ANSI S1.11-1986
(ASA 65-1986)
[Revision of
ANSI S1.11-1966(R1976)]

Standards Secretariat
Acoustical Society of America
335 East 45th Street
New York, New York 10017-3483

AMERICAN NATIONAL STANDARD
Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters

ABSTRACT

This standard provides performance requirements for fractional-octave-band bandpass filters, including, in particular, octave-band and one-third-octave-band filters. Basic requirements are given by equations with selected empirical constants to establish limits on the required performance. The requirements are applicable to passive or active analog filters that operate on continuous-time signals, to analog and digital filters that operate on discrete-time signals and to fractional-octave-band analyses synthesized from narrow-band spectral components. Filter designs are described by an Order number which is usually related to the number of poles in the analog prototype low-pass filter or the number of pole pairs in the analog prototype bandpass filter. The overall accuracy of a filter set is described by a Type number, which is determined by the accuracy of a measurement of a white noise signal, and a required Sub-Type letter, which is determined by the accuracy of the measurement of signals with moderate spectral slopes. Four accuracy grades are allowed: the most accurate for precise analog and digital filters; the next for filters achievable with the technology of the 1980s. The two least accurate grades describe filters which meet the requirements of S1.11-1966. An Appendix is included for reference to terminology used in digitial signal processing.

Published by the American Institute of Physics for the Acoustical Society of America

AMERICAN NATIONAL STANDARDS ON ACOUSTICS

The Acoustical Society of America holds the Secretariat for Accredited Standards Committees S1 on Acoustics, S2 on Mechanical Shock and Vibration, S3 on Bioacoustics, and S12 on Noise. Standards developed by these committees, which have wide representation from the technical community (manufacturers, consumers, and general-interest representatives), are published by the American Institute of Physics for the Acoustical Society of America as American National Standards after approval by their respective standards committees and the American National Standards Institute.

These standards are developed as a public service to provide standards useful to the public, industry, and consumers, and to Federal, State, and local governments.

This standard was approved by the American National Standards Institute as ANSI S1.11-1986 on 16 July 1986.

An American National Standard implies a consensus of those substantially concerned with its scope and provisions. An American National Standard is intended as a guide to aid the manufacturer, the consumer, and the general public. The existence of an American National Standard does not in any respect preclude anyone, whether he has approved the standard or not, from manufacturing, marketing, purchasing, or using products, processes, or procedures not conforming to the standard. American National Standards are subject to periodic review and users are cautioned to obtain the latest editions.

Caution Notice: An American National Standard may be revised or withdrawn at any time. The procedures of the American National Standards Institute require that action be taken to affirm, revise, or withdraw this standard no later than five years from the date of publication.

FOREWORD

[This Foreword is for information only and is not a part of American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters, S1.11-1986 (ASA Catalog No. 65-1986).]

This standard was developed under the jurisdiction of Accredited Standards Committee S1 using the American National Standards Institute (ANSI) Standards Committee Procedure. The Acoustical Society of America holds the Secretariat for Accredited Standards Committee S1. This standard was approved for publication by Accredited Standards Committee S1 and by the American National Standards Institute.

This standard is a revision of ANSI S1.11-1966 (R1976), Octave, Half-Octave, and Third-Octave Band Filter Sets. This standard differs from the 1966 version in several ways principally related to advances in the state-of-the-art. The design complexity of a bandpass filter is now described by an Order number instead of a Class number. The Order number is usually related to the number of poles in the prototype analog low-pass filter design. This standard also introduces Type and Sub-Type designations to describe the accuracy of a measurement for signals having several spectral slopes, including white noise as a reference. As in the previous standard, performance requirements are established by equations with selected empirical constants to establish limits. Two type numbers are established to cover octave-band and one-third-octave-band filters which meet the requirements of S1.11-1966. Digital as well as analog filters are covered by this standard.

Accredited Standards Committee S1, under whose jurisdiction this standard was developed, has the following scope:

> Standards, specifications, methods of measurement and test, and terminology, in the fields of physical acoustics, including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to safety, tolerance, and comfort.

At the time this standard was submitted to Accredited Standards Committee S1, for final approval, the membership was as follows:

E. H. Toothman, *Chairman* D. L. Johnson, *Vice-Chairman* A. Brenig, *Secretary*

Acoustical Society of America • T. F. Embleton, D. L. Johnson (*Alt*)
Air-Conditioning and Refrigeration Institute • R. Harold, H. C. Skarbek (*Alt*)
American Industrial Hygiene Association • C. C. Bohl
American Iron & Steel Institute • E. H. Toothman, P. A. Hernandez (*Alt*)
AT&T • S. R. Whitesell, R. M. Sachs (*Alt*)
Audio Engineering Society • L. W. Sepmeyer, M. R. Chial (*Alt*)
Bruel & Kjaer Instruments, Inc. • G. C. Michel
Computer & Business Equipment Manufacturers Association • L. F. Luttrell
Exchange Telephone Group Committee (ETGC) • J. L. Sullivan, O. J. Gusella (*Alt*)
Larson-Davis Laboratories • D. L. Johnson, L. Davis (*Alt*)
National Bureau of Standards • S. L. Yaniv
National Council of Acoustical Consultants • A. P. Nash, G. W. Kamperman (*Alt*)
Scantek, Inc. • R. J. Peppin
U.S. Army Aeromedical Research Laboratory • B. Mozo, J. H. Patterson, Jr. (*Alt*)
U.S. Army Communication Electronics Command • M. S. Mayer
U.S. Army Human Engineering Laboratory • J. Kalb, G. Garinther (*Alt*)
U.S. Department of the Air Force • R. McKinley
U.S. Department of the Army, Environmental Office • P. D. Schomer, R. Raspet (*Alt*)

Individual experts of the S1 Committee were:

L. Batchelder
S. L. Ehrlich
K. M. Eldred
D. R. Flynn
R. S. Gales
W. J. Galloway
E. E. Gross, Jr.
R. M. Guernsey
R. K. Hillquist
R. Huntley
W. W. Lang
G. C. Maling, Jr.
A. H. Marsh
A. P. G. Peterson
L. W. Sepmeyer
W. R. Thornton
H. E. von Gierke
G. S. K. Wong
R. W. Young

The membership of Working Group S1-5 Octave-band and Fractional-octave-band Analog and Digital Filters, which assisted the Accredited Standards Committee S1 in the preparation of this standard, had the following membership:

L. W. Sepmeyer, *Chairman*

D. Cox
A. H. Gray, Jr.
F. J. Harris
J. F. Kaiser
A. H. Marsh
G. C. Michel
A. P. Nash
B. E. Walker
G. S. K. Wong

Suggestions for improvement of this standard will be welcomed. They should be sent to the Standards Manager, Standards Secretariat, Acoustical Society of America, 335 East 45th Street, New York, NY 10017-3483.

CONTENTS

FIGURES

TABLES

American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters

0 INTRODUCTION

0.1 General Objective

The spectral distribution of the power in sound and vibration signals is determined for various purposes. Those purposes include scientific, technical, legal, and artistic requirements. The types of signals involved cover wide variations of waveform, amplitude, frequency content, duration, coherence, etc. Suitable standards are required for spectrum analysis systems so that satisfactorily uniform results can be obtained from any analyzer that meets the standard for its Type.

0.2 Spectrum Analysis

Frequency-selective networks used in spectrum analyzers fall into two broad classes: (1) constant-bandwidth filters where the difference between the upper and lower bandedge frequencies remains constant over the tuning range of the analyzer; and (2) constant-percentage-bandwidth filters, where the ratio of the upper bandedge frequency to the lower bandedge frequency is constant over the tuning range, e.g., fractional-octave-band filters.

0.3 Selection of Frequency Bands

An octave-band or fractional-octave-band filter has bandedge frequencies that have a fixed relationship to the geometric-mean frequency of the passband. For convenience, filters are identified by the standard preferred frequencies, or, for octave-band and one-third-octave-band filters, by the standard band numbers specified in ANSI S1.6-1984.

The performance requirements of this standard, while directed mainly at octave-band and one-third-octave-band filters may be applied to any fractional-octave-band filter. For analog filters, it is often desirable to base the filter design on geometric mean frequencies that are related by fractional powers of ten, while for digital filters and switched capacitor filters, geometric mean frequencies based on powers of two may be more convenient.

0.4 Designation of Filter Sets

For many purposes, it is sufficient for a filter set to contain a particular number of filters covering the usual audio-frequency range. However, other applications may require fewer or additional filters to cover a range less than or greater than the usual range. To permit this standard to cover a wide range of applications, filter sets of three designations have been delineated for use in the audio-frequency range: i.e., Restricted Range, Extended Range, and Optional Range.

0.5 Designation of Filter Characteristics

During the 1970s, advances in technology and the operation of the market place resulted in what was then known as the Class II octave-band filter and Class III third-octave-band filter becoming *de facto* industry standards. During that time, few, if any, Class I octave-band filters were offered for sale. Also, the half-octave-band filter became extinct. Since the design goal characteristics in S1.11-1966 for both the Class II octave-band and Class III third-octave-band filters were based on the third-order Butterworth or maximally flat filter characteristic, this standard bases filter complexity designations on the order of the prototype low-pass filter design or the number of pole pairs or resonators in the bandpass filter design. Thus the one-third-octave-band filter designation number remains the same.

However, to avoid possible confusion, the term Class was changed to Order, the term commonly used in filter design engineering. Thus, a third-order Butterworth octave-band filter that was previously called a Class II octave-band filter is now designated as an Order 3 octave-band filter. Stated in another way, the filter Order designation is directly correlated with the number of resonators or pole pairs in an analog filter design including the prototype for an infinite impulse response or recursive digital filter. Hence, a single tuned circuit would be designated an Order 1 filter while an ideal, or brick-wall, filter would be designated an Order infinity filter. Also, roman numerals have been replaced by arabic numerals for the Order number.

The choice of filter Order to be used for a given measurement depends on the accuracy required. For many applications, any set of Order 3 filters is sufficient. For some applications (e.g., determining the pressure spectrum level of a short duration impulsive sound or the true one-third-octave-band sound-pressure levels of a noise signal that has propagated a moderate distance through an absorptive medium), spectrum slopes ranging from -30 to -90 decibels per octave can occur and filters of Order greater than 3 may be warranted. For general usage with sound level meters, octave-band and one-third-octave-band filters should be not less than Order 3.

0.6 Specification of Filter Characteristic Shape

Specification of the attenuation characteristic of a filter by a small number of straight-line segments often complicates the design of economical real filters which do not have attenuation characteristics approximating long straight-line segments. For this standard, the limiting attenuation characteristics are specified by mathematical expressions based upon the design formulas of maximally flat bandpass filters. This choice makes it easier for the designer to provide filters that meet the requirements for attenuation characteristics, effective bandwidth, and nominal midband frequency.

It was considered feasible for the tolerance limits in this standard to be more stringent than in the 1966 version. However, for those applications where technical and economic considerations do not require increased accuracy, the previous 10% bandwidth error limit is retained and three additional categories of effective bandwidth accuracy are established. In conformance with the usage of American National Standard Specification for Sound Level Meters, ANSI S1.4-1983, the basic accuracy of the response of a filter to white noise (relative to the response of a corresponding ideal filter) is designated by a Type number, with increasing error being designated by a larger number.

The bandwidth error of a filter depends upon its attenuation at the bandedges, the slope of the attenuation characteristic in the transition bands, and the spectrum slope of the input noise spectrum. To accommodate various practical applications, an additional component has been added to the bandwidth error Type designator to apprise the user of the inherent bandwidth error for analysis of sound or vibration signals having moderately steep spectral slopes.

A mathematical equation is given to specify the design reference attenuation characteristic for any filter. That equation, together with equations defining filter bandwidth and filter bandwidth quotient, constitutes the basic performance parameters for any filter covered by this standard. The equations are unambiguous and not subject to errors of curve plotting, interpretation, or reproduction.

0.7 Phase and Transient Response

For most sound and vibration measurements, a time-averaged mean-squared voltage is measured as the analog of the time-mean-square sound pressure or vibration signal. Since that quantity is independent of the relative phase among spectral components, the phase response characteristic of a filter set is of no concern and therefore has not been specified. Also, deviations from the inherent phase and transient responses of an analog filter will result from the allowed manufacturing tolerances on component values, making it impractical to specify closely filter characteristics, particularly the phase response.

Distortion of the waveform of a transient signal introduced by phase nonlinearity of a filter should have negligible effect on the indicated filtered level. However, phase distortion can affect the measurement of the peak level of a complex waveform. When a filter of any practical design is excited by a short-duration transient and the output is displayed on an oscilloscope, a transient response will be observed in the form of damped oscillations called "ringing." Limits are placed on the magnitude and duration of such oscillations.

0.8 Influence of External Conditions

When filter sets made according to this standard are to be used with portable instruments such as sound level meters, the filter set should be designed to meet the same environmental conditions specified for those instruments. The range of environmental conditions specified in ANSI S1.4-1983 is incorporated in this standard.

0.9 Extension to Infrasonic and Ultrasonic Frequencies

The 1966 version of this standard was intended to apply primarily to measurement of sound or vibration signals with a frequency content within the usual range of human audibility, commonly called the audio-frequency range. For many applications, it is desirable to extend the frequency range downward to infrasonic frequencies or upward to ultrasonic frequencies.

This standard permits the use of either of two geometric series for specifying midband frequencies. One series is a fractional-octave series based on powers of two; it is referred to as the base two system. The other series is based on powers of ten; it is referred to as the base ten system.

In the base ten system, the midband frequencies included within any 10:1 frequency range are the same as within any other 10:1 frequency range except for the position of the decimal point. In the base two system, the midband frequencies are unique and do not have values which repeat.

For measurement of the spectral content of a signal within the audio-frequency range, each of the two series of midband frequencies includes 1 kilohertz exactly. [The standard reference frequency of 1 hertz (see ANSI S1.6-1984) is included in the base ten series and in the base two series for filter sets designed for analyses of signals in the infrasonic-frequency range.]

As the range of midband frequencies is extended from the audio-frequency range to infrasonic frequencies, midband frequencies calculated by the base ten system become increasingly greater than those calculated by the base two system, and vice versa for extension to ultrasonic frequencies.

This standard recommends the use of different reference frequencies to minimize the differences between exact midband frequencies calculated by the base two and base ten systems in the infrasonic- and ultrasonic-frequency regions.

1 PURPOSE

The purpose of this American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters is to specify the geometric mean frequencies, bandedge frequencies, bandwidths, attenuation charcteristics, bandwidth error, and other pertinent design parameters for constant-percentage-bandwidth bandpass filters so that spectral analyses made with filters conforming to the specified performance requirements will be consistent within known tolerance limits.

2 SCOPE

The scope of this standard specification includes bandpass filter sets suitable for analyzing electrical signals as a function of frequency. The bandwidth of the filters is a constant percentage of the midband frequency of each filter band. The scope includes passive, active, and sampled-data bandpass filters obtained by any design realization procedure. All filters, including fractional-octave-band filters synthesized by Discrete Fourier Transform (FFT) techniques shall meet all electrical requirements of the standard. Three frequency ranges of filter sets are described for use in the audio-frequency range where the reference frequency is one kilohertz. The filters in a filter set may be of any filter design Order. Four filter Types and five Sub-Types are established based on the amount of passband ripple and on the bandwidth error for both white noise and random noise having specified moderately sloping spectral distibutions.

3 APPLICATIONS

Bandpass filters designed in conformance with this standard are well suited to the spectral analysis of any electrical signals having energy distributed over a broad frequency range. The resolution of discrete frequency components in a signal depends on the Order and bandwidth of the filter and the position of the components within the passband of the filter. The frequency range of filters covered by the requirements of this standard may be extended to as low or as high a frequency as desired.

NOTE: Existing filter sets designed to the requirements of earlier versions of this standard may be shown to comply with the applicable requirements by performing the specified tests and computations.

4 STANDARDS REFERRED TO IN THIS DOCUMENT

4.1 American National Standards

[When the following American National Standards are superseded by a revision approved by the American National Standards Institute, Inc., the revision shall apply.]

(1)American National Standard Acoustical Terminology (Including Mechanical Shock and Vibration), ANSI S1.1-1960 (R1976).

(2) American National Standard Specification for Sound Level Meters, ANSI S1.4-1983.

(3) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements, ANSI S1.6-1984.

4.2 International Standards

[When the following publications are superseded by an approved revision, the revision shall apply.]

(1) "Octave, half-octave, and third-octave-band filters intended for the analysis of sounds and vibration," International Electrotechnical Commission Recommendation, Publication 225(1966).

(2) "Acoustics and electroacoustics," Chapter 801 of the International Electrotechnical Vocabulary, International Electrotechnical Commission Publication IEC 50(801)-1984.

(3) "Preferred frequencies for acoustical measurements," International Organization for Standardization Recommendation R 266(1975).

5 DEFINITIONS

[Definitions of certain terms are included here for the purposes of this standard. The definitions are consistent with those in ANSI S1.1-1960 (R1976) and those in IEC Publication 50(801)-1984.]

5.1 wave filter (filter): a transducer for separating waves on the basis of their frequency.

5.2 bandpass filter: a wave filter with a single transmission band (passband) extending from a lower bandedge frequency greater than zero to a finite upper bandedge frequency.

5.3 bandedge frequencies: the upper and lower cutoff frequency of an ideal bandpass filter. In this standard, the ratio of the upper to the lower cutoff frequency is specified as a fractional power of two.

5.4 filter bandwidth: difference between the upper and lower bandedge frequencies, and, hence, the width of the passband.

5.5 filter bandwidth quotient: figure-of-merit measure of the relative bandwidth, or sharpness, of a bandpass filter and described by the ratio of the midband frequency to the bandwidth.

5.6 spectrum: description of the resolution into components of a function of time, each component having a different frequency and (usually) different amplitude and phase. A *continuous spectrum* has spectral components continuously distributed over a range of frequencies. A *white noise spectrum* has a power spectral density (mean-square amplitude per unit frequency) essentially independent of frequency over a specified frequency range. A *pink noise spectrum* has a spectrum level slope of − 3 dB per octave; hence, equal power per fractional-octave band.

5.7 attenuation: in decibels, reduction in the level of some characteristic of a signal between two stated points in a transmission system relative to a reference attenuation.

NOTES:

(1) Since this standard is concerned mainly with active analog and digital filters, matched source and load impedances are usually not necessary. Then, attenuation, in decibels, is the difference between the level of a signal applied to the input of a filter and the level of the signal delivered by the filter to its output.

(2) **transmission loss** is used synonymously with attenuation, as defined above, in connection with filter characteristics.

(3) **insertion loss** is a term also frequently used in connection with filters. The insertion loss, in decibels, resulting from insertion of a transducer in a transmission system is 10 times the logarithm to the base 10 of the power delivered to that part of the system that will follow the transducer, before insertion, to the power delivered to the same part of the system after its insertion. For passive filters operated between resistive terminating impedances, the insertion loss characteristic, employing the minimum insertion loss value, as referent, is the same as the attenuation characteristic.

5.8 terminating impedance: impedances of the external input and output electrical circuits between which the filter is connected.

5.9 peak-to-valley ripple: difference in decibels between the extremes of a series of minima and maxima of ripple in the attenuation in the passband. See Appendix E, Fig. 14 and paragraph IV (29).

5.10 dynamic range: in decibels, difference between the output level when the maximum-rated band-centered sinusoidal input signal is applied and the wideband (e.g., 5 Hz to 100 kHz) output noise level when the input is terminated with rated impedance and no input signal is applied.

5.11 terminology used in digital signal processing: see Appendix E for additional information on terms used to specify digital filter performance requirements.

6 REQUIREMENTS

6.1 Filter Sets

6.1.1 An audio-frequency filter set shall provide a number of filter bands according to the schedules listed in Table 1, and shall bear the appropriate Range designation: Restricted Range, Extended Range, or Optional Range.

6.1.2 Octave-band and one-third-octave-band filters shall be identified, or labeled, as shown in Table 1 by the preferred frequencies of ANSI S1.6-1984 and ISO Recommendation R266(1975).

6.1.3 Octave-band analyses may be determined by combining the squared outputs of adjacent fractional-octave-band filters.

6.2 Reference Frequencies

When the range of the filter set is predominantly in the audio-frequency range, the reference frequency shall be 1 kilohertz. When the predominant frequency range of the filter set is infrasonic, i.e., preferred frequencies less than 20 hertz, the reference frequency of 1 hertz is recommended. When the predominant frequency range of the filter set is ultrasonic, i.e., preferred frequencies greater than 31.5 kilohertz, the reference frequency of 1 megahertz is recommended.

6.3 Exact Midband Frequencies

6.3.1 This standard permits two systems for specifying the exact geometric-mean frequencies. One system is based on powers of ten and is referred to as the base ten system. The other system is based on powers of two and is referred to as the base two system.

6.3.2 The exact value of the midband or geometric-mean frequency of a bandpass filters of any bandwidth shall be based on the applicable reference frequency chosen in accordance with 6.2.

6.3.3 Exact values of midband or geometric-mean frequencies f_m shall be determined from the reference frequency f_r, and the frequency ratio U, according to

$$f_m = (f_r)(U^k), \qquad (1)$$

where the frequency ratio is that applicable to the base two system or the base ten system and k is an integer having positive, negative, or zero value such that $f_m = f_r$ when $k = 0$.

NOTE: For the audio-frequency range of 6.1.1 and the base ten system, k is related to the band number N in Table 1 or ANSI S1.6-1984 by $k = N - 30$.

TABLE 1. Listing of filter bands to be provided by a set of filters for measurement of the spectral content of signals over the frequency range of the audio frequency band.

Band number, N	Preferred frequency, hertz	Octave-band range: Restricted	Octave-band range: Extended	One-third-octave-band range: Restricted	One-third-octave-band range: Extended	Optional range
14	25				x	
15	31.5		x		x	
16	40				x	
17	50				x	
18	63		x		x	
19	80				x	A
						S
20	100			x	x	
21	125	x	x	x	x	S
22	160			x	x	P
						E
23	200			x	x	C
24	250	x	x	x	x	I
25	315			x	x	F
						I
26	400			x	x	E
27	500	x	x	x	x	D
28	630			x	x	
						B
29	800			x	x	Y
30	1000	x	x	x	x	
31	1250			x	x	M
						A
32	1600			x	x	N
33	2000	x	x	x	x	U
34	2500			x	x	F
						A
35	3150			x	x	C
36	4000	x	x	x	x	T
37	5000			x	x	U
						R
38	6300				x	E
39	8000		x		x	R
40	10 000				x	
41	12 500				x	
42	16 000		x		x	
43	20 000				x	

6.3.4 For midband frequencies in the base two system, the ratio of a pair of midband frequencies in two adjacent filter bands shall be

$$U_2 = f_{(i+1)}/f_i = 2^b, \quad (2)$$

where f_i is a midband frequency in the ith filter band, and $f_{(i+1)}$ is the midband frequency of the next higher band. The symbol b is the bandwidth designator for the particular fractional-octave-band filter of interest. Thus $b = 1$ for octave-band filters, $b = 1/3$ for one-third-octave-band filters, $b = 1/6$ for one-sixth-octave-band filters, etc.

NOTE: The frequency ratio of Eq. (2) is expressed in b octaves by

$$b = \mathrm{lb}(f_{(i+1)}/f_i) = \mathrm{lb}(U_2), \quad (3)$$

where lb is the symbol for base 2 (binary) logarithms.

6.3.5 For midband frequencies in the base ten system, the ratio of an adjacent pair of frequencies shall be

$$U_{10} = f_{(i+1)}/f_i = 10^{(3b/10)}. \quad (4)$$

6.4 Bandedge Frequencies

For any fractional-octave-band filter, the lower and upper bandedge frequencies shall be determined from

$$f_1 = 2^{-b/2} f_m, \quad (5)$$

$$f_2 = 2^{b/2} f_m, \quad (6)$$

where b is the bandwidth designator described in 6.3.4. Table 2 gives the formulas and numerical values of f_1 and f_2 for Order 3 octave-band ($b = 1$) and Order 3, one-third-octave-band ($b = 1/3$) filters.

6.5 Reference-Filter Bandwidths and Bandwidth Quotients

6.5.1 The reference bandwidth B_r of any fractional-octave-band filter shall be determined from the difference between the upper and lower bandedge frequencies as

$$B_r = (2^{b/2} - 2^{-b/2}) f_m. \quad (7)$$

6.5.2 For any fractional octave-band-band filter, the reference bandwidth quotient Q_r for a filter of any Order shall be determined from

TABLE 2. Nominal bandedge frequency ratios and filter bandwidth ratios for Order 3 octave-band and one-third-octave-band filters.

Formula	Octave-band Base 2 or Base 10	One-third-octave-band Base 2 or Base 10
f_1/f_m	$2^{-1/2}$	$2^{-1/6}$
f_2/f_m	$2^{1/2}$	$2^{1/6}$
Numerical value		
f_1/f_m	0.70711	0.89090
f_2/f_m	1.41421	1.12246
Reference bandwidth ratio		
$(f_2 - f_1)/f_m$	0.70711	0.23156
Reference Q_r		
Eq. (8)	1.41421	4.31847
Butterworth design Q		
Eq. (9) for $n = 3$	1.48096	4.52229

$$Q_r = f_m/B_r = 1/(2^{b/2} - 2^{-b/2}). \quad (8)$$

Table 2 gives the numerical values for the reference bandwidth and reference bandwidth quotient for Order 3 octave-band and one-third-octave-band filters.

6.5.3 In order for a maximally flat Butterworth design bandpass filter to transmit the same white-noise power as an ideal (Order infinity) bandpass filter [i.e., to have zero error in effective bandwidth, see 6.7], the actual bandwidth of the filter needs to be smaller than the reference bandwidth given by Eq. (7). Appendix A discusses the analytical background for this determination.

6.5.4 As shown in Appendix A, the amount by which the design bandwidth quotient Q_d is larger than Q_r for Butterworth filters is given by

$$Q_d = [(\pi/2n)/\sin(\pi/2n)] Q_r, \quad (9)$$

where n is the order of the filter design and Order of the filter. Table 2 gives the numerical values for the design bandwidth quotients for Order 3 [$n = 3$ in Eq. (9)] octave-band and one-third-octave-band Butterworth filters.

6.6 Attenuation Characteristics of Individual Filters

6.6.1 For each filter characteristic, attenuation is specified with respect to the reference attenuation, see Eqs. (17) and (18), in the frequency range f_1 to f_2

from Eqs. (5) and (6). Attenuation characteristics are defined according to the order of the filter design (i.e., the number of poles in the low-pass prototype or the number of pole pairs or resonant circuits in the analog prototype bandpass filter). Filter sets shall be marked with the appropriate Order number.

6.6.2 The mathematical statement shall be the governing consideration for the design attenuation characteristic specified below. When tested, the actual filter characteristic shall meet the requirements on effective bandwidth (see 6.7) and passband uniformity (see 6.8 and 6.9).

6.6.3 The design reference attenuation A_d, in decibels, for any frequency and any filter Order, shall be that provided by the maximally flat (Butterworth) characteristic as given by Eq. (10):

$$A_d = 10\lg\{1 + Q_d^{2n}[(f/f_m) - (f_m/f)]^{2n}\}, \quad (10)$$

where n is the order of the design and Order of the filter, and Q_d is given by Eq. (9).

6.6.3.1 The stopband attenuation shall be not less than 65 decibels. [See Appendix E, Fig. 13 and paragraphs IV(27) to (30).]

6.6.3.2 For a Butterworth filter of particular Order and Type number, the value of Q_d calculated by Eq. (9) for use in Eq. (10) shall be between 1.023 Q_d and 0.977 Q_d for Type 0-X filters, 1.059 Q_d and 0.944 Q_d for Type 1-X filters, and 1.100 Q_d and 0.900 Q_d for Type 2-X and Type 3-X filters. See Appendix C for values of A_d obtained by this procedure for Type 1-X and Type 2-X Order 3 Butterworth filters (see 6.7.1.2 for the significance of Sub-Type letter "X").

6.6.3.3 The design attenuation of a non-Butterworth filter, at any frequency in the transition bands less than f_1 and greater than f_2, shall be equal to or greater than the attenuation of a Butterworth filter of the same Order and Type designation, calculated with the lower value of the allowable range of Q_d given in 6.6.3.2.

6.7 Effective Bandwidth

6.7.1 Bandwidth Error Designators

For each filter in the set, the white noise power and sloping spectrum power passed by the filter relative to that which would be passed by an ideal filter with the exact midband frequency specified by Eq. (1) and Eq. (2) or Eq. (4) in 6.3 shall determine the filter Type classification which shall have two components. The first, an arabic numeral, shall be the primary designator as determined by the relative white noise bandwidth error of the filters. The second component, an English alphabet character, shall be determined by the relative bandwidth error for sloping spectra in accordance with the procedure specified below.

NOTE: See Appendix A for a discussion of the influence of Butterworth filter Order number on bandwidth error for various spectral slopes.

6.7.1.1 The Type number designator based on the bandwidth error for white noise shall be determined as specified in Table 3.

6.7.1.2 For any Type number, the Sub-Type alphabet character designator based on the composite bandwidth error (see 6.7.6) shall be determined as specified in Table 4.

TABLE 3. Criteria for selecting Type number.

White noise bandwidth error, millibel	Type number
= or < than 10	0
= or < than 25	1
= or < than 41	2 or 3*

*depends on passband ripple.

TABLE 4. Criteria for selecting Sub-Type letter.

Composite bandwidth error, millibel	Sub-Type letter
= or < than 13	AA
= or < than 25	A
= or < than 50	B
= or < than 100	C
greater than 100	D

6.7.2 Normalized Sloping Power Spectral Density

To determine the bandwidth error for a filter when the input signal has a nonzero spectrum slope (i.e., not white noise), an equation is needed to calculate the power spectral density, PSD, as a function of frequency.

6.7.2.1 For computation purposes, the continuous PSD is made nondimensional by normalizing with respect to the PSD at the exact midband frequency f_m. For the purpose of this standard, the normalized PSD shall be given by

$$[S_g(f/f_m)]/S_m = \{[(f/f_m)^g]C + 1\}/[(f/f_m)^g + C], \quad (11)$$

where $S_g(f/f_m)$ is the PSD of the signal at any frequency ratio f/f_m; g is the constant nondimensional slope of the normalized PSD function; S_m is the PSD at midband frequency $f=f_m$; and C is a nondimensional constant equal to 2000.

NOTES:

(1) If the normalized PSD in decibels is plotted versus the base 10 logarithm of the frequency ratio f/f_m, then slope g equals the slope in decibels per octave divided by 10 lg(2).

(2) The constant C is included in Eq. (11) to provide low-frequency and high-frequency asymptotes of constant normalized PSD to limit the calculated value of normalized wideband equivalent power (NWEP) passed by a bandpass filter when the input signal has a nonzero slope.

6.7.2.2 For octave-band filters, the NWEP transmitted by a practical filter shall be determined for spectral slopes of $g = -5.0$, 0.0, and +3.0.

6.7.2.3 For fractional-octave-band filters, the NWEP transmitted by practical filter shall be determined for values of $g = -12.0$, 0.0, and +10.0.

6.7.3 Normalized Equivalent Power of Noise Transmitted by an Ideal Filter

For a random noise signal having a spectrum slope g not equal to -1 (for $g = -1$, see Ref. A1), the NWEP, W_{ig}/W_m, transmitted by an ideal filter, with bandwidth designator b, shall be calculated from

$$W_{ig}/W_m = [2^{0.5(g+1)b} - 2^{-0.5(g+1)b}]/(g+1), \tag{12}$$

where W_{ig} is the equivalent power (EP) transmitted by an ideal filter for input signal of spectral slope g and W_m is the EP transmitted by the ideal filter at frequency $f=f_m$ and equal to the product $S_m f_m$.

NOTES:

(1) Since the transmission of an ideal filter is unity between bandedges f_1 and f_2 and is zero for all other frequencies, Eq. (12) may be obtained by the integral of $[S_g(f/f_m)]/S_m = (f/f_m)^g$ over the frequency range from f_1 and f_2.

(2) For octave-band filters, $W_{ig}/W_m = 0.707107$ for $g=0$ and 0.937500 for $g = +3$ or -5. For one-third-octave-band filters, $W_{ig}/W_m = 0.231563$ for $g=0$ and 0.298453 for $g = +10$ or -12.

6.7.4 Normalized Equivalent Power of Noise Transmitted by a Realizable Filter

6.7.4.1 For a random noise signal, the total NWEP transmitted by a realizable filter shall be computed from

$$\frac{W_{tg}}{W_m} = \int_0^\infty \left(\frac{S_g(f/f_m)}{S_m}\right)\left[\left|H_p\left(\frac{f}{f_m}\right)\right|^2\right] d\left(\frac{f}{f_m}\right), \tag{13}$$

where $[S_g(f/f_m)]/S_m$ is given by Eq. (11) and $|H_p(f/f_m|^2$ is the measured or calculated squared magnitude transfer function of the filter under consideration and is equal to $10^{-A/10}$, where A is the attenuation in decibels.

6.7.4.2 The minimum attenuation in the passband shall be the reference for determining the attenuation characteristic as a function of frequency ratio (see 6.6.1 and 8.1.1).

6.7.4.3 For analog filters, the Type and Sub-Type designation shall be determined for the attenuation characteristic corresponding to the worst-case combination of component tolerances. For a digital filter, the effect of the antialias filter shall be included. Where several different filters are combined to obtain one octave band, the largest composite error shall be used to determine the Sub-Type designation.

6.7.4.4 The integral in Eq. (13) may be evaluated by any suitable method; numerical integration is recommended. The range of values used for f/f_m during an evaluation shall extend below and above 1.0 such that additional contributions to the value of the summation do not change the calculated composite bandwidth error by more than 1 millibel to be consistent with the requirements of Table 4.

6.7.5 Determination of Filter Bandwidth Error

The bandwidth error, in millibels, shall be calculated from

$$\begin{aligned} E_g &= 1000\,\mathrm{lg}[(W_{tg}/W_m)/(W_{ig}/W_m] \\ &= 1000\,\mathrm{lg}(W_{tg}/W_{ig}). \end{aligned} \tag{14}$$

NOTE: The quantity $W_m = S_m f_m$ that is used as a normalizing equivalent power in Eqs. (12) and (13) appears in both the numerator and denominator of the ratio in the middle term of Eq. (14). Hence, the actual value of the product is irrelevant to the determination of the bandwidth error for any spectral slope g and may, if desired, be assigned an arbitrary value such as 1.0.

6.7.6 Composite Bandwidth Error

The composite bandwidth error E_c to determine the Sub-Type designation for octave-band and fractional-octave-band filters, shall be determined from

$$E_c = 3|E_{g=0}| + |E_{g=-5}| + |E_{g=+3}|, \tag{15a}$$

$$E_c = 3|E_{g=0}| + |E_{g=-12}| + |E_{g=+10}|, \tag{15b}$$

respectively. The result shall be rounded to the nearest millibel and the Sub-Type determined in accordance with Table 4.

6.7.7 Filter Designation

To meet the requirements of this standard, a filter set or equivalent filtering system shall be marked to include the applicable Order and Type designations. No filter set or equivalent shall be stated to be in accord with this standard unless its Order, Type, and Sub-Type designations are given. (Example: One-Third-Octave-Band Filter Set, Order 5, Type 0-A, Extended Range, per ANSI S1.11-1986).

6.8 Passband Uniformity

The peak-to-valley ripple within the passband of each filter in the set, whether by design choice or effect of component tolerance, shall not exceed 10 millibels for Type 0-X filters (see 7.2.2 for type 0-X digital filters), 25 millibels for Type 1-X filters, or 50 millibels for Type 2-X filters. Filters having 60 to 200 millibels peak-to-valley ripple shall be designated as Type 3-X filters.

6.9 Variation of Reference Passband Attenuation

The reference passband attenuation, Eq. (17), of any filter band in a set shall not differ from the reference passband attenuation of any other filter band in the set by more than 10 millibels for Type 0-X, 30 millibels for Type 1-X, 100 millibels for Type 2-X or 200 millibels for Type 3-X filters.

6.10 Removal of Filters From Circuit

If means are incorporated in the filter set to remove all filter bands from the circuit, the manufacturer shall explicitly state the frequency characteristics of the substituted broadband circuit. Over a frequency range extending 1 octave below and above the range of the filter set, the frequency response of the broadband circuit shall have uniform (flat) response over the operating range and shall not droop below the midband flat response by more than 1 decibel at frequencies 1 octave below and above the analysis range of the filter set.

6.11 Terminating Impedances

6.11.1 The input and output terminating impedances necessary to ensure proper operation of filter sets shall be purely resistive. The necessary terminating impedances shall be explicitly labeled on the filter set.

6.11.2 Active filters shall be buffered, if necessary, to ensure that their operation is substantially independent of the terminating impedances between which they are connected.

6.12 Maximum Input Signal

The manufacturer shall state the maximum root-mean-square (rms) input wideband white or pink noise voltage and the maximum midband sinusoidal rms input voltage at which each filter in the filter set will meet the performance requirements of this standard. Filter sets not dedicated to a specific instrument should be capable of accepting a sinusoidal rms input voltage of at least 1 volt.

6.13 Linearity

For any steady input signal and for any input signal level within the specified operating range of the filter set, the transfer gain (output-to-input ratio) of each filter in a set shall not vary by more than 10 millibels for Type 0-X and Type 1-X filters or 30 millibels for Type 2-X and Type 3-X filters.

6.14 Nonlinear or Harmonic Distortion

The manufacturer shall specify the total harmonic distortion in the output signal of a filter when a sinusoidal signal of frequency f_m at maximum rated voltage is applied to the input of any filter in the set. He shall also state how the distortion varies as the input voltage is reduced to zero.

6.15 Transient Response

When a sinusoidal signal of frequency f_m is suddenly applied to the input of a filter, the maximum of the envelope of the signal voltage appearing at the output shall not exceed the steady state output voltage by more than a factor of 1.26 (2 dB) and the duration, in milliseconds, of the ringing, defined as the time required for the output to settle to within 10 millibels of the steady state value, shall not exceed 2000 divided by the filter reference bandwidth in hertz [Eq. (7)].

6.16 Phase and Group Delay Response

See Appendix B for the phase and group delay characteristics of an exact reference-design Order 3 Butterworth filter. For filter designs different from the Order 3 Butterworth shown in Appendix B, the manufacturer shall include the design-goal normalized phase and group delay response of the filters in the Instruction Manual for the filter set.

6.17 Dynamic Range

The dynamic range of analog filters shall be not less than 80 decibels. If the dynamic range is less than 80 decibels, the manufacturer shall specify the value. (See 7.2.3 for minimum dynamic range requirement for digital filters.)

6.18 Analysis of Nonstationary Signals

6.18.1 With the filter set or filter system incorporated within a suitable measurement system that includes appropriate squaring and time-averaging capabilities (e.g., an integrating-averaging sound level meter or a fast-Fourier-transform analyzer), and with the voltage of the input signal constant at the manufacturer-specified maximum sinusoidal input voltage (see 6.12), for at least three filters in a filter set the difference between the time-period-average level L_c of a continuous sinusoidal signal and the time-period-average level L_b of a burst of a sinusoidal signal that starts and stops at a zero crossing shall not differ from the theoretical difference by more than 10, 30, 50, and 100 millibels for type 0, 1, 2, or 3 filters, respectively, but not to exceed the tolerance limits specified by the manufacturer for the overall instrument accuracy. The complete measurement system shall be specified by the manufacturer in the Instruction Manual.

6.18.2 The theoretical difference, in millibels, between the two time-period-average levels of a continuous sinusoidal signal and a sinusoidal tone burst of the same amplitude and frequency shall be determined from

$$L_c - L_b = 1000 \lg(T_{av}/T_b), \tag{16}$$

where T_{av} is the averaging time in seconds and T_b is the duration, in seconds, of the tone burst from the first to the last axis crossing.

6.18.3 The duration of the averaging time period T_{av} shall be at least 8 seconds for each test signal. For each filter tested, the length of the tone burst N_b shall be equal to 32 complete cycles at the nominal midband frequency in order to maintain a constant value of 7.41 for the product $N_b B_r$.

6.18.4 The requirement in 6.18.1 shall apply to tone bursts occurring at any time during the duration of the total averaging time period, without use of internal triggering.

6.18.5 For filters intended for the audio-frequency range (see 6.1.1), the test frequencies shall include 125, 1000, and 8000 hertz. Manufacturers of filter sets intended for applications to other frequency ranges shall specify the test frequencies in the Instruction Manual.

6.19 Sensitivity to External Conditions

6.19.1 Temperature

The attenuation characteristics of the filter set shall conform to the applicable requirements of this standard over the temperature range from $-10\,°C$ to $+50\,°C$. If the influence of changes in ambient temperature causes the tolerance limits on attenuation or effective bandwidth to be exceeded, conformance with this standard may be established by determining the influence of the environmental change by measurement to a precision of 10 millibels and making the information available to the user of the filter set. The manufacturer shall indicate the ambient temperature limits and exposure time, which, if exceeded, may cause permanent damage to the filter set.

6.19.2 Humidity

The manufacturer shall state the range of relative humidity between which the filters will function correctly together with the corresponding permissible exposure periods within the rated operating temperature range. At any temperature between $-10\,°C$ and $+50\,°C$, the filters should operate indefinitely within the various tolerance limits over a relative humidity range extending from 10% to 90% without condensation. If operation within the tolerance limits is not possible over that combined range of temperature and humidity, the manufacturer shall state the amount and kind of degradation as a function of relative humidity, temperature, and exposure time.

6.19.3 Electromagnetic Fields

The effects of electromagnetic fields on the operation of the filters shall be reduced as far as practical. The filters shall be tested in a magnetic field of strength 80 amperes per meter at 50 and 60 hertz and for the orientation which gives maximum response to the radiation. The manufacturer shall state the rms output voltage. For filters designed for use with a specific type of instrument, e.g., a sound level meter, the output produced by the test field shall be given in terms of the output indication for each band affected.

6.19.4 Vibration

Portable filter sets should be designed and built to withstand transportation shock and vibration accelerations.

6.19.5 Sound

The rms voltage at the output of each filter in a set of filters for the audio-frequency range resulting from exposure to a sinusoidal sound field with a level of 110 decibels at the position of the filter set before its insertion and swept from 31.5 hertz to 8 kilohertz at a rate not greater than 0.1 octave per second shall be at least 80 decibels below the output voltage produced by a sinusoidal signal of maximum input voltage specified for the filter in accordance with 6.12.

7 SAMPLED DATA SYSTEMS

7.1 Analog Filters

Analog filters that are sampled in time are subject to all of the requirements of Sec. 6. Appendix D describes the general characteristics of one implementation of a sampled analog filter.

7.2 Digital Filters

7.2.1 Except as specified below, digital filters and all numerically synthesized fractional-octave-band filters shall conform with all requirements of Sec. 6.

7.2.2 Antialias filters, analog and digital, shall provide attenuation such that the sum of the guard filter attenuation plus the level of possible aliased spectra shall be equal to or greater than the dynamic range of the digital filter. Passband ripple in the guard filter may contribute to the passband ripple of the bandpass filter. For Type 0-X filters, the total passband ripple specified in 6.8 may be increased to 15 millibels owing to this contribution.

7.2.3 The cutoff attenuation in the upper transition band of a digital bandpass filter shall not be increased by the attenuation of an antialias filter, either analog or digital, by an amount greater than that produced by placing the 3-dB-down frequency of such filter at the upper 40-dB attenuation frequency of the digital filter.

7.2.4 The minimum dynamic range of a digital filter shall be not less than 72 decibels.

7.2.5 The sampling rate of the analog-to-digital converter shall be sufficiently high not to affect signal components which are within the dynamic range and analysis band of the highest frequency filter.

7.2.6 The bilinear z transform [see Par. IV(31) in Appendix E] is the preferred method for realizing a digital bandpass filter design from an analog prototype bandpass filter. If the bilinear transformation is used, then the inverse transformation (prewarping) shall be applied to the midband and bandedge frequencies of the prototype analog filter to maintain the applicable relationships specified in Sec. 6.

7.2.7 Deadband effect or limit cycle resulting from truncation error in recursive filter designs shall be ameliorated by appropriate means. See Par. V(9) in Appendix E.

8 METHOD OF TEST

Tests described in this section apply to analog, digital, or numerically synthesized filters, as appropriate. When the digital data are available from a digital filter, before squaring and time averaging, all of the tests specified for an analog filter may be made by the addition of a digital-to-analog converter and a post filter. In all cases, the manufacturer shall include in the Instruction Manual the procedures by which a user may verify the proper operation of the filters.

8.1 Attenuation Characteristic

8.1.1 The attenuation characteristic of each filter band and, if available, the broadband circuit of Sec. 6.10 shall be measured according to the following basic procedure and the general arrangement of Fig. 1. The input terminals of the filter set shall be connected to a variable frequency sine-wave generator of zero equivalent source impedance (if necessary) in series with an input terminating impedance of the value specified by the manufacturer. The signal generator output voltage V_1 shall be measured by a voltmeter with adequate accuracy and frequency range and set to the manufacturer specified maximum filter input voltage. The output terminals of the filter shall be connected to an output terminating impedance of value specified by the manufacturer, and the output voltage V_2 across this impedance measured with a second similar voltmeter. The ratio V_1/V_2 shall be determined at appropriate frequencies throughout the frequency range necessary to demonstrate compliance with this standard. For reference, note the maximum value of V_2 in the passband and compute A_{ref} from Eq. (17).

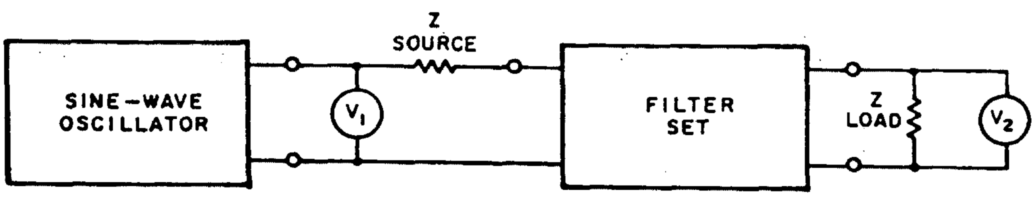

FIG. 1. Schematic of two-voltmeter test arrangement.

$$A_{\text{ref}} = 20\,\lg(V_1/V_2)_{\min} \tag{17}$$

and the filter transmission loss A at any frequency is

$$A = 20\,\lg(V_1/V_2) - A_{\text{ref}}. \tag{18}$$

For a constant input voltage V_1, Eq. (18) reduces to

$$A = 20\,\lg(V_{\text{ref}}/V_2). \tag{19}$$

NOTE: The purpose of 8.1.1 is not to specify the only way in which filter characteristics and performance may be determined, but a basic way which does not require special purpose instruments. For the user of a sound level meter and fractional-octave-band analyzer who wishes to check his filter set, the obvious choice for the output measuring device is the sound level meter itself. For best accuracy, a digital display is preferred.

8.1.2 When the attenuation is being measured at frequencies below the lower bandedge frequency, a suitable technique shall be employed to remove the effects of oscillator harmonics from the apparent response of the filter. A tuned voltmeter (wave analyzer) shall not be used to remove those effects, since such a device at the output of the filter would simultaneously remove any distortion or noise introduced by the filter set, which properly should be ascribed to analysis error of the filter set (see 6.14).

8.1.3 In establishing compliance with this standard, the attenuation characteristic of a filter shall be measured using the maximum input signal voltage specified by the manufacturer according to 6.12 and also at voltages that are smaller by factors of 0.32 and 0.032 (i.e., 10 decibels and 30 decibels below the maximum input level).

8.2 Dynamic Range

8.2.1 Analog Filters

With the filter set properly terminated at its input and output (see 6.11), and no other signal applied, measure the level of the time-averaged squared wideband noise voltage appearing at the output of each filter in the set. The difference between the maximum rated output level corresponding to the maximum rated sinusoidal input (see 6.12) and the noise level is the dynamic range. The frequency range for the instrument used to measure the wideband output noise voltage shall be the same as for the measuring device with which the filter set would normally be used, e.g., the electronic circuits of a sound level meter.

8.2.2 Digital Filters

The dynamic range of digital filters shall be determined from the overall system response from analog input to digital or analog output. The internal noise floor of the filter shall be determined by applying a full scale sinusoidal signal at nominal midband frequency to one filter and measuring the level of the time-averaged squared output signal in another filter band sufficiently removed in frequency so that its attenuation at the frequency of the sinusoid is equal to or greater than the expected dynamic range. The dynamic range, in decibels, is the difference in level between full scale input voltage and the noise floor.

8.3 Phase-Response Characteristic (Optional)

The typical phase response of a filter may be measured in the same manner as the attenuation characteristic. A phase meter is preferred for these measurements, although an x–y oscilloscope displaying a Lissajous figure may be used. The reference phase shall be that of the signal generator output. All requirements on signal characteristics stated in Sec. 8.1 above apply to phase measurements. Since the phase shift of bandpass filters is zero at the midband frequency and since the rate of change of phase with frequency is large at this point, a sensitive means for determining the actual midband frequency is provided by phase measurement (see Appendix B). The phase response may be determined at the same time the attenuation characteristic is measured.

APPENDIX A: PERFORMANCE OF BUTTERWORTH BANDPASS FILTERS

[This Appendix is not a part of American National Standard Specification for Octave-Band and Fractional-Octave Band Analog and Digital Filters, S1.11-1986, but is included for information purposes only.]

A1. Traditionally, filter bandwidths have been expressed in terms of the half-power or the 3 dB-down frequencies of the filter. However, when random noise is analyzed, the power which is transmitted by a bandpass filter depends not only on the filter bandwidth but also on the steepness of the attenuation in the transition band and the slope of the spectrum being analyzed. For the 1966 version of this standard, the Writing Group examined many representative sound spectra and found spectrum level slopes ranging from +6 to −21 decibels per octave. Since then, even steeper spectrum level slopes have been found to occur in practice.

A2. The bandwidth error for a number of filter attenuation characteristics and spectrum level slopes was first computed for an ideal filter (see Ref. A1) then for a filter having a maximally flat (Butterworth) attenuation characteristic (see Ref. A2). In addition, for Butterworth filters, the bandedge attenuation required to give zero bandwidth error was computed for white and pink noise signals. The results of the latter two investigations are given in Table A1.

A3. Theoretical bandwidth error curves for fractional-octave-band filters are symmetric about the pink noise spectrum-level slope of −3 dB per octave. If a signal that has a wideband pink noise power spectral density is applied to a series of contiguous constant-percentage-bandwidth filters, the filtered spectrum has the same power in each frequency band. Table A1 reveals that a very small difference in filter bandwidth is involved to correct for zero bandwidth error for pink or white noise signals. In addition, it was found that the analytical solution for zero error for a white noise signal gave the same bandedge attenuation for all bandwidths of the same filter design complexity. However, the zero error adjustment for a pink noise signal produced a slightly different bandedge attenuation for each filter bandwidth of a given filter design complexity. For one-third-octave-band filters, the two results differed by only 1 millibel. Owing to the greater simplicity in carrying out the integration to determine the effective bandwidth for white noise and the slight shifting in the axis of symmetry for zero bandwidth error, it was decided to use white noise as the referent for the Type number specification in 6.7.1.1 of the standard.

A4. By transforming the Butterworth bandpass attenuation equation to the equivalent lowpass coordinate system, the white noise power in the passband may be calculated directly by integration. The ratio of the effective bandwidth to the 3-dB bandwidth is given by $(\pi/2n)/\sin(\pi/2n)$; see 6.5.4.

TABLE A1. Performance of Butterworth bandpass filters: n = filter Order or number of pole-pairs; γ = spectrum level slope in dB per octave; H_1 = bandwidth error for filter 3 dB down at bandedge frequencies; H_2 = bandwidth error for filter adjusted for zero bandwidth error for pink noise; A_1 = bandedge attenuation required to give zero bandwidth error for pink noise; A_2 = bandedge attenuation required to give zero bandwidth error for white noise.

	n	γ dB/octave		H_1 mB	H_2 mB	A_1 dB	A_2 dB
For one-octave bandwidth	2		−3	37	0	3.84	
		−6	0	46	7		4.02
		−9	+3	78	32		
		−12	+6	155	92		
	3		−3	16	0	3.56	
		−6	0	20	3		3.65
		−9	+3	31	11		
		−12	+6	50	26		
	4		−3	8	0	3.42	
		−6	0	11	1.4		3.48
		−9	+3	17	5.6		
		−12	+6	26	12.8		
For one-third-octave bandwidth	2		−3	44	0	4.03	
		−6	0	46	1.1		4.02
		−9	+3	51	4.7		
		−12	+6	61	12		
	3		−3	19	0	3.64	
		−6	0	20	0.3		3.65
		−9	+3	21	1.4		
		−12	+6	24	3.2		
	4		−3	10	0	3.48	
		−6	0	11	0.1		3.48
		−9	+3	12	0.7		
		−12	+6	13	1.6		

A5. A closed form integration of Eq. (13) for positive, even and unbounded spectrum slopes g was achieved by changes of variable (Ref. A3). However, removal of the asymptotes introduced by the constant C in Eq. (11) may cause the resulting calculated values of E_g to be slightly larger than those found by numerical integration of Eq. (13) in conjunction with use of Eq. (11) for the normalized spectrum slope function. The normalized equivalent noise power passed by an analog Butterworth fractional-octave-band filter is given by

$$\frac{W_{tg}}{W_{ig}} = \frac{1}{Q_d} \sum_{k=0}^{g/2} [A_{2k}(g)] \frac{(1/2Q_d)^{2k}(\pi/2n)}{\sin[(2k+1)\pi/2n]}, \tag{A1}$$

TABLE A2. Bandwidth error for analog Butterworth bandpass filters. {Note: Spectral slope in dB per octave equals $g[10\lg(2)]$}.

Bandwidth	Order	Spectral slope, g	Ratio of limit Q to design Q_d 0.910	0.944	0.977	1.000	1.023	1.059	1.100
			E_g bandwidth error in millibels						
Octave									
		−5	113.8	85.8	60.1	43.1	26.7	2.1	−24.2
		−3	54.8	36.4	19.2	7.7	−3.4	−20.4	−38.9
	3	0	40.9	25.0	10.1	0	−9.8	−24.8	−41.3
		1	55.0	36.5	19.3	7.8	−3.3	−20.3	−38.8
		3	115.5	87.3	61.5	44.3	27.8	3.1	−23.3
		Sub-Type	D	D	D	C	C	C	D
One-third-octave									
		−12	127.2	97.5	70.5	52.6	35.5	10.1	−17.0
		−6	56.0	37.4	20.2	8.6	−2.6	−19.6	−38.1
	3	0	40.9	25.0	10.1	0	−9.8	−24.8	−41.3
		4	56.1	37.5	20.2	8.7	−2.5	−19.5	−38.1
		10	133.7	103.1	75.3	57.0	39.5	13.5	−14.1
		Sub-Type	D	D	D	D	D	C	D
		−12		62.6	38.2	22.0	6.4	−16.7	
		−6		32.4	15.5	4.1	−6.9	−23.6	
	4	0		25.0	10.1	0	−9.8	−24.8	
		4		32.4	15.5	4.1	−6.8	−23.6	
		10		62.8	38.4	22.2	6.6	−16.5	
		Sub-Type		D	D	B	B	D	
		−12		51.9	28.3	12.6	−2.4	−25.0	
		−6		30.5	13.7	2.4	−8.5	−25.1	
	5	0		25.0	10.1	0	−9.8	−24.8	
		4		30.5	13.7	2.5	−8.4	−25.1	
		10		52.0	28.4	12.7	−2.3	−24.9	
		Sub-Type		D	C	A	B	D	
		−12			23.7	8.3	−6.5		
		−6			12.9	1.6	−9.2		
	6	0			10.1	0	−9.8		
		4			12.9	1.6	−9.2		
		10			23.8	8.3	−6.5		
		Sub-Type			C	A	B		
		−12			21.2	5.6	−8.8		
		−6			12.4	1.1	−9.7		
	7	0			10.1	0	−9.8		
		4			12.4	1.2	−9.7		
		10			21.3	5.9	−8.8		
		Sub-Type			C	AA	B		

where

$$A_{2k}(g) = [(g/2)+k]!\ 2^{2k}/\{[(g/2)-k]!\ 2k!\} \quad \text{(A2)}$$

for g even and = or < zero. See 6.5.4 for values of Q_d.

A6. Table A2 presents the bandwidth errors for the limit conditions specified for Butterworth bandwidth quotient in 6.6.3.2 for Order 3 octave-band and for one-third-octave-band filters of Order 3 through 7. The resulting Sub-Type designation in accordance with 6.7.6 and Table 4 is also given for each case.

A7. Close empirical approximations for the bandwidth error of Butterworth filters using the values of Q_d given by Eq. (9) were determined by curve fitting procedures. For both octave-band and one-third-octave-band filters, the approximate bandwidth error is given by

$$E_g \simeq m^2(g+2)^2 \quad \text{(A3)}$$

for negative values of spectrum slope g.

For octave-band filters of Order n = or > 2,

$$m \simeq 4.5/\{n[1+(1/n)+(2/n^2)]\}. \quad \text{(A4)}$$

For one-third-octave-band filters of Order n = or > 3,

$$m \simeq 1.5/\{n[1+(1/n)+(1/n^2)]\}. \quad \text{(A5)}$$

A8. References

[A1] L. W. Sepmeyer, "Bandwidth error of symmetrical bandpass filters used for analysis of noise and vibration," J. Acoust. Soc. Am. **34**, 1653 (1962).

[A2] L. W. Sepmeyer, "On bandwidth error of Butterworth bandpass filters," J. Acoust. Soc. Am. **35**, 404 (1963).

[A3] J. Kalb (private communication).

APPENDIX B: PHASE AND GROUP DELAY RESPONSE OF BUTTERWORTH BANDPASS FILTERS

[This Appendix is not a part of American National Standard Specification for Octave-Band and Fractional-Octave Band Analog and Digital Filters, S1.11-1986, but is included for information purposes only.]

B1. The amplitude and phase response of a minimum-phase network are related by the Hilbert trans-

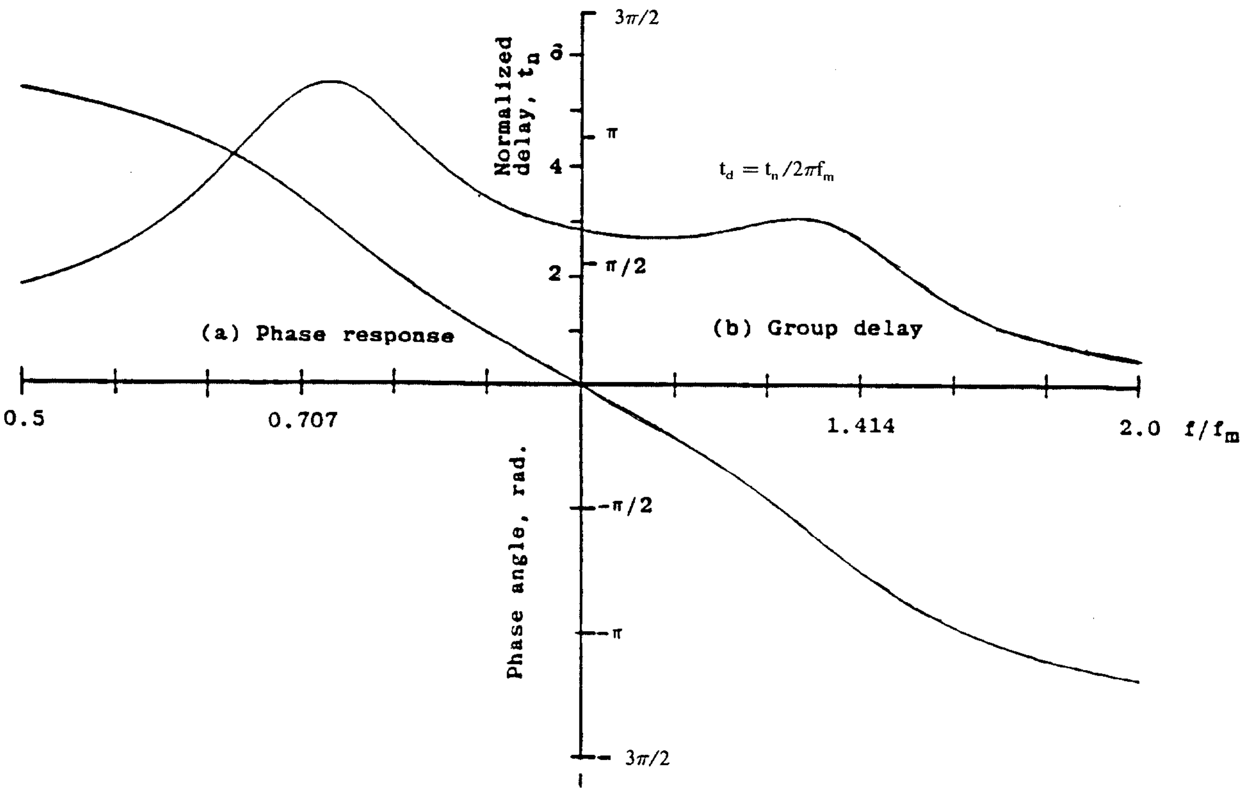

FIG. B1. Phase response and normalized group delay for an Order 3 reference Butterworth octave-band filter.

form. By "minimum phase" is meant that for a given attenuation characteristic the network produces the least possible shift in phase of the output signal relative to the input signal as a function of frequency. That is, the network contains no allpass sections. The phase shift of a network is given by the arctangent of the ratio of the imaginary part to the real part of its complex transfer function.

For an Order 3 (third-order) Butterworth bandpass filter, the normalized complex transfer function, $H(jF)$, is given by

$$H(jF) = (1 - 2Q^2F^2) + j(2QF - Q^3F^3), \quad \text{(B1)}$$

where $F = (f/f_m - f_m/f)$ and Q is the filter bandwidth quotient. Note that at $f = f_m$, $F = 0$ and $H(jF) = 1$.

Attenuation A is given by

$$A = 10 \lg|H(jF)|^2$$
$$= 10 \lg\{[\mathrm{Re}H(jF)]^2 + [\mathrm{Im}H(jF)]^2\}, \quad \text{(B2)}$$

and the phase by arctan $[\mathrm{Im}(H)/\mathrm{Re}(H)]$. Phase response plots for octave-band and one-third-octave-band Order 3 filters are shown in Figs. B1(a) and B2(a). Note that the expansion of Eq. (B1) in Eq. (B2) leads directly to Eq. (10) for $n = 3$.

B2. The group delay of a filter is given by the derivative of the phase response with respect to the angular frequency. Figures B1(b) and B2(b) show the group delays corresponding to the phase responses in

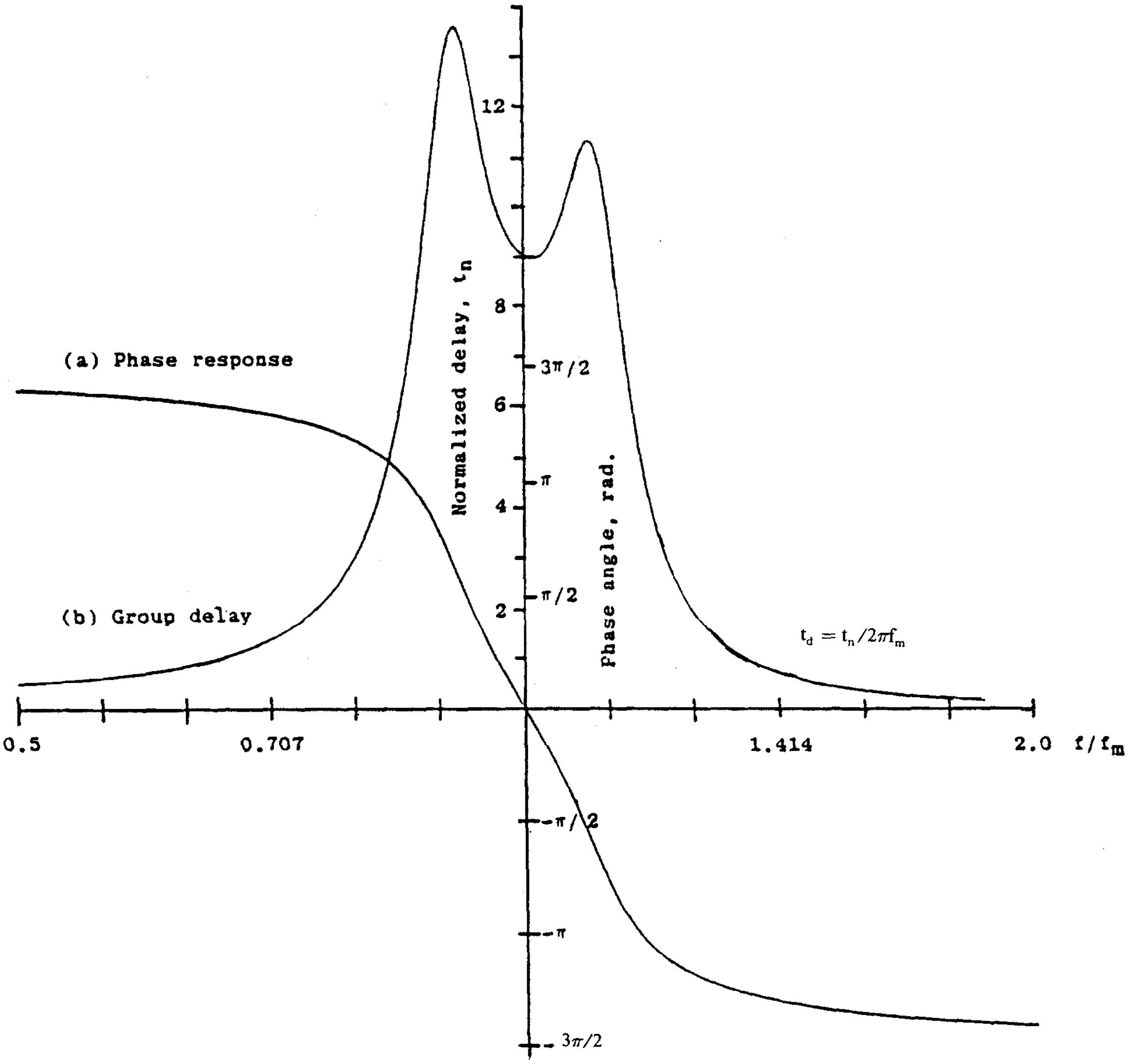

FIG. B2. Phase response and normalized group delay for an Order 3 reference Butterworth one-third-octave-band filter.

Figs. B1(a) and B2(a). Note that the group delay is not constant in the passband of the filter, nor is it symmetric with respect to midband frequency. In Figs. B1(a) and B1(b) the group delays are presented as normalized time delays t_n, which are related to actual group delay time t_d by $t_n = (t_d)(2\pi f_m)$.

B3. For some applications, it may be desirable to have uniform delay in the filter passband. In those cases, allpass networks may be added to the filter in order to adjust the phase response so that sufficiently uniform delay is achieved. For most acoustical measurements a uniform delay is not important since the time-averaged squared-magnitude of the filter output is measured.

APPENDIX C: ATTENUATION FOR ORDER 3 OCTAVE-BAND AND ONE-THIRD-OCTAVE-BAND FILTERS

[This Appendix is not a part of American National Standard Specification for Octave-Band and Fractional-Octave Band Analog and Digital Filters, S1.11-1986, but is included for information purposes only .]

C1. Table C1 presents values of attenuation calculated by use of Eq. (10) for selected values of f/f_m or f_m/f. The calculations were done for Order 3 ($n = 3$) octave-band ($b = 1$) and one-third-octave-band ($b = 1/3$) Butterworth filters. For each filter, five values of bandwidth quotient were used, first, for Q_d as given by Eq. (9), then, for the extremal values specified in 6.6.3.2 for Type 1-X and Type 2-X filters.

TABLE C1. Attenuation for Order 3 Type 1-X and Type 2-X octave-band and one-third-octave-band Butterworth filters.

	Octave-band attenuation, dB						One-third-octave-band attenuation, dB				
f/f_m or f_m/f		Type 1-X		Type 2-X		f/f_m or f_m/f		Type 1-X		Type 2-X	
	Q_d	Q_{max}	Q_{min}	Q_{max}	Q_{min}		Q_d	Q_{max}	Q_{min}	Q_{max}	Q_{min}
1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1.20	0.11	0.16	0.08	0.19	0.06	1.08	0.47	0.65	0.33	0.80	0.26
1.26	0.45	0.62	0.31	0.76	0.24	1.10	1.50	2.01	1.09	2.39	0.86
1.41	3.52	4.43	2.70	5.07	2.21	1.12	3.37	4.26	2.58	4.89	2.11
1.60	10.03	11.42	8.60	12.32	7.65	1.16	8.44	9.77	7.10	10.64	6.21
1.80	16.04	17.53	14.48	18.48	13.39	1.18	10.99	12.41	9.53	13.32	8.54
2.00	20.83	22.34	19.24	23.30	18.12	1.22	15.59	17.07	14.03	18.02	12.95
2.25	25.64	27.16	24.03	28.12	22.91	1.26	19.49	21.00	17.90	21.96	16.79
2.50	29.57	31.09	27.96	32.05	26.83	1.33	25.06	26.57	23.45	27.53	22.32
3.15	37.36	38.88	35.75	39.85	34.62	1.41	30.06	31.58	28.45	32.54	27.32
3.35	39.30	40.82	37.69	41.79	36.56	1.50	34.57	36.09	32.96	37.06	31.83
4.00	44.67	46.19	43.06	47.16	41.93	1.65	40.44	41.96	38.83	42.93	37.70
5.00	51.11	52.63	49.49	53.59	48.36	1.80	45.02	46.54	43.41	47.50	42.27
5.60	54.28	55.80	52.67	56.76	51.53	2.00	49.89	51.41	48.27	52.37	47.14
6.30	57.53	59.05	55.92	60.01	54.78	2.25	54.72	56.24	53.11	57.20	51.97
7.10	60.79	62.30	59.17	63.27	58.04	2.60	60.05	61.57	58.44	62.53	57.30
8.00	64.01	65.53	62.40	66.49	61.26	3.00	64.88	66.40	63.27	67.36	62.13
10.00	69.97	71.49	68.36	72.45	67.23	3.50	69.75	71.26	68.13	72.23	67.00
12.50	75.88	77.40	74.27	78.36	73.13	4.25	75.54	77.06	73.93	78.02	72.80
16.00	82.38	83.90	80.77	84.86	79.63	5.00	80.20	81.71	78.58	82.68	77.45

TABLE C2. Frequency ratios for Order 3 Type 1-X and Type 2-X octave-band and one-third-octave-band Butterworth filters.

	Octave-band f/f_m or f_m/f						One-third-octave-band f/f_m or f_m/f				
		Type 1-X		Type 2-X				Type 1-X		Type 2-X	
A, dB	Q_d	Q_{max}	Q_{min}	Q_{max}	Q_{min}	A, dB	Q_d	Q_{max}	Q_{min}	Q_{max}	Q_{min}
0.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00
1.00	1.31	1.29	1.33	1.27	1.34	1.00	1.09	1.09	1.10	1.08	1.10
2.00	1.36	1.33	1.38	1.32	1.40	2.00	1.11	1.10	1.11	1.10	1.12
3.65	1.41	1.39	1.44	1.37	1.47	3.65	1.12	1.12	1.13	1.11	1.14
5.00	1.46	1.43	1.49	1.41	1.51	5.00	1.13	1.13	1.14	1.12	1.15
10.00	1.60	1.56	1.64	1.54	1.68	10.00	1.17	1.16	1.18	1.16	1.19
15.00	1.76	1.71	1.82	1.68	1.86	15.00	1.21	1.20	1.23	1.19	1.24
20.00	1.96	1.90	2.04	1.86	2.09	20.00	1.27	1.25	1.28	1.24	1.30
25.00	2.21	2.13	2.31	2.08	2.38	25.00	1.33	1.31	1.35	1.30	1.37
30.00	2.53	2.43	2.65	2.36	2.74	30.00	1.41	1.38	1.44	1.37	1.46
35.00	2.93	2.80	3.08	2.72	3.19	35.00	1.51	1.48	1.55	1.46	1.58
40.00	3.43	3.26	3.61	3.17	3.75	40.00	1.64	1.60	1.69	1.57	1.72
45.00	4.04	3.84	4.27	3.72	4.44	45.00	1.80	1.75	1.86	1.71	1.91
50.00	4.81	4.56	5.09	4.41	5.30	50.00	2.01	1.94	2.08	1.90	2.14
55.00	5.75	5.44	6.09	5.26	6.35	55.00	2.27	2.18	2.36	2.13	2.44
60.00	6.90	6.52	7.32	6.30	7.63	60.00	2.60	2.49	2.72	2.42	2.81
65.00	8.30	7.85	8.82	7.57	9.20	65.00	3.01	2.88	3.17	2.79	3.28
70.00	10.01	9.46	10.64	9.12	11.10	70.00	3.53	3.36	3.72	3.26	3.87
75.00	12.09	11.42	12.85	11.01	13.42	75.00	4.17	3.96	4.41	3.84	4.59
80.00	14.62	13.80	15.54	13.30	16.23	80.00	4.97	4.71	5.26	4.55	5.48

C2. Table C2 is similar to Table C1 except here the independent variable is the frequency ratio. By following the procedures stated in Sec. 8, either Table C1 or C2 may be used to determine if the attenuation characteristic of any Order 3 filter meets the requirements in 6.6 of this standard.

APPENDIX D: SWITCHED CAPACITOR FILTERS

[This Appendix is not a part of American National Standard Specification for Octave-Band and Fractional-Octave Band Analog and Digital Filters, S1.11-1986, but is included for information purposes only.]

The name "Switched Capacitor Filter" applies to those active filters in which resistors are replaced by capacitors whose state of charge is periodically sampled, thereby simulating the function of a resistor. By using that technique, the attenuation characteristic of a filter is determined by capacitor ratios thus permitting the production of filters that are completely integrated (i.e., entirely contained on a tiny silicon chip) and requiring no external components. Since the filter midband frequency is a function of the switching frequency (or clock rate) applied to the filter, a large range of midband frequencies may be realized with one filter chip and an adjustable clock generator.

APPENDIX E

[This Appendix is not a part of American National Standard Specification for Octave-Band and Fractional-Octave Band Analog and Digital Filters, S1.11-1986, but is included for information purposes only.]

Terminology in Digital Signal Processing

LAWRENCE R. RABINER, JAMES W. COOLEY, HOWARD D. HELMS, LELAND B. JACKSON, JAMES F. KAISER, CHARLES M. RADER, RONALD W. SCHAFER, KENNETH STEIGLITZ, and CLIFFORD J. WEINSTEIN

Abstract—The committee on Digital Signal Processing of the IEEE Group on Audio and Electroacoustics has undertaken the project of recommending terminology for use in papers and texts on digital signal processing. The reasons for this project are twofold. First, the meanings of many terms that are commonly used differ from one author to another. Second, there are many terms that one would like to have defined for which no standard term currently exists. It is the purpose of this paper to propose terminology which we feel is self-consistent, and which is in reasonably good agreement with current practices. An alphabetic index of terms is included at the end of the paper.

Introduction

As an aid to classifying the different types of terms to be defined, we have placed each term in one of the following groups:

1) Introductory Terms—General Definitions
2) Discrete Systems—Block Diagram Terminology
3) Relations Between Discrete and Continuous Signals
4) Theory and Design of Digital Filters
5) Finite Word Length Effects—A/D, D/A Conversion
6) Discrete Fourier Transforms and the FFT
7) Discrete Convolution and Spectrum Analysis.

In the above mentioned sections of this paper we will be discussing terminology related to the processing of one-dimensional signals. For convenience, we will assume that this dimension is time—although the definitions apply equally well to any single dimension.

I. Introductory Terms—General Definitions

1) In discussing waveform processing problems, the distinctions analog versus digital and continuous time versus discrete time are often made. Although they are often used interchangeably, different meanings should be attributed to the two sets of terms.

2) The term *analog* generally describes a waveform that is continuous in time (or any other appropriate independent variable) and that belongs to a class that can take on a continuous range of amplitude values. Examples of *analog waveforms* or *analog signals* are those derived from acoustic sources. Such signals are represented mathematically as functions of a continuous variable. The functions sin (ωt) and the step function $au_{-1}(t)$ are examples of common mathematical functions that could describe "analog signals." The use of the term "analog" in this context appears to stem from the field of analog computation, where a current or voltage waveform serves as a physical analog of some variable in a differential equation.

3) The term *continuous time* implies that only the independent variable necessarily takes on a continuous range of values. In theory the amplitude may, but need not, be restricted to a finite or countable infinite set of values (i.e., the amplitude may be quantized). Therefore, analog waveforms are continuous-time waveforms with continuous amplitude. In practice, however, "continuous-time waveforms" and "analog waveforms" are equivalent. Since most signal processing problems have nothing to do with analogs as such, the use of the term analog waveform is often ambiguous at the least and may in fact be misleading. Thus, the term continuous-time waveform is preferable.

4) *Discrete time* implies that time (the independent variable) is quantized. That is, discrete-time signals are defined only for discrete values of the independent variable. Such signals are represented mathematically as *sequences* of numbers. Those discrete-time signals that take on a continuum of values are referred to as *sampled-data* signals.

5) The term *digital* implies that both time and amplitude are quantized. Thus a *digital system* is one in which signals are represented as sequences of numbers which take on only a finite set of values. Thus one uses *digital* when discussing actual physical realizations (as hardware or programs) of discrete-time signal processing systems, whereas the term *discrete time* is a better modifier when considering mathematical abstractions of such systems in which the effects of amplitude quantization are ignored. A *digital signal* or *digital waveform* is a sequence produced, for example, by digital circuitry or by an analog-to-digital converter which is sampling a continuous-time waveform. In digital signal processing these terms are commonly shortened to *signal* or *waveform*. Sometimes the term signal is restricted to being a desirable component of a sequence instead of being used interchangeably with waveform. *Noise* is either defined as a) an undesirable component of a sequence, or b) a sequence of random variables.

6) (*Digital*) *simulation* is the exact or approximate representation of a given system (discrete or contin-

Manuscript received August 1, 1972.
L. R. Rabiner, J. F. Kaiser, and R. W. Schafer are with Bell Telephone Laboratories, Inc., Murray Hill, N. J. 07974.
J. W. Cooley is with the IBM T. J. Watson Research Center, Yorktown Heights, N. Y. 10598.
H. D. Helms is with Bell Telephone Laboratories, Inc., Whippany, N. J. 07981.
L. B. Jackson is with Rockland Systems Corporation, West Nyack, N. Y. 10994.
C. M. Rader and C. J. Weinstein are with M.I.T. Lincoln Laboratory, Lexington, Mass. 02173. (Operated with support from the U. S. Air Force.)
K. Steiglitz is with Princeton University, Princeton, N. J. 08540.

uous) called the source system, by a (digital) system called the *object system.*

7) *Next-state simulation* is a method of digital simulation whereby the values of the digital system signals are represented by nodes in a block diagram representation. Usually, there is a close correlation between blocks in the object system and elements of the source system. The method entails ordering the calculations in the digital system so that all the inputs to each block at a given sample time are computed before the output is computed.

8) A *real-time process* is one for which, on the average, the computing associated with each sampling interval can be completed in a time less than or equal to the sampling interval. A program running in 100 times real time requires 100 times as long to process the same number of samples; i.e., it is 100 times too slow for real time operation. A program ten times as fast as it needs to be could be said to run in 1/10 real time. Obviously, the extra speed can only be used if other computing can be done in the interstices, or if the complete sequences to be processed have been stored beforehand.

9) *Throughput rate* is the total rate at which digital information is processed by a discrete-time system, measured in bits per second or samples per second. In a multiplexed system, where several signals are processed, we may refer to the *throughput rate per signal*, measured in bits per second per signal or samples per second per signal. Thus, a multiplexed system which processes 10 signals, each at 1000 bits/s, has a throughput rate of 10 000 bits/s, and a throughput rate/signal of 1000 bits/s/signal.

10) A *multirate system* is a discrete-time system in which there are signals sampled at different intervals which are usually integer multiples of some basic or fundamental interval.

II. Discrete Systems

1) The *z transform* plays a role in discrete-time system theory analogous to that of the Laplace transform in continuous-time system theory. Two view points regarding the z transform are common. One is based on what may be termed the one-sided z transform, which is defined as

$$X(z) = \sum_{n=0}^{\infty} x(n)z^{-n} \tag{1}$$

regardless of the value of $x(n)$ for $n<0$.[1] One application of the *one-sided z transform* is in the solution of linear difference equations with constant coefficients. Solutions are obtained for the interval $0 \leq n < \infty$ subject to prescribed initial conditions. These solutions are obtained with the aid of the equations

[1] The notation $x(n)$ is used rather than x_n or $x(nT)$ to denote a sequence because of the ease of handling complicated indices, e.g., $x(N-1/2)$.

$$y(n) \equiv x(n-m), \qquad m > 0 \tag{2}$$

$$Y(z) = z^{-m}\left[X(z) + \sum_{i=1}^{m} x(-i)z^{i}\right]. \tag{3}$$

In (3), $X(z)$ appears multiplied by z^{-m}. This result for $m=1$ accounts for the fact that z^{-1} is often termed the *unit delay operator*, since a delay of the sequence by one sample is equivalent to multiplication of the z transform by z^{-1}. (Similarly, z is often called the *unit advance* operator.)

2) In many cases, sequences are defined over both positive and negative values of n. In such cases, a somewhat more general point of view is called for. In general, the z transform is written as

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}. \tag{4}$$

It should be noted that a common usage is to call (1) simply the z transform, and (4) the *two-sided z transform.* Since (4) is most general, it would seem preferable to refer to (4) as the z transform, and the special case, (1), as the *one-sided z transform.*

3) It is possible to think of the z transform as simply a formal series whose properties can be tabulated, and which never need be summed. However, it is generally preferable to realize that if certain convergence conditions are met, both (1) and (4) are Laurent series in the complex variable z. As such, all the properties of the Laurent series apply. For example, if the series in (1) converges, it must converge in a region $|z| > R_+$. If the series of (4) converges, it must converge in an annular region $R_+ < |z| < R_-$, where R_+ may be zero and R_- may be infinity. The coefficients of a Laurent series are determined by an integral relationship. In the context of the z transform, this relation is

$$x(n) = \frac{1}{2\pi j}\oint_C X(z)z^{n-1}\,dz, \tag{5}$$

where C is a closed contour inside the region of convergence of the power series and enclosing the origin. Equation (5) is referred to as the *inverse z transform.*

4) In the region of convergence of the series, both (1) and (4) represent analytic functions of the complex variable z. These functions can often be extended by analytic continuation everywhere except at certain singular points (poles). Since these singularities of the z transform are characteristic of the particular sequence, it is common to plot their locations in the *z plane*, (i.e., the complex plane determined by the real and imaginary parts of z). It should be noted that it is often convenient, because of the special functional form which characterizes exponential sequences, to plot singularities in the *z^{-1} plane.* Furthermore, some authors define the z transform as

$$\tilde{X}(z) = \sum_{n=-\infty}^{\infty} x(n)z^{n}. \tag{6}$$

Clearly, (6) is related to our definition by

$$\tilde{X}(z) = X(z^{-1}). \tag{7}$$

If either (6) or the z^{-1} plane is encountered, it is a simple matter to replace z with z^{-1} in order to relate the z-transform definitions and to note that the region inside the unit circle of the z plane corresponds to the region outside the unit circle of the z^{-1} plane.

5) A *discrete-time impulse at* $k = k_0$ is a discrete-time signal $x(k)$ such that $x(k) = 0$ unless $k = k_0$, in which case $x(k) = 1$. This is an analogy with an *impulse* at time t_0 in the continuous-time case, where $x(t) = \delta(t - t_0)$, the Dirac delta function. The response of a digital filter to a *discrete-time impulse* at $k = 0$ is called its *impulse response*, or sometimes, the *unit sample response*. Other terms generally used for *digital impulse* are *unit sample, unit pulse*, or simply *impulse*.

6) A *sample value* is the value or number associated with one member of a sequence that represents a discrete-time signal. This term is generally used regardless of whether or not the value represents a sample of a continuous-time signal.

7) A *discrete-time linear time-invariant system* or a *discrete-time linear filter* is characterized by its impulse-response sequence $h(n)$.

8) *Discrete-time convolution* is the operation on a signal or sequence $g(n)$ by the impulse response sequence $h(n)$ to yield a digital signal (or sequence) $f(n)$; the operation is defined by the expression

$$f(n) = \sum_{k=-\infty}^{\infty} h(k)g(n-k). \tag{8}$$

This expression is the discrete-time counterpart of the convolution integral for continuous-time systems.

9) An alternate characterization of a discrete-time linear system is the z transform of $h(n)$:

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n}. \tag{9}$$

The complex function $H(z)$ is called the *system function* or transfer function. The values taken by $H(z)$ when evaluated on the unit circle in the z plane give the *frequency response*. Each point on the unit circle, characterized by its angle only, corresponds to a particular frequency.[2] Several different units of frequency are in common use. Some authors express frequency in the conventional units of Hz, kHz, etc. Others use a system of rad/s. Still other authors use a normalized frequency with each frequency expressed as a fraction of the sampling frequency (f/f_s) or half the sampling frequency $(f/(f_s/2))$. Finally, some authors express the normalized frequency in rad/sample. The following table relates the units of frequency to the corresponding angle in the z-plane ($T = 1/f_s$ is the sampling period).

Unit of Frequency	z-Plane Substitution	Range of Frequency Around Unit Circle
f in Hz	$z = e^{i2\pi fT}$	$0 \leq f \leq f_s$
ω in rad/s	$z = e^{i\omega T}$	$0 \leq \omega \leq 2\pi f_s$
$\hat{f} = (f/f_s)$, a fraction of the sampling frequency	$z = e^{i2\pi \hat{f}}$	$0 \leq \hat{f} \leq 1$
$\tilde{f} = (f/0.5f_s)$, a fraction of half the sampling frequency	$z = e^{i\pi \tilde{f}}$	$0 \leq \tilde{f} \leq 2$
θ in rad/sample	$z = e^{i\theta}$	$0 \leq \theta \leq 2\pi$

When the frequency response is expressed in polar form, its magnitude as a function of frequency is called the *amplitude response*, and its angle as a function of frequency is called the *phase response*.

10) One class of linear time-invariant discrete-time filters is characterized by system functions of the form

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 + \sum_{k=1}^{N} a_k z^{-k}}. \tag{10}$$

Such filters have a *recursive realization* in the form of the difference equation

$$y(n) = -\sum_{k=1}^{N} a_k y(n-k) + \sum_{k=0}^{M} b_k x(n-k), \tag{11}$$

where y is the output sequence and x is the input sequence. The system function is generally written in terms of powers of z^{-1} as in (10) because it places in evidence the form of the difference equation given in (11), i.e., in terms of delays. $H(z)$ can also be written as

$$H(z) = \frac{z^{N-M} \sum_{k=0}^{M} b_k z^{M-k}}{z^N + \sum_{k=1}^{N} a_k z^{N-k}} \tag{12}$$

where it is assumed that the zeros of the numerator [zeros of $H(z)$] are distinct from the zeros of the denominator [poles of $H(z)$]. This form places in evidence the fact that regardless of the relative values of M and N, $H(z)$ has the same number of poles as zeros. In most cases—especially digital filters derived from analog designs—M will be less than or equal to N, and there will be at least $N - M$ zeros at $z = 0$. Systems of this type are called Nth-order systems. When $M > N$, the order of the system is no longer unambiguous. Here, N gives the order as the term is used to characterize the mathematical properties of the difference equations. M gives the order used to characterize the complexity of a realization of the system. There is no general agreement as to which of M or N best characterizes the system when $M > N$.

[2] As will be clear from the discussion in Section III, each point on the unit circle, in fact, corresponds to an infinite set of uniformly spaced frequencies that are indistinguishable in a digital system.

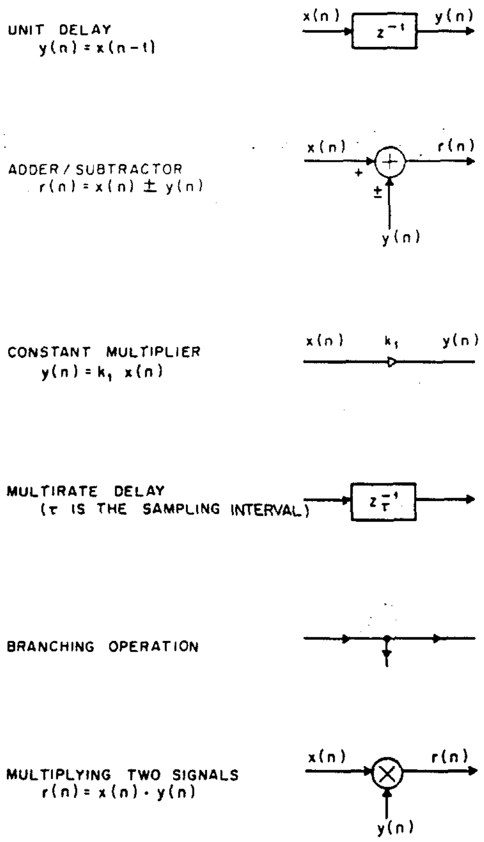

Fig. 1. Recommended terminology for use in block diagrams of digital systems.

11) Since the purpose of a *block diagram* is to graphically depict the way in which a particular system is realized, the terminology shown in Fig. 1 is recommended.

III. Relations Between Discrete and Continuous Signals

1) If a sequence arises as the result of periodic sampling of a continuous-time signal $x_c(t)$, i.e., $x(n)=x_c(nT)$ where T is the sampling period, then $X(z)$, the z transform of $x(n)$, is related to $X_c(s)$, the Laplace transform of $x_c(t)$ by the relationship

$$X(z)\big|_{z=e^{sT}} = \sum_{n=-\infty}^{\infty} x_c(nT)e^{-snT}$$
$$= \frac{1}{T}\sum_{k=-\infty}^{\infty} X_c\left(s+j\frac{2\pi}{T}k\right). \qquad (13)$$

This implies that the s plane and the z plane are related. It should be emphasized that (13) makes clear the fact that the s plane is *not* mapped into the z plane in a one-to-one manner. The actual relationship of $X(z)$ to $X_c(s)$ is depicted in Fig. 2, which shows the s plane divided into an infinite number of horizontal strips of

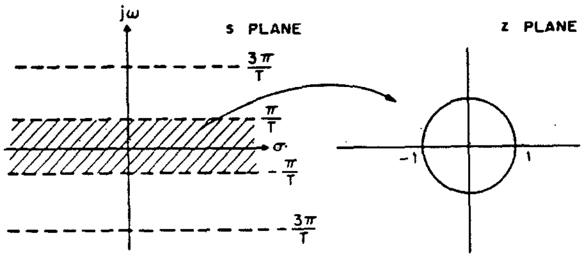

Fig. 2. The mapping of the s plane to the z plane implied by sampling a continuous-time signal.

width $2\pi/T$, each of which maps into the entire z plane. The contributions from each strip are added to produce $X(z)$.

2) The j axis of the s plane corresponds to the unit circle of the z plane. For this reason, the point $z=1$ is often casually referred to as the *DC point* since it corresponds to the point $s=0$ of the s plane. The z transform evaluated on the *unit circle* ($|z|=1$) is of particular interest in digital filtering of sampled signals. For example, there are some sequences for which the z transform does not converge (does not exist) except on the unit circle, e.g., the ideal lowpass digital filter and the ideal digital differentiator for band-limited waveforms. The z transform evaluated on the unit circle is called the *Fourier transform* of the sequence. This definition is consistent with the classical terminology for continuous-time systems. Evaluating (4) and (5) of Section II on the unit circle, yields

$$X(e^{j\theta}) = \sum_{n=-\infty}^{\infty} x(n)e^{-jn\theta} \qquad (14)$$

$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\theta})e^{jn\theta}\,d\theta. \qquad (15)$$

This pair of equations is a Fourier transform pair for the sequence $x(n)$. Alternatively, any of the frequency units of the table in Section II can be used in place of θ.

3) If a continuous-time waveform, $x_c(t)$, is band-limited to a frequency f_0, i.e., $X_c(j2\pi f)$, the Fourier transform of $x_c(t)$, is zero for $|f|>f_0$, then $x_c(t)$ can be recovered exactly from its samples, i.e., $x_c(nT)$, $-\infty<n<\infty$ if $T<1/2f_0$. This is clear from (13) where we can see that

$$X(e^{j2\pi f T}) = \frac{1}{T}X_c(j2\pi f), \qquad -\frac{1}{2T}<f<\frac{1}{2T} \text{ if } T<1/2f_0.$$

$f_N=2f_0$ is called the *Nyquist rate*, and is the lowest rate at which $x_c(t)$ can be sampled, and still be recovered.[3] $T_N=1/f_N=1/2f_0$ is often called the *Nyquist interval*. Additional terms which are often used are *sampling frequency* or *sampling rate* for f_s, the rate at which $x_c(t)$ is actually sampled in a particular system; and $T=1/f_s$

[3] This discussion ignores the possible reduction in sampling rate which can be obtained for bandpass signals.

326 IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS, DECEMBER 1972

is called the *sampling interval*. The highest frequency present in $x_c(t)$, (defined above) f_0, is called the *Nyquist frequency*. The Nyquist frequency is sometimes called the *folding frequency*. It is recommended that the term Nyquist frequency be avoided because of the general confusion with the term Nyquist rate. Furthermore, we recommend that the term folding frequency refer to half of the actual sampling frequency (see Fig. 3).

4) The relationship (13) between the Fourier transform of a sequence of samples $x_c(nT)$ and the Fourier transform of the continuous time signal $x_c(t)$ is depicted in Fig. 4. Part (a) of this figure depicts a band-limited Fourier transform $X_c(j2\pi f)$. In Fig. 4(b) and (c) the sampling rate is greater than or equal to the Nyquist rate and we note that the form of $X_c(j2\pi f)$ is preserved to within a constant multiplier $1/T$ in the frequency range $-1/2T<f\leq 1/2T$. However, in Fig. 4(d) the signal $x_c(t)$ is *undersampled*, i.e., sampled at a rate below the Nyquist rate. In this case the Fourier transform of the sequence obtained by sampling is not equal to $X_c(j2\pi f)/T$ due to the fact that some of the other terms in (13) such as $X_c(j2\pi f-j(2\pi/T))$ are nonzero in the frequency range $-1/2T\leq f\leq 1/2T$. One way of viewing this is to say that a set of frequencies in $X_c(j2\pi f)$ is indistinguishable from a different set of frequencies in $X_c(j2\pi f-j2\pi/T)$. These frequencies are called *aliases* of one another and the process of confounding frequencies as in Fig. 4(d) is called *aliasing*.

5) Suppose we have a sampled waveform $x(n)$ with z transform $X(z)$. We define a new sampled waveform $y(n)$ using one of every M samples as the samples of the new waveform, i.e., $y(n)=x(Mn)$, with M any positive integer. Clearly, this process is equivalent to sampling at a lower rate, and it is to be expected that aliasing may occur. When the aliasing occurs due to "sampling" a discrete-time signal it is called *digital aliasing*. It is readily shown that $Y(z)$ can be written in terms of $X(z)$ as

$$Y(z) = \frac{1}{M}\sum_{l=0}^{M-1} X(z^{1/M}e^{-j(2\pi/M)l}). \tag{16}$$

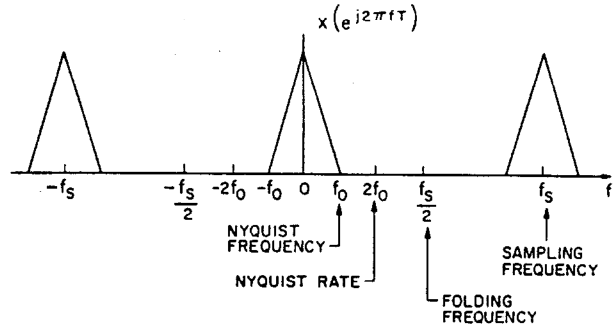

Fig. 3. Labeling of terminology concerned with frequencies related to the sampling process.

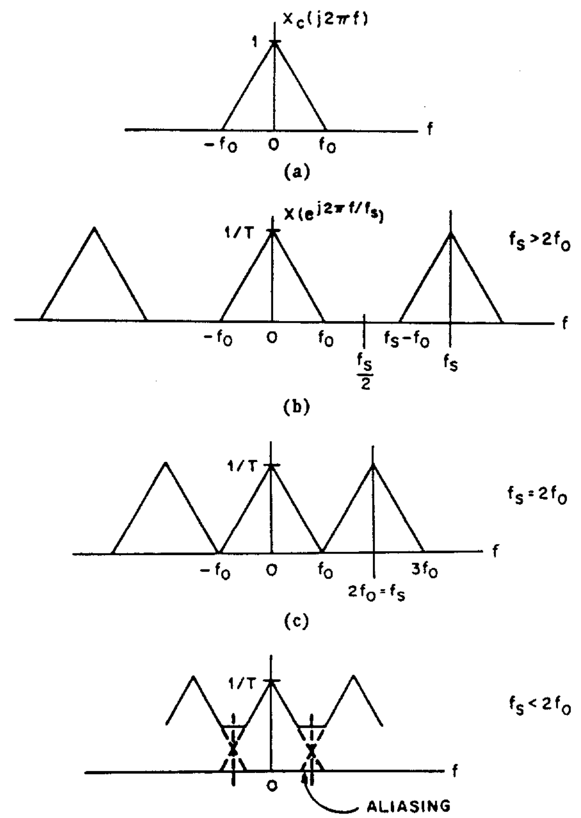

Fig. 4. An example of the effects of various sampling frequencies on the frequency response of the digital signal.

IV. Theory and Design of Digital Filters

1) *Discrete filters* may be divided into two classes on the basis of whether the signal values can take on a continuum of values (sampled-data filters) or a finite set of values (digital filter). Thus we have the following.

a) A *sampled-data filter* is a computational process or algorithm by which a sampled-data signal acting as an input is transformed into a second sampled-data signal termed the output. The sampled-data signal is considered only at a set of points (usually equally spaced in time or space as the independent variable); at these points the signal can take on a continuum of values.

b) A *digital filter* is a computational process or algorithm by which a digital signal or sequence of numbers (acting as input) is transformed into a second sequence of numbers termed the output digital signal. The numbers are limited to a finite precision. The algorithm may be implemented in software as a computer subroutine for a general-purpose machine or in hardware as a special-purpose computer. The term digital filter is then applied to the specific routine in execution or to the hardware.

2) Further complexity of filtering action may be obtained by switching. Thus, a *switched filter* is one in

which the input and output are simultaneously switched in a definite pattern among a group of input and output ports. The filter being switched may be either of the continuous or discrete types. Examples of switched filters are commutating filters or n-path filters.

3) A *multiplexed filter* is a restricted form of a switched filter; commonly a single discrete filter which by means of a switching action is made to perform the function of several discrete filters virtually simultaneously. The multiplexing is most commonly done in a time-division manner whereby the input to the discrete filter is sequentially switched from a number of input signals and the filter output sequentially switched in synchronism to a corresponding set of output signal lines. Thus a single filter may be made to do the work of many filters by this time division multiplexing.

4) A *recursive filter* is a discrete-time filter which is realized via a recursion relation, i.e., the output samples of the filter are explicitly determined as a weighted sum of past output samples as well as past and/or present input samples. For example, $y(n)=b_0x(n)+b_1x(n-1)+b_2x(n-2)-a_1y(n-1)-a_2y(n-2)$.

5) A *nonrecursive filter* is a discrete-time filter for which the output samples of the filter are explicitly determined as a weighted sum of past and present input samples only. For example, $y(n)=b_0x(n)+b_1x(n-1)+b_2x(n-2)$.

6) A *finite impulse response* (FIR) filter is a filter whose impulse response $h(n)$ is zero outside some finite limits, i.e., $h(n)=0$, for $n>N_1$ and $n<N_2$ with $N_1 \geq N_2$.

7) An *infinite impulse response* (IIR) filter is a filter for which either $N_1=\infty$ or $N_2=-\infty$ or both, in 6). Thus the duration of the filter's impulse response is infinite.

8) It should be noted that the poles of FIR filters are restricted to $z=0$ or $z=\infty$, whereas there are no such restrictions on the positions of either the poles or zeros of IIR filters.

9) The terms recursive and nonrecursive are recommended as descriptions for how a filter is *realized* and not whether or not the filter impulse response is of finite duration. (Although IIR filters are generally realized recursively, and FIR filters are generally realized nonrecursively, IIR filters can be realized nonrecursively and FIR filters can be realized recursively.)

10) A *transversal filter* is a filter (either continuous or discrete) in which the output signal is generated by summing a series of delayed versions of the input signal weighted by a set of weights termed the *tap gains*. If the signal delays are accomplished by a tapped delay line then the filter is termed a *tapped delay line* filter.

11) A *comb filter* is a filter comprised of the sum or difference of input and output of a delay of M units and unit gain yielding a transfer characteristic $H(z)=1\pm z^{-M}$ (see Fig. 5); this filter has M zeros of transmission equally spaced on the unit circle in the z plane thus giving rise to a frequency characteristic having M equal peaks and M real frequency zeros.

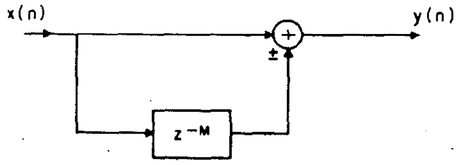

Fig. 5. Block diagram representation of comb filter.

12) A *frequency-sampling filter* is an FIR filter which is designed by varying one or more of its DFT[4] coefficients (called *frequency samples*) to minimize some aspects of the filter's frequency response. For example, the DFT coefficients of a frequency-sampling lowpass filter are 1.0 in the passband, 0.0 in the stopband, and variable in the transition band. One design criterion would be to choose the variable coefficients to minimize peak stopband ripple.

13) An *extraripple filter* (also called *maximal ripple filter*) is an FIR filter whose frequency response is equiripple in both the passband and stopband, and whose frequency response contains the maximum possible number of ripples.[5] There is no general agreement as to the appropriateness of this term, and as such, no recommendation as to its usage is made.

14) An *equiripple* (*optimal*) *filter* is an FIR filter which is the unique best approximation in the minimax sense to some specified frequency response characteristic over any closed subset of the frequency interval. For the lowpass filter case the optimal filter may be an extraripple filter, an equiripple filter with one less than the maximum possible number of ripples, or a filter with the maximum possible number of ripples all except one of which are of equal amplitude.

15) A *frequency-sampling realization* is a means of realization of an FIR filter of duration N samples as a cascade of a comb filter and a parallel bank of N complex pole resonators. The filter output is obtained as a weighted sum of the outputs of each of the parallel branches; the multiplier on the kth branch being the kth DFT coefficient of the filter impulse response.

16) A *Kalman filter* (*discrete time*) is a linear, but possibly time-varying discrete-time filter with the property that it provides a least mean-square error estimate of a (possibly vector-valued) discrete-time signal based on noisy observations. The statistical description of the problem is such that the Kalman filter has a recursive implementation, using a linear combination of new observations and old estimates. The filter design may be based on a more general error criterion, using a nonquadratic loss function. Its essential features are that its design is based on a statistical criterion in the time domain, and that it is, in general, time varying. If the filter is further restricted to be time invariant it becomes the *Wiener filter*.

[4] See Section VI-1 for a definition of DFT.
[5] See Section IV-29 for a definition of ripple.

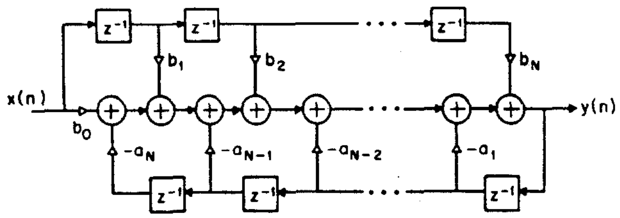

Fig. 6. Block diagram representation of direct form 1 for an Nth-order system.

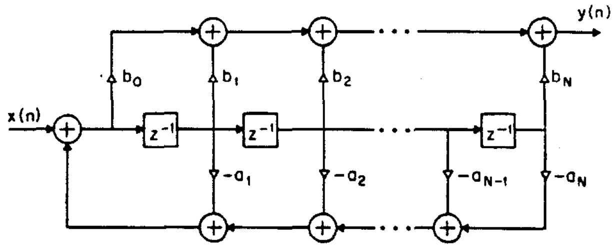

Fig. 7. Block diagram representation of direct form 2 for an Nth-order system.

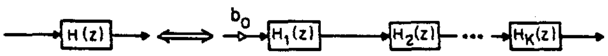

Fig. 8. Block diagram representation of the cascade form.

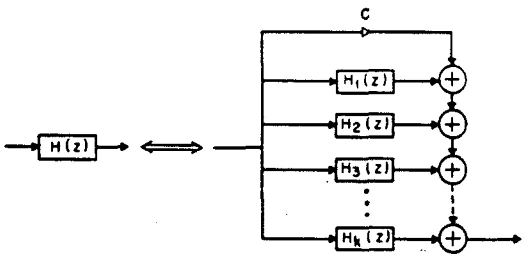

Fig. 9. Block diagram representation of the parallel form.

17) The forms for realizing digital filters include the following:

a) *Direct form 1* (shown in Fig. 6) where

$$H(z) = \frac{\sum_{i=0}^{N} b_i z^{-i}}{\sum_{i=0}^{N} a_i z^{-i}}, \qquad a_0 = 1. \tag{17}$$

For convenience in showing the realization, the order of the numerator and denominator are set to be the same. Direct form 1 uses separate delays for both the numerator polynomial and the denominator polynomial. In certain cases, e.g., floating-point additions, the results may depend on the exact ordering in which the additions are performed.

b) *Direct form 2* is shown in Fig. 7. Direct form 2 has been called the *canonic form* because it has the minimum number of multiplier, adder, and delay elements, but since other configurations also have this property, this terminology is not recommended.

c) *Cascade canonic form* (or *series form*), which is shown in Fig. 8, where

$$H(z) = b_0 \prod_{i=1}^{K} H_i(z) \tag{18}$$

and $H_i(z)$ is either a *second-order section*, i.e.,

$$H_i(z) = \frac{1 + b_{1i}z^{-1} + b_{2i}z^{-2}}{1 + a_{1i}z^{-1} + a_{2i}z^{-2}}, \tag{19}$$

or a *first-order section*, i.e.,

$$H_i(z) = \frac{1 + b_{1i}z^{-1}}{1 + a_{1i}z^{-1}}, \tag{20}$$

and b_0 is implicitly defined in (17), where K is the integer part of $(N+1)/2$.

d) *Parallel canonic form*, which is shown in Fig. 9, where

$$H(z) = C + \sum_{i=1}^{K} H_i(z) \tag{21}$$

where $H_i(z)$ is either a *second-order section*, i.e.,

$$H_i(z) = \frac{b_{0i} + b_{1i}z^{-1}}{1 + a_{1i}z^{-1} + a_{2i}z^{-2}}, \tag{22}$$

or a *first-order section*, i.e.,

$$H_i(z) = \frac{b_{0i}}{1 + a_{1i}z^{-1}}, \tag{23}$$

and K = integer part of $(N+1)/2$ and C is proportional to b_N as defined in (17).

18) The individual second- and first-order sections of the cascade and parallel forms are generally realized in one of the direct forms.

19) *Transpose configurations* for all of the above forms can be obtained by reversing the directions of all signal flow (i.e., by reversing the directions of all arrows) and by interchanging all branch nodes and summing junctions. The resulting circuits have the same transfer functions but different roundoff noise and overflow properties.

20) When the transfer function of a high-order filter is decomposed into a cascade connection of lower order filter sections by distributing the pole and zero factors among the lower order sections, then *pairing* refers to the associating of a specific zero factor with a specific pole factor to form an elemental or individual section. *Ordering* refers to the sequence or order in which the individual sections are connected in cascade to form the composite higher order filter. Varying the pairing and ordering can dramatically change the noise properties and dynamic range of both discrete and continuous filters. As an example, if

$$H(z) = \frac{\prod_{j=1}^{l} N_j(z)}{\prod_{j=1}^{m} D_j(z)} \tag{24}$$

and $l = m = 5$ then a possible detailed realization would be

$$H(z) = \frac{N_1(z)}{D_2(z)} \times \frac{N_3(z)}{D_1(z)} \times \frac{N_2(z)}{D_5(z)} \times \frac{N_4(z)}{D_3(z)} \times \frac{N_5(z)}{D_4(z)} \quad (25)$$

where the pairing is N_1 with D_2, N_3 with D_1, N_2 with D_5, N_4 with D_3, and N_5 with D_4. The implied ordering is N_1/D_2 first, followed by N_3/D_1, N_2/D_5, N_4/D_3, and finally N_5/D_4.

21) Two important properties of digital filters are *stability* and *causality*. The definition of *stability* most often used in digital filtering is as follows: a system is stable if every bounded (finite) input produces a bounded (i.e., finite) output. For linear time-invariant digital filters, a necessary and sufficient condition for stability is

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty. \quad (26)$$

22) A system is said to be *causal* if the output for $n = n_0$ is dependent only on values of the input for $n \leq n_0$. For linear time-invariant digital filters, this implies that the unit sample response sequence (i.e., the impulse response) is zero for $n < 0$. For the case of most interest, i.e., causal linear time-invariant filters with rational transfer functions, stability implies that all the poles of $H(z)$ must be inside the unit circle in the z plane.

23) The *gain of a discrete filter* is the steady-state ratio of the peak magnitude (or any other consistent measure like root-mean-square, for example) of the output to the peak magnitude (or other consistent measure) of the input signal to the discrete filter. The usual input signals are either periodic sequences, e.g., sine waves, or pseudorandom sequences.

24) The *frequency-scale factor* is the factor by which all the poles and zeros of a normalized filter (cutoff frequency of 1 rad/s) must be multiplied to yield the actual filter pole and zero values, i.e., the ratio of the unnormalized to the normalized frequency scale of a filter.

25) The *filter bandwidth* is the width, in units of frequency, between the two points that define the edges of the passband of the frequency characteristics of a filter. The frequency points are usually defined as those values of frequency at which the attenuation or loss is a specified amount and beyond which the essential filter characteristic changes from pass (small attenuation) to stop (larger attenuation).

26) A commonly specified frequency point is the 3-dB or half-power point. For the elliptic and Chebyshev filters the frequency points are the highest and lowest frequencies at which the filter attenuation satisfies the equiripple passband attenuation limits. For other filters the frequency points may be defined in terms of the asymptotic intersections of the passband and stopband logarithmic asymptotes. Some examples of typical filter characteristics are shown.

27) Typical magnitude-square characteristics for several of the standard forms of continuous-time filters are given below using the following terminology.

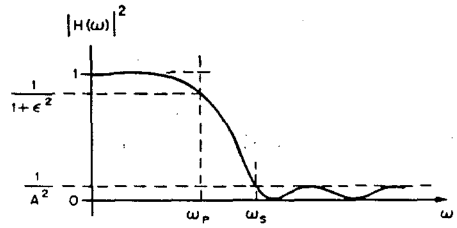

Fig. 10. An example of the magnitude-squared characteristics of a typical filter.

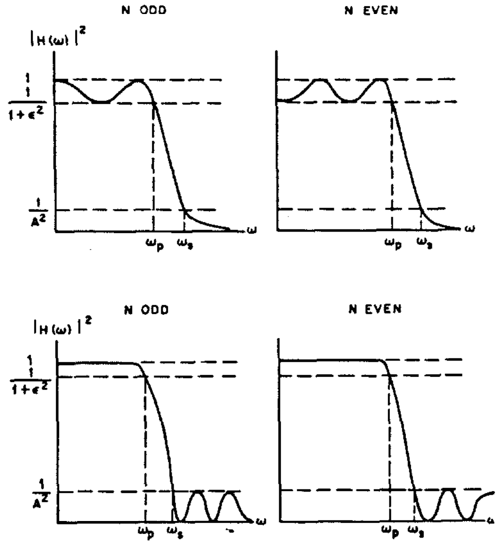

Fig. 11. The magnitude-squared characteristics for even and odd order Chebyshev types I (top) and II (bottom) filters.

$|H(\omega)|^2 \triangleq$ Magnitude-squared characteristic (frequency in rad/s).
ω_p Passband edge frequency.
ω_s Stopband edge frequency.

A typical response is shown in Fig. 10.

a) *Butterworth filter*: Maximally flat magnitude at $\omega = 0$

$$|H(\omega)|^2 = \frac{1}{1 + (\omega/\omega_p)^{2N}}. \quad (27)$$

b) *Chebyshev filters*:

Type I—Equiripple passband, monotone stopband:

$$|H(\omega)|^2 = \frac{1}{1 + \epsilon^2 C_N^2(\omega)}. \quad (28)$$

Type II—Equiripple stopband, passband maximally flat at $\omega = 0$:

$$|H(\omega)|^2 = \frac{1}{1 + \epsilon^2 \left[\frac{C_N(\omega_s)}{C_N(\omega_s/\omega)}\right]^2} \quad (29)$$

where the $C_N(\omega)$ are the Chebyshev polynomials. Fig. 11

330 IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS, DECEMBER 1972

shows the response for the two types of Chebyshev filters for N odd and even.

c) *Elliptic (Cauer) filters*—equiripple passband and stopband:

$$|H(\omega)|^2 = \frac{1}{1 + \epsilon^2 \psi_N^2(\omega)} \tag{30}$$

where the $\psi_N(\omega)$ is a rational Chebyshev function involving elliptic functions. Fig. 12 shows the response of elliptic filters for N both odd and even.

28) The term *transition band* is used to describe an interval of frequencies where a filter characteristic changes from one kind of behavior to another, one example being the transition band from a pass to a stop characteristic. The *transition ratio* is a relative measure of the passband width to the sum of the widths of the passband and the adjacent transition band(s). It can also be defined for a single-transition band-passband pair provided the width of the passband is defined. For the filter shown in Fig. 13 the transition ratios are defined as

$$\text{transition ratio} = \frac{\omega_c - \omega_{l2}}{\omega_c - \omega_l} \text{ (lower region)} \tag{31}$$

$$\text{transition ratio} = \frac{\omega_{u1} - \omega_c}{\omega_u - \omega_c} \text{ (upper region)} \tag{32}$$

where ω_c may be defined as either the arithmetic mean of the band-edge frequencies, i.e.,

$$\omega_c = \frac{\omega_{u1} + \omega_{l2}}{2} \tag{33}$$

or as the geometric mean of these same two frequencies, i.e.,

$$\omega_c = \sqrt{\omega_{u1}\omega_{l2}}. \tag{34}$$

Thus the transition ratio is bounded on the upper side by unity. Transition ratios near unity imply sharp cutoff filters.

29) The nature of a filter's response characteristic that approximates a desired characteristic by being alternatively greater than and less than the desired response as the independent variable is increased is called the *ripple*. The ripple may be expressed as the ratio of the maximum to the minimum of the response in a specified range, e.g., the passband of a filter. In this case, the ripple is usually expressed in percent or in decibels by taking 20 $\log_{10}$ of the ratio. Alternatively, the ripple may be expressed relative to some specified level of response such as plus or minus a fixed number of units. For example, consider the magnitude response shown in Fig. 14, where *passband ripple* $= 2.268/2.160 = 1.05$ which implies a $(2.268)/\sqrt{2.268 \times 2.160} = 1.0247$ or ± 2.47 percent variation about the geometric mean; thus passband ripple expressed in dB (=) 20 $\log_{10}(1.05)$

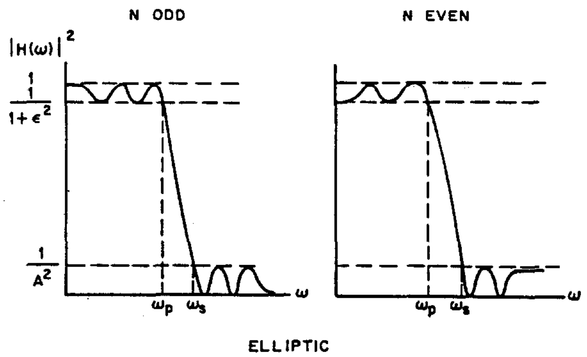

Fig. 12. **The magnitude-squared characteristics for even and odd order elliptic filters.**

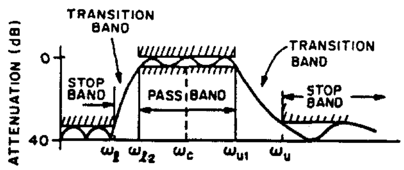

Fig. 13. **The attenuation characteristics of a typical bandpass filter showing passband, stopbands, and transition bands.**

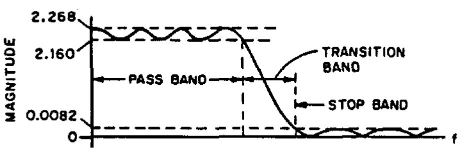

Fig. 14. **The magnitude characteristic of a typical filter showing its ripple characteristics.**

$= 0.424$ dB overall or ± 0.212 dB ripple about the geometric mean.

30) The passband ripple is also termed the *in-band ripple*. The terms *stopband ripple* and *out-of-band ripple* have also been used when the out-of-band frequency response has the characteristic of a ripple; numerically this has been used to express the ratio of the minimum out-of-band attenuation to the mean in-band attenuation. We recommend that this ratio be termed ***minimum stopband attenuation*** and that the terms stopband ripple and out-of-band ripple not be used except qualitatively. In the example, minimum stopband attenuation $= 0.0082 (=) -41.7$ dB and the relative minimum stopband attenuation $= (0.0082)/\sqrt{2.268 \times 2.160} = 0.003705$ $(=) -48.6$ dB.

Methods for Designing Digital Filters

31) An important class of techniques for designing infinite impulse response filters to be realized recursively is based on a transformation of a continuous-time filter. This class consists of at least three techniques.

a) *Impulse invariance* (also called the standard z transformation or standard z) is a technique in which the impulse response of the derived digital filter is identical to the sampled impulse response of a continuous-time filter. If the continuous-time filter has a transfer function[6]

$$H_c(s) = \frac{\sum_{k=0}^{M} d_k s^k}{\sum_{k=0}^{N} c_k s^k} = \sum_{k=1}^{N} \frac{A_k}{s+\alpha_k}, \qquad M < N \quad (35)$$

then the requirement that

$$h(n) = h_c(t)\big|_{t=nT}, \qquad 0 \le n \le \infty \quad (36)$$

implies that $H(z)$ is obtained from the partial fraction expansion of $H_c(s)$ by the substitution

$$\frac{1}{s+\alpha_k} \rightarrow \frac{1}{1-e^{-\alpha_k T}z^{-1}}. \quad (37)$$

It can be shown that

$$H(z)\big|_{z=e^{sT}} = \frac{1}{T}\sum_{k=-\infty}^{\infty} H_c\left(s + j\frac{2\pi}{T}k\right). \quad (38)$$

Thus, impulse invariance is only satisfactory when $H_c(s)$ is band limited. If as in most instances, $H_c(s)$ is not sufficiently band limited, $H(z)$ is an aliased version of $H_c(s)$. Therefore, this technique is primarily used for narrowband filter designs or else the transformation is applied to the cascade combination of a *guard filter* and $H_c(s)$.

Another important point is clear from (38). Due to the $1/T$ multiplier, digital filters derived by impulse invariance have a gain approximately $1/T$ that of the continuous-time filter. This is generally compensated by multiplying each factor in the partial fraction expansion by T, so that the digital filter will have approximately the same gain as the continuous-time filter from which it was derived.

b) *Bilinear transformation* (also called the bilinear z transform, the bilinear z transformation or z form) is a technique used to circumvent the aliasing problem of the impulse invariant technique. This approach uses the algebraic transformation

$$s = \frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}, \quad (39)$$

to derive the system function of the digital filter as

$$H(z) = H_c(s)\big|_{s=(2/T)(1-z^{-1})/(1+z^{-1})}. \quad (40)$$

This transformation has the effect of mapping the entire s plane into the z plane in such a way that the left-half s plane maps into the inside of the unit circle and the right-half s plane maps to the outside of the unit circle. This results in a nonlinear warping of the frequency scale according to the relation

$$\frac{\omega_c T}{2} = \tan\frac{\omega_d T}{2} \quad (41)$$

where ω_c is the continuous-time frequency variable and ω_d is the discrete-time frequency variable. Because of this warping of the frequency scale, this design technique is most useful in obtaining digital designs of filters whose frequency response can be divided into a number of pass and stop bands in which the response is essentially constant. Generally it is necessary to take appropriate account of the warping of the frequency scale.

c) *Matched z transform* (also called the matched z transformation, mapping poles and zeros, or matched z) is a technique based on mapping the poles and zeros of the continuous-time filter by the substitution

$$s - s_i \rightarrow 1 - e^{s_i T}z^{-1}. \quad (42)$$

This means that the poles of $H(z)$ will be identical to those obtained by impulse invariant method, however the zeros will not correspond.

32) In the context of designing a discrete-time system and especially a digital filter, an *optimization technique* is a procedure for minimizing a prescribed performance function based on design requirements. An example is the design of a discrete-time filter to have the minimum mean-square deviation from a desired frequency-domain characteristic. An *iterative optimization technique* is a procedure for generating successive approximations converging (hopefully) to an optimum. This is opposed to an *analytical design technique*, which yields a closed form solution, such as the Chebyshev design for a lowpass filter.

V. Finite Word Length Effects—A/D, D/A Conversion

1) A *digital-to-analog* (*D/A*) *converter* is a device which operates on a digital input signal $s(nT)$ to produce a continuous-time output signal $s(t)$ ideally defined by

$$s(t) = \sum_n s(nT)h(t-nT) \quad (43)$$

where $h(t)$ characterizes the particular converter. For example, $h(t)$ is a square pulse of duration T for a zero order hold D/A converter. The D/A converter is usually followed by a linear time-invariant low-pass continuous-time filter called a *postfilter*. The combination of D/A converter and postfilter is called a *reconstruction device* or *reconstruction filter*.

2) An *analog-to-digital* (*A/D*) *converter* is a device which operates on a continuous-time waveform to pro-

[6] This formulation assumes that all poles are distinct. Appropriate modifications can be made to deal with multiple order poles.

duce a digital output consisting of a sequence of numbers each of which approximates a corresponding sample of the input waveform. Expressing the numerical equivalent of each sample by a finite number of bits (instead of the infinite number required to completely specify each sample) is the *quantizing* inherent in the conversion process. The error produced by quantizing is called *quantizing noise* or *A/D conversion noise.*

Representation of Numbers

3) Various systems are used to represent the numbers in a digital filter. In *fixed-point number* representation, the position of the binary (or decimal) point is assumed fixed. The bits to the right of the (fixed) binary (or decimal) point represent the fraction part of the number and the bits to the left represent the integer part. For example, the binary number 011.001 has the value $0\times2^2+1\times2^1+1\times2^0+0\times2^{-1}+0\times2^{-2}+1\times2^{-3}$.

4) A *floating-point number* is formed by two fixed point numbers, the *mantissa*[7] and the *exponent.* The floating-point number is equal to the product of the mantissa with the quantity resulting when a given *base* is raised to the power denoted by the exponent. The base is the same for every floating-point number in the digital filter. Consequently, the numerical value of an entry in a specified position in the mantissa is determined by the exponent. The mantissa is generally normalized to be as large as possible but less than some number (e.g., 1.0). For example, 0.1×10^2 is legitimate, whereas 0.01×10^3 and 10.0×10^0 are usually considered to be illegitimate floating-point decimal representations of the number 10. The most commonly used base is two (*binary representation*). The base 16 (*hexadecimal representation*) is used in some general purpose computers. The base 8 is called *octal representation.*

5) The representation of *block floating-point numbers* is determined by examining all numbers in a block (i.e., array). The largest number is represented as an ordinary floating-point number with a normalized mantissa. The remaining numbers in the block use the exponent associated with this largest number. This use of a single exponent for the whole block saves memory. This type of arithmetic is popular in realizations of the fast Fourier transform.

Representation of Negative Numbers

6) The discussion so far has dealt with the representation of nonnegative numbers. There are three common systems used for representing signed numbers. The representation of positive numbers is the same in these three systems. The first, and most familiar, is *sign and magnitude*, i.e., the magnitude (which is, of course, positive) is represented as a binary number and the sign is represented by an additional binary digit in the leading position which, if 0 corresponds to a + and if 1 corresponds to a − (or vice versa). Thus, for example, in sign and magnitude 0.0011 represents 3/16 and 1.0011 represents −3/16. Two related representations of signed numbers are ones complement and twos complement. In each of these systems a positive number is represented as in sign and magnitude. For *twos-complement* representation the negative of a particular positive number is obtained by complementing all the bits and adding one unit in the position of the least significant bit. For example, −(0.0110) would be represented in twos complement as (1.1001)+(0.0001)=1.1010. A carry out of the sign bit is neglected in the addition, so that −(0.0000)=(1.1111)+(0.0001)=0.0000. For *ones-complement* representation the negative of a given positive number is obtained simply by complementing all the bits.

7) The choice of representation for negative numbers in a particular system is based almost entirely on hardware considerations. With ones-complement and twos-complement numbers, subtraction can be performed conveniently with an adder. For example, in twos complement, the difference $A-B$ is formed by simply adding to A the twos complement of B.

Finite Word Length Effects

8) Even though the input to a digital filter is represented with finite word length (e.g. through A/D conversion), the result of processing will naturally lead to values requiring additional bits fo. their representation. For example, a b-bit data sample multiplied by a b-bit coefficient results in a product which is $2b$ bits long. If in a recursive realization of a filter we do not quantize the result of arithmetic operations, the number of bits required will increase indefinitely, since after the first iteration $2b$ bits are required, after the second iteration $3b$ bits are required, etc. Two common methods are used to eliminate the lower order bits resulting from arithmetic operations in a digital filter.

a) *Truncation* is accomplished by discarding all bits (or digits) less significant than the least significant bit (or digit) which is retained.

b) *Rounding* of a number to b bits, when the number is initially specified to more than b bits, is accomplished by choosing the rounded result as the b-bit number closest to the original unrounded quantity. When the unrounded quantity lies equidistant between two adjacent b-bit numbers, a random choice ought to be made as to which of these numbers to round to. For example, 0.01011 rounded to three bits would be 0.011; but 0.01010 rounded to three bits can be chosen as either 0.011 or 0.010, and the choice should be random. In many situations, however, one can choose to always round up in this midway situation with negligible effect on the accuracy of the computation.

9) *Roundoff error* (or *roundoff noise*) or *truncation error* (or *truncation noise*) is caused by rounding off or truncating the products formed within the digital filter.

[7] The term mantissa as defined here is unfortunately not the same as the term mantissa commonly used in logarithm tables. The definition presented here is dictated by its extensive occurrence in the literature.

The roundoff or truncation error is sometimes well modeled as a random process. On the other hand, if the data sequence to a recursive realization of a digital filter consists of constants (e.g., zero) or some other periodically repeating samples, the roundoff or truncation error is periodic and causes a *deadband effect* or *limit cycle* in the filter output. A common type of limit cycle is a *zero-input limit cycle* where the output of a digital filter remains periodic and nonzero, after the input has been set to zero. *Dither* is a sequence of numbers that is added to the input to a recursive digital filter to ameliorate the deadband effect. Even though dither increases the mean-square error in the output, it can disrupt the pattern of roundoff errors causing the deadband effect, thereby permitting the output to return to zero.

10) *Overflow* occurs when a digital filter computes a number that is too large to be represented in the arithmetic used in that filter. If no compensation is made for the overflow then large errors in the filter output can result either in the form of transients or of *overflow oscillations.* A technique used to compensate in part for overflow is *saturation arithmetic* where a sum that is too large to be represented is set equal to the largest representable number in the filter.

11) *Dynamic range* is the ratio between the largest and smallest signals which can be represented in the filter with a given fidelity criterion. Unfortunately the fidelity criterion is often vague or unspecified.

12) The roundoff or truncation noise introduced within a digital filter produces a resultant noise at the output of the digital filter. A *signal-to-noise ratio* can be defined in this context, for example, as the ratio of the ideal mean-squared output signal (filter output in the absence of any rounding) to the mean-squared output noise due to rounding or truncation. Expressing this ratio in bits, as (1/2) $\log_2$ of the ratio, gives an approximate indication of the number of accurate bits in the filter output. Different definitions of signal-to-noise ratio may be appropriate in different contexts.

13) Another effect of finite word length is *coefficient quantization error* (or *parameter quantization error*), which occurs when the coefficients of a digital filter, initially specified with unlimited accuracy, are quantized by rounding or truncation. Coefficient quantization error appears as error in the digital filter's response (e.g., impulse response, transfer function, frequency response, etc.).

VI. Discrete Fourier Transforms and the FFT

1) For a sequence of N numbers, possibly complex, the *discrete Fourier transform* (*DFT*) is another sequence of exactly N numbers which are the values of the z transform of the original finite sequence for N values of z, specifically

$$z = e^{j(2\pi/N)k}, \qquad k = 0, 1, \cdots, N-1. \tag{44}$$

Discrete Fourier transformation is the operation of computing, or otherwise forming, the discrete Fourier transform of the sequence.

2) From the definition, the sequence $\{f(0), f(1), \cdots, f(N-1)\}$ has the DFT[8] $\{F(0), F(1), \cdots, F(N-1)\}$

$$F(k) = \sum_{n=0}^{N-1} f(n)e^{-j(2\pi/N)nk}. \tag{45}$$

3) It is possible to recover the original sequence from its DFT by the operation

$$f(n) = \frac{1}{N}\sum_{k=0}^{N-1} F(k)e^{j(2\pi/N)nk}, \tag{46}$$

giving a sequence of N samples $\{f(0), f(1), \cdots, f(N-1)\}$ as the inverse discrete Fourier transform of $\{F(0), F(1), \cdots, F(N-1)\}$. The operation is called ***inverse discrete Fourier transformation*** or IDFT, and is remarkably similar in form to the discrete Fourier transformation.

4) Some authors have defined the DFT in related but different ways, involving $e^{j(2\pi/N)nk}$, or a multiplicative factor of $1/N$ or $1/\sqrt{N}$. By considering the expressions for the DFT and IDFT it is evident that the constants $1/N$ or $1/\sqrt{N}$ and the possible use of $e^{j(2\pi/N)nk}$ in other definitions of the DFT can easily be compensated for in the other definitions of the IDFT.

5) Suppose for an N-point sequence we are interested in computing its DFT, and suppose N is a composite integer

$$N = r_1 \times r_2 \times \cdots \times r_\mu \tag{47}$$

where the r_i are a set of factors of N, not necessarily prime factors. Of the various algorithms for computing such a DFT, some require a number of operations ***proportional*** to $N\sum_{i=1}^{\mu} r_i$ (since the word proportional allows considerable latitude, it is not necessary to be too specific about the meaning of "operation"); such algorithms are called *fast Fourier transforms* (FFT).[9] An important special case is when

$$r_1 = r_2 = \cdots = r_\mu = 2$$

so that

$$\sum_{k=1}^{\mu} r_k \rightarrow 2 \log_2 N.$$

For fast Fourier transforms in this case, the proportionality is to $N \log_2 N$.

6) A subclass of FFT algorithms is known which use high speed convolution techniques to compute the DFT of a sequence through a formula in which it is expressed as a convolution. Examples of such algorithms are the chirp z transform and the prime algorithm.

7) In order to classify different FFT algorithms and

[8] For convenience, the notation of $F(k)$ is used to denote DFT coefficients rather than $F(e^{j(2\pi/N)k})$.

[9] The word transform instead of algorithm is embedded ineradicably in the literature.

to relate them to one another it is useful to consider the following procedure applied to a sequence $f(n)$ of length N where N has a factorization $P \times Q \times R$. The reader should generalize to more complicated factorizations. Let us replace the index n by the triplet (n_0, n_1, n_2) where

$$(n_0, n_1, n_2) = n = n_0 + Rn_1 + QR\,n_2 \tag{48}$$

and

$$0 \le n_0 < R$$
$$0 \le n_1 < Q$$
$$0 \le n_2 < P.$$

Similarly, we replace the index k by the triplet (k_0, k_1, k_2) where

$$(k_0, k_1, k_2) = k = k_0 + Pk_1 + PQ\,k_2 \tag{49}$$

and

$$0 \le k_0 < P$$
$$0 \le k_1 < Q$$
$$0 \le k_2 < R.$$

Then (45) can be manipulated into the form

$$F(k_0, k_1, k_2) = \sum_{n_0=0}^{R-1} \left(\left\{ \sum_{n_1=0}^{Q-1} \left(\left\{ \sum_{n_2=0}^{P-1} f(n_0, n_1, n_2) \cdot e^{-j(2\pi/P)n_2k_0} \right\} \phi_A \right) e^{-j(2\pi/Q)n_1k_1} \right\} \phi_B \right) e^{-j(2\pi/R)n_0k_2}. \tag{50}$$

Here ϕ_A and ϕ_B are of unit magnitude and have arguments dependent on the indices. Equation (50), if followed as a recipe, suggests a way of computing an N-point DFT as a collection of smaller DFT's. There are: QR DFT's of P-point sequences; PR DFT's of Q-point sequences; and PQ DFT's of R-point sequences.

The only other operations called for in (50) are the multiplications by ϕ_A, ϕ_B. These have been called *twiddle factors*, *phase factors*, and *rotation factors* by various authors.

8) To save multiplications (50) is commonly modified in one of two ways. The first way is to combine the factors ϕ_A and $e^{-j(2\pi/P)n_2k_0}$ inside the sum over n_2 and the factors ϕ_B and $e^{-j(2\pi/Q)n_1k_1}$ inside the sum over n_1. The algorithm so produced has been called a *decimation-in-frequency* algorithm or a *Sande–Tukey* algorithm.

9) A second way to save multiplications is to combine the factors ϕ_A and $e^{-j(2\pi/Q)n_1k_1}$ inside the sum over n_1 and the factors ϕ_B and $e^{-j(2\pi/R)n_0k_2}$ inside the sum over n_0. The resulting form of the algorithm is called a *Cooley–Tukey* or *decimation-in-time* algorithm.

10) An algorithm with the twiddle factors explicitly present, as separate multiplications, is neither Cooley–Tukey nor Sande–Tukey.

11) There is a third way to save multiplications, which works only when the factors P, Q, R are relatively prime. By permuting the data sequence $f(n)$ and accepting a permuted transform sequence, a formula like (50) can be derived in which $\phi_A = \phi_B = 1$. This algorithm is called the *prime factor algorithm.*

12) Any of the basic forms can be programmed so that each summation is computed and the result stored in the memory formerly occupied by its input data as soon as that input data is no longer needed. In the case of (50), the amount of extra storage over the N cells needed for the sequence itself is only the greater of P, Q, or R cells.

A programmed version of an algorithm which takes advantage of this possibility is called *in place.* Unfortunately, when storage use is minimized in this way, either the input sequence or the output sequence must be unusually ordered. An example of this unusual order might be that

$$F(k_0 + Pk_1 + PQk_2) \tag{51}$$

ends up in storage location $k_2 + Rk_1 + QRk_0$. The unusual ordering may also take place in the use of weights in the algorithm. All such effects are called "digit reversal." If N is a power of two, and $P = Q = R = 2$, the effect is called *bit reversal.*

13) If the factors of N are equal, say to r, the algorithm is called a *base-r* or *radix-r* algorithm and, if the factors are different, it is called a *mixed-radix* algorithm.

14) A common notation is to let W or W_N represent the reciprocal of the Nth principle root of unity,

$$W = W_N = \exp\left(-j\frac{2\pi}{N}\right). \tag{52}$$

Some authors have used

$$W = \exp\left(j\frac{2\pi}{N}\right). \tag{53}$$

15) For the base-2 Cooley–Tukey form of the FFT algorithm, the most fundamental operation is of the form

$$X = A + W^kB$$
$$Y = A - W^kB. \tag{54}$$

The Sande–Tukey form can be obtained by solving for A and B in terms of X and Y,

$$A = 0.5(X + Y)$$
$$B = 0.5(X - Y)W^{-k}. \tag{55}$$

This gives an elementary operation of an inverse transform. The Sande–Tukey form of the forward transform (DFT) is obtained by eliminating the 1/2 and replacing W^{-k} by W^k. From the appearance of the system flow graph for these operations, each operation is called a *butterfly*.

16) For diagramming the flow of processing data for the FFT, a flow graph notation is used. The flow graph for a trivial eight-point FFT appears as follows:

$$F(k) = \sum_{n=0}^{7} f(n)W^{nk}, \qquad k = 0, 1, \cdots, 7 \tag{56}$$

where

$$W = \exp\left(-j\frac{\pi}{4}\right). \tag{57}$$

The flow graph is shown in Fig. 15.

17) Each node represents a variable and the arrows terminating at that node originate at the nodes whose variables contribute to the value of the variable at that node. The contributions are additive, and the weight of each contribution, if other than unity, is indicated by the constant written close to the arrowhead of the contribution. Each node is assigned a pair of indices n, L. Variables at nodes in row n replace each other as they are computed and are stored in the cell with index n. All nodes in a column L are computed on iteration L. In this form of the algorithm, the exponent of W on the line entering node $(n; L)$ is $n \cdot 2^{3-L}$ (mod 8). It can also be seen that for each pair of operands, the second W is $W^4 = -1$ times the first and that the "butterfly" operation is indeed given by (54).

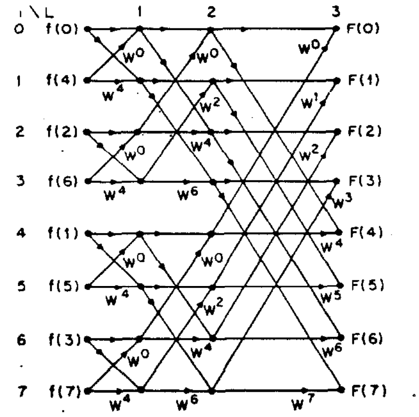

Fig. 15. The flow graph for an eight-point FFT.

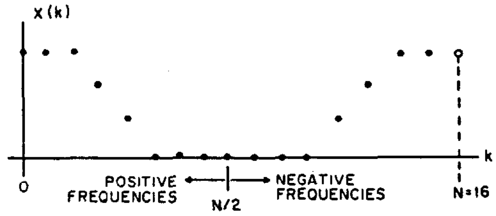

Fig. 16. The locations (in frequency) of the DFT points for a 16-point transform.

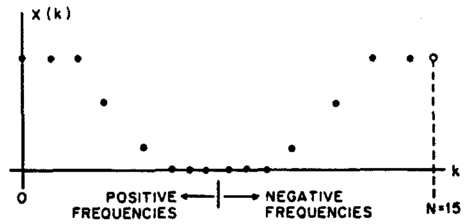

Fig. 17. The locations (in frequency) of the DFT points for a 15-point transform.

VII. Discrete Convolution and Spectrum Analysis

1) The Fourier transform (i.e., the z transform evaluated for $z = e^{j2\pi fT}$) of a sampled waveform is periodic in frequency, i.e., $X(e^{j2\pi T(f+f_s)}) = X(e^{j2\pi Tf})$ where $f_s = 1/T$ = sampling frequency. For this reason it is convenient to represent a *negative frequency* as the equivalent *positive frequency* below the sampling frequency, i.e., $f = -f_s/20$ is equivalent to $f = -f_s/20 + f_s = 19/20\, f_s$. In this manner one need only describe the spectrum in the positive frequency range of $0 \le f < f_s$. For the DFT this convention also holds. Thus a typical spectrum of a 16-point DFT is shown in Fig. 16. For the above example the number of DFT points was even (16). The first DFT point, $X(0)$, corresponds to the Fourier transform evaluated at 0 frequency. The $(N/2+1)$st DFT point, $X(N/2)$, corresponds to the Fourier transform evaluated at half the sampling frequency. The index k corresponds to a frequency $f = k/TN$ in $X(e^{j2\pi Tf})$.

2) If the number of points in the DFT were odd, say $N = 15$, a typical spectrum would be as shown in Fig. 17. In this case there is no DFT point which corresponds to evaluating the Fourier transform at half the sampling frequency.

3) The *discrete convolution* of two sequences can be computed from the inverse discrete Fourier transform of the product of the discrete Fourier transforms of the two sequences. Thus, if $X(k)$ and $Y(k)$ are the discrete Fourier transforms of $x(n)$ and $y(n)$, the inverse discrete Fourier transform of the product of these discrete Fourier transforms produces a *periodic discrete convolution*, also called a *cyclic or circular discrete convolution* or simply a cyclic convolution.

4) The cyclic convolution can be written algebraically as

$$\sum_{m=0}^{N-1} x((n-m))y((m)) \tag{58}$$

where $((m))$ means the index is taken modulo N. Equation (58) can be written as

$$\sum_{m=0}^{n} x((n-m))y((m)) + \sum_{m=n+1}^{N-1} x((n-m))y((m)), \tag{59}$$

or, alternately,

$$\sum_{m=0}^{n} x((m))y((n-m)) + \sum_{m=n+1}^{N-1} x((m))y((n-m+N)). \tag{60}$$

The simultaneous presence of both summations is generally undesirable for computing ordinary convolutions. For example, if the first summation in (60) is chosen to represent the desired convolution, then the second summation represents an error term. By augmenting both sequences with zeros so that they have the same length N, which is at least as great as one less than the sum of the lengths of the two sequences, cyclic convolution can be made to yield the same result as ordinary convolution.

5) This use of the fast Fourier transform to compute discrete convolutions is sometimes called *fast convolution* or *FFT convolution*. This technique can be easily adapted for computing acyclic (i.e., nonperiodic) correlation functions. In this form, it is called *fast correlation or FFT correlation*.

6) If one of the two sequences is much shorter than the other, the longer sequence can be *sectioned* into pieces whose discrete convolutions can be computed separately. These discrete convolutions can be combined to produce the discrete convolution of the whole sequence. (*Sectioning* is used because it reduces the required amounts of computation and memory.) One of these sectioning techniques (*overlap-save* or *select-save*) involves computing the inverse DFT of the product of the DFT's of a) N samples of the input sequence, and b) the shorter sequence augmented with a sufficient number of zeros so that its sequence contains N samples. (Usually N is at least twice as large as the length of the shorter sequence.) Some of the members of the sequence resulting from the inverse DFT are members of the sequence formed by the desired acyclic (i.e., nonperiodic) convolution. (This number of members equals one more than the number of zeros originally augmenting the shorter sequence.) The longer original sequence is advanced by this number of members. Iterating this process gives the whole convolution. The *overlap-add* technique for sectioning uses a similar technique but additionally requires adding shifted sequences of partial convolutions.

7) A *window* is a finite sequence, each element of which multiplies a corresponding element of the main sequence. (This is called *windowing*.) The sequence of products formed by this element-by-element multiplication is often more useful than the main sequence. The Fourier transform of a typical window (sometimes called the *spectral window*) consists of a *mainlobe*, which usually contains a large percentage of the energy in the window, and *sidelobes* which contain the remaining energy in the window.

8) Windows can be used in estimating power spectra. In the *direct method*, the power spectrum is estimated by computing the square of the absolute value of the DFT of the windowed sequence. The DFT of the windowed sequence is the convolution of the DFT's of the window and the original sequence. This convolution smoothes the input power spectrum, consequently values of the power spectrum at frequencies separated by less than the width of the mainlobe of the spectral window cannot be *resolved*. In addition to this limit on *resolution*, the estimate of the power spectrum may contain significant *leakage*, i.e., erroneous contributions from components of the power spectrum at frequencies possibly distant from the frequency of interest because of the nonzero energy in the spectral window sidelobes.

9) Windows are useful in determining the coefficients of a finite impulse response digital filter. In this case, the original sequence consists of samples of the impulse response corresponding to a transfer function which is approximated by the Fourier transform of the sequence of pairwise products; the product sequence is used as the coefficients of the finite impulse response digital filter.

10) Windows are used also in the *indirect method* of computing a power spectrum. In this method, the sequence consisting of samples of the autocorrelation function is multiplied by the window. The DFT of the resulting sequence is an estimate of the power spectrum.

11) Windows can be used also in estimating cross spectra where the estimates are obtained by multiplying the products of the DFT's of two or more distinct sequences.

12) The determination of a finite impulse response described by an ordinary convolution is called *deconvolution* or *FIR identification*.

Acknowledgment

The authors would like to acknowledge the comments and criticisms of this paper provided by G. D. Bergland, C. H. Coker, D. L. Favin, B. Gold, O. Hermann, S. Lerman, A. V. Oppenheim, H. O. Pollak, and H. F. Silverman.

Alphabetic Index of Terms

ACOUSTICAL SOCIETY OF AMERICA STANDARDS SECRETARIAT

The Standards Program of the Acoustical Society of America is the responsibility of the ASA's Committee on Standards (ASACOS) and is executed by ASA's Standards Secretariat headed by its Standards Manager.

The Acoustical Society of America holds the Secretariat for four accredited standards committees of the American National Standards Institute (ANSI); S1 on Acoustics, S2 on Mechanical Shock and Vibration, S3 on Bioacoustics, and S12 on Noise; and provides the United States input to ISO/TC 43 on Acoustics, for which S1 and S3 serve as Technical Advisory Groups, and S12 serves as the Technical Advisory Group for ISO/TC43/SC1 Noise. ASA also administers the international secretariat of ISO/TC 108 on Mechanical Vibration and Shock (on behalf of the American National Standards Institute) and provides the U.S. input via the Technical Advisory Group for ISO/TC 108, which is Standards Committee S2 (Standards Committee S3 is the Technical Advisory Group for ISO/TC108/SC4 Human Exposure to Mechanical Vibration and Shock).

Standards are produced in four broad areas: physical acoustics, mechanical shock and vibration, bioacoustics, and noise, and are reaffirmed or revised every five years. The latest information on current ANSI standards as well as those under preparation is available from the Standards Secretariat.

When ordering ANSI S1.11-1986, please also use ASA Catalog No. 65-1986 for easy reference. This standard and other standards listed in the ASA Standards Catalog may be ordered from the following address:

Publication Sales, Dept. STD
American Institute of Physics
335 East 45th Street
New York, New York 10017
Telephone: (516) 349-7800

Standards listed in the ASA Standards Catalog can be ordered by out-of-state customers by calling the American Institute of Physics WATS number, 1-800-AIP-PHYS Toll Free. Also, AIP accepts major credit cards (American Express, VISA, MasterCard) with a minimum order of $10.00.

If you wish to have further information on the Standards Publication Program of the Acoustical Society of America, address your inquiries to

Standards Manager
Standards Secretariat
Acoustical Society of America
335 East 45th Street
New York, New York 10017
Telephone: (212) 661-9404