fraction of a per cent are all feasible.

To achieve such values the audio signal to be digitally processed must be represented by numbers ('words') containing a sufficiently large number of bits. Words of 16 bits are usual at the beginning and end of the digital path, but during the signal processing the number of bits required is generally much larger. This has an immediate impact on the design of the circuits and on their size, of course. If the number of bits per word is too small it can give rise to additional distortion of various kinds. If the number of bits is larger than strictly required, the circuit is made unnecessarily complicated. During the design of a circuit it is therefore important at an early stage to have some understanding of the effect of the number of bits per word on the output signal.

Ir L. D. J. Eggermont is with Corporate Product Development Coordination, Philips International B.V., Eindhoven; Dr P. J. Berkhout is with Philips Research Laboratories, Eindhoven.

It is a great advantage for the designer to be able to simulate the intended functions with a computer. A complete algorithm or a single logic element can be programmed, so that simulation is also possible at the level of the logic operations in an integrated circuit — and ICs are usually the ultimate objective of the design. Parameters relating both to the algorithm and to the circuits can then be varied in the program, so as to arrive at an optimum result in a flexible, interactive method of design. If in addition the simulation can be performed in 'real time', this may often be a further argument for dispensing with 'breadboards', i.e. trial circuits with discrete components, thus considerably shortening the design time.

For evaluation of a digital audio circuit criteria are necessary. Objective criteria, such as the frequency-response characteristic, signal-to-noise ratio and distortion, can sometimes be calculated in advance at the design stage; later they can then be measured with the aid of test signals. For music and speech, however, such measurements may not provide the necessary information, since the original recording may already have contained imperfections. Nor is it always possible to calculate the effect of a finite word length analytically. What is more, it is not possible to predict

[*] Philips Tech. Rev. 40, 149-180, 1982.

how combinations of different types of distortion will be perceived by the listener. So tests with real listeners are essential for an evaluation of the quality of a digital audio circuit.

At Philips Research Laboratories in Eindhoven a computer system has been installed for the specific purpose of simulating digital audio signal processing, where possible in real time. An adjoining room has been fitted out for listening to the results (see *fig. 1*). We shall now take a closer look at both of these installations.

The computer system

The computer system (see the block diagram in *fig. 2*) is grouped around a 32-bit minicomputer, which has a main memory of 2.5 Mbyte (1 byte = 8 bits). A large memory capacity is necessary for processing audio signals: 1 second of digitized audio (stereo) occupies 200 kbyte ($2 \times 50\,000$ samples of 16 bits). Four disc systems are therefore provided as on-line random-access mass storage. One disk will store 16 minutes of stereo music. The other disks are used for temporary storage of provisional results of signal processing not performed in real time, and also for storing programs. There is also a magnetic-tape storage facility for signal archiving and program exchange.

Fig. 1. Listening room and adjoining computer room. The listening room is well insulated against extraneous noise (background sound level 18 dB(A)). A computer terminal provides the listener with a direct control of the digital audio signal processing.

The *data input/output* block (fig. 2) deals with the input and output of analog or digital audio signals either through on-line interfaces or from an off-line data-collection system. Two audio signals can be digitized or converted back into analog signals with an accuracy of 16 bits at a maximum sampling rate of 100 kHz for each channel. The off-line data-collection system consists of a digital magnetic-tape recorder, developed for this application, which is used for recording digitized music and also for mass storage.

During listening tests a person can enter his assessment into the computer from an input terminal. Properties of the audio signal heard in the listening room can be shown on a display, permitting a direct comparison of visual and auditive presentations. It is possible, for example, to display the frequency spectrum or the waveform of a selected audio fragment. The output equipment also includes a plotter and a printer.

Digital signal processing may sometimes require a very large number of arithmetical operations such as multiplications and additions. The computer is therefore connected to an array processor, designed to perform high-speed vector operations with a maximum of 16 000 numbers. This array processor can perform 12 million floating-point operations per second. Therefore processing digital audio signals in real time is possible for many algorithms.

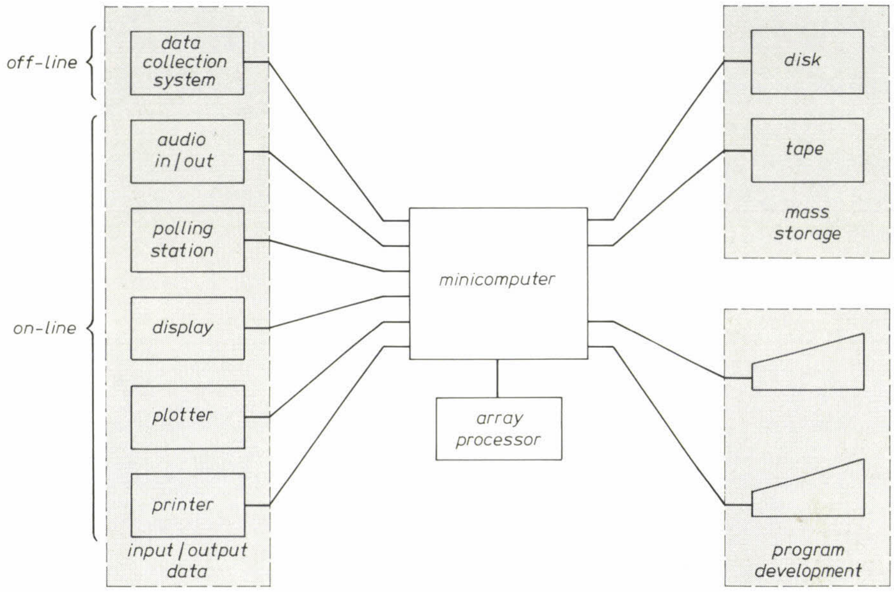


Fig. 2. Block diagram of the computer system for simulating digital audio signal-processing operations. Some of the peripherals grouped around a 32-bit minicomputer were developed for this application. The *data-collection system* is a digital tape recorder, the *polling station* enables the listener to give his assessment as direct computer input. The *array processor* performs vector operations on 16 000 numbers, enabling the digital audio signals to be processed in real time.

A software package in a high-level language is available for the analysis of original and processed signals. The package includes modular programs for interactive signal processing and synthesis. A well-organized file-management system stores and retrieves the data required for the programs. Also contained in the software package are programs for simulating various digital operations on the audio signals, such as encoding, equalization and 'teeth' filters for digital reverberation.

The listening room

The listening room, where the effect of the signal operations can be evaluated without disturbance from extraneous sounds, has a volume of about 45 m^3. One of the main requirements to be met by such a room is a very low interference noise level. This is because of the very wide dynamic range made possible by digital audio technology; we have already mentioned a signal-to-noise ratio of more than 90 dB. If a person listening to audio programmes with such a dynamic range is to be able to perceive weak noise and distortion without having to increase the gain to such an extent that signal peaks approach the threshold of pain (about 120 dB above the threshold of audibility, i.e. above an acoustic power of 10^{-12} W/m^2 carried by the sound wave), then the background sound level in the listening room must be low.

In our listening room this level is 18 dB(A). (The 'A' indicates that the frequencies in the measurement were weighted with respect to a standard curve A, which approximates to the sensitivity curve of the human ear at low levels.) This level is only found in good recording studios. Such a low level can only be reached by careful sound insulation: a 'floating floor' is fitted, i.e. a floor with a resilient connection to the concrete floor of the building structure, and the walls, which stand completely free from the concrete structure, are weighted with lead.

The intention is that the sound reaching the listener from the loudspeakers should be affected as little as possible by the acoustics of the room. The walls are therefore made highly sound-absorbent, so that the listener perceives the sound as coming mainly from the loudspeakers. This gives an exceptionally short reverberation time, between 0.2 and 0.3 seconds.

As mentioned earlier, there is a display in the listening room on which frequency spectra, waveforms etc.

can be shown. One can vary the signal-processing algorithm, as well as the parameters involved, from the listening room, so that the design procedure can be interactive.

Two examples of the design of a signal-processing algorithm

Physiological volume control

The human ear does not have a flat frequency response. Moreover, its response depends on the sound level: the weaker the sound, the more the frequency-response curve differs from a straight line. At low levels there is a pronounced maximum in the sensitivity between 2000 Hz and 5000 Hz (see *fig. 3*).

This has its effect on the fidelity of the sound reproduction. If a piece of music is played back at a level different from the level at which it was recorded — the playback level will usually be lower — the tonal balance perceived will not be the same as during the recording.

This was realized long ago, and since then efforts have been made to apply a correction by connecting a simple filter network to a tap on the volume control. A more accurate correction is possible, however, by building a digital filter whose coefficients vary with the position of the volume control. *Fig. 4* shows the corrections that are necessary at different sound levels to maintain the tonal balance, taking a reference sound level of 60 dB where the filter curve is flat.

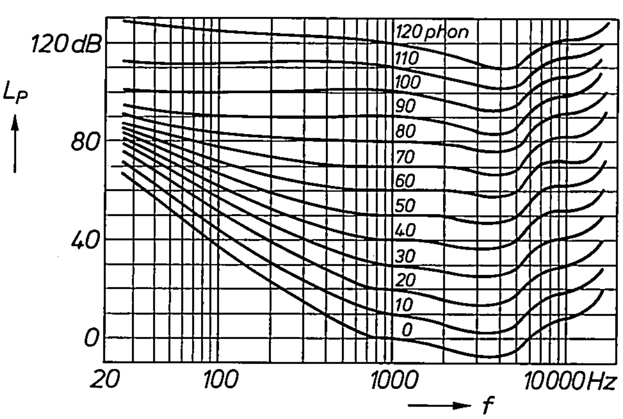


Fig. 3. The human ear does not have a flat frequency response. In the above graph (after H. Fletcher and W. A. Munson [1]) each curve connects sinusoidal tones of different frequency f that sound equally loud to the human ear even though the power level L_P is different: the level of subjective loudness (expressed in phons) is constant on each curve. The curve for 0 phon is the audibility threshold; it gives the power level of the tones that are only just audible, i.e. have a loudness level of 0 phon. The deviation from a flat response differs for different sound levels and increases as the level decreases. When music recorded on tape or disc is played back at a level different from the recording level, the subjective balance between low and high tones is different from that at the original level. Attempts are made to correct for this with a filter whose characteristic depends on the volume setting (physiological volume control).

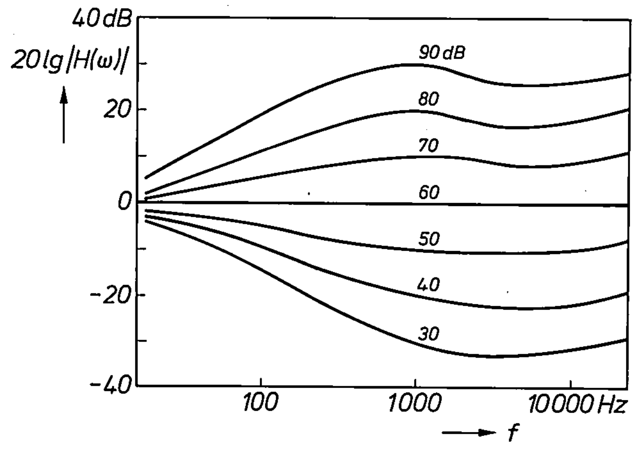


Fig. 4. Various correction curves for the frequency response during sound reproduction at different levels (*30 dB ... 90 dB*; physiological volume control). The transfer function $H(\omega)$, obtained by means of a digital filter, is adjusted to the position of the volume control.

The curves in fig. 4 could not be obtained by using standard programs for the design of filters, since these were written for designing highpass, lowpass or bandpass filters. The characteristics of fig. 4 were therefore modelled interactively with a minimum number of poles and zeros. The result was visually compared with the curves indicated by fig. 3, and assessed by listening tests.

A problem encountered when translating the computer-simulated curves of fig. 4 into hardware concerns the effect of finite word length. Digital words sometimes have to be rounded off to a given number of bits, or are truncated. A theoretical analysis was made of the effects of quantization and overflow and a computer simulation was performed. Listening tests were made to give a comparison with a simulation using the full computer word length.

It turns out that at least three poles and zeros are necessary to give a good approximation to the curves in fig. 4. The general form of the filter transfer function is then

$$H(z) = \frac{a + bz^{-1} + cz^{-2} + dz^{-3}}{1 + ez^{-1} + fz^{-2} + gz^{-3}}.$$

To find the frequency response it is necessary to calculate the absolute value of this transfer function for $z^{-1} = e^{-j\omega T}$, where ω is the angular frequency and T the sampling period. For an approximation to the upper curve (90 dB) in fig. 4 the coefficients are found to be: $a = 23.9397$, $b = -43.2$, $c = 24.417$, $d = -4.1406$, $e = -1.7721$, $f = 0.7838$ and $g = 0$. With these coefficients an accuracy of 12 bits gives a satisfactory sound quality, as confirmed in comparisons with a simulation using the full computer word length in listening tests.

For analysing the effects of quantization the impulse response is calculated from the point where the quantization is introduced as far as the output. 'Overflow', i.e. the result of exceeding the signal amplitude that can be expressed by the given number of bits, is detected by applying a pulse to the input and seeing whether the response at a particular point exceeds a certain amplitude. To avoid overflow 10 extra bits are required for the internal representation and the word length at the output must be 6 bits longer than at the input. With a 16-bit input signal the internal word length must thus be 26 bits and the word length at the output 22 bits. If the output signal is subsequently quantized to 16 bits, this final quantization step will determine the signal-to-noise ratio at the output.

Reverberation

In connection with an investigation of electronic reverberation systems being carried out elsewhere in the laboratory, simulations of digital reverberation circuits were made. Reverberation is the sum of a large number of delayed versions of the sound signal, each with its own delay. In digital audio technology reverberation is produced by means of delay lines. The output of such a delay line can be fed back through a filter to the input, so that the filtered signal travels along the delay line again, and so on (*fig. 5*). We prefer to call the circuit thus produced a 'teeth filter' [2]. In this way the long delays necessary for reverberation can be built up.

These long delays create problems in an investigation of the consequences of finite word length and overflow, however, because it is difficult to find a closed-form expression for the desired impulse response. In this way it is only possible to estimate the number of bits required in the worst case.

The transfer function of the teeth filter in fig. 5 is

$$\frac{Y(z)}{X(z)} = \frac{z^{-m}}{1 - z^{-m}H(z)},$$

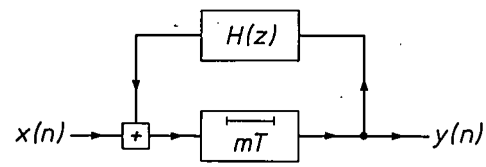
where m is the number of signal samples stored in the delay line and $H(z)$ is the transfer function of the filter in the feedback loop. This filter was introduced to give the reverberation a natural tonal colour, i.e. a natural frequency response. The number of poles ranges from 2000 to 4000 at sampling rates of 40 to 50 kHz; this large number makes the analysis difficult. An analysis has been given of the effects of finite word length and overflow for the case where $H(z)$ is a first-order recursive filter [3]. However, the results of this analysis only apply directly to a signal whose amplitude does not change significantly during the entire duration of the impulse response (1-5 s), and this is not generally the case with music.

[1] H. Fletcher and W. A. Munson, J. Acoust. Soc. Am. 5, 82, 1933/34.

[2] This filter is not really the same as the well-known *comb* filter. If $H(z)$ is a constant its frequency characteristic consists of a succession of real-frequency poles, whereas the frequency characteristic of a comb filter consists of a succession of real-frequency zeros. See L. R. Rabiner *et al.*, Terminology in digital signal processing, IEEE Trans. **AU-20**, 322-337, 1972.

[3] P. J. Berkhout and L. D. J. Eggermont, Some design issues in digital signal processing for digital-audio systems, Proc. ICASSP 82, Paris 1982, pp. 81-84.

A digital reverberation system will have not just one teeth filter but a combination of teeth filters, and they will also be connected to other processing modules. To limit the word length it is necessary to scale the amplitude of the signals between the different operations. The scale factors are obtained in part from a statistical analysis, but with a signal that varies as greatly as music the scale factors have to be verified in



$$\frac{Y(z)}{X(z)} = \frac{z^{-m}}{1-z^{-m}H(z)}$$

Fig. 5. Teeth filter for artificial reverberation. The digital input signal $x(n)$ (where n is the serial number of the signal sample) is stored in a delay line for m sampling periods, each of duration T, and then passed as $y(n)$ to the output. The output signal $y(n)$ is then added to the input signal, via a filter H, and delayed once again. If in a hypothetical case the input signal consisted of one very short pulse (one sample differs from zero, all the others are zero), then this pulse would appear at the output after mT seconds and again after each successive period of mT seconds, although affected by the filter H. The representation of the transfer function of the feedback system in the frequency domain has a series of maxima at regular intervals, which suggests the name 'teeth filter'.

listening tests. It has been found that a word length of more than 22 bits is necessary in the filters for 16-bit quality to be maintained from input to output, corresponding to a signal-to-noise ratio of more than 90 dB.

The example of digital reverberation, like that of physiological volume control, shows that it is a great advantage at the design stage if computer simulation is available, preferably interactive, and with real-time signal processing capability. The listening tests that are indispensable in such a design process can then be made immediately. The facilities outlined here meet these requirements and considerably reduce the design time for digital audio circuits.

Summary. At Philips Research Laboratories, Eindhoven, facilities are available for rapidly and flexibly designing the algorithms used in the digital processing of audio signals. The facilities consist of a computer system and a well-insulated listening room in which the background level is 18 dB(A). The algorithms can be simulated by the computer system and immediately applied to music signals stored in digital form on magnetic disks or tapes. While listening to the results the listener can vary parameters and can also see the resulting spectrograms or waveforms on a visual display. Algorithms that have been developed for physiological volume control and for artificial reverberation are discussed as examples of the capabilities of the system; the number of bits of word length that should be used in the signal processing if the operations are not to impair the sound quality has been indicated.

Scientific publications

These publications are contributed by staff of laboratories and plants that form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the research laboratories mentioned below. The publications are listed alphabetically by journal title.

Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 5100 Aachen, Germany	*A*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels, Belgium	*B*
Philips Natuurkundig Laboratorium, Postbus 80000, 5600 JA Eindhoven, The Netherlands	*E*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*
Philips Research Laboratories, Cross Oak Lane, Redhill, Surrey RH1 5HA, England	*R*
Philips Research Laboratories Sunnyvale P.O. Box 9052, Sunnyvale, CA 94086, U.S.A.	*S*

A. Daniels, M. Gasser* & A. Sherman* (* *Goddard Space Flight Center, Greenbelt, MD*) *N*	Magnetically suspended Stirling cryogenic space refrigerator: status report	Advances in cryogenic engineering, Vol. 27, ed. R. W. Fast, Plenum, New York	711-719	1982
H. M. Meehan & R. C. Sweet *N*	Novel titanium-aluminum joints for cryogenic cold finger structures	Advances in cryogenic engineering, Vol. 27, ed. R. W. Fast, Plenum, New York	721-726	1982
A. K. Niessen *E*	Note on the prediction of the enthalpy of formation of solid compounds of the transition and noble metals with chalcogens and halogens	High Temp. — High Pressures **14**	649-651	1982
S. Hoekstra, M. A. Munnig Schmidt-van der Burg, H. Galenkamp & H. van Wijngaarden *E*	Quantitative determination of stress and strain distributions during high axial compression of aluminium with the aid of small ruby particles	J. Appl. Mech. **105**	194-198	1983
S. Colak & W. K. Zwicker *N*	Transition rates of Tb^{3+} in TbP_5O_{14}, $TbLiP_4O_{12}$ and $TbAl_3(BO_3)_4$: an evaluation for laser applications	J. Appl. Phys. **54**	2156-2166	1983
O. Boser *N*	Internal friction due to hysteretic dislocation motion in solid solution crystals	J. Appl. Phys. **54**	2338-2343	1983
J. M. F. van Dijk & M. F. H. Schuurmans *E*	On the nonradiative and radiative decay rates and a modified exponential energy gap law for *4f-4f* transitions in rare-earth ions	J. Chem. Phys. **78**	5317-5323	1983
P. C. M. Gubbens*, A. M. van der Kraan* (* *Interuniv. Reactor Inst., Delft*) & K. H. J. Buschow *E*	First order transition and magnetic structure of $TmCo_2$	J. Magn. & Magn. Mater. **29**	113-116	1982
A. L. J. Burgmans & A. H. M. Smeets *E*	The ionisation coefficient in Ar-Hg mixtures	J. Phys. D **16**	755-762	1983
R. Boom (*Hoogovens, IJmuiden*), F. R. de Boer (*Univ. Amsterdam*), A. K. Niessen & A. R. Miedema *E*	Enthalpies of formation of liquid and solid binary alloys based on 3d metals, III	Physica **115B**	285-309	1983
A. G. van Nie & G. Kersuzan (*TRT, Le Plessis-Robinson*) *E*	Accurate 'fixed-angle' models for calculating the temperature rise of rectangular heat sources in hybrid circuits	Proc. 4th Eur. Hybrid Microelectron. Conf., Copenhagen 1983	523-530	1983
S. Nakahara (*Bell Labs, Murray Hill, NJ*) & F. J. A. den Broeder *E*	Diffusion-induced grain boundary migration in heated Cu-Ni alloy targets during sputtering	Scr. Metall. **17**	607-610	1983
P. K. Larsen & J. F. van der Veen (*FOM, Amsterdam*) *E*	Photoemission from MBE grown III-V surfaces and interfaces	Surf. Sci. **126**	1-19	1983

Volume 41, 1983/84, No. 3 pages 73-104 Published 6th January 1984