

A Hierarchical Approach for Audio Capture, Archive, and Distribution

J. ROBERT STUART,¹ *AES Life Fellow*, AND PETER G. CRAVEN,² *AES Life Member*
(jrs@mqa.co.uk) (peter@algol.co.uk)

¹*MQA Ltd., Huntingdon, PE29 6YE, UK*
²*Algol Applications Ltd., BN44 3QG, UK*

Recent interest in high-resolution digital audio has been accompanied by a trend to higher and higher sampling rates and bit depths, yet the sound quality improvements show diminishing returns and so fail to reconcile human auditory capability with the information capacity of the channel. We propose an audio capture, archiving, and distribution methodology based on sampling kernels having finite length, unlike the “ideal” *sinc* kernel that extends indefinitely. We show that with the new kernels, original transient events need not become significantly extended in time when reproduced. This new approach runs contrary to some conventional audio desiderata such as the complete elimination of aliasing. The paper reviews advances in neuroscience and recent evidence on the statistics of real signals, from which we conclude that the conventional criteria may be unhelpful. We show that this proposed approach can result in improved time/frequency balance in a high-performance chain whose errors, from the perspective of the human listener, are equivalent to those introduced when sound travels a short distance through air.

0 SETTING THE SCENE

By considering “High Resolution” to be an attribute of a *complete system* (from microphone to loudspeaker), rather than of the signal or a specific technology, we introduce a hierarchical method by which high resolution can be delivered efficiently.

0.1 Outline of This Paper

Sec. 1 introduces the digital audio chain as a transmission channel with analog converters at each end. These converters have properties that are described and placed in context of the debate on high resolution.

In Sec. 2 we consider the listener. By bringing together ethological and neuroscientific insights we derive a plausible framework which proposes that a “natural” and resolving playback chain need only introduce errors equivalent to those introduced by air. This takes us beyond lossless considerations into a framework where there is a noise-floor based on acoustic Brownian motion, where we can propose time-domain constraints based on the known human timescales for inter- and intra-aural time windows for correlation and where we can argue for causality.

In Sec. 3 we examine the signal. An analysis of a large corpus of stereo recordings shows us that the peak information content rarely exceeds 1.3 Mbps and can often be as low as 1 Mbps, allowing that even when a recording uses a

high sample rate, we could convey the information at much lower rates than currently.

In Sec. 4 we take a fresh look at digital sampling and introduce a hierarchical approach based on B-splines.

In Sec. 5 we describe a hierarchical delivery coding framework that can provide a transmission function modeled in the frequency, time and amplitude domains to be similar to a short column of air and which can accept all the information of the source recording and deliver it in a hierarchical manner to a variety of playback devices.

Although the quantity of data are irrelevant in an archive, efficiency can be critical in distribution, particularly when listening on the move.

This paper presents a conservative approach, based on the measured statistics of music and the physics of sound transmission, without recourse to adaptive processing or a varying noise floor and not implausibly pre-judging human auditory capability.

Due to the wide scope of the paper, many topics are introduced in references, which are grouped by topic in Sec. 9.

1 DIGITAL AUDIO

Recording preserves an analogy of the music waveform. Early recordings were mechanical and analog magnetic tape

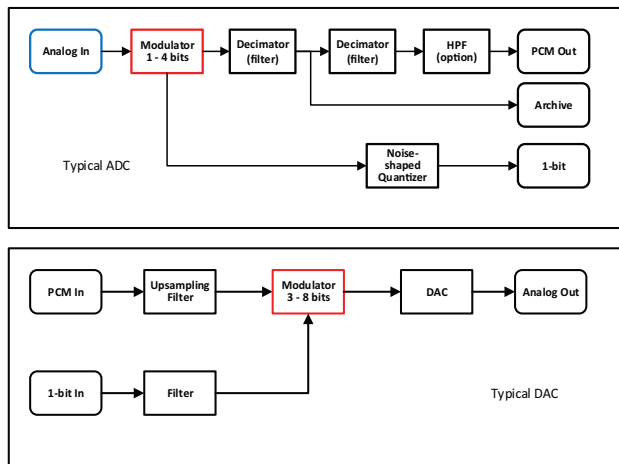


Fig. 1. Internal blocks in typical integrated-circuit converters; A/D (upper) and D/A (lower).

followed. More recently the signal has been brought into a digital representation.

Once the audio information is contained within digital data it can be transmitted through time or space losslessly and playback can be substantially repeatable, although careless coding or incautious signal processing may introduce distinctive problems; a topic tackled in some detail in [3]. However, the most critical steps remain at the analog-digital (A/D) and digital-analog (D/A) gateways and in the compromises or permanent limitations made at these points.

1.1 Technology and Limitations

Early converters operated at the base sample rate, delivering directly the final PCM stream. Fig. 1 illustrates typical internal architectures of delta-sigma converters which have been used widely for three decades. Oversampling delta-sigma structures permit simplified analog filtering and have the potential for the highest performance when using dither in a small-word-size hardware quantizer/modulator.¹ These concepts are explained in [5–7].

Even though it has significant problems as a release or distribution code, the direct output of the A/D modulator would be, from a perfectionist point of view, a more appropriate “archive” than either the decimated multi-bit PCM output or the noise-shaped, quantized single-bit stream, both of which require some processing as shown in Fig. 1.

In fact, an ideal system might connect the A/D modulator output directly to its counterpart in the D/A converter, as shown in Fig. 2, although the data rate would be high.²

The modulators operate at a high sample rate chosen to optimize their performance, while for the PCM signal

¹ In its most extreme form the modulator is 1 bit and the converter can sample or reconstruct at 64 or more times F_s [4]

² Modern converters rarely use 1-bit modulators. 384 kHz and 8 bits might also be a candidate [22], particularly if the sampling kernel had been chosen according to the principles shown in Sec. 4 of this paper.

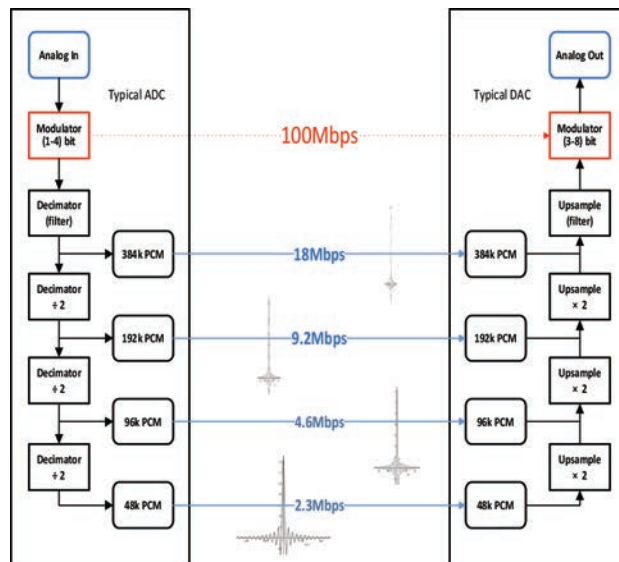


Fig. 2. The converters of Fig. 1 are redrawn (A/D left and D/A right). For convenience, the decimation (A/D) and upsampling (D/A) filters are shown as cascades enabling different PCM transmission rates. The lower the transmission rate, the more signal processing intervenes between the two modulators.

passed from the A/D to the D/A, the sample rate may be chosen arbitrarily.

The properties of the decimation and upsampling filters can significantly impact sound quality, and a considerable part of research into high-resolution audio has centered on these filters and on varieties of dither [8–20].

1.2. High Resolution

The term “high resolution” has a visual analogy.³ In optics, resolution or resolving power is the ability of a device to produce *separate* images of closely spaced objects; a high-resolution image has clarity, depth, absence of filtering or coding artifacts, little blur, and is rapidly assimilated. In an image we can measure resolution of details and the impact of coding or transmission, e.g., via a sensor or lens. *We perceive resolution as definition.*

In audio, high-resolution sound should also resemble real life: sounding natural; objects having clear locations (position and distance) and separate readily into perceptual streams (through absence of noise, distortion, time-smearing or modulation effects), particularly where environmental reverberation causes multiple arrivals closely separated in time—temporal resolution of microstructure in sound being somewhat analogous to spatial resolution in vision. With this perspective, high resolution should be considered an attribute of a *complete chain* and therefore more correctly described in the analog domain [21].

When considering the frequency and time responses of an end-to-end distribution channel, we must bear in mind

³ Many of the problems arising when high-resolution audio is discussed arise from poor definition or agreement on the terms [21–23].

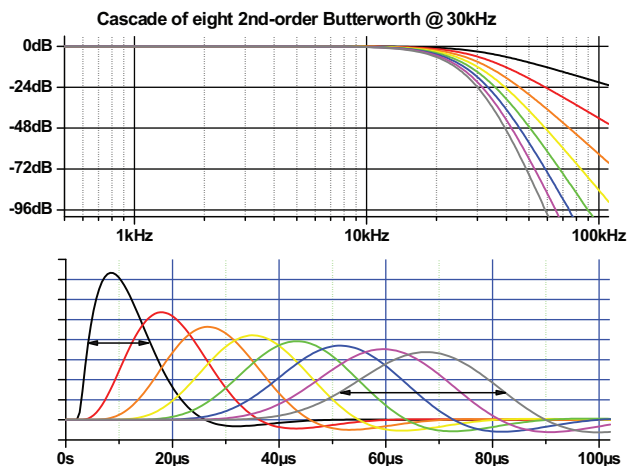


Fig. 3. Showing the frequency and impulse responses of a cascade of eight 2nd-order Butterworth low-pass filters.

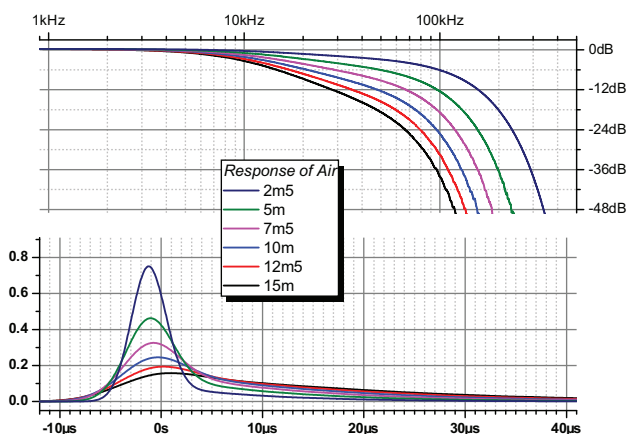


Fig. 4. Attenuation of sound in air at STP (standard temperature and pressure) and 30% RH (relative humidity); frequency and impulse responses varying with distance. Based on model and data from [33, 147].

that temporal-dispersion can build up through a cascade of otherwise blameless components.

Fig. 3 illustrates the response of such a cascade built up of eight stages, each with a 2nd-order roll-off at 30 kHz, plausibly representing a microphone, preamplifier, mixer, converter pre- and post-filters, replay pre- and power amplifier, and transducer. Such a *chain* may disguise aspects that a wider-band replay system might reveal and could confuse listening tests [24, 25].

A more severe viewpoint, based on ethological considerations, proposes that an ideal high-performance chain would be one whose “errors,” from the perspective of the human listener, are equivalent to those introduced by sound traveling a short distance through air. Within reasonable limits, air does not introduce distortion or modulation noise, but it does blur sound, progressively attenuating higher frequencies and we readily adapt to its effect. See Fig. 4 [21, 33].

A system having similar properties, if placed between the listener and the performer, might not be noticed. Similarly,

we should carefully consider the inclusion of a component whose transmission errors might be considered “unnatural.”

In the last decade it has become more common for recording professionals to self-select higher sample- or data-rate formats to improve sound quality. It’s not uncommon to find recordings being laid down at 192 or even 352.8 kHz in 24- or 32-bit precision; through this, data rate has unfortunately become a proxy for resolution.

Higher-than-CD data rate doesn’t guarantee improved sound quality, but doubling or quadrupling sample rate from 44.1 or 48 kHz has shown incremental improvements [26–32].

It is now accepted that one benefit of higher sample rates isn’t conveying spectral information beyond human hearing but the opportunity to modify the dispersive properties of filtering. Wider-transition anti-alias and reconstruction filters provide opportunity for short impulse response and there is also opportunity to apodize [11, 12] to remove extended pre- and post-rings.

Providing the sampling kernel⁴ is not too extended and that each quantization is properly dithered, then transient events can be accurately located in time [6, 7].

We must bear in mind that, even though higher sampling rates can convey frequencies above 20 kHz, this does not necessarily mean that such frequencies directly benefit our impression. As will be shown later, higher sample rates do allow shorter details to be captured and enable encoding kernels that provide much less uncertainty of an event’s duration, but again, we do not listen to impulses and must keep a clear line between our engineering descriptions and the listener’s experience.

2 THE LISTENER

The quality of an audio channel can only be finally judged in its intended use: “conveying meaningful content to human listeners.” The auditory sciences (psychoacoustics and neuroscience) help us to bridge listeners’ impressions and engineering.

2.1 Psychoacoustics and Modeling

With care to context, psychoacoustics can help us estimate the audible consequence of imperfect “conveying,” allowing errors arising in the recording chain to be ranked. Essentially any change can be isolated and modeled to estimate its impact in context; a special case is to estimate when channel errors might be inaudible.

Fundamental characteristics of the hearing system are complexity and non-linearity. To the listener, sounds have pitch and loudness rather than frequency and intensity, and the relationships between these measures are non-linear. Some non-linearities are extreme, such as: thresholds; detectability or loudness of a stimulus incorporating adjacent frequency elements; and masking by components slightly further away in time or frequency.

⁴ As will be explained, sampling is modeled by convolution with a kernel such as a sinc function, followed by instantaneous sampling.

Psychoacousticians have designed auditory experiments that explore the limits of the human hearing system as a receiver—and which, in general, attempt to minimize the impact of cognition [36, 45].

However, it is important to consider the higher-level process of cognition—where sounds take on meaning. In cognition, higher-level processes modify the listener’s ability to discriminate more, less or differently than indicated by the perceptual model. This process, in which, in the ascending neural pathway, elements of the arriving sound are grouped, for example by envelope, pitch movement, correlated timing, location, memory, and expectation is very complex, mathematically non-linear, and confounding to simple experiments [34–41].

In the cognitive process we hear “objects” rather than “stimuli” and we distinguish “what” from “where.” Mechanisms such as auditory streaming exploit similarity, contrast, and other cues to modify the basic percepts; so there is a risk that system errors that correlate to the signal, for example modulation or quantization noise, can attach to and modify “perceived objects” [42].

2.2 Neuroscience and Modeling

Recently there has been considerable progress towards understanding how we hear, in particular, in the related disciplines of neuroscience and computational neuroscience; introductory texts include [43] and [44].

Neuroscience provides a second framework and the approach tends to be different. Rather than devise archetypical experiments to select between alternatives [45], it is sometimes more useful to consider how neurons respond to the complexity of the natural world in which stimuli are not known in advance but might instead be partitioned into “scenes” and “objects” chosen from large but representative sets.

Regarding natural auditory stimuli, three important classes are the background sounds of the environment, animal vocalizations, and speech. In ensembles, all three exhibit self-similarity and a general spectral tendency for amplitude to fall with frequency; environmental sounds show a $1/f$ trend, see Fig. 5 [63–67].

Hearing is important for survival and we can’t wait too long to make a decision. Steady-state signals are not normal; an averaging detector might take too long. So, a better model is of a “running commentary” guided by attention and memory; trying to seek out or make sense of the sounds as they arrive. To parse this running commentary, we can’t always “rewind” into the short-term auditory memory and so strategies that robustly extract acoustic features in the presence of noise or interference have evolved.

Our ability to rapidly externalize objects or to follow speech or a melody is amazingly robust, and we can understand an extensively modified or damaged stream of sound and even induce missing fragments, although degraded stimuli probably increase cognitive load. However, in the current context, we want to avoid straying into the area where meaning survives but subtlety and ultimate realism do not.

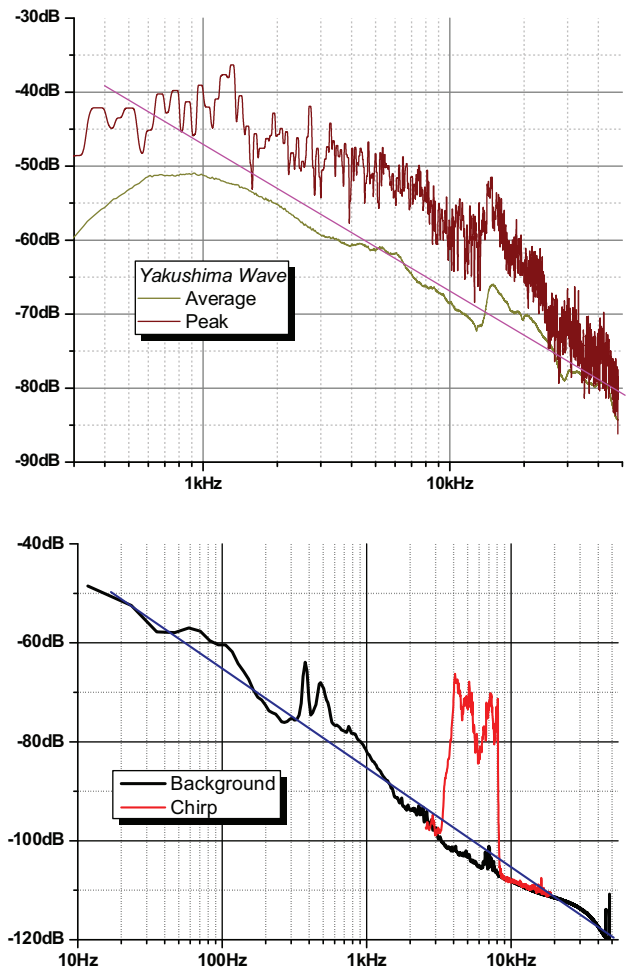


Fig. 5. Upper: spectral analysis of a recording of waves breaking on a beach [62]. Lower: Environmental sound and bird call (Chirp). Included in both are lines at -20 dB/decade showing how the spectra follow the expected $1/f$ trend.

When we listen, it isn’t the acoustic waveform or spectrum that we interpret but the spikes from around 30,000 afferent inner-hair-cell cochlear neurons—whose actions, in turn, are ultimately modified by a similar number of efferent (descending) neurons, some of which connect to the cochlear outer hair cells.

As the signals travel through the brain stem, the mid-brain, and on to the auditory cortex (wherein finally, we “hear”), tonotopically organized neurons, initially coding for level, spectrum, modulations, onset, and offset pass through complex combining structures that exchange, encode or extract a variety of temporal, spectral, and ethological features [46–49, 53–56].

By exploiting population coding, temporal resolution can approach $8 \mu\text{s}$ and this precision reflects neural processing, rather than being strictly proportional to our 18 kHz bandwidth (an estimate of the upper “bin” of the cochlea and upper limit of pitch perception). The role of the descending neurons is not yet completely understood. At a simplified level they are implicated in gain control, in modifying feature extraction through attention, and conscious and unconscious control of the outer-hair-cell active process which is

responsible for mechanical gain and “filter width” implied in basilar-membrane motion. This idea that auditory-filter width can be responsive to attention and context has profound implications for detection and masking models.

In an important set of papers, Lewicki showed computational neural models proposing efficient auditory coding using kernels tuned to ensembles of natural sounds [57–59]. His models evolved highly efficient, “auditory filters” adapted to the three classes of natural sounds mentioned earlier and showed how each sound-class benefits from a different time-frequency balance and therefore filter bandwidth.

Filters adapted to animal vocalizations preferred fine frequency resolution. Speech drives fine frequency resolution in the region of 500 Hz but selects for temporal resolution above 1.5 kHz, whereas environmental sounds preferred fine temporal discrimination—particularly at high frequencies. Although a model, these findings augment our understanding and reinforce that “listening for” or “attending to” “objects” or “streams” might indeed involve direct control of both the cochlea and the ascending neural pathway [34].

Rieke et al. [60] describe neurons that respond to higher moments of the stimulus; e.g., high-frequency auditory neurons which are not sensitive to phase but instead encode the envelope of the sound-pressure [40, 47, 50, 52].

These findings in neuroscience guide us to speculate that audio can be more efficiently transmitted if the channel coding is optimized for natural sounds rather than specified with independent “rectangular” limits for frequency and amplitude ranges.

2.3 Temporal Limits

Human hearing is exquisitely complex and capable. At a basic engineering level, we have to consider it to be nonlinear, since our response to sound is the rapid perception of objects, the process of which involves not only iterative grouping and assignment of extracted features in the sound, but also that both cochlear and brainstem processes are directable interactively by both attention and the cortex [51–54, 69, 70].

Recent studies have highlighted aspects of this “non-linearity” with trained and untrained listeners’ ability to apparently exceed the Fourier uncertainty limit for time/frequency judgment and in a manner that depends on presentation order [71–74].

We routinely function in reverberant surroundings: sounds arrive many times, including with fine structure from the pinnae, yet we readily fuse these into coherent percepts, especially if the temporal-fine-structure is preserved.

In certain circumstances the human hearing system is highly sensitive to temporal features or closely located sound elements. Since, in principle, higher sample rates permit finer-grained details to be resolved it is important to understand where the limits for transparency may lie. In fact, transparency may not be the only measure; there has been occasional reported evidence that changing certain properties of the playback chain could lead to a more

involving or relaxed enjoyment of the music, sometimes over long intervals [75, 77].

This controversial topic has seen early hints of objective evidence in studies using EEG [75–86].

For the audio distribution channel, we can consider temporal resolution in two aspects: (i) its ability to maintain separation between closely spaced events (and not blur them together) and (ii) its ability to maintain a precise unquantized time-base within and between channels.

The first aspect can apply to all sound transmission systems. Low-pass filtering may ultimately impact the separation of nearby events (hinted at in Fig. 3) while filters in the digitizing process, that are sharper in the frequency domain and therefore more extended in the time-domain, may also bring uncertainty to transient events and smear backwards or forwards in time.

Our ability to localize sounds swiftly and accurately is vital for survival. Sound intensity and arrival time provide important binaural cues and humans can discriminate interaural time differences as low as 10 μ s for frequencies below 1.5 kHz [87–91] (we are most sensitive in the region 0.8–1 kHz [51, 52, 92–97, 99–101]) and as low as 6 μ s for sounds with ongoing disparities, such as in reverberation [40, 89, 93, 98, 102–104].

Other mechanisms have been investigated that hint to similar discrimination limits within a channel, i.e., monaurally, including temporal fine structure in pitch perception; the comprehension of speech against a fluctuating background [103, 104].

It has been suggested by Kunchur [106–108] that listeners can discriminate timing differences of the order of 7 μ s.

Woszczyk has also provided a convenient review of psychophysical and acoustic temporal factors [28].

In light of these psychophysical data, even though one limit on resolving events will always be the microphone system bandwidth, it would seem prudent to provide for an archive that can resolve 3 μ s. On the other hand, based on current recordings we have analyzed, and bearing in mind the response of microphones currently favored by recording engineers, a sensible target for today’s distribution system would be of the order of 10 μ s.

2.4 Spectral and Amplitude Limits

The standard hearing threshold for pure tones is shown in Fig. 6. This minimum audible field has a standard deviation of around 10 dB and individuals are to be found whose thresholds are as low as –20 dB SPL at 4 kHz. Although the high-frequency response cut-off rate is always rapid, some can detect 24 kHz at high intensity [113–120].

There are some fundamental limitations in analog electronics (such as thermal and shot noise) and in the air itself.

3 THE SIGNAL

3.1 Spectral Content of Music

There is significant content above 20 kHz in many types of music, as an analysis of high-rate recordings summarized

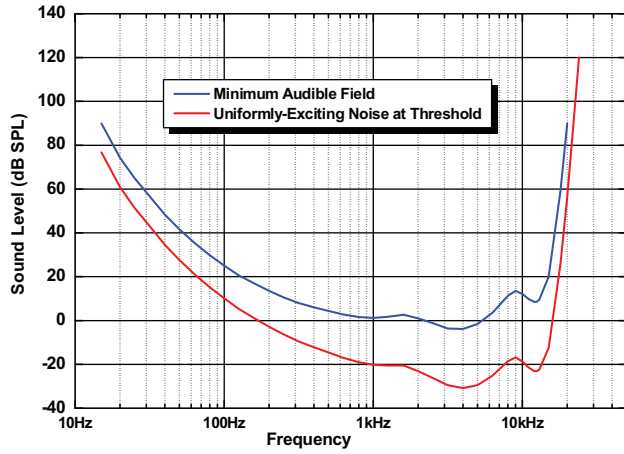


Fig. 6. The upper curve is the minimum audible field threshold for pure tones [110–112]. For evaluating noise spectra, the lower curve is uniformly exciting noise at threshold (in 1 Hz bandwidth), from [109].

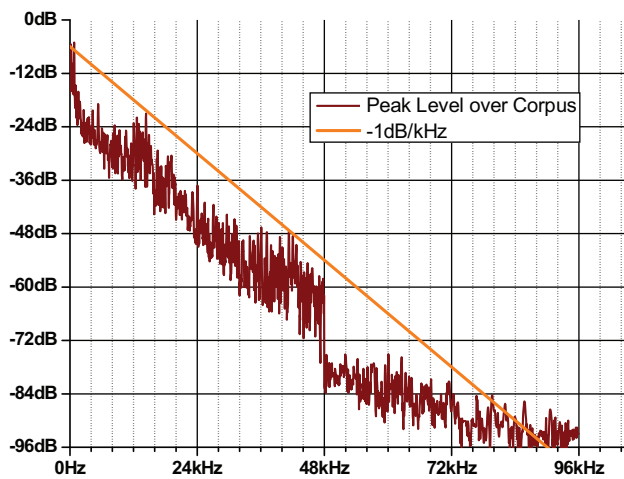


Fig. 7. Peak spectral level gathered over a corpus of 96- and 192-kHz recordings that have not clipped in the digital domain.

in Fig. 7 has revealed. One notable and common characteristic of musical instrument spectra is that the power declines, often significantly, with rising frequency.

Even though some musical instruments produce sounds above 20 kHz [121, 122] it does not necessarily follow that a transparent system needs to reproduce them; what matters is whether or not the means used to reduce the bandwidth can be detected by the human listener [123–126].

3.2 Noise in Recordings

Fig. 8 (upper) shows measurements of noise in a range of early analog tape and modern digital recordings. Obviously, these analyses embody the microphone and room noise of the original venue. The lower plot in Fig. 8 shows a summary of the lowest spectral noise in an analysis of commercial 24-bit recordings. This noise is expressed as TPDF (triangular probability density function) dither level, i.e., the bit-depth of a triangularly dithered quantization at 192 kHz having an equivalent noise spectral density [150].

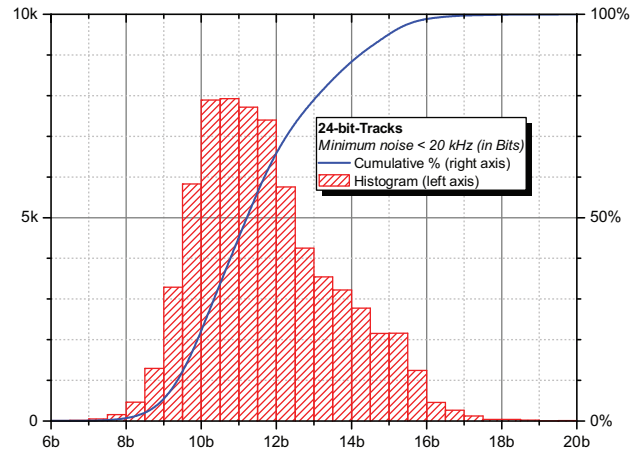
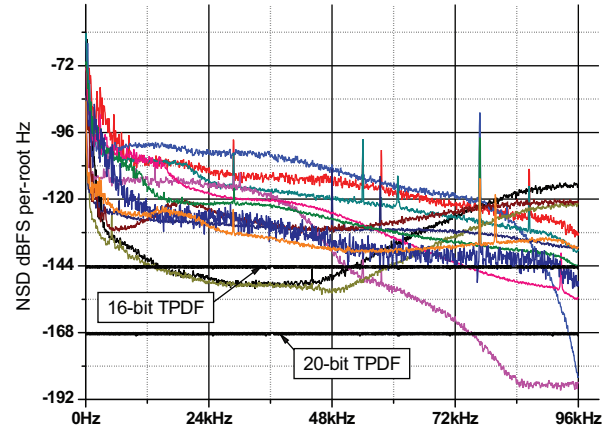


Fig. 8. Upper: Examples of background noise in 192 kHz 24-bit commercial releases. Also shown is TPDF dither noise for 192-kHz 16- and 20-bit quantization. Curves plotted as NSD (noise-spectral-density in 1-Hz bandwidth). Lower: An analysis of 100,000 commercial 24-bit recordings with sampling rates between 88.2 and 192 kHz, where the lowest spectral noise level, up to 20 kHz, is expressed as TPDF-bits.

It is worth noticing that a 20-bit PCM channel is adequate to contain these recordings and consequently 32-bit precision offers no clear benefit in this regard.

3.3 Environment and Microphones

Fellgett [128] derived the fundamental limit for microphones based on detection of thermal noise, shown for an omnidirectional microphone at 300° K in Fig. 9.

Cohen and Fielder included useful surveys of the self-noise for several microphones [127]. Their data showed one microphone with a noise-floor 5 dB below the human hearing threshold, but other commonly used microphones show mid-band noise 10 dB higher in level than just-detectable noise.

Inherent noise is less important if the microphone is close to the instrument, but for recordings made from a normal listening position then the microphone is a limiting factor on dynamic range—more so if several microphones are mixed. This further suggests that those recordings can be entirely distributed in 192 kHz channels using 18–20 bits.

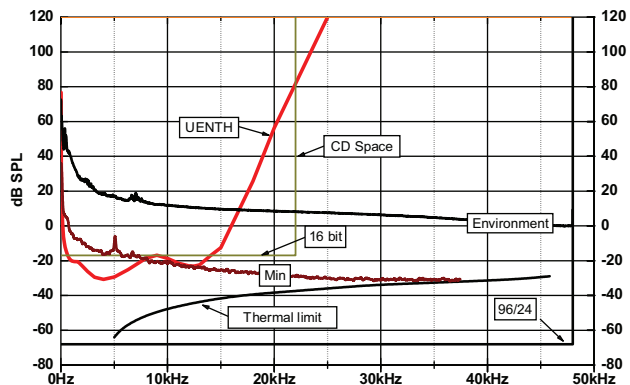


Fig. 9. Showing the noise-spectral density of the lowest-background recording analyzed (Min) set to a high replay gain of 0 dBFS = 120 dB SPL and in context of UENTH (uniformly exciting noise at threshold) from Fig. 6. Also shown are the thermal-noise microphone limit, the environmental background noise from Fig. 5, and coding spaces for CD and 96-kHz 24-bit PCM.

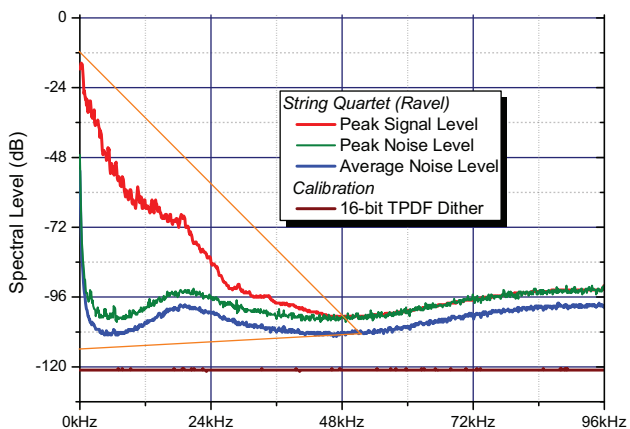


Fig. 10. Showing the peak spectral level and background noise in a 192-kHz 24-bit recording of the Guarneri Quartet playing Ravel’s String Quartet in F, 2nd movement [149].

3.4 Properties of Music

Content of interest to human listeners has temporal and frequency structure and never fills a coding space specified with independent “rectangular” limits for frequency and amplitude ranges. As noted in Sec. 2.2, environmental sounds show a 1/f spectral tendency. Ensembles of animal vocalizations and speech have self-similarity which leads to spectra that decline steadily with frequency. Music is similar, as seen in Fig. 7.

Fig. 10 shows peak spectral level and background noise for one recording; some significant features are apparent. First, the declining trend of peak level with frequency is typical, as is the background noise spectrum. We see here that at around 52 kHz the curves converge and above that region we must assume that noise will obscure any higher-frequency details of the content.

This picture of the content occupying a “triangular” space is common in more than 100,000 24-bit recordings we have analyzed and the converging point is usually below 48 kHz, with the highest so far being at 60 kHz.

From this spectral viewpoint we could deduce that the information content relating to the original signal in the channel occupies a triangular space (within the 192-kHz 24-bit outer envelope) equivalent to that of a stream having a peak data-rate of 960 kb/s/channel.⁵ The question is whether we can restrict the capture to just that signal-related information without disturbing the sound.

This insight has profound implications for the design of an efficient coding scheme.

4 A HIERARCHICAL APPROACH

When converting analog audio to a digital representation, the waveform is quantized in time and amplitude. Amplitude quantization using dither has been well described in the literature [5–10, 150], while system performance consequences are previously covered in [3, 21], so here we concentrate on time discretization using sampling and subsequent reconstruction to continuous time (analog). Sampling captures amplitude and timing information present in the original continuous time signal, while reconstruction presents that information in a form accessible to the ear.⁶

4.1 Sampling

In the several decades since both Shannon [129] and Nyquist [130], there has been considerable development in understanding of sampling theory [129–144]. Shannon’s theorem shows how appropriate band-limiting allows repeated resampling of a signal without build-up of alias products. If linear-phase brickwall filters are used throughout, a communications system can then be characterized by a single number, its bandwidth, which is the narrowest bandwidth of any of the filters or subsystems that have been used in cascade.

Digital audio has thus inherited the notion that “brick-wall” band-limiting is the ideal, thence the common specifications of passband, stopband, and transition band that measure the deviation of anti-alias filters from that ideal.⁷

⁵ In a Shannon diagram, the area between a signal’s peak spectrum and its background noise represents information content, excluding data that merely allows one to accurately reconstruct noise and other processing artefacts. When losslessly compressed, the bit-rate needed to convey the signal might halve.

⁶ More precisely, reconstruction to continuous time is the first step in rendering to an acoustic signal, for we are not proposing to present samples to the brain directly. Were we to do so, it would be arguable that the sampling kernel should mimic the cochlear kernel, which has a finite width [34]. For acoustic rendering, the requirement is that the total effect of sampling, reconstruction, and rendering plus the cochlear kernel should not be significantly different from that of the cochlear kernel alone. This unfortunately places a tighter time constraint on sampling and reconstruction processes.

⁷ Linear-phase brick-wall filters are critical for frequency-division multiplex transmission links, where packing of channels and low cross-talk are more important than subtle audio problems. Audio converters tended to use linear-phase filtering, not because the human is sensitive to relative phase at frequencies above 1 kHz (where wavelength approximates head-width), but because such

The impulse response of such an “ideal” Shannon-sampled system is a *sinc* function which has a fairly sharp central pulse but also a pre-ring and a post-ring, which build up and die away slowly, giving an extended time-response. In traditional communications practice this property has been accepted as a small price to pay for the system cascability conferred by the *sinc* filter.⁸ In high-resolution audio however we would hope to avoid arbitrary resampling: we wish to obtain the best sound from a single sampling process and a single reconstruction process. Cascability is thus not paramount and the time-response can be given some serious consideration.

Some may wonder how a time-domain analysis can tell us anything different from a more conventional frequency-domain analysis, since it is known that the frequency-domain and time-domain descriptions of a linear system are completely equivalent.

One answer is to consider that a Fourier analyzer uses a window that extends both forwards and backwards in time. Thus, although the two descriptions are equivalent if one considers the global signal, the frequency-domain description is very unhelpful in thinking about the situation at a particular point in time when the future of the signal is not known. Neurons in the brain-stem and cortex must make decisions to fire on the basis of the signals and correlations they see *now*.

Another interesting question is whether sampling can convey time differences that are shorter than the periods between successive samples. The answer depends on the sampling method. Supposing the input to be an analog impulse, it is clearly unsatisfactory to use instantaneous sampling because there is a danger that details that lie between the sampling points will be missed altogether.

The next possibility is to derive the samples by averaging the continuous-time signal over each sample interval, so all detail is certain to be “seen”. See Fig. 11 (upper), illustrating the use of an (analog) impulse as probe. Unfortunately, the averaged values will be the same wherever within a given interval the probe impulse lies and the answer to the above question is “no”.

We note that such averaging is equivalent to convolving the input with a rectangular sampling *kernel* followed by instantaneous sampling.⁹ In this case the rectangle spans one sampling period as shown in Fig. 11 (upper). Had we used a triangle as a kernel, as in Fig. 11 (lower), the impulse would almost always register at two sampling points. The ratio of the two sample values would indicate unambiguously where the impulse lay between the two sample points and the answer to the question would be “yes.”

filters can use less silicon area and power than minimum-phase designs. However, in many chip converters, the ideal is further compromised by use of a half-band filter.

⁸ Mathematically the *sinc* filter is idempotent: once it has been applied, further applications make no difference. Here however we are considering the total end-to-end processing, which is not cascaded so different considerations apply.

⁹ Linear-phase brickwall filtering is similarly described as convolution with a *sinc* kernel.

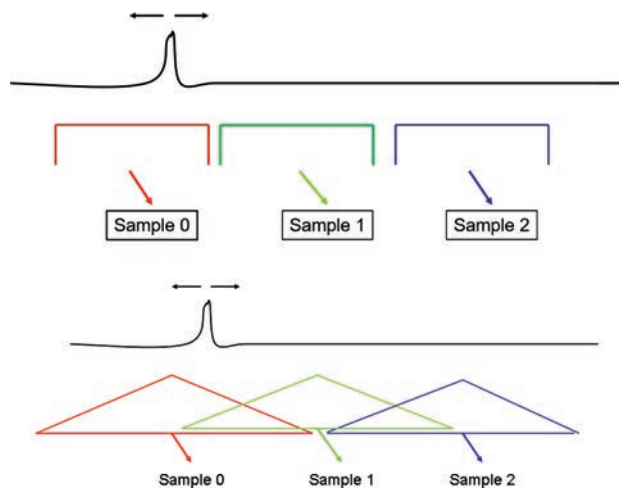


Fig. 11. Illustrating the kernels described in the text. Upper: rectangular kernels, where each sample is obtained by averaging over one sample period. Lower: where each sample is obtained by triangle-weighted integration kernels. For clarity, the triangles are shown as slightly separated. In reality, the first and last should touch.

The rectangle and triangle are examples of *B-splines*, a class of functions that we now explore.

Splines are made by joining polynomial segments, respecting some kind of continuity condition at each “knot” where one polynomial takes over from another. They are frequently used for interpolation, including interpolation between unequally-spaced data points.

4.2 B-Spline basics

In audio we generally assume that the knots will lie on a uniform sampling grid, so the knots are equally spaced. A general spline can be obtained by adding B-splines (“basis splines”), each of which takes the value zero outside a certain range. B-splines with equally-spaced knots (“cardinal splines” [148]) can be generated by convolving a Dirac δ -function with a rectangle “box” function zero or more times.

In Table 1 the order is the number of times the δ -function has been convolved and is also the total length of the spline (in units of the width of the rectangle).

For some purposes it may be more natural to refer to the degree of the polynomials making up the segments of a spline, as for example in the “cubic spline,” widely used for interpolation. The degree is one less than the order.

The Fourier transform of a rectangle function is the well-known *sinc*. In the case of a rectangle occupying one sample period, the corresponding *sinc* has a zero at the sampling frequency and all its multiples, the intervening peaks having amplitudes dying away in a manner proportional to $1/f$. For the triangle the die-away is proportional to $1/f^2$, while for general *order* it is proportional to $1/f^{\text{order}}$.

Thus, in the context of sampling, higher B-spline orders provide greater suppression of high frequencies. As the order tends to infinity, the B-spline approaches a Gaussian.

Table 1. B-splines of order 0, 1, 2, 3, and 4.






Order	Degree	Name		Fourier Transform
0	-	Dirac δ -function		1
1	0	Rectangle		sinc
2	1	Triangle		sinc^2
3	2	Quadratic B-spline		sinc^3
4	3	Cubic B-spline		sinc^4
Etc.				

Table 2. Binomial coefficients for the “two-scale relations” that allow a B-spline to be synthesized from two or more “cardinal” B-splines of half the width.

Name	Coefficients
Rectangle	(1, 1)/2
Triangle	(1, 2, 1)/4
Quadratic B-spline	(1, 3, 3, 1)/8
Cubic B-spline	(1, 4, 6, 4, 1)/16

The same principle can be extended to the higher-order splines. The convolution uses binomial coefficients in each case as shown in Table 2; illustration and formal equations for these *two-scale relations* may be found at [148]. It will be evident that this procedure of halving the sampling rate can be applied repeatedly to allow the rate to be reduced by any power of two. It follows that a spline-sampled signal at a low sample rate (for example 48 kHz) can be derived precisely from a spline-sampled signal archived at any power-of-two multiple of the low rate (e.g., 96, 192, 384 or 768 kHz, etc.).

The freedom to choose the archival rate in a manner that is transparent as far as a final distribution format is concerned is key to the *hierarchical* nature of the methodology. One can have precisely the same result whether the final signal has been spline-sampled directly from the analog signal, or whether it has been through one or more intermediate archives.

The “two-scale relations” can be used for both down-sampling and up-sampling, and it is envisaged that they could be used for the “Decimator $\div 2$ ” and the “Upsample $\times 2$ ” units in Fig. 2.

4.4 Sampling-Spline Order

In the framework presented here, instantaneous sampling corresponds to a B-spline of order zero. As we have already noted, that will not be satisfactory (unless there is pre-filtering in the analog domain). Order 1, the rectangle, does not allow time resolution better than one sample period. In contrast, the triangle (and also higher-order splines, the sinc and other sampling schemes), allows an arbitrarily small displacement of an impulse to be detected on the basis of waveform comparison, assuming one has sufficient signal-to-noise ratio.

Higher order splines permit the possibility of determining the positions and amplitudes of two or more impulses that might land in the same sample period, as explained in [134].¹⁰ This may not be directly relevant since the mathematics required to perform such a determination is not plausibly within the ear’s capability. Determining the position of a single isolated impulse is vastly more straightforward: for material sampled using B-splines of order 2 or higher it is merely a matter of determining the center-of-gravity of the subsequently reconstructed pulse. These determinations

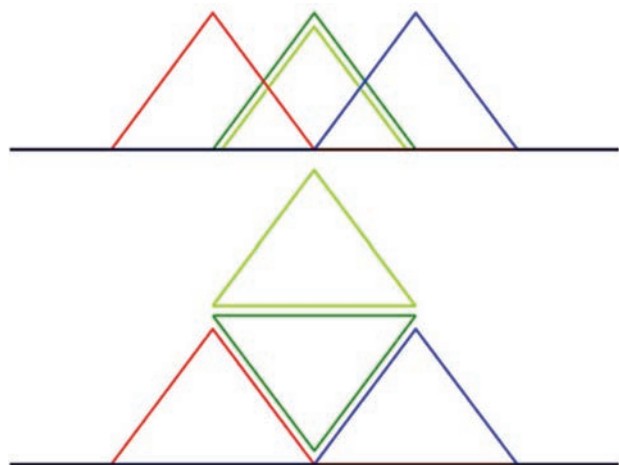


Fig. 12. Upper: Three triangular kernels centred on consecutive sample points, the central one shown doubled (dark green and light green) to represent the weight in the $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ convolution. Lower: the same rearranged to show the equivalence to a single triangle of twice the width.

4.3 Hierarchical Sampling Using B-Splines

Consider that two rectangles placed side-by-side and touching are equivalent to a single rectangle of twice the width. It follows that if a signal is sampled at a particular sampling rate using a rectangular kernel, simple averaging of pairs of the samples to furnish a stream at half the sampling rate will give the same result as sampling the signal directly at the half rate using a double-width rectangular kernel. This pairwise averaging is equivalent to convolving the original samples with the sequence $(\frac{1}{2}, \frac{1}{2})$ and selecting alternate convolved samples.

Similarly, if the signal is sampled using a triangular kernel, convolving with the sequence $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ and selecting alternate convolved samples will provide a half-rate stream identical to a stream produced by sampling at the half rate using a double-width triangle. Fig. 12 provides visual support for the equivalence of the two procedures.

¹⁰ This paper is one of several that highlight possibilities for non-traditional sampling methods, listed in 9.17.

Table 3. First four rows: sampling and reconstruction both of order two, showing the reduction of 20-kHz droop with increasing flattener order. The extent increases but the 20–80% step response becomes faster. Last four rows: showing increasing droop and extent with increasing sampling spline order.

Sampling spline order	Flattener order	20-kHz droop ($f_s = 96$ kHz)	Extent (samples)	20–80% step response (samples)
2	0	2.5 dB	4	1.00
2	1	0.61 dB	4.5	0.78
2	2	0.14 dB	5	0.70
2	3	0.03 dB	5.5	0.66
3	3	0.05 dB	6.5	0.72
4	3	0.09 dB	7.5	0.78
5	3	0.13 dB	8.5	0.83
6	3	0.19 dB	9.5	0.88

provide relative timings that are precisely accurate, independently of the exact details of the reconstruction [145].

Another consideration is vulnerability to high frequency noise on the input signal. If the input has a white noise spectrum, instantaneous sampling will pick up an infinite amount of noise unless there is an analog bandwidth limitation. With rectangular sampling the noise is finite but still with a significant contribution from downward-aliased components. Triangular (spline order 2) sampling reduces this contribution to insignificance given a white noise spectrum, however a rising input noise spectrum may suggest using a spline of order 3 or higher [134, 135, 139–142].

A particular case of a steeply rising input spectrum is the output of a noise-shaped A/D modulator. It is standard practice [146] to perform at least the initial stages of down-sampling using CIC (cascaded-integrator-comb) filters. An integrator-plus-comb combination implements a discrete approximation to a rectangular kernel, so a cascade of n integrator-combs can be used to implement a spline kernel of order n . A B-spline kernel of order 5 or 6 is normally sufficient to control the noise from a modulator of order 4 or 5.

The “hierarchical” methodology can be applied to the whole chain, from the A/D modulator to the PCM signal presented to the listener. Thus, using the hierarchical methodology, *the sampling frequency at the output of the A/D converter becomes somewhat arbitrary.*

4.5 Reconstruction

Reconstruction can be regarded as the dual of sampling. It is not recommended to present unfiltered Dirac spikes to subsequent equipment. Even if each sample is presented as a rectangle having a width of one sample period, the slew-rates at the transitions will theoretically be infinite.

We advocate B-splines as a means of softening the transitions between samples while not extending the impulse response more than necessary. A B-spline of order 2 (triangle of width two sample periods), equivalent to linear interpolation between samples, thus appears to be the least

that is needed to reconstruct an analog signal that can be handled satisfactorily¹¹ [133, 141].

If sampling uses a triangular kernel and reconstruction also presents each sample as a triangle, then the combined analog-to-analog impulse response¹² is a 4th-order B-spline, of total width four sampling periods (42 μ s at 96 kHz).

Such a 4th order B-spline impulse response implies a droop in frequency response of 2.5 dB at 20 kHz if sampling at 96 kHz.

This droop can be corrected without introducing pass-band ripples or pre-responses using a maximally-flat minimum-phase FIR flattening filter. The filter will inevitably extend the total impulse response but this effect can be minimized by running it at a higher (integer-multiple) sample rate. Table 3 illustrates the case of transmission at 96 kHz but with the filter running at 192 kHz. The extent of the impulse response is increased by one half of a transmission sample period for each order of flattening, hence a total extent of 5.5 samples for 2nd order spline sampling and 3rd order flattening as shown in the table, reducing the 20 kHz droop to 0.03 dB.

For sources with steeply rising high-frequency content a higher-order sampling spline may be used: performance is compromised only slightly, as exemplified by the 5th order sampling spline in Table 3. Faster rise-times may be obtained by running at a higher transmission rate as shown in Fig. 13 (lower); doubling to 192 kHz (from 96 kHz) brings us closer to the 3 μ s future target suggested in Sec. 2.3.

4.6 Transparency

Continuing the argument from Sec. 3.4, we can infer from Fig. 10 that the noise-floor of the recording is prolifically described by a 24-bit channel.

¹¹ Particular situations may argue for higher-order splines giving smoother reconstruction but, in this paper, we assume second-order reconstruction throughout.

¹² This is the response averaged over all the possible positions of a test impulse relative to the sampling points.

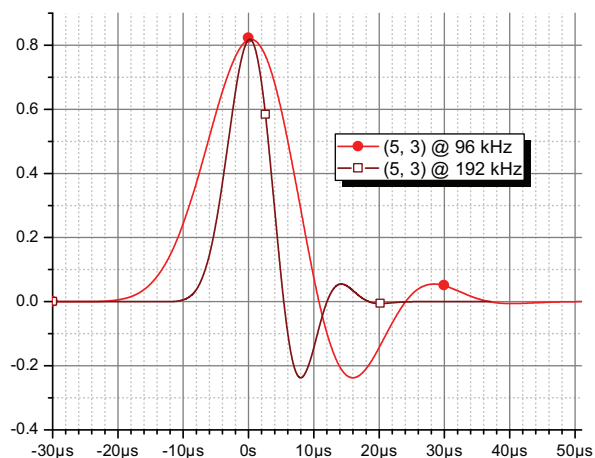
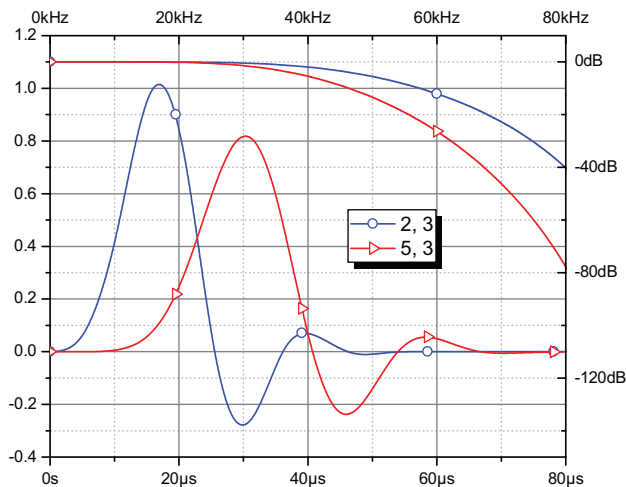


Fig. 13 Upper: comparing analog-to-analog frequency and impulse responses of two systems transmitting at 96 kHz, both using 3rd-order flattening with either 2nd- or 5th order splines (from Table 3). Lower: comparing the 5th order system when operated at 96- and 192 kHz.

Since in a dithered system, word-size provides a measure of dynamic range rather than of precision or resolution, with care, the word-size could be reduced for distribution with no audible impact [150].

Psychoacoustic modeling and listening tests show us that, providing the noise from re-quantization stays more than 12 dB below the original noise spectral density at frequencies below 15 kHz, there is no audible consequence [3, 34, 44, 109].

Fig. 14 shows spectra relating to the same 192-kHz recording as in Fig. 10.

The red (open squares) curve shows peak spectral density after convolving with a filter related to the 5th order B-spline, which attenuates higher frequency components above 48 kHz. When resampled at 96 kHz, frequencies that lie above 48 kHz in the filtered spectrum fold back to mirror-image positions below 48 kHz in the down-sampled spectrum, as shown in brown (filled squares).

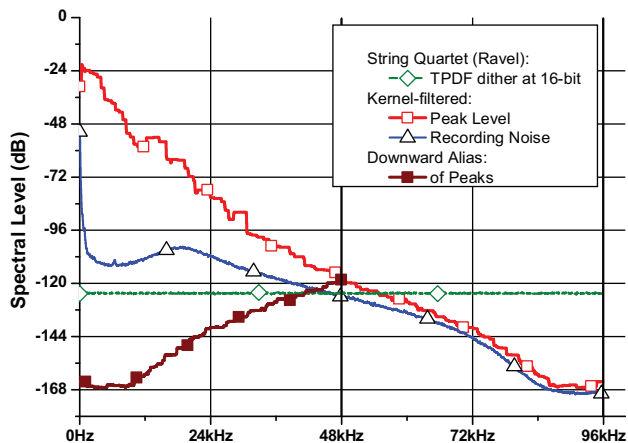


Fig. 14. Showing the kernel-filtered noise and peak spectrum along with aliasing, as described in the text.

Loss of signal information is minimal. From Fig. 10 we deduce that nearly everything above 48 kHz is noise from the recording system without tonal qualities.

We cannot chop off this content without introducing pre-responses or increasing blur: the sampling process *replicates* the content and at the frequency where the replica is reproduced it is less than the kernel-filtered noise from the original recording and at least 40 dB below for image frequencies under 20 kHz.

Fig. 14 also shows that the resampling could be benignly quantized to around 16 bits, preferably selecting appropriate dither with possibly mild noise-shaping, whereupon these aliased components would be covered with an inaudible noise.¹³ We therefore assert that the audible effect of these aliased images is minuscule—an assertion supported by very detailed listening with recording professionals who have helped us confirm this coding paradigm. (See Sec. 7.)

Aliasing in the frequency domain is equivalent to the time-domain phenomenon of an impulse response that depends on where, relative to the sampling instants, the original stimulus was presented. However, as noted in Sec. 4.4, the first moment (or “center-of-gravity”) of a transient event will be correctly reconstructed and determine its perceived position. Alternatively, in the frequency-domain, the downward aliased components are concealed by original noise, while for the upward aliased components we rely on plausibility arguments, verified by listening to the final result, that these alias products lying above 48 kHz, are inherently inaudible and low enough in level to avoid slew-rate problems or to protrude above the replay noise floor.

Of course, there is also blur caused by the kernel filter and it might be supposed that the sampling and reconstruction filter would inevitably degrade the sound to some small extent. However, this blur is less than conventional methods and listening tests using commercial 192-kHz material consistently show very positive results.

¹³ At any sensible acoustic gain that dither would be below the threshold of hearing.

While these concepts might surprise some, the theory of sampling has evolved considerably since Shannon and Nyquist. Moreover, in several other disciplines, such as image processing or astronomy, it has been found that under-sampling can increase resolution with careful application-specific thinking [131–144].

4.7 Real-World Performance

Using filters similar to the spline filters described here, we have been able to take a 192-kHz sampled signal, re-sample to 96 kHz for more economical transmission to a listener, then resample again to 192 kHz in order to optimally feed a D/A converter in the listener’s decoder.

The downsampling filter has six taps at 192 kHz and the upsampling filter (which includes flattening as described above) also has six taps at 192 kHz, giving a combined response of 11 taps.

Assuming that a preceding A/D samples using a triangular kernel and that a following D/A reconstructs also with a triangular kernel, the end-to-end response will introduce considerably less blur than transmission at 96 kHz using conventional filters, as shown in Fig. 15 (middle and lower).

The end-to-end response can also be compared with air, as shown in Fig. 16.¹⁴

We thus have recipes for downward and upward conversion within a hierarchy of rates such as 44.1, 88.2, 176.4, and 352.8 kHz, however these methods do not provide satisfactory conversion from, for example, 96 kHz to 88.2 kHz. This is a reason why it is not recommended that the down-sampled signal be stored in the archive.

Even if it sounds wonderful it is “locked” into its own sample-rate family and cannot be transported to another without some loss.

If a recording has been archived at 192 kHz and it is required to produce an 88.2-kHz version, a suitable procedure would be first to convert the sample rate to 176.4 kHz by conventional means, using severe filtering to suppress aliases, and then to convert to 88.2 kHz using the methods described here. The filtering implied by this second conversion can be expected to provide substantial suppression of ringing and other artifacts near 88.2 kHz caused by the first sample rate converter.

5 DISTRIBUTION SYSTEM

The methodology described above can be used to design a hierarchical audio digitization, archival, and distribution system.

If we are starting with an analog signal, many commonly-used A/D chips employ CIC filtering in the first decimation stage (e.g., see Fig. 2) to reduce the high rate modulator output, for example 5.6448 MHz, to a more manageable frequency such as 705.6 kHz. In the limit of high sampling frequency, the CIC filter response becomes a close approximation to a B-spline kernel.

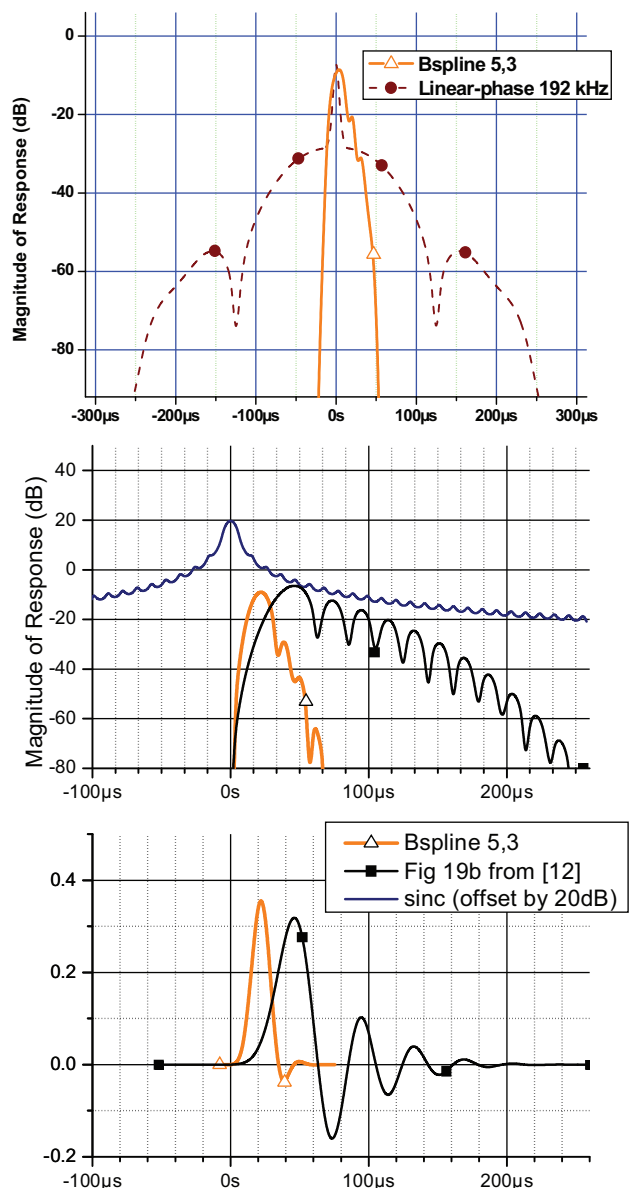


Fig. 15. Impulse responses: Upper: envelope of the 5,3 B-spline method at 96 kHz compared with a typical linear-phase cascade at 192 kHz, as dB magnitude. Middle: envelope of the 5,3 B-spline method at 96 kHz compared with a 96-kHz apodized design from [12] (Fig. 19b as squares). Lower: as middle but plotted as waveform (arbitrary units) not dB envelope. The new method shows substantially improved temporal fidelity over the earlier designs, even when run at half the sample rate.

For simplicity the A/D converter output rate should be a power-of-two multiple of the final rate delivered to the consumer. Archiving and mastering can then be done at any intermediate power-of-two multiple, the rate conversions being performed using the two-scale relations discussed in Sec. 4.3. With this methodology the final delivery is conceptually a B-spline-sampled signal at the final rate, with the freedom to insert an archiving step at an intermediate sample rate without making any change to the final result.

For compatibility with existing playback equipment, the signal’s frequency response may be flattened using an

¹⁴ Calculated from the equations in Appendix A of [147] on the assumption of minimum phase.

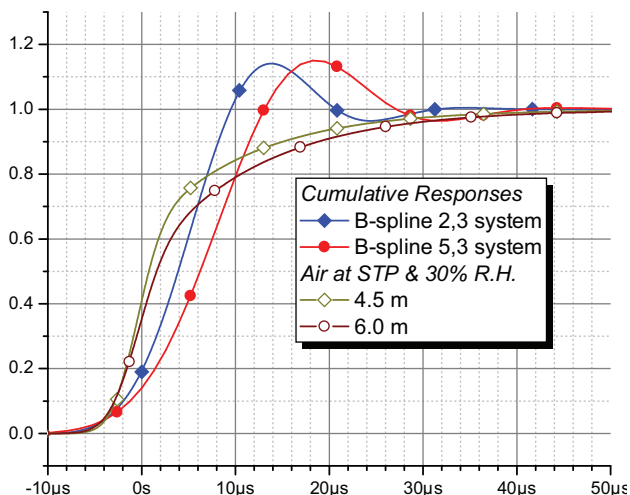


Fig. 16. Showing the cumulative end-to-end impulse responses for 2,3 and 5,3 B-spline systems. Also shown are the equivalent responses for 4.5 and 6 m of air at STP and 30% RH.

invertible filter prior to consumer delivery. New equipment should invert the flattening filter to recover the B-spline representation before re-constructing to analog using the method outlined in Sec. 4.5.

Alternatively, starting from an existing high-rate PCM or DSD encoding, the sample rate can be brought down using the same methods, ignoring the fact that the starting point is not truly a B-spline sampled signal; the difference may be not important if the original sampling rate is high enough.

Using the coding concepts described in Sec. 4, it is possible to re-code a signal presented originally as PCM so as to preserve both spectral and temporal features in a smaller coding space. The encoding kernel may be chosen separately for each song (track) on the basis of signal analysis but should be kept constant for that segment to avoid correlated noise modulation.

The receiver (decoder) should implement an appropriate up-sampling reconstruction, a flattening filter matching the chosen encoding kernel, and a platform-specific D/A manager.

Conceptually, we are trying to connect the A/D and D/A modulators together with a signal that encapsulates the entire sound of the original—but without artefacts that imply lack of resolution—and to package it for efficient distribution.

When starting from PCM, it is not necessary to reconstruct the original sample rate, the correct choice depending on the hardware available. For example, a higher overall performance, avoiding quantization steps, may result by feeding a D/A converter at the highest rate it will accept; this can be understood with reference to the processing blocks in Fig. 2.

This method is also efficient. For example, the Ravel segment illustrated in Fig. 10 can be encapsulated into a distribution file containing all the relevant spectral and temporal information of the 192-kHz 24-bit original (9.2 Mbps) using an average data rate of 930 kbps.

6 CONCLUDING REMARKS

It is a fact of modern digital-audio life that some signals are not band-limited, some are taken “outside Nyquist” by compression or overload, anti-alias filters are not ideal, and quantizations are not always dithered. However, in the context of distribution, we show that self-similarity in signals allow us to employ innovation-rate concepts while optimizing for temporal accuracy—appropriate for separating and locating environmental and music sounds.

Using insights from the auditory sciences we review targets for dynamic range, frequency response, and time response. We propose that for digital distribution, overall analog-to-analog temporal “blur” makes a better performance metric than sample rate; an upper limit of 10 μ s blur should ensure transparency.

We advocate distribution using hierarchical up/down-sampling, lossless compression, and lossless processing.

We suggest that for the current music archive, an efficient distribution channel-coding may be based on spline kernels that provide a music-appropriate coding, this being paired with complementary reconstruction at playback. Resampling should preferably be avoided, except within the same sample rate family and performed using the “hierarchical” methods described here. The aim is to ensure that degradations introduced by the analog–digital–analog signal path are comparable to those of sound passing a short distance through air.

This approach to re-coding can result in superior sound and significantly lower data-rate when compared to unstructured encoding and playback.

To potentiate archives, we recommend that modern digital recordings should employ a wideband coding system that places specific emphasis on time and frequency and sampling at no less than 352.8 kHz.

7 ACKNOWLEDGMENTS

This paper covers a sustained enquiry and the authors are particularly grateful to our co-workers in MQA and Algol; in particular to Spencer Chrislu, Hiroaki Suzuki, Trefor Roberts, Alan Wood, Michael Capp, Meredydd Luff, Richard Hollinshead, Malcolm Law, and Cosmin Frateanu.

Our enquiry involved many listening sessions and in-studio comparisons of sources and coding options that could not have been accomplished without particular help from Warner Music (especially Craig Kallman, Mike Jbara, George Lydecker, Scott Levitin, Justin Smith, and Craig Anderson) and Universal Music (Barak Moffet, Pat Kraus, and Tadashi Takagi). Sony Music also enabled in-studio listening and we are very grateful to Steve Berkowitz and to Brooke Eittlee and Mark Wilder of Battery Studios.

George Lydecker, Mark Wilder, and Morten Lindberg helped us to capture detailed measurements and characterizations of various analog tape recorders, desks, and AD/DA converters and Thomas Bårdsen helped with digital recorder measurements.

Many mastering engineers participated in recording, mastering, and listening sessions including Bob

Ludwig, George Massenberg, Bruce Botnik, Mandy Parnell, Morten Lindberg, Gonzalo Nonque, Ian Sheppard, Mick Sawaguchi, Reiji Asakura, and members of JPRS. We would also like to thank Keith Johnson, Peter McGrath, David Chesky, and many others.

Thanks also for many stimulating discussions on auditory science with Profs. Brian Moore, Hiroshi Nittono, Wieslaw Woszczyk, and Tsutomu Oohashi.

8 NEUTRALITY STATEMENT

This paper describes a target for a transparent recording chain for music and natural sounds intended for human listeners. It covers a theoretical basis and illustrates a general framework for a transmission method that achieves the target parameters. Whereas the authors have been involved in the development of a practical system based on this general thinking, this paper is not a description of any particular system.

9 REFERENCES

References are grouped into topic groups in approximate order of introduction in the paper.

9.1 The audio channel

[1] J. R. Stuart and P.G. Craven, “A Hierarchical Approach to Archiving and Distribution,” presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), convention paper 9178. Open Access: <http://www.aes.org/e-lib/browse.cfm?elib=17501>

[2] S. Peus, “Measurements on Studio Microphones,” presented at the *103rd Convention of the Audio Engineering Society* (1997 Sep.), convention paper 4617. <http://www.aes.org/e-lib/browse.cfm?elib=7162>

[3] J. R. Stuart “Coding for High-Resolution Audio Systems,” *J. Audio Eng. Soc.*, vol. 52, pp. 117–144 (2004 Mar.). <http://www.aes.org/e-lib/browse.cfm?elib=12986>

9.2 Digital audio filters and quantization

[4] S. P. Lipshitz and J. Vanderkooy, “Why 1-Bit Sigma-Delta Conversion Is Unsuitable for High-Quality Applications,” presented at the *110th Convention of the Audio Engineering Society* (2001 May), convention paper 5395. <http://www.aes.org/e-lib/browse.cfm?elib=9903>

[5] B. Widrow and I. Kollár, *Quantization Noise: Round-off Error in Digital Computation, Signal Processing, Control, and Communications* (CUP, Cambridge, UK, 2008).

[6] S. P. Lipshitz and J. Vanderkooy, “Pulse-Code Modulation—An Overview,” *J. Audio Eng. Soc.*, vol. 52, pp. 200–214 (2004 Mar.). <http://www.aes.org/e-lib/browse.cfm?elib=12991>

[7] J. Vanderkooy and S. P. Lipshitz, “Digital Dither: Signal Processing with Resolution Far Below the Least Significant Bit,” presented at the *AES 7th International Conference: Audio in Digital Times* (1989 May), conference paper 7-014. <http://www.aes.org/e-lib/browse.cfm?elib=5482>

[8] P. G. Craven and M. A. Gerzon, “Compatible Improvement of 16-Bit Systems Using Subtractive Dither,” presented at the *93rd Convention of the Audio Engineering Society* (1992 Oct.), convention paper 3356. <http://www.aes.org/e-lib/browse.cfm?elib=6777>

[9] M. A. Gerzon and P. G. Craven, “Optimal Noise Shaping and Dither of Digital Signals,” presented at the *87th Convention of the Audio Engineering Society* (1989 Oct.), convention paper 2822. <http://www.aes.org/e-lib/browse.cfm?elib=5872>

[10] M. A. Gerzon, P. G. Craven, J. R. Stuart, and R. J. Wilson, “Psychoacoustic Noise Shaped Improvements in CD and Other Linear Digital Media,” presented at the *94th Convention of the Audio Engineering Society* (1993 Mar.), convention paper 3501. <http://www.aes.org/e-lib/browse.cfm?elib=6647>

[11] P. G. Craven, “Controlled Pre-Response Antialias Filters for Use at 96 kHz and 192 kHz,” presented at the *114th Convention of the Audio Engineering Society* (2003 Mar.), convention paper 5822. <http://www.aes.org/e-lib/browse.cfm?elib=12588>

[12] P. G. Craven, “Antialias Filters and System Transient Response at High Sample Rates,” *J. Audio Eng. Soc.*, vol. 52, pp. 216–242 (2004 Mar.). <http://www.aes.org/e-lib/browse.cfm?elib=12992>

[13] J. R. Stuart and R. J. Wilson, “Dynamic Range Enhancement Using Noise-Shaped Dither at 44.1, 48, and 96 kHz,” presented at the *100th Convention of the Audio Engineering Society* (1996 May), convention paper 4236. <http://www.aes.org/e-lib/browse.cfm?elib=7538>

[14] Acoustic Renaissance for Audio, “DVD: Pre-emphasis for Use at 96 kHz or 88.2 kHz” (1996 Nov.) (described in [3]).

[15] J. Dunn, ‘Anti-Alias and Anti-Image Filtering: The Benefits of 96 kHz Sampling Rate Formats for Those Who Cannot Hear above 20 kHz,’ presented at the *104th Convention of the Audio Engineering Society* (1998 May), convention paper 4734. <http://www.aes.org/e-lib/browse.cfm?elib=8446>

[16] J. R. Stuart and R. J. Wilson, “A Search for Efficient Dither for DSP Applications,” presented at the *92nd Convention of the Audio Engineering Society* (1992 Mar.), convention paper 3334. <http://www.aes.org/e-lib/browse.cfm?elib=6799>

[17] J. R. Stuart and R. J. Wilson, “Dynamic Range Enhancement Using Noise-shaped Dither Applied to Signals with and without Pre-emphasis,” presented at the *96th Convention of the Audio Engineering Society* (1994 Feb.), convention paper 3871. <http://www.aes.org/e-lib/browse.cfm?elib=6361>

[18] M. Akune, R. M. Heddle, and K. Akagiri, “Super Bit Mapping: Psychoacoustically Optimized Digital Recording,” presented at the *93rd Convention of the Audio Engineering Society* (1992 Oct.), convention paper 3371. <http://www.aes.org/e-lib/browse.cfm?elib=6762>

[19] R. Lagadec, “New Frontiers in Digital Audio,” presented at the *89th Convention of the Audio Engineering Society* (1990 Sep.), convention paper 3002. <http://www.aes.org/e-lib/browse.cfm?elib=5691>

[20] J. R. Stuart, “Auditory Modeling Related to the Bit Budget,” presented at the *AES UK 9th Conference: Managing the Bit Budget* (1994 May), conference paper MBB-18. <http://www.aes.org/e-lib/browse.cfm?elib=6110>

9.3 High resolution

[21] J. R. Stuart, “Soundboard: High-Resolution Audio,” *J. Audio Eng. Soc.*, vol. 63, pp. 831–832 (2015 Oct.). Open Access <http://www.aes.org/e-lib/browse.cfm?elib=18046>

[22] ADA, “Proposal of Desirable Requirements for the Next Generation’s Digital Audio,” *Advanced Digital Audio Conference*, Japan Audio Society (1996 Apr.).

[23] JEITA “On the Designation of High-Res Audio, 25 JEITA-CP No 42 (2014 Mar.).

9.4 Listening tests and evaluation

[24] J. D. Reiss, “A Meta-Analysis of High Resolution Audio Perceptual Evaluation,” *J. Audio Eng. Soc.*, vol. 64, pp. 364–379 (2016 Jun.). <https://doi.org/10.17743/jaes.2016.0015>

[25] H. M. Jackson, M. D. Capp, and J. R. Stuart, “The Audibility of Typical Digital Audio Filters in a High-Fidelity Playback System,” presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), convention paper 9174. <http://www.aes.org/e-lib/browse.cfm?elib=17497>

[26] A. Pras and C. Guastavino, “Sampling Rate Discrimination: 44.1 kHz vs. 88.2 kHz,” presented at the *128th Convention of the Audio Engineering Society* (2010 May), convention paper 8101. <http://www.aes.org/e-lib/browse.cfm?elib=15398>

[27] T. Nishiguchi and K. Hamasaki, “Differences of Hearing Impressions among Several High Sampling Digital Recording Formats,” presented at the *118th Convention of the Audio Engineering Society* (2005 May), convention paper 6469. <http://www.aes.org/e-lib/browse.cfm?elib=13185>

[28] W. Wozczyk, “Physical and Perceptual Considerations for High-Resolution Audio,” presented at the *115th Convention of the Audio Engineering Society* (2003 Oct.), convention paper 5931. <http://www.aes.org/e-lib/browse.cfm?elib=12372>

[29] M. Story, “A Suggested Explanation for (Some of) the Audible Differences between High Sample Rate and Conventional Sample Rate Audio Material,” <http://www.cirilca.com/include/aes97ny.pdf> (1997 Sep.).

[30] S. Yoshikawa, S. Noge, M. Ohsu, S. Toyama, H. Yanagawa, and T. Yamamoto, “Sound Quality Evaluation of 96-kHz Sampling Digital Audio,” presented at the *99th Convention of the Audio Engineering Society* (1995 Oct.), convention paper 4112. <http://www.aes.org/e-lib/browse.cfm?elib=7654>

[31] B. Leonard, “The Downsampling Dilemma: Perceptual Issues in Sample Rate Reduction,” presented at the *124th Convention of the Audio Engineering Society*

(2008 May), convention paper 7398. <http://www.aes.org/e-lib/browse.cfm?elib=14528>

[32] G. H. Plenge, H. Jakubowski, and P. Schone, “Which Bandwidth Is Necessary for Optimal Sound Transmission,” presented at the *62nd Convention of the Audio Engineering Society* (1979 Mar.), convention paper 1449. <http://www.aes.org/e-lib/browse.cfm?elib=2905>

9.5 Acoustics

[33] G. W. C. Kay and T. H. Laby, “Tables of Physical and Chemical Constants,” section 2.4.1, online at NPL, <http://bit.ly/1rkaKGv>.

9.6 Human auditory perception

[34] C. J. Plack (ed.), *The Oxford Handbook of Auditory Science: Hearing*, 3rd ed. (OUP, 2010).

[35] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press, 1990).

[36] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models* (Springer, Berlin, 1990), vol. 22.

9.7 Auditory neuroscience

[37] J. M. Oppenheim et al., “Minimal Bounds on Nonlinearity in Auditory Processing,” q-bio.NC. arXiv:1301.0513 (2013 Jan.).

[38] T. Deneux et al., “Temporal Asymmetries in Auditory Coding and Perception Reflect Multi-Layered Nonlinearities,” *Nature Communications*, Article number: 12682 (2016 Sep.). <https://dx.doi.org/10.1038/ncomms12682>

[39] C. A. Atencio et al., “Cooperative Nonlinearities in Auditory Cortical Neurons,” *Neuron*, vol. 58, pp. 956–966 (2008 Jun.). <https://doi.org/10.1016/j.neuron.2008.04.026>

[40] P. X. Joris, “Envelope Coding in the Lateral Superior Olive. II. Characteristic Delays and Comparison with Responses in the Medial Superior Olive,” *J. Neurophysiol.*, vol. 76, no. 4, pp. 2137–2156 (1996 Oct.). <https://doi.org/10.1152/jn.1996.76.4.2137>

[41] D. A. Hall and D. R. Moore, “Auditory Neuroscience: The Saliency of Looming Sounds,” *Current Biology*, vol. 13, R91–R93 (2003 Feb.). [https://doi.org/10.1016/S0960-9822\(03\)00034-4](https://doi.org/10.1016/S0960-9822(03)00034-4)

[42] W. A. Yost et al., “Auditory Perception of Sound Sources,” *Springer Handbook of Auditory Research*, vol. 29 (Springer Science+Business Media, 2008).

[43] J. Schnupp et al., *Auditory Neuroscience: Making Sense of Sound* (MIT Press, 2011).

[44] A. Rees and A. R. Palmer (eds.), *The Oxford Handbook of Auditory Science: The Auditory Brain*, 2nd ed. (OUP, 2010).

[45] D. M. Green and J. A. Swets *Signal Detection Theory and Psychophysics* (Wiley: New York, 1966).

[46] T. M. Talavage et al., “Tonotopic Organization in Human Auditory Cortex Revealed by Progressions of Frequency Sensitivity,” *J. Neurophysiology*, vol. 91, pp. 1282–1296 (2004). <https://doi.org/10.1152/jn.01125.2002>

[47] D. Oertel et al., “Integrative Functions in the Mammalian Auditory Pathway,” *Springer Handbook of Auditory Research*, vol. 15 (Springer Verlag, 2002).

9.8 Auditory coding of space and time

[48] J. Ahvenin, N. Kopco, and I. P. Jääskeläinen, “Psychophysics and Neuronal Bases of Sound Localization in Humans,” *Hearing Research*, vol. 307, pp. 86–97 (2014). <https://doi.org/10.1016/j.heares.2013.07.008>

[49] A. J. King, J. W. H. Schnupp, and T. P. Doubell, “The Shape of Ears to Come: Dynamic Coding of Auditory Space,” *Trends in Cognitive Sciences*, vol. 5, no. 6, pp. 261–270 (2001 Jun.).

[50] A. Brand, O. Behrend et al., “Precise Inhibition Is Essential for Microsecond Interaural Time Difference Coding,” *Nature*, vol. 417, pp. 543–547 (2002 May). <https://doi.org/10.1038/417543a>

[51] I. Siveke, S. Ewert, B. Grothe, and L. Wiegand, “Psychophysical and Physiological Evidence for Fast Binaural Processing,” *J. Neuroscience*, vol. 28, no. 9, pp. 2043–2052 (2008 Feb.). <https://doi.org/10.1523/JNEUROSCI.4488-07.2008>

[52] T. T. Takahashi, “The Neural Coding of Auditory Space,” *J. Exp. Biol.*, vol. 146, pp. 307–322 (1989 Sep.).

[53] T. Deneux et al., “Temporal Asymmetries in Auditory Coding and Perception Reflect Multi-Layered Non-linearities,” *Nature Communications* (Sep. 2016). DOI: <https://doi.org/10.1038/ncomms12682>

[54] G. Chechik et al., “Reduction of Information Redundancy in the Ascending Auditory Pathway,” *Neuron* vol. 51, pp. 359–368 (2006 Aug.). <https://doi.org/10.1016/j.neuron.2006.06.030>

[55] D. A. Abrams, “Population Responses in Primary Auditory Cortex Simultaneously Represent the Temporal Envelope and Periodicity Features in Natural Speech,” *Hearing Research*, vol. 348, pp. 31–43 (2017). <https://doi.org/10.1016/j.heares.2017.02.010>

[56] T. Maekawa, M. Honda, E. Nishina, N. Kawai, and T. Oohashi, “Structural Complexity of Sounds Necessary for the Emergence of the Hypersonic Effect: Estimation of Autocorrelation Order,” *Asiagraph J.*, vol. 8, no. 2, pp. 35–40 (2013).

9.9 Natural sounds and ethology

[57] M. S. Lewicki, “Efficient Coding of Natural Sounds,” *Nature Neurosci.*, vol. 5, pp. 356–363 (2002). <https://doi.org/10.1038/nn831>

[58] E. C. Bluvas, and T. Q. Gentner, “Attention to Natural Auditory Signals,” *Hearing Research*, vol. 305, pp. 10–18 (2013). <https://doi.org/10.1016/j.heares.2013.08.007>

[59] E. C. Smith and M. S. Lewicki, “Efficient Auditory Coding,” *Nature*, vol. 439, pp. 978–982 (2006 Feb.). <https://doi.org/10.1038/nature04485>

[60] F. Rieke et al., *Spikes: Exploring the Neural Code* (MIT Press, 1997).

[61] E. Simoncelli and B. Olshausen, “Natural Image Statistics and Neural Representations,” *Annual*

Rev. Neuroscience, vol. 24, pp. 1193–1216 (2001). <https://doi.org/10.1146/annurev.neuro.24.1.1193>

[62] M. Sawaguchi, “Yakushima Wave,” 192 kHz field recording, private communication.

[63] R. F. Voss and J. Clarke “1/f Noise in Music and Speech,” *Nature*, vol. 258, pp. 317–318 (1975).

[64] M. N. Geffen et al., “Auditory Perception of Self-Similarity in Water Sounds,” *Frontiers in Integrative Neuroscience*, vol. 5, art. 15, pp. 1–11 (2011 May). <https://doi.org/10.3389/fnint.2011.00015>

[65] J. A. Garcia-Lazaro, B. Ahmed, and J. W. H. Schnupp, “Tuning to Natural Stimulus Dynamics in Primary Auditory Cortex,” *Current Biology*, vol. 16, pp. 264–271 (2006 Feb.). <https://doi.org/10.1016/j.cub.2005.12.013>

[66] B. De Coensel et al., “1/f Noise in Rural and Urban Soundscapes,” *Acta Acustica*, vol. 89, pp. 287–295 (2003).

[67] M. Yang et al., “Presence of 1/f Noise in the Temporal Structure of Psychoacoustic Parameters of Natural and Urban Sounds,” *J. Acoust. Soc. Amer.*, vol. 138, pp. 916 (2015). <https://doi.org/10.1121/1.4927033>

[68] R. F. Potter and P. D. Bolls, *Psychophysiological Measurement and Meaning: Cognitive and Emotional Processing of Media* (Routledge: New York, 2012)

9.10 Time vs Frequency

[69] D. Gabor, “Theory of Communication,” *J. Inst. Electr. Eng’rs*, vol. 93, no. III, p. 429 (1946 Nov.). <https://doi.org/10.1049/ji-1.1947.0015>

[70] D. Gabor, “Acoustical Quanta and the Theory of Hearing,” *Nature*, vol. 159, pp. 591–594 (1947). <https://doi.org/10.1038/159591a0>

[71] J. M. Oppenheim and M. O. Magnasco, “Human Time-Frequency Acuity Beats the Fourier Uncertainty Principle,” *Phys. Rev. Lett.*, vol. 110, 044301 (2013). <https://doi.org/10.1103/PhysRevLett.110.044301>

[72] J. M. Oppenheim et al., “Degraded Time-Frequency Acuity to Time-Reversed Notes,” *PLOS ONE*, vol. 8, pp. 1–6 (2013 Jun.). <https://doi.org/10.1371/journal.pone.0065386>

[73] M. Majka, P. Sobieszczyk, et al., “Hearing Overcomes Uncertainty Relation and Measure Duration of Ultrashort Pulses,” *Euro Physics News*, vol. 46, no. 1, pp. 27–31 (2015). <https://doi.org/10.1051/epn/2015105>.

[74] T. J. Gardner and M. O. Magnasco, “Sparse Time-Frequency Representations,” *Proc. Natl. Acad. Sci.*, vol. 103, pp. 6094–6099 (2006). <https://doi.org/10.1073/pnas.0601707103>

9.11 EEG response to sound

[75] R. Kuribayashi and H. Nittono, “High-Resolution Audio with Inaudible High-Frequency Components Induces a Relaxed Attentional State without Conscious Awareness,” *Frontiers in Psychology*, vol. 8, article 93 (2017 Feb.). <https://doi.org/10.3389/fpsyg.2017.00093>

[76] R. Kuribayashi, R. Yamamoto, and H. Nittono, “High-Resolution Music with Inaudible High-Frequency Components Produces a Lagged Effect

on Human Electroencephalographic Activities,” *Clinical Neuroscience NeuroReport*, vol. 25, no. 9 (2014). <https://doi.org/10.3389/fpsyg.2017.00093>

[77] A. J. Blood, and R. J. Zatorre, “Intensely Pleasurable Responses to Music Correlate with Activity in Brain Regions Implicated in Reward Emotion,” *Proc. Natl. Acad. Sci. USA*, vol. 98, pp. 1181–11823 (2001). <https://doi.org/10.1073/pnas.191355898>

[78] T. Kawai, “Hyper-Sonic Effect Study on Sound Environment Comfort by Brain Activation,” *Hypersonic Scenic Science*, vol. 83, pp. 290–295 (2013).

[79] T. Oohashi et al., “Inaudible High-Frequency Sounds Affect Brain Activity: Hypersonic Effect,” *J. Neurophysiol.*, vol. 83, pp. 3548–3558 (2000). <https://doi.org/10.1152/jn.2000.83.6.3548>

[80] N. Emi, “Hypersonic Effect and its Mechanism of Emergence,” *J. Acoust. Soc. Japan*, vol. 65, no. 1, pp. 40–45 (2009). https://doi.org/10.20697/jasj.65.1_40

[81] T. Oohashi et al., “Multidisciplinary Study on the Hypersonic Effect,” *International Congress Series*, vol. 1226, pp. 27–42 (2002). [https://doi.org/10.1016/S0531-5131\(01\)00494-0](https://doi.org/10.1016/S0531-5131(01)00494-0)

[82] S Nakamura et al. “Electroencephalographic Evaluation of the Hypersonic Effect,” *Soc. Neuroscience Abstract*, vol. 752, no.1 (2004).

[83] T. Harada, S Ito, et al., “Effect of High-Resolution Audio Music Box Sound on EEG,” *International Music J.*, vol. 23, no. 1, pp 1–3 (2016 Apr.).

[84] S. Ito et al., “Effects on the Autonomic Nervous System Function by the High-Resolution Music Box Sound,” *Medical Treatment and New Medicine*, vol. 52, no. 2, pp. 382–386 (2015 Feb.).

[85] S. Ito et al., “Effects of Differences in the Number of Quantization Bits of High-Resolution Music Box Sound on Autonomic Nervous System Function,” *Medical Treatment and New Medicine*, vol. 54, no. 2, pp. 137–140 (2017).

[86] H. Schulze and G. Langner, “Auditory Cortical Responses to Amplitude Modulations with Spectra above Frequency Receptive Fields: Evidence for Wide Spectral Integration,” *J. Comp. Physiol. A*, vol. 185, pp. 493–508 (1999). <https://doi.org/10.1007/s0035900504>

9.12 Audio temporal discrimination

[87] G. B. Henning, “Detectability of Interaural Delay in High-Frequency Complex Waveforms,” *J. Acoust. Soc. Amer.*, vol. 55, no. 1, pp. 84–90 (1974). <https://doi.org/10.1121/1.1928135>

[88] R. G. Klump and H. R. Eady, “Some Measurements of Interaural Time Difference Thresholds,” *J. Acoust. Soc. Amer.*, vol. 28, no. 5, pp. 859–860 (1956). <https://doi.org/10.1121/1.1908493>

[89] K. E. Hancock and B. Delgutte, “A Physiologically Based Model of Interaural Time Difference Discrimination,” *J. Neuroscience*, vol. 24, no. 32, pp. 7110–7117 (2004 Aug.). <https://doi.org/10.1523/JNEUROSCI.0762-04.2004>

[90] F. L. Wightman and D. J. Kistler “The Dominant Role of Low Frequency Interaural Time Differences

in Sound Localization,” *J. Acoust. Soc. Amer.*, vol. 91, pp. 1648–1661 (1992).

[91] W. A. Yost “Discrimination of Interaural Phase Differences,” *J. Acoust. Soc. Amer.*, vol. 55, pp. 1299–1303 (1974).

[92] J. O. Nordmark, “Binaural Time Discrimination,” *J. Acoust. Soc. Amer.*, vol. 35, no. 4, pp. 870–880 (1976).

[93] W. M. Hartmann and E. J. Macauley, “Anatomical Limits on Interaural Time Differences: An Ecological Perspective,” *Front. Neurosci.* (2014 Feb.). <https://doi.org/10.3389/fnins.2014.00034>

[94] A. Brughera, L. Dunai and W. M. Hartmann ‘Human Interaural Time Difference Thresholds for Sine Tones: The High-Frequency Limit,’ *J. Acoust. Soc. Amer.*, vol. 133, pp. 2839–2855 (2013). <https://doi.org/10.1121/1.4795778>

[95] L. R. Bernstein, “Auditory Processing of Interaural Timing Information: New Insights,” *J. Neurosci. Res.*, vol. 66, no. 6, pp. 1035–1046 (2001 Dec.). <https://doi.org/10.1002/jnr.10103>

[96] W. M. Hartmann et al., “Interaural Time Difference Thresholds as a Function of Frequency,” *Adv. Exp. Med. Biol.*, vol. 787, pp. 239–246 (2013). https://doi.org/10.1007/978-1-4614-1590-9_27

[97] R. C. G. Smith and R. R. Price, “Modeling of Human Low Frequency Sound Localization Acuity Demonstrates Dominance of Spatial Variation of Interaural Time Difference and Suggests Uniform Just-Noticeable Differences in Interaural Time Difference,” *PLoS One*, vol. 9, no. 2, pp. e89033 (2014). <https://doi.org/10.1371/journal.pone.0089033>

[98] H. Gaskell and G. B. Henning, “Forward and Backward Masking with Brief Impulsive Stimuli,” *Hearing Research*, vol. 129, pp. 92–100 (1999). [https://doi.org/10.1016/S0378-5955\(98\)00228-7](https://doi.org/10.1016/S0378-5955(98)00228-7)

[99] T. D. Rossing and A. J. M Houtsma, “Effects of Signal Envelope on the Pitch of Short Sinusoidal Tones,” *J. Acoust. Soc. Amer.*, vol. 79, pp. 1926–1933 (1986). <https://doi.org/DOI%10.1121/1.393199>

[100] M. Majka, P. Sobieszczyk, R. Gębarowski, and P. Zieliński, “Sub-Millisecond Acoustic Pulses: Effective Pitch and Weber-Fechner Law in Discrimination of Duration Times,” arXiv:1404.6464v2 [physics.class-ph] (2015).

[101] N. Kraus and T. White-Schwoch, Timescales of Auditory Processing,” *Hearing Matters*, vol. 69, no. 1, pp. 36–37 (2016 Jan.).

[102] G. B. Henning, “Monaural Phase Sensitivity with Ronken’s Paradigm,” *J. Acoust. Soc. Amer.*, vol. 70, no. 6, pp. 1669–1673 (1981 Dec.). <https://doi.org/10.1121/1.387231>

[103] B. J. C. Moore, “The Role of Temporal Fine-Structure Processing in Pitch Perception, Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People,” *J. Ass’n ‘Research’ Otolaryngology*, vol. 9, pp. 399–406 (2008). <https://doi.org/10.1007/s10162-008-0143-x>

[104] K. Krumbholz, R. D. Patterson, et al., “Microsecond Temporal Resolution in Monaural Hearing without Spectral Cues,” *J. Acoust. Soc. Amer.*, vol. 113, no. 5, pp. 2790–2800 (2003). <https://doi.org/10.1121/1.1547438>

[105] P. Heil and H. Neubauer, “A Unifying Basis for Auditory Thresholds Based on Temporal Summation,” *PNAS*, vol. 100, no. 10, pp. 6151–6156 (2003). <https://doi.org/10.1073/pnas.1030017100>

[106] M. N. Kunchur, “Temporal Resolution of Hearing Probed by Bandwidth Restriction,” *Acta Acustica*, vol. 94, pp. 594–603 (2008). <https://doi.org/10.1121/1.1547438>

[107] M. N. Kunchur, “Audibility of Temporal Smearing and Time Misalignment of Acoustic Signals,” <http://www.ejta.org/en/kunchur1> *Electronic Journal Technical Acoustics ISSN 1819-2408*, <http://www.ejta.org>, 17 (2007).

[108] M. N. Kunchur, “Auditory Mechanisms that Can Resolve ‘Ultrasonic’ Timescales,” presented at the *128th Convention of the Audio Engineering Society* (2010 May), Workshop 6; <http://boson.physics.sc.edu/~kunchur/papers/Auditory-mechanisms-that-can-resolve-ultrasonic-time-scales.pdf>.

9.13 Hearing thresholds

[109] J. R. Stuart, “Noise: Methods for Estimating Detectability and Threshold,” *J. Audio Eng. Soc.*, vol. 42, pp. 124–140 (1994 Mar.). <http://www.aes.org/e-lib/browse.cfm?elib=6959>

[110] D. W. Robinson and R. S. Dadson, “Acoustics—Expression of Physical and Subjective Magnitudes of Sound or Noise in Air,” ISO131 (1959).

[111] D. W. Robinson and R. S. Dadson, “A Redetermination of the Equal-Loudness Relations for Pure Tones,” *Brit. J. Appl. Physics*, vol. 7, pp. 166–181 (1956 May). <https://doi.org/10.1088/0508-3443/7/5/302>

[112] R. S. Dadson and J. H. King, “A Determination of the Normal Threshold of Hearing and its Relation to the Standardization of Audiometers,” *J. Laryngol. Otol.*, vol. 66, pp. 366–378 (1952).

[113] I. M. de Castro Silva and M. A. G. Feitosa, “High-Frequency Audiometry in Young and Older Adults when Conventional Audiometry Is Normal,” *Brazilian J. Otorhinolaryngology*, vol. 72, no. 5, pp. 665–672 (2006 Sep./Oct.).

[114] K. Kurakata, T. Mizunami, K. Matsushita, and K. Ashihara, “Statistical Distribution of Normal Hearing Thresholds under Free-Field Listening Conditions,” *Acoust. Sci. & Tech.*, vol. 26, no. 5, pp. 440–446 (2005). <https://doi.org/10.1260/026309208785844149>

[115] G. G. Harris, “Brownian Motion in the Cochlear Partition,” *J. Acoust. Soc. Amer.*, vol. 44, no. 1, pp. 176–186 (1968). <https://doi.org/10.1121/1.1911052>

[116] R. J. Pumphrey “Upper Limit of Frequency for Human Hearing,” *Nature*, vol. 166, p. 571 (1950).

[117] S. Buus et al., “Tuning Curves at High-Frequencies and Their Relation to the Absolute Threshold Curve,” in B. C. J. Moore and R. D. Patterson (eds.), *Auditory Frequency Selectivity* (Plenum Press, 1986)

[118] M. J. Shailer, B. C. J. Moore, B. R. Glasberg, N. Watson, and S. Harris “Auditory Filter Shapes at 8 and 10 kHz,” *J. Acoust. Soc. Amer.*, vol. 88, pp. 141–148 (1990).

[119] K. Ashihara, K. Kurakata, T. Mizunami, and K. Matsushita, “Hearing Threshold for Pure Tones above 20 kHz,” *Acoust. Sci. & Tec.*, vol. 27, no. 1, pp. 12–19 (2006).

[120] T. Fujioka, R. Kakigi A. Gunji, and Y. Takeshima “The Auditory Evoked Magnetic Fields to Very High Frequency Tones,” *Neuroscience*, vol. 112, pp. 367–381 (2002). [https://doi.org/10.1016/S0306-4522\(02\)00086-6](https://doi.org/10.1016/S0306-4522(02)00086-6)

9.14 Spectra of music and instruments

[121] R. Kuribayashi and H. Nittono, “Music Instruments that Produce Sounds with Inaudible High-Frequency Components,” *Studies in Human Sciences*, vol. 10, pp. 35–41 (2016) (in Japanese).

[122] J. Boyk, “There’s Life above 20 kilohertz! A Survey of Musical Instrument Spectra to 102.4 kHz,” <http://www.cco.caltech.edu/~boyk/spectra/spectra.htm> (2000).

9.15 Effects of high-frequency sounds

[123] Y. Tohkura, T. Oohashi, and N. Koizumi, “Discussions on the Effect of Ultra-High Frequency Sound,” *J. Acoust. Society Jap.*, vol. 62, no. 12, pp. 891–896 (2006 Jun.). https://doi.org/10.20697/jasj.62.12_891

[124] T. Nishiguchi, K. Hamasaki, K. Ono, M. Iwaki, and A. Ando, “Perceptual Discrimination of Very High Frequency Components in Wide Frequency Range Musical Sound,” *Applied Acoustics*, vol. 70, pp. 921–934 (2009). [doi:https://doi.org/10.1016/j.apacoust.2009.01.002](https://doi.org/10.1016/j.apacoust.2009.01.002)

[125] R. Yagi, E. Nishina, M. Honda, and T. Oohashi “Modulatory Effect of Inaudible High-Frequency Sounds on Human Acoustic Perception,” *Neuroscience Letters*, vol. 351, pp. 191–195 (2003). <https://doi.org/10.1016/j.neulet.2003.07.020>

[126] R. Yamamoto, N. Kanetada and M. Mizumachi, “Evaluation of Sound Quality of High Resolution Audio,” *J. Industrial Applic. Engineers*, Japan, vol. 1, no. 2, pp. 52–57 (2013 Sep.).

9.16 Criteria for noise floor

[127] E. A. Cohen and L. D. Fielder, “Determining Noise Criteria for Recording Environments,” *J. Audio Eng. Soc.*, vol. 40, pp. 384–402 (1992 May). <http://www.aes.org/e-lib/browse.cfm?elib=7049>

[128] P. B. Fellgett, “Thermal Noise Limits of Microphones,” *J. IERE*, vol. 57, no. 4, pp. 161–166 (1987).

9.17 Sampling theory (ancient and modern)

[129] C. E. Shannon, “Communication in the Presence of Noise,” *Proc. IRE*, vol. 37, no. 1, pp. 10–21 (1949 Jan.).

[130] H. Nyquist, “Certain Topics in Telegraph Transmission Theory,” *Trans. Amer. IEE*, vol. 47, pp. 617–644 (1928).

[131] M. Unser, “Sampling—50 Years after Shannon,” *Proc. IEEE*, vol. 88, no. 4, pp. 569–587 (2000 Apr.). DOI: <https://doi.org/10.1109/5.843002>

[132] P. P. Vaidyanathan, “Generalizations of the Sampling Theorem: Seven Decades after Nyquist,” *IEEE Trans. Cir. Sys.*, vol. 48, no. 9, pp. 1094–1109 (2001 Sep.).

[133] P. L. Dragotti, M. Vetterli, and T. Blu, “Sampling Signals with Finite Rate of Innovation,” *IEEE Trans. Sig. Proc.*, vol. 50, no. 6, pp. 1417–1428 (2007 May).

[134] P. L. Dragotti, Vetterli, et al., “Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang–Fix,” *IEEE Trans. Sig. Proc.*, vol. 55, pp. 1741–1757 (2007 May).

[135] M. Vetterli et al., “Sampling Signals with Finite Rate of Innovation,” *IEEE Trans. Sig. Proc.*, vol. 50, no. 6, pp. 1417–1428 (2002 May).

[136] J. Oñativia and P. L. Dragotti, “Sparse Sampling: Theory, Methods and an Application in Neuroscience,” *Biol. Cybern.*, vol. 109, pp. 125–139 (2015). <https://doi.org/10.1007/s00422-014-0639-x>

[137] P. L. Butzer and R. L. Stens, “Sampling Theory for Not Necessarily Band-Limited Functions: A Historical Review,” *SIAM Review*, vol. 34, no. 1, pp. 40–53, (1992 Mar.). <http://www.jstor.org/stable/2132784>

[138] A. Urigüen, Y. C. Eldar, P. L. Dragotti, and Z. Ben-Haim, “Sampling at the Rate of Innovation: Theory and Applications,” in *Compressed Sensing: Theory and Applications* (Cambridge University Press, 2012).

[139] Y. C. Eldar, and T. Michaeli, “Beyond Bandlimited Sampling: Nonlinearities, Smoothness and Sparsity,” CCIT Report no. 698 (2008 Jun.). arXiv:0812.3066 [cs.IT]

[140] F. Gensun, “Whittaker-Kotelnikov-Shannon Sampling Theorem and Aliasing Error,” *J. Approx. Theory*, vol. 85, pp. 115–131 (1996).

[141] C. Herley and P.W. Wong, “Minimum Rate Sampling and Reconstruction of Signals with Arbitrary Frequency Support,” *IEEE Trans. Information Theory*, vol. 45, no. 5, pp. 1555–1564 (1999 Jul.).

[142] V. Pohl, F. Yang, and H. Boche, “Causal Reconstruction Kernels for Consistent Signal Recovery,” *EU-SIPCO*, Bucharest, pp. 1174–1178 (2012).

[143] M. Unser and A. Aldroubi, “A General Sampling Theory for Non-Ideal Acquisition Devices,” *IEEE Trans. Signal Processing*, vol. 42, pp. 2915–2925 (1994 Nov.).

[144] J. A. Urigüen, P. L. Dragotti, and T. Blu, “On the Exponential Reproducing Kernels for Sampling Signals with Finite Rate of Innovation,” *Proceedings of Sampling Theory and Applications* (SampTA), vol. 21 (2011).

9.18 Miscellaneous

[145] G. Strang and G. Fix, “Fourier Analysis of the Finite Element Variational Method,” in *Constructive Aspect of Functional Analysis* (Springer Edizioni Cremonese, Rome, Italy, 1971), pp. 796–830.

[146] S. R. Norsworthy, R. Schrier, and G. C. Temes (editors), *Delta-Sigma Data Converters: Theory, Design and Simulation* (IEEE Press, 1997), pp. 416–418.

[147] Engineering Acoustics: Outdoor Sound Propagation https://en.wikibooks.org/wiki/Engineering_Acoustics/Outdoor_Sound_Propagation.

[148] https://en.wikipedia.org/wiki/Spline_wavelet

[149] M. Ravel, “String Quartet in Fmajor: 2,” *Guarneri String Quartet*, SBE (2001).

[150] J. R. Stuart and P. G. Craven, “The Gentle Art of Dither,” *J. Audio. Eng. Soc.*, vol. 67 (2019 May) (this issue) Open access: DOI: <https://doi.org/10.17743/jaes.2019.0011>

THE AUTHORS



J. Robert (Bob) Stuart

J. Robert (Bob) Stuart was born in 1948. He studied electronic engineering and acoustics at the University of Birmingham and took an M.Sc. in operations research at Imperial College, London. While at Birmingham he studied psychoacoustics under Professor Jack Allison, which began a lifelong fascination with the subject.

In 1977 he co-founded Meridian Audio and served as CTO until early 2015. In 2014 he founded MQA Ltd. where he is currently full time as Chairman and CTO.

At the request of Hiro Negishi and Raymond Cooke, Bob chaired the advocacy group Acoustic Renaissance for Audio between 1994 and 2002.

In the 1990s he worked with Michael Gerzon and Peter Craven on lossless compression and was instrumental in its adoption for optical discs.

Bob has contributed to DVD-Audio and BluRay standards and has served on the technical committees of the National Sound Archive, JAS and the ADA (Japan).

Bob's professional interests are the furthering of analog and digital audio and developing understanding of human auditory perception mechanisms relevant to live and recorded music. His specialties include the auditory sciences and the design of analog and digital electronics, loudspeakers, audio coding, and signal processing.



Peter Craven

Bob joined AES in 1971, has been a fellow since 1992, and is a member of ASA, IEEE, and the Hearing Group at Cambridge.

Bob has a deep interest in music and spends a good deal of time listening to live and recorded material.

Peter Craven was born in 1948. He studied maths and then astrophysics at Oxford University while collaborating with Michael Gerzon and others at the Oxford University Tape Recording Society on making live recordings and on audio quality topics generally, including the ideas (1972) leading to the Ambisonic Soundfield Microphone.

After some years in research and academia, Peter became independent in 1981, specializing in digital signal processing for audio. Further collaborations with Gerzon resulted in a seminal paper on noise-shaping in 1989. Work on room equalization with B&W loudspeakers was followed by collaboration with Bob Stuart on noise shaping and buried data, leading eventually to the MLP lossless compression system and, more recently, to the MQA music delivery system.

A collector of pre-war coarse-groove 78 r.p.m. records, Craven seeks a synergy between modern high definition multichannel technology and the simpler recording techniques of yesteryear.

Bob and Peter have worked together on audio topics since they first met in 1975.