

Transfer-Function Measurement with Sweeps

**DIRECTOR'S CUT INCLUDING PREVIOUSLY UNRELEASED MATERIAL
AND SOME CORRECTIONS**

THE ORIGINAL HAS BEEN PUBLISHED IN J.AES, 2001 JUNE, P.443-471

SWEN MÜLLER, AES member,

Institut für Technische Akustik, RWTH, 52056 Aachen, Germany

AND

PAULO MASSARANI

Acoustic Testing Laboratory, INMETRO, Xerém, Duque de Caxias (RJ), Brazil

Compared to using pseudo-noise signals, transfer function measurements using sweeps as excitation signal show significantly higher immunity against distortion and time variance. Capturing binaural room impulse responses for high-quality auralization purposes requires a signal-to-noise ratio of >90 dB which is unattainable with MLS-measurements due to loudspeaker non-linearity but fairly easy to reach with sweeps due to the possibility of completely rejecting harmonic distortion. Before investigating the differences and practical problems of measurements with MLS and sweeps and arguing why sweeps are the preferable choice for the majority of measurement tasks, the existing methods of obtaining transfer functions are reviewed. The continual need to use pre-emphasized excitation signals in acoustical measurements will also be addressed. A new method to create sweeps with arbitrary spectral contents, but constant or prescribed frequency-dependent temporal envelope is presented. Finally, the possibility of simultaneously analysing transfer function and harmonics is investigated.

0 INTRODUCTION

Measuring transfer functions and their associated impulse responses (IRs) is one of the most important daily tasks in all areas of acoustics. The technique is practically needed everywhere. A loudspeaker developer will check the frequency response of a new prototype many times before releasing it for production. As the on-axis response does not sufficiently characterize a loudspeaker, a full set of polar data requiring many measurements is needed. In room acoustics, the IR plays a central role, as many acoustical parameters related to the perceived quality can be derived from it. The room transfer function obtained by Fourier-transforming the RIR may be useful to detect modes at low frequencies. In building acoustics, the frequency dependent insulation against noise from outside or other rooms is a common concern. In vibroacoustics, the propagation of sound waves in materials and radiation from their surface is a vast field of simulation and verification by measurements with shakers. Profiling by detection of reflections (sonar, radar) is another area closely linked to the measurement of IRs.

While many of these measurement tasks do not require an exorbitant dynamic range, the situation is different when it comes to acquiring room impulse responses (RIRs) for use in convolutions with dry anechoic audio material. Because of the wide dynamic range of our auditory system and the logarithmic relationship between sound pressure level (SPL) and perceived loudness, any abnormalities in the reverberant tail of a RIR are easily recognizable. This is especially apparent when speech, with its long intermediate pauses, is used for convolution and when the auralization results are monitored with headphones, as required for virtual reality based on binaural responses. As today's digital recording technology offers signal-to-noise ratios (SNR) in excess of 110 dB, it does not seem too daring to demand an SNR for measured RIRs that is at least equivalent to the 16 bit CD standard.

This work has been fueled by the constant disappointment from maximum-length-sequence (MLS) based measurements of RIRs. Even under optimal conditions, with a supposed absence of time variance, very little background noise, appropriate pre-emphasis and an arbitrary number of synchronous averages, it seems impossible to achieve a dynamic range superior to that of, say, an analogue tape recorder. The reason for this is that in any measurement using noise as the excitation signal, distortion (mainly induced by the loudspeaker) spreads out over the whole period of the recovered IR. The ensuing noise level can be reduced using longer excitation signals, but it can never be isolated entirely. Although distortion can be reduced using lower volume, this leads to more background noise which contaminates the results. Hence, some compromise level must be carefully chosen for each measurement site [34], often leaving the power capabilities of the driving amplifier and the speaker largely unexploited.

In contrast, using sweeps as excitation signals relieves the engineer to a great extent from these limitations. Using a sweep somewhat longer than the RIR to be measured allows the exclusion of all harmonic distortion products, practically leaving only background noise as the limitation for the achievable SNR. The sweep can thus be fed with considerable more power to the speaker without introducing artifacts in the acquired RIR. Moreover, in anechoic conditions, the distortion can be classified into

single harmonics related to the fundamental, allowing for a simultaneous measurement of transfer function and frequency-dependent distortion. This possibility already anticipated by Griesinger [1], Norcross/Vanderkooy [42] and eventually described in Farina [2] will be further examined in section 5.

Sweep-based measurements are also considerably less vulnerable to the deleterious effects of time variance. For this reason, they are sometimes the only option in long-distance outdoor measurements in windy weather conditions or for measurements of analogue recording gear.

1 EXISTING METHODS

Quite a number of different ways to measure transfer functions have evolved in the past century. Common to all of them is that an excitation signal (stimulus) containing all the frequencies of interest is used to feed the device under test (DUT). The response of the DUT is captured and in some way compared with the original signal. Of course, there is always a certain amount of noise, reducing the certainty of a measurement. Therefore, it is desirable to use excitation signals with high energy so as to achieve a sufficient SNR in the whole frequency range of interest. Using gating techniques to suppress noise and unwanted reflections further improves the SNR. In practice, there is always a certain amount of non-linearity, and time-variances are also commonplace in acoustical measurements. We will see that the different measurement methods react quite differently to these kinds of disturbances.

1.1 The Level Recorder

One of the oldest methods of bringing a transfer function onto paper already involved sweeps as excitation signal. The DUT's response to a sweep generated by an analogue generator is rectified and smoothed by a low-pass filter. The resulting voltage is input to a differential amplifier whose other input is the voltage derived from a discrete precision potentiometer which is linked mechanically to the writing pen. The differential amplifier's output controls the writing pen, which is swept over a sheet of paper with the appropriate scale printed on it. The potentiometer may be either linear or logarithmic to produce amplitude or dB readings on the paper.

Obviously, this method does not need any digital circuitry and for many years, it used to be the standard in frequency response testing. Even today, the famous old B&K level writers can be seen in many laboratories, and due to their robustness they may continue so in the future.

The excitation signal being used is a logarithmic sweep, which means that the frequency increases by a fixed factor per time unit (for example, it doubles every second). As the paper is moved with constant speed under the writing pen, the frequency scale on the paper is correspondingly logarithmic. The FFT spectrum of such a logarithmic sweep declines by 3 dB/octave. Every octave shares the same energy, but this energy spreads out over an increasing bandwidth. Therefore the magnitude of each frequency component decreases. We will later see that this excitation signal, which has already been in use for such a long time, has some unique properties that keep it attractive for use in the digital world of today. One of these properties is that the spectral distribution

is often quite well adapted to the ambient noise, resulting also in a good SNR at the critical low end of the frequency scale.

While the level recorder cannot really suppress neither noise or reflections, a smoothing effect is obtained by reducing the velocity of the writing pen. The ripple in a frequency response caused by a reflection as well as any irregular movement induced by noise can be “flattened out” by this simple means. If the spectral details to be revealed are too blurred by the reduced responsiveness of the writing pen, reducing the sweep rate helps to reestablish the desired spectral resolution. This way, measurement length and measurement certainty can be compromised, just as with the more modern methods based on digital signal processing.

The evident shortcomings of level recorders are that they do not show phase information and the produced spectra reside on a sheet of paper instead of being written to a hard disc for further processing. Clearly, the “horizontal” accuracy of displayed frequencies cannot match the precision offered by digital solutions deploying quartz-based clocking of AD- and DA converters. On the vertical scale, the resolution of the dB or amplitude readings is restricted due to the discrete nature of the servo potentiometer, which is composed of discrete precision-resistors.

1.2 Time Delay Spectrometry (TDS)

TDS is another method to derive transfer functions with the help of sweeps. Devised by Heyser [3-6] especially for the measurement of loudspeakers, it is also applicable for room acoustic measurements or any other LTI system in general. The principal functionality of a TDS analyzer is shown in Fig. 1.

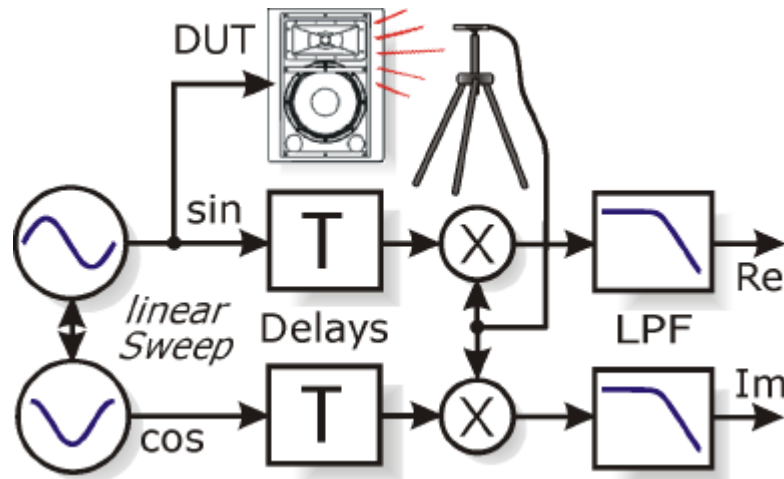


Fig. 1. TDS signal processing.

The analyzer features a generator that produces both a swept sine and, simultaneously, a phase-locked swept cosine. The sine is fed to the loudspeaker under test (LUT), and its captured response is multiplied separately by both the original sine (to get the transfer function’s real part) and the 90° phase-shifted cosine (to get the imaginary part). The multiplier outputs are filtered by a low-pass with fixed cut off frequency. The multipliers work similar to the mixers used in the intermediate-frequency stages of HF receivers (superhet principle), producing the sums and differences of the input frequencies. The sum terms of both multiplier outputs must be rejected by the low-pass

filters, whereas the difference terms may pass, depending on their frequency. If both the generated and the captured frequencies are almost equal, the output difference frequency will be very low and thus not be attenuated by the low-pass filters.

As the sound that travels from the LUT to the microphone arrives with a delay, its momentary frequency will be lower than the current generator signal. This causes a higher mixer output difference frequency that, depending on the cutoff frequency, will be attenuated by the low-pass filters. For this reason, the generated signal must be “time-delayed” by an amount equivalent to the distance between loudspeaker and microphone before being multiplied with the LUT’s response. This way, the difference frequency will be near DC. In contrast, reflections will always take a longer way than the direct sound and thus arrive with a lower instantaneous frequency, causing higher frequency components in the multiplier outputs, which will be attenuated by the low-pass filters.

With proper selection of the sweep rate and the low-pass filter cutoff frequency, simulated quasi-free-field-measurements are possible with TDS. In addition to the attenuation of unwanted reflections, distortion products are also suppressed very well. Distortion products arrive with a higher instantaneous frequency and thus cause high mixer output frequencies. They too will be strongly attenuated by the filters, thus excluding the disturbing influence of the harmonics from the measurement. Likewise, extraneous noise in the wide band above the filter cutoff frequency will be rejected.

The controlled suppression of reflections is the motivation why TDS analyzers utilize a linear sweep ($df/dt = \text{constant}$) as the excitation signal. The frequency difference between incoming direct sound and reflection will thereby stay constant over the whole sweep range, keeping the attenuation of each reflection frequency-independent.

If a logarithmic sweep were used instead, the low pass filters would have to increase their cutoff frequency by a constant factor per time to avoid a narrowing of the imposed equivalent time window. (the impact of the low-pass filter is indeed similar to the windowing of IRs in the FFT methods described later.) On the other hand, the higher frequency components of a typical loudspeaker IR will decay faster than the lower ones. Thus, a narrowing of the window at higher frequencies (corresponding to “adaptive windowing” proposed by Rife to process IRs) should even be desirable for many measurement scenarios. It would increase the SNR at high frequencies without corrupting the IR more than at low frequencies.

There are a number of drawbacks associated with TDS measurements. The most serious is the fact that TDS uses linear sweeps and hence a white excitation spectrum. In most measurement setups, this will lead to SNR at low frequencies. If the whole audio range from 20 Hz to 20 kHz is swept through in 1 second, then the subwoofer range up to 100 Hz will only receive energy within 4 ms. This most often is insufficient in a frequency region where the output of a loudspeaker decreases while ambient noise increases. To overcome the poor spectral energy distribution, the sweep must be made very long or the measurement split into two ranges (for example one below and one above 500 Hz). Both methods extend the measurement time far beyond of what would be needed physically to perform a measurement of the particular spectral resolution.

Another problem is ripple, which occurs at low frequencies. As mentioned previously, the multipliers produce sum and difference terms of the “time-delayed” excitation signal

and the incoming response. At higher instantaneous frequencies, the sum is sufficiently high to be attenuated by the output low-pass filter. But at the low end of the sweep range, when the sum is close to or lower than the low pass cutoff frequency, “beating” will appear in the recovered magnitude response. To remedy this, the sweep can be made very long and the low-pass cutoff frequency reduced by the same factor. The better method, however, is to repeat the measurement with a “mirrored” setup, that is, exciting the DUT with a cosine instead of a sine and treat the captured signal as depicted by the dotted lines in Fig 1b. The real part of the complex result of this second measurement is added to the real part obtained by the previous measurement, while the imaginary part is subtracted. The effect of this operation is that the sum terms of the mixer output will be cancelled, as shown in [7-9]. As a consequence, thanks to the absence of the interfering sum terms over the whole sweep range, the low-pass filters following the multiplier stages may be omitted. In fact, they have to be omitted if a full IR is to be recovered, a case in which obviously the low-pass filter impact of attenuating reflections is not desired. Indeed, the measurement of room acoustics, which always involves acquiring lengthy IRs, is only feasible with the double excitation method. If, however, a loudspeaker is the object of interest, it is worth keeping the low-pass filters inserted in order to reject reflections, noise and harmonics.

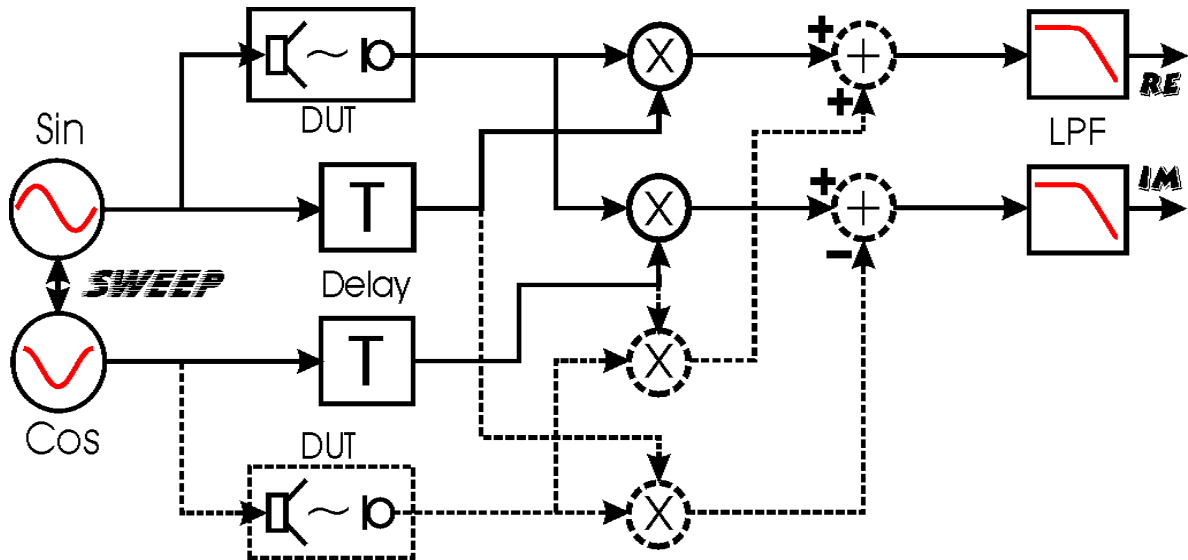


Fig 1b. Full TDS signal processing. The dotted signal paths are used in the second run of a double excitation measurement.

Even with the sum-term-canceling double excitation method, some ripple might still appear at the very beginning and at the end of the sweep frequency range because of the sudden onset of the linear sweep. According to system theory, this switched sine produces a corrugated spectrum near the initial frequency (see Fig. 15). The switching corresponds to *multiplying* a continuous time signal with a rectangular window. In the frequency domain, this corresponds to *convoluting* the spectrum of the sweep by the rectangular window’s spectrum (that is, the $\sin(x)/x$ function). A common way to circumvent this problem is to let the excitation sweep start well below the lowest frequency of interest. This might entail starting the sweep at “negative” frequencies, which in practice means starting at the corresponding positive frequency, then lowering down to 0 Hz and from there increasing the frequency normally [8]. A better possibility would be to formulate the excitation sweep in the spectral domain to create a signal that does not suffer spectral leakage (see Fig. 16), as will be demonstrated in section 4.2.

Of course, the necessity to use the double excitation method to recover a full RIR further extends the time needed to complete a TDS-measurement. On the other hand, FFT- or MLS-based methods using periodic stimuli in practice also require emitting the excitation signal twice to recover the periodic IR. With these methods, the DUT's periodic response is captured and processed only in the second run after stabilizing. In contrast, the TDS double excitation method uses both passes, which endows it with an additional advantage of 3 dB in SNR over MLS, given the same excitation length (both linear sweep and MLS have a white spectrum). With the absence of the output bw-pass filters, the spectral resolution of a TDS measurement is just as high as with periodic excitation of same period length. So a TDS double excitation measurement requires just about the same time as an MLS measurement to achieve the same spectral resolution. This contrasts to some TDS-MLS comparisons in the literature [13] in which a certain TDS low-pass filter frequency is assumed and a correspondent frequency resolution calculated, in an attempt to prove that an exorbitant sweep length would be required to achieve the resolution of a non-windowed MLS measurement. But of course, the impact of the TDS low-pass filter is equivalent to the application of a window to the captured IR, and any windowing reduces the spectral resolution. The possibility of performing simulated free-field measurements and the associated smoothing effect that occurs when attenuating reflections with either method is normally very desirable for loudspeaker measurements (at least in the higher octave bands).

However, the relation between TDS sweep rate, low-pass cutoff frequency, and achieved attenuation of a delayed reflection is not evident immediately, albeit not too difficult to calculate [10]. But it is more intuitive to inspect the full IR as derived by MLS and FFT measurements (or dual excitation TDS) and to position a window whose right leg ends just in front of the first annoying reflection. In this way, all subsequent reflections are muted entirely. In contrast, even after laborious adjustment of the cutoff frequency, TDS is not capable of complete reflection suppression due to the limited steepness of the low-pass filters. Their smoothing effect on the captured transfer functions is not very well defined, while windowing offers a well-explained [33] compromise of main-lobe broadening and side-lobe suppression.

Using sweeps purely in conjunction with FFT analysis, without multipliers to produce intermediate results, obviates many of the problems inherent in TDS, especially insufficient energy at low frequencies and long measurement cycles. However, some advantages of TDS measurements over measurements with noise signals such as MLS should not be ignored. The higher achievable SNR in a full double-excitation TDS measurement can be augmented further when taking advantage of the low crest factor of only 3 dB inherent in a swept sine. In practice, MLSs have a crest factor of at least 8 dB, as will be revealed later. TDS measurements should also offer higher tolerance against time variance and better rejection of harmonic distortion that can be filtered out along with noise and reflections.

1.3 Dual-Channel FFT-Analysis

A review of transfer function measurement methods would be incomplete without mentioning dual-channel FFT analysis. It is as old as the first FFT analyzers and in the past years has passed through a certain revival due to the omnipresence of stereo sound boards in PCs, although it neither boasts speed nor precision.

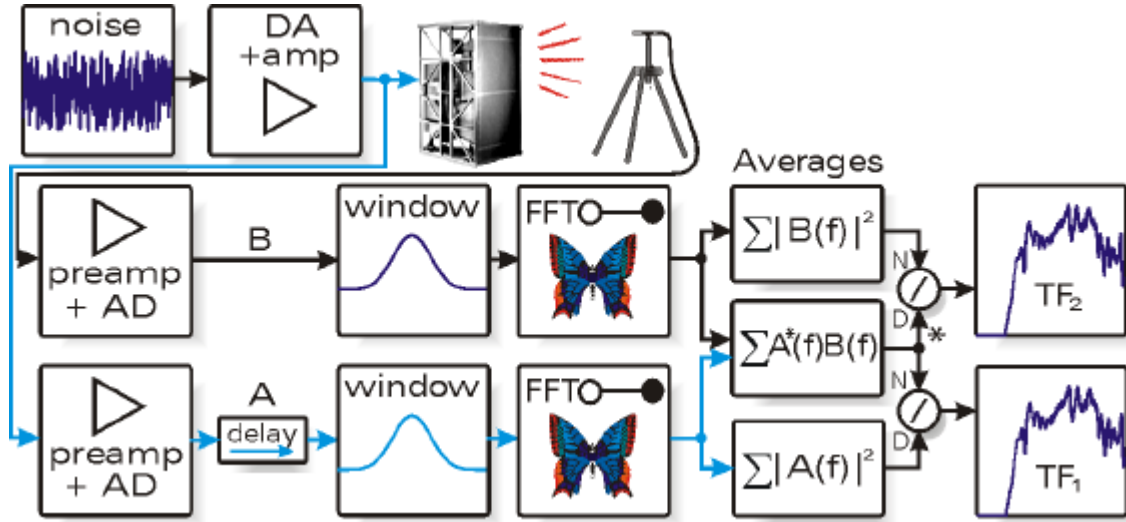


Fig. 2. Signal processing steps for classical 2-channel FFT analysis with asynchronous noise used as excitation signal.

A dual channel analyzer captures both the input (channel A in Fig. 2) and the output signal of the DUT (channel B in Fig. 2).

The signal is cut in contiguous or overlapping segments which are windowed and then transformed to the spectral domain via FFT. The two resulting spectra are involved in three averaging processes: Two for the autospectra

$$G_{AA}(f) = \frac{1}{n} \sum |A(f)|^2 \quad \text{and} \quad G_{BB}(f) = \frac{1}{n} \sum |B(f)|^2 \quad (n = \text{number of averages}) \quad (1.1)$$

which are the squared modulus of the spectra for channel A and B, and one for the cross-spectrum

$$G_{AB}(f) = \frac{1}{n} \sum A^*(f) \cdot B(f) \quad (1.2)$$

which is the complex conjugate of the spectrum for channel A (spectrum A with its phases negated) multiplied with the spectrum for channel B. While the two autospectra are real-valued and add up both signal and noise power in every measurement run, the cross spectrum is complex and uncorrelated noise with its random phase spectrum tends to average out with increasing number of averages [40].

The complex transfer function defined by

$$H(f) = \frac{B(f)}{A(f)} \quad (1.3)$$

can now be computed in two ways. Multiplying both the numerator and the denominator of the right side of the above equation with $A^*(f)$ yields:

$$H_1(f) = \frac{A^*(f)B(f)}{A^*(f)A(f)} = \frac{G_{AB}(f)}{G_{AA}(f)} \quad (1.4)$$

so $H(f)$ can be estimated by the division of the cross spectrum with the autospectrum of the DUT's input. On the other hand, multiplying both the numerator and the denominator of the right side with $B^*(f)$ yields:

$$H_2(f) = \frac{B^*(f)B(f)}{B^*(f)A(f)} = \frac{G_{BB}(f)}{G_{AB}^*(f)} \quad (1.5)$$

so $H(f)$ can also be computed by the division of the autospectrum of the DUT's output with the complex conjugated cross spectrum. When no noise is present neither on channel A nor on channel B, both methods obviously lead to identical results. With noise, however, the two results differ. $H_1(f)$ is a better estimate for the true TF when output noise prevails (the more typical case), while $H_2(f)$ comes closer to reality when input noise dominates.

To determine the amount of uncorrelated noise in the measurement, the two estimates of the transfer function can be divided:

$$v^2(f) = \frac{H_1(f)}{H_2(f)} = \frac{|G_{AB}|^2}{G_{AA} \cdot G_{BB}} \quad (1.6)$$

This *coherence function* assumes values between 1 and 0. When no uncorrelated noise is present, the coherence function will become 1, as $H_1(f)$ and $H_2(f)$ are identical. When there is only uncorrelated noise without signal, the coherence function drops to near 0 because the noise tends to average out in the cross spectrum, while it is being added up energetically in the autospectra.

With the help of the coherence function, the power of the autospectrum G_{BB} (derived from the DUT's output) can be split into the coherent power $v^2(f) \cdot G_{BB}(f)$ which originates from the excitation signal passed through the DUT, and the non-coherent power $(1 - v^2(f)) \cdot G_{BB}(f)$ originating from uncorrelated noise.

The signal to noise ratio becomes thus

$$SNR = \frac{v^2(f)}{1 - v^2(f)} \quad (1.7)$$

and it is an interesting feature of the dual channel analysis technique that this value is available without turning off the excitation signal.

Traditionally, FFT analyzers are operated with asynchronous noise sources. However, the use of an asynchronous excitation signal has some implications:

The FFT yields correct results only for signals repeated with a period equal to the FFT block length. For non-periodic signals, start point and end point of the analyzed signal section generally do not match. This discontinuity introduces a considerable error, the famous *leakage*, which has to be lowered by windowing the analyzed sections prior to the FFT. However, windowing has a smoothing effect which reduces the spectral resolution, especially at low frequencies, and introduces a DC bias error.

The asynchronous noise source commonly employed in 2 channel analyzers has a white (or pink or other specified) spectrum when averaged over a long time, but a single snapshot of the noise signal has a very corrugated spectrum that suffers from deep magnitude dips. Thus, a dual channel analyzer must always average over several individual measurements before a reliable result can be obtained.

Inadequate SNR can be detected by means of the coherence function [12, 13, 40]. Due to the necessity to average many blocks of data to achieve a consistent display, the responsiveness of dual-channel analysis is very poor and makes it unattractive for adjustment purposes.

The convergence of the frequency response could be improved considerably by generating the cross-correlation function in every single measurement cycle by IFFT of the cross spectrum, window it, and back-transform it to the spectral domain via FFT. Windowing the cross correlation function would offer the crucial freedom to control the amount of reflections entering into the result and to mute the noise outside the windowed interval, thus speeding up the convergence process. However, classical dual-channel analyzers do not seem to incorporate this very useful feature.

In acoustic measurements, the precise delay of the acoustical transmission path must be known, as the direct signal has to be delayed by exactly this amount of time to ensure that the same parts of the excitation signal will be analyzed on both channels. This is a major nuisance because the estimation of the delay requires a separate preparative measurement of the cross-correlation function. The location of its peak indicates the approximate delay to be used in the subsequent two-channel analysis. Changing the measurement distance inflicts repeating the whole procedure.

There is one well-known application for dual-channel analysis that no other measurement technology offers: The possibility to measure sound systems unobtrusively during a performance, using the program material itself as the excitation signal. However, music with its erratic spectral distribution is usually a much worse excitation signal than uncorrelated noise sources and requires even longer averaging periods to achieve a reliable result, if at all possible. Thus, when the unobtrusiveness is not needed, it is advisable to use a noise generator as source. Far better results in far shorter time, however, can be achieved with custom-tailored synchronous deterministic excitation signals.

1.4 Stepped Sine

Probably the most time consuming method of acquiring a transfer function is exciting the DUT step by step with pure tones of increasing frequency. The DUT's response to this steady-state excitation can be either analyzed by filtering and rectifying the fundamental, or by performing an FFT and retrieving the fundamental from the spectrum. The latter method requires the use of a sine that is exactly periodic within the bounds of one FFT block-length to avoid spectral leakage. In practice, this can only be realized generating the sine digitally and emitting it via a DA converter that is synchronized to the capturing AD converter. As a big plus, the FFT method allows for the complete suppression of all other frequencies and thus is the preferable method over analysis in the time domain, which involves band-pass filters with restricted selectivity and precision.

After each single measurement, the excitation sweep's frequency is raised by a value according to the desired spectral resolution. In acoustic measurements, the frequency will usually be incremented by multiplying the previous value with a fixed factor to obtain a logarithmic spacing. Clearly, the spectral resolution in stepped sine measurements is much lower at high frequencies compared to what could be achieved using a broad-band excitation signal with FFT analysis. But this is not necessarily a disadvantage as the frequency-linear resolution of FFT-spectra often yields unnecessary fine frequency steps in the HF region while sometimes lacking information in the LF region, which occurs when the time interval used for the FFT is too small. The possible logarithmic spacing of the stepped sine measurements results in much smaller data records than those obtained by FFT, but this is not a serious advantage in the age of gigabyte hard disks.

The biggest advantage of the stepped-sine method is the enormous signal-to-noise ratio that can be realized in a single measurement. All energy is concentrated at a single frequency, and the feeding sine wave has a low crest factor of only 3 dB. Thus, the measurement certainty and repeatability for a single frequency can be very high compared to broad-band excitation, especially when using the synchronous FFT technique.

Thus despite of the considerable amount of time needed for the complete evaluation of a transfer function, this method is still popular for precision measurement and calibration of electronic equipment or acoustic transducers such as microphones. Stepped sines are also the established method when it comes to precise distortion measurements. Every harmonic can be picked up easily and with high precision from the FFT spectra.

However, when only the transfer function is of interest, the stepped-sine method is everything but elegant. Only part of the energy emitted by the DUT can be used for the analysis, since after each switching to a new frequency, one must wait until the DUT settles to a steady state. Especially when high-Q resonances (corresponding to large IRs) exist, this settling time must be made very long to reduce errors to a negligible level. On the other hand, when the system is very noisy and many synchronous averages must be executed to achieve an acceptable measurement certainty, the settling time plays only a minor role [14].

In acoustic measurements with pure tones, gating out reflections is only possible when the difference in the time-of-flight between direct sound and reflection is longer than the analysis interval. This is a clear disadvantage compared to the methods that recover IRs.

The unmatched accuracy achievable with single-tone measurements must also be cast in doubt. It should be clear that the same accuracy could be achieved with broad-band measurements in lesser time. While stepped-sine measurements deliver single-frequency magnitudes with very high SNR, a broad-band measurement yields many values in the particular frequency interval. Each of them clearly has a lesser certainty, but by performing a spectral smoothing over a width corresponding to the frequency increment used in the pure tone measurements, the random noise should decrease to comparable values. This, however, assumes that the harmonic distortion products can be excluded entirely from the broad-band measurement, a condition that can only be fulfilled by sweep measurements, as will be revealed later.

1.5 Impulses

Using an impulse as excitation signal is the natural way to obtain an IR and also the most straightforward approach to performing FFT-based transfer function measurements. The impulse can be created by analogue means, or preferably sent out by a DA-converter and amplified. It feeds the DUT, whose response is captured by the microphone, amplified and digitized by an AD converter (Fig. 3). As the name implies, this captured response already is the desired IR, provided that a Dirac-style pulse with its associated linear frequency response has been used. To increase the SNR, the pulse can be repeated periodically and the responses of each period added. This leads to the periodic IR (PIR) which is practically equal to the non-periodic IR if it is shorter than the measurement period (in practice: if the IR has vanished in the noise floor before the end of the period). As is well known, such synchronous averaging leads to a reduction in uncorrelated noise by 3 dB relative to the IR for each doubling of the number of averages.

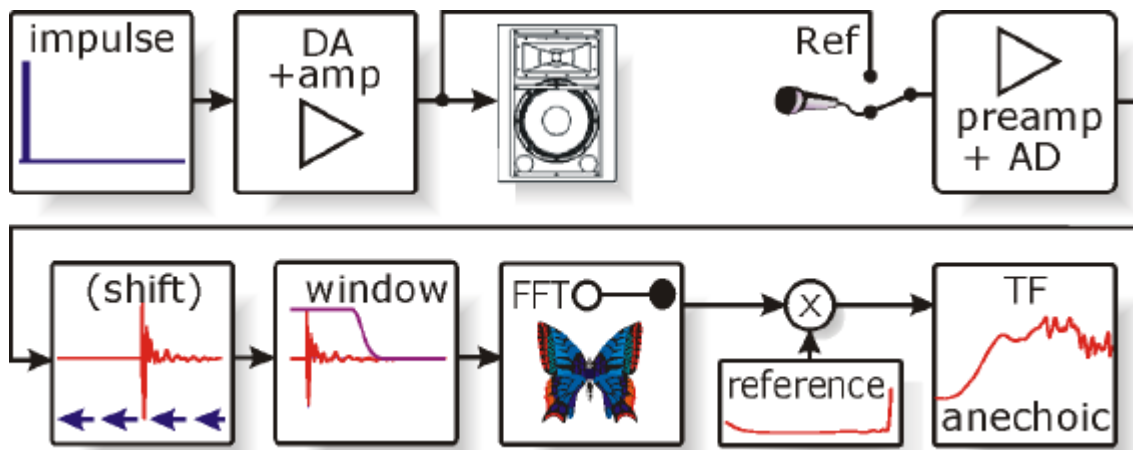


Fig. 3. Signal-processing for transfer-function measurement with impulses.

The IR may optionally be shifted to the left (or the PIR shifted in a cyclic fashion) to compensate for the delay introduced by the propagation time between loudspeaker and microphone in an acoustical measurement. Windowing then mutes unwanted reflections and increases the SNR.

The IR can then be transformed into the transfer function by FFT. To increase the precision of the measurement considerably, the result should be multiplied by a reference spectrum. This reference spectrum is obtained by linking the output and the input of the measurement system and inverting the measured transfer function. Applying this technique (independently of the kind of excitation signal) offers the crucial freedom of pre-emphasizing the excitation signal to adapt it to the spectral contribution of background noise. This pre-emphasis will automatically be removed from the resulting transfer function by applying the obtained reference spectrum in all subsequent measurements.

Impulses are a simple and viable choice when the measurement is purely electrical (no acoustic path in the measurement chain) and when the measurement should be as fast as possible. However, they require a low noise floor of the DUT to achieve reasonable measurement certainty. When measuring low-noise audio equipment, this requirement is easily fulfilled. Despite of their far from optimal SNR performance, impulses can

even be useful in acoustics. In an anechoic chamber where ambient noise is typically very low at high frequencies, tweeters can be measured with reasonable SNR. Because of their short duration, pulses can be fed with very high voltage without the risk of overheating the voice coil. Care must be taken, however, not to cause excursion into the non-linear range of the speaker (although this will hardly be provoked with a very narrow pulse [15]), as this will make the amplitude smaller than expected and hence lead to an apparent loss of sensitivity [16]. In general, all distortion in a pulse measurement occurs simultaneously with the IR and, hence, cannot be separated from it. To increase the SNR of a tweeter measurement, the impulses can be repeated and averaged in fairly short intervals, as the IR to be recovered is very short and the required linear frequency resolution quite low.

Impulse testing does not allow identifying distortion, but is pretty immune to the detrimental effects of time variance that frequently haunt MLS- or noise-based outdoor measurements. It is simple, does not require sophisticated signal processing, and works very well for some measurement tasks. Consequently, it has been a popular method for quite a while [15, 17]. When amplifier power is available in abundance, the increase of SNR of measurements using excitation signals stretched out in time compared to single-pulse measurement is not as large as one might expect, because a loudspeaker can generally be fed with pulses of very high voltage.

1.6 Maximum length sequences (MLS)

MLS are binary sequences that can be generated very easily with an N -staged shift register and an XOR-gate (with up to four inputs) connected with a shift register in such a way that all possible 2^N states, minus the case “all 0”, are run through [18]. This can be accomplished by hardware with very few simple TTL-ICs or by software with less than 20 lines of assembly code.

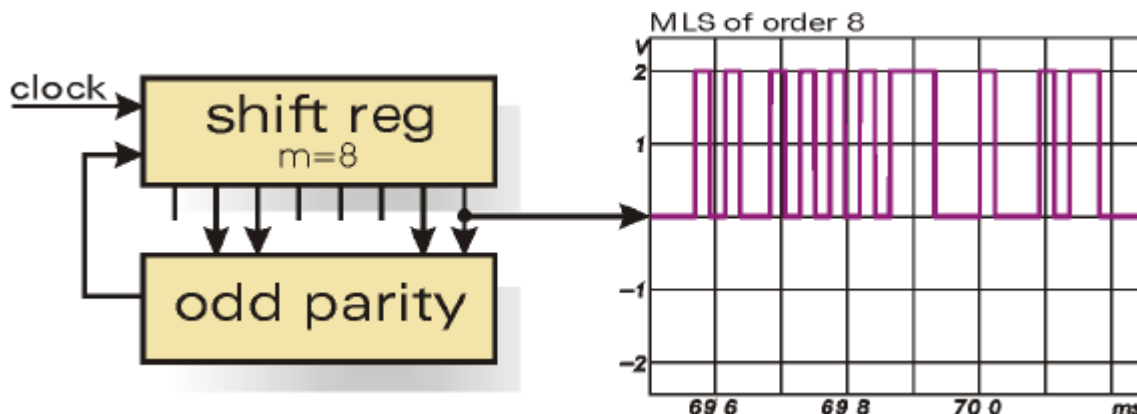


Fig.4a. Generation of MLS with shift register fed back over odd parity generator.

During the time MLSs grew popular, the possibility to create the sequences by hardware alleviated from memory constraints. In the 1980s, the maximum memory deployed by an 8088-based IBM PC was 640 KB. To circumvent the need to store the excitation signal permitted larger data arrays to capture and process the DUT's response. Today, this advantage has totally vanished and it is more cost-effective and flexible to create MLSs by software and output them from memory via the DA converter of a measurement system.

As the case “all zeros” is excluded from the sequence, the length of an MLS is $2^N - 1$. MLSs have some unique properties that make them suited for transfer function measurements. Their auto-correlation comes close to a Dirac pulse, indicating a white spectrum. Repeated periodically as a pulse train, all frequency components have indeed exactly the same amplitude, meaning their spectrum is perfectly white. Compared to a pulse of same amplitude, much more energy can be fed to the DUT as the excitation signal is now stretched out over the whole measurement period. This means increased SNR.

Normally, an MLS is not output as a pulse train, as this would mean feeding very little power to the subsequent DUT. Instead of this, the output of a hardware-MLS generator is usually kept constant between two clock pulses. This first-order hold function leads to a $\text{sinc}(x)$ aperture loss, which reaches almost 4 dB at $f_s/2$ and therefore must be compensated. In contrast, when the MLS is output by an oversampling audio DA converter, as is standard today, the spectrum will be flat up to the digital filter’s cutoff frequency. In the case of cheaper codecs, a noticeable ripple might be introduced over the whole passband. These frequency-linear undulations are always present to a certain extent in oversampling audio converters. They originate from the linear-phase FIR anti-alias filters. These usually trade passband ripple against stop band attenuation by means of the Parks-McClellan algorithm [19]. Furthermore, they are practically always designed as half-band filter, which halve the required calculation power, but only exhibit an unsatisfactory attenuation at $f_s/2$. This is why a small invalid aliasing region always exists near the Nyquist frequency when measuring with audio converters. The anti-alias filter also induces a hefty overshoot of the output MLSs which means that they cannot be fed with full level. Section 2.2 will clarify the issue.

Excitation signals with white spectrum allow the use of the cross-correlation between output and input of the DUT to retrieve its IR. While normally a cross correlation is most efficiently performed in the spectral domain by complex-conjugate multiplication, the well-known fast Hadamard transform (FHT) can perform this task for MLSs without leaving the time domain. We will omit the presentation of the theory behind the FHT, as it has been thoroughly explained many times (see for example [18], [20-23]). The butterfly algorithm employed in the FHT only uses additions and subtractions and can operate on the integer data delivered by the AD converter. In former times, this meant calculation times that were much shorter than that of an FFT of similar length, but today this difference has shrunk a lot. Modern processors as those of the Pentium II/III family are able to perform floating-point multiplications, additions and subtractions as fast as the respective integer operations.

Again back in the 1980s, the time-saving property of the Hadamard transform was very welcome, as the calculation of a long broad-band RIR still took many seconds. The advantage became especially prominent when the IR alone, and not its associated transfer function, was of interest. This, for example, holds for the evaluation of reverberation times by backward integration of RIRs [24]. In such cases, just one FHT is required to transform the MLS response captured by the microphone into the desired PIR. And this FHT is faster than an FFT. In contrast, using arbitrary noise signals or sweeps as stimuli requires at least one FFT and one IFFT to retrieve the IR. However, the processing times are no longer of concern, as the transformations with today’s more powerful processors can be performed much faster than real time.

For example, processing an MLS of degree 18 (with a period length of six seconds at 44.1 kHz sampling rate, a typical length for broad-band measurements in quite reverberant ambiances) is completed in only 138 ms on a Pentium III/500, using a 32-bit-integer Radix-4 MMX-FHT (making use of the eight 64-bit-wide MMX registers) partitioned into sub-chunks accommodated to the sizes of 1st and 2nd level cache. This encompasses the permutation needed before and after the butterfly algorithm and a peak search. A real-valued FFT [11] for the same length, also using nested sub-chunks that can be processed entirely in the caches, terminates in roughly double the time (280 ms), still a lot less than the measurement period. So regardless of the measurement principle, today it is possible to transmit the excitation signal continuously and to complete calculation and display updating within every period, even for two or more input channels. Shorter measurement periods even achieve a higher real-time score as they are handled entirely in the caches. Of course, the number of operations per output sample also decreases slightly in the ratio of the degrees (for example, an FFT of degree 12 only needs two thirds of the operations per output sample that one of degree 18 needs).

In an MLS based measurement, the FHT is the first signal-processing step after digitization by the AD converter (Fig. 4). The resulting IR can be shifted in a cyclic fashion and windowed, as with simple impulse testing. If the transfer function is the objective, an additional FFT must be performed. But as MLSs have a length of $2^n - 1$, one sample must be inserted to patch the IR to full 2^n length. While this may be a trivial action computationally, care must be exercised regarding where to place this sample. It must be in a region where the IR has decayed to near zero to avoid gross errors.

When using a window, the sample can be placed in the muted area. The acquired transfer function again can and should be corrected by multiplication with a reference spectrum obtained previously by a self-response measurement.

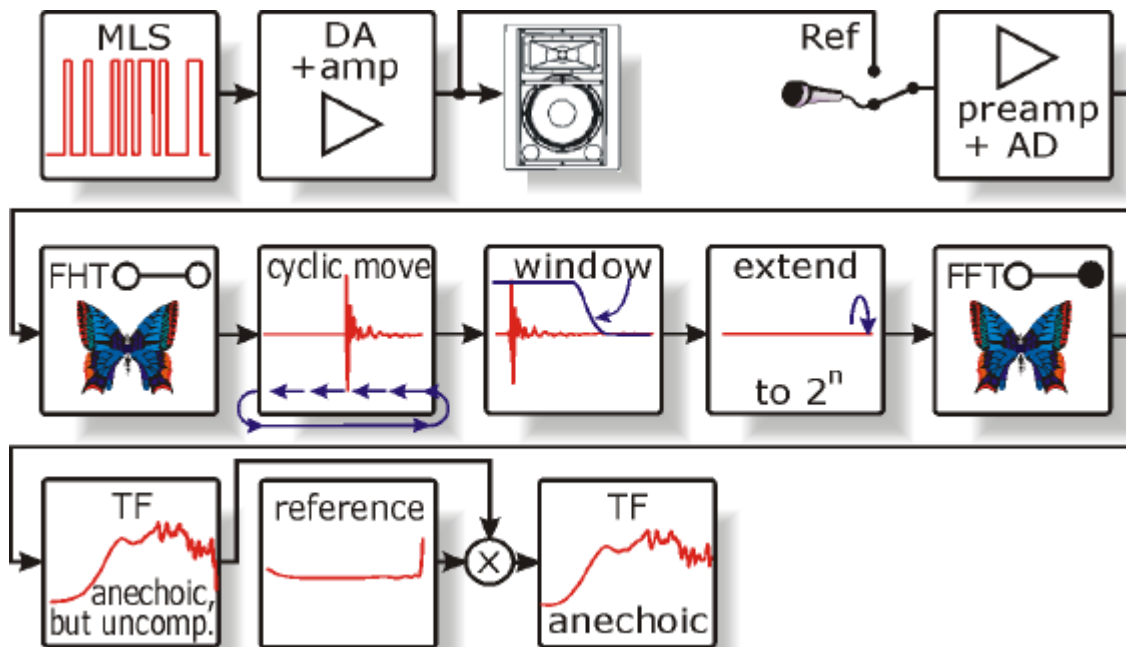


Fig. 4: Signal-processing stages for transfer-function measurements with MLS.

MLS measurements have proved quite popular in acoustics, but have several drawbacks. Along with a high vulnerability to distortion and time variance (these will

be compared directly to sweep measurements in section 2) the most undesired property of MLSs is their white spectrum. As will be advocated further in section 3, a non-white spectrum is desirable for almost all types of acoustical measurements. This requirement can be achieved by coloring the MLS with an appropriate emphasis. Clearly, the MLS will lose its binary character by pre-filtering. Thus, this technique is only viable if the pre-filtered MLS is output by a true DA converter, not just a one bit switching stage as used in some old-fashioned hardware-based MLS analyzers. The latter ones are restricted to analogue post-filters to emphasize the MLS, but these do not offer the versatility of FIR filters, such as linear phase or compensation of the measurement system's self-response [25].

Creating an emphasized MLS can be done most efficiently by means of the inverse fast Hadamard transform (IFHT) [25, 26, 39]. The IFHT simply consists of time inverting the IR of the desired emphasis filter (curtailed to 2^N-1 samples), applying a normal FHT on the inverted IR and then time inverting the result again. This will yield an MLS periodically convolved with the emphasis filter. Due to the periodicity, every discrete frequency component of the former MLS can be influenced independently in both amplitude and phase.

When an emphasized instead of a pure MLS is being used as stimulus, obviously the IR obtained by FHT will consist of the IR of the DUT convolved with the IR of the emphasis filter. For acoustical measurements, it is meaningful to give the excitation signal a strong bass boost of maybe 20 or 30 dB, as will be illustrated later. In this case, the recovered IR may become much broader than the one of the DUT alone. This broadening often constraints windowing, especially when reflections that are to be muted are in close proximity of the main peak. In these cases, applying a window “pre-comp” to the non-equalized IR will noticeably attenuate the low frequency energy spread out in time. Thus, it is better to perform the windowing “post-comp”, that is, after transforming the uncorrected IR into the spectral domain, then multiplying it with the inverse emphasis frequency response, and eventually back-transforming it into the time domain. This will yield the true IR of the DUT alone, which can now be windowed with lesser low frequency energy loss.

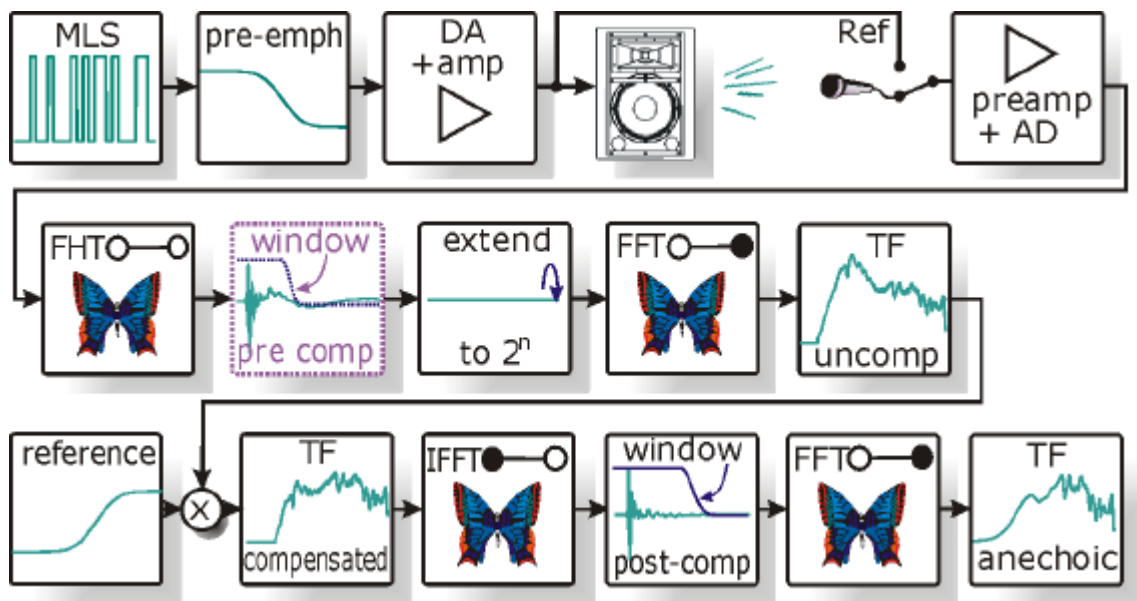


Fig. 5. Signal-processing stages for TF measurements with pre-emphasized MLS.

Instead of multiplying just with the inverse emphasis spectrum, it can be interesting to “over-compensate” the frequency response by even stronger attenuating the low frequencies. After the IFFT-window-FFT operation, the frequency response is corrected by dividing it by the product of emphasis spectrum with the over-compensating reference spectrum. This way, the window takes out even less low frequency energy, which extends the validity of quasi-anechoic measurements towards lower frequencies [41].

If the transfer function is the desired result of the measurement, the total number of transformations becomes one FHT and three FFTs when measuring transfer functions with pre-emphasized MLSs and applying “post-comp”-windowing of the IR.

1.7 Periodic Signals of length 2^N

A thorough examination of the MLS measurement setup in Fig. 5 reveals that in fact the use of MLSs and the application of the FHT are pretty superfluous. If the excitation signal had 2^N samples instead of the odd 2^N-1 of an MLS, the DUT response could be transformed directly to the spectral domain, omitting the FHT. There, it could simply be multiplied by the reference spectrum (the inverse of the product of the excitation signal’s spectrum and the measurement system’s frequency response). As this multiplication is a complex operation, not only the magnitude, but also the phases are thereby corrected to yield the true complex transfer function of the DUT, regardless of the excitation signal’s nature. Performing an IFFT on this compensated spectrum will produce the correspondent true IR of the DUT (Fig. 6).

When comparing this deconvolution technique (FFT, compensation, IFFT) to the FHT, it becomes clear that it is far more powerful and flexible, allowing the use of arbitrary signals of length 2^N . The FHT, being a cross-correlation algorithm, is able to merely reshuffle the phases of a special class of excitation signals, namely, MLS. Its operation is “pulse-compressing” the MLS by means of correlation with the correspondent “matched filter”.

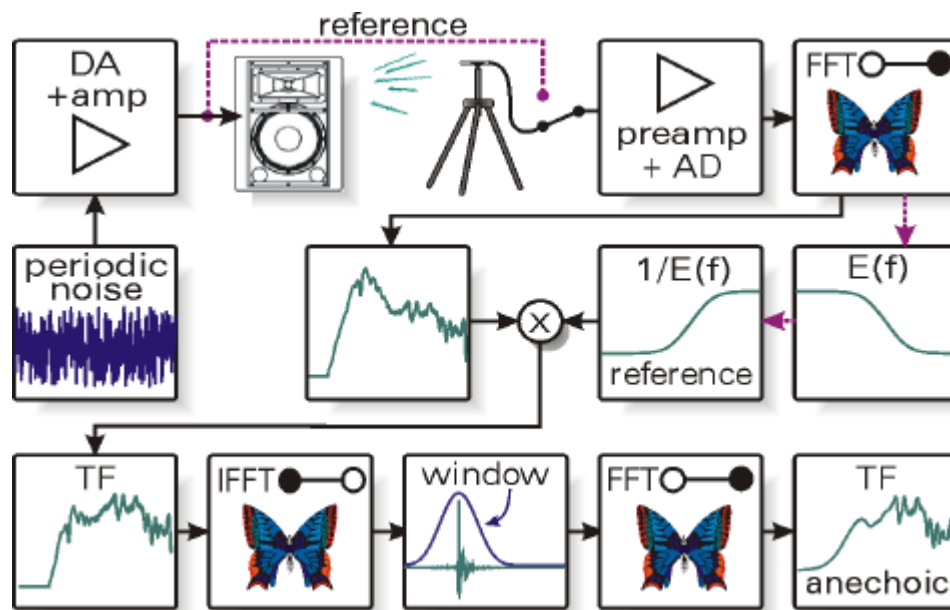


Fig. 6. Signal-processing stages for TF measurements with any deterministic signal.

In contrast, the FFT approach compensates the phase *and the magnitude* of any excitation signal, be it noise, sweeps, or even short chunks of music. This operation is sometimes referred to as “mismatched filtering”, a misleading name, as in reality, the filter precisely can be matched to every excitation signal. In contrast to “matched filters”, it is not restricted to white excitation signals. The only obvious restriction is that the excitation signal must have enough signal energy over the whole frequency range of interest to avoid noisy parts in the transfer function obtained.

Clearly, performing two FFTs consumes more processing time than one single FHT. But with today’s more powerful microprocessors, this disadvantage is insignificant. As we have seen, using pre-emphasized MLSs and “post-comp” windowing (Fig. 5) even leads to slightly longer calculation times than using arbitrary 2^N -signals.

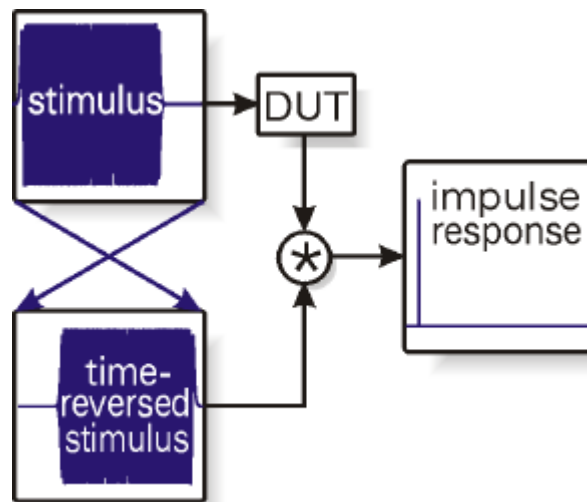


Fig 6a. In the case of white excitation signals (here, a linear sweep has been used as example), the DUT’s impulse response can be calculated by performing a *cross-correlation*, which is the *convolution* (denoted by the * symbol) with the *time-reversed* excitation signal. This would generally take a long time in the time domain, but the FHT does it in a very fast way for MLS.

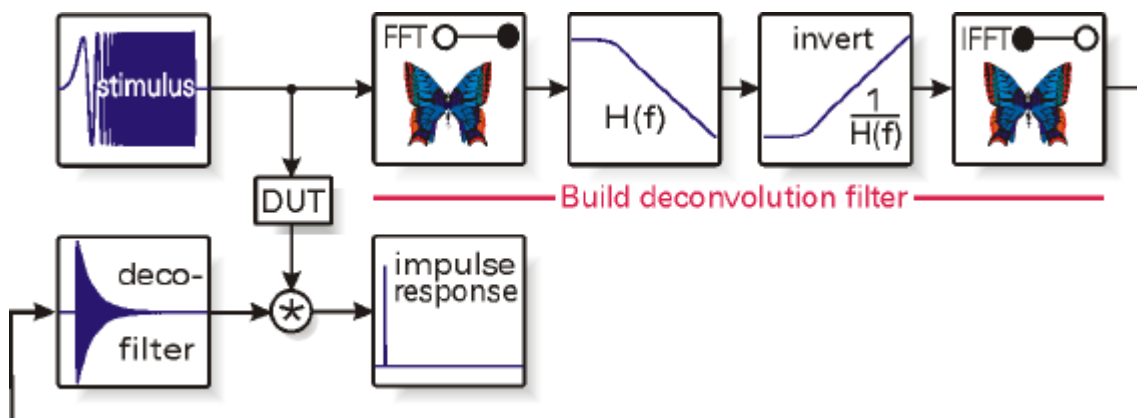


Fig. 6b. For any non-white excitation signal, the evaluation of the impulse response in the time domain could only be done by convolution with the inverse filter, which itself has to be constructed in the frequency domain. Convolution in the time domain corresponds to multiplication in the frequency domain, which can be performed much faster. That’s why the deconvolution is usually performed in the frequency domain.

The technique resembles good old dual channel FFT analysis, but differs from it in that the excitation signal is known in advance. Hence its spectrum needs to be calculated only once and can be used in all subsequent measurements. This removes the need of a second channel, or if present anyway, allows for the analysis of two inputs simultaneously. An accompanying benefit is the fact that the achievable precision of a “single channel FFT analyzer” outperforms that of any dual channel analyzer. With the latter, any difference in the frequency response of the two input channels will be reflected in the DUT’s measured frequency response. Of course, manufacturers of pricey dual channel analyzers endeavor to make these differences as small as possible. However, creating a reference file by replacing the DUT with a wire makes the excitation signal pass over exactly the same stages and guarantees higher precision. Even with consumer equipment, a certainty of 1/1000 dB or better can be achieved without large efforts in purely electrical measurements. The reference voltage sources included in modern AD and DA converters are stable enough to permit this for a certain while (when slow drift due to heat-up occurs, the reference measurement can quickly be repeated). However, sources of error to guard against are the impedances of the analogue input and output stages. The DA output impedance should be as low as possible to prevent a drop of the generated voltage when connecting the DUT, whereas the input impedance should be sufficiently high to avoid influencing the DUT’s output voltage. Clearly, these conditions are rarely fulfilled when using simple soundboards without buffering amplifiers.

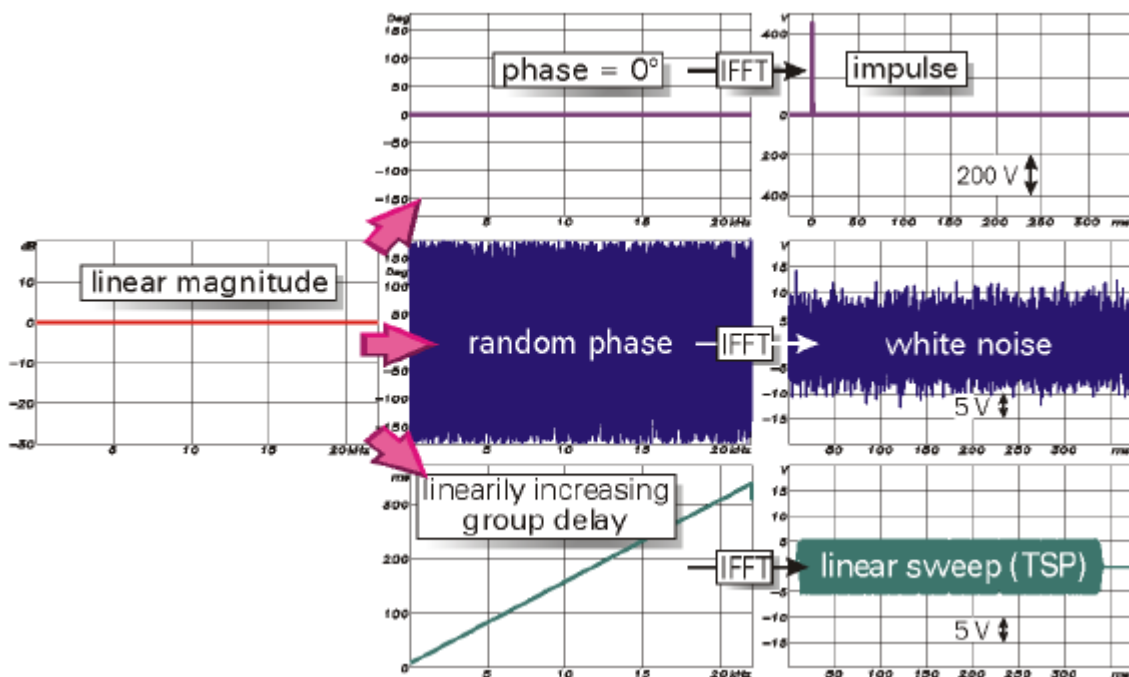


Fig. 6c. Three white excitation signals: impulse, noise and sweep, generated by synthesis in the frequency domain. All have the same amplitude spectrum, but their phase is obviously very different. For the impulse, the phase has to be set to 0° which corresponds to an equal arrival time for all frequencies. The phase spectrum for the noise has to be set to random values. To create a white sweep, the group delay (which is proportional to the derivative of the phase) has to increase proportionally with the frequency.

The three excitation signals shown here are normalized to have identical energy. The sweep has the lowest Crest factor of all, 6 dB lower than white noise.

The impulse needs an amplitude of several hundred volts to concentrate the same energy. This, of course, restricts its practical usefulness.

An even bigger asset of the “single channel analysis” is the use of a deterministic signal in contrast to the uncorrelated noise sources normally used in dual-channel FFT analyzers. As stated before, the latter have a steady spectrum when averaged over a long time, but a single snapshot of the noise signal suffers from deep magnitude dips. Thus, a dual-channel analyzer must always average over many single measurements before being able to present a reliable result. In contrast, the deterministic stimulus used in the method of Fig. 6 can be custom tailored by defining an arbitrary magnitude spectrum free of dips, adapted to the prevailing noise floor, to accomplish a frequency-independent SNR. According to the desired signal type, the corresponding phase spectrum can then be constructed in quite different manners.

A noise signal can be generated easily by setting its phases to random values. The excitation signal is obtained by IFFT. Repeated periodically, it will have exactly the magnitude spectrum previously defined (for example, flat or pink). Noise signals have similar properties as MLSs, especially concerning their vulnerability to distortion and time variance. Some people refer to noise signals as “multi-sine signals”, but of course, just any non-pure tone is a multi-sine signal.

As the predefined spectrum of a noise signal is only valid for periodic repetition, the measurement cannot be started immediately after turning on the stimulus. A time corresponding to at least the length of the IR must elapse for the DUT response to stabilize. As the length of the IR is not always known in advance, it is practical to simply eject two periods of the excitation signal. Signal acquisition only starts in the second run, just as with MLS measurements. Similarly, only half of the emitted energy is used for analysis in a single-shot measurement. On the other hand, the periodicity again allows manipulating every frequency component completely independently. For example, single frequencies can be selectively muted or enhanced to reduce or improve the signal energy in particular frequency bands.

1.8 Non periodic sweeps

Instead of randomizing the phases to obtain a noise signal with the desired spectral shape, the phase spectrum can also be adjusted to yield an increasing group delay (the group delay is proportional to the negative derivative of the phase). The IFFT will then reveal a sweep instead of a noise signal. For several reasons, sweeps are a far better choice for transfer-function measurements than noise sequences. First, in contrast to the latter, the spectrum of a non-repeated single sweep is almost identical to that of its periodic repetition. This means that it is not necessary to emit the excitation signal twice to establish the expected spectrum. The sweep must be sent out only once and the DUT’s response can be captured and processed immediately. Thus the measurement duration is cut in half, maintaining the same spectral resolution and signal-to-noise ratio as in a measurement with the stimulus periodically repeated. The minor remaining differences in the spectrum of the repeated and the non-periodic sweep do not matter, as they are reflected and later canceled by the reference measurement that also uses the non-repeated sweep.

The other enormous advantage of a sweep measurement is the fact that the harmonic distortion components can be isolated entirely from the acquired IR. These appear at negative times relative to the direct sound where they can be separated completely from the actual IR. Thus, the IR remains untouched from distortion energy. In contrast,

measurements using noise as stimulus unavoidably lead to the distribution of the distortion products over the whole period.

The reason for the distortion-rejecting property can be explained easily with a small example: Consider a sweep that glides through 100 Hz after 100 ms and reaches 200 Hz at 200 ms. To compress this excitation signal to a Dirac pulse, the reference spectrum needs to have a correspondent group delay of -100 ms at 100 Hz and -200 ms at 200 Hz. When the instantaneous frequency is 100 Hz and the DUT produces second order harmonics, a 200 Hz component with the same delay as the 100 Hz fundamental will be present in the DUT's response. This 200 Hz component will then be treated with the -200 ms group delay of the reference spectrum at 200 Hz and hence appear at -100 ms after the deconvolution process. Likewise, higher-order harmonics will appear at even more negative times.

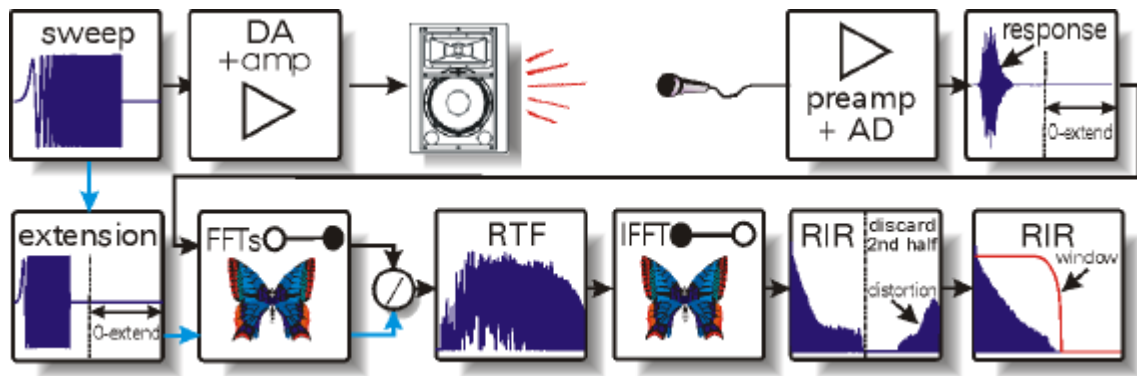


Fig. 6d. For the non-periodic sweep, the most appropriate way to obtain the IR is a linear (i.e. non-cyclic) deconvolution. This will indeed pull all harmonic distortion products to negative times relative to the direct sound, where it is a simple act to discard them.

The linear deconvolution can be accomplished most simply by extending both the excitation sweep and the recorded sweep response with zeros to double their previous size. Both are then submitted to an FFT and the spectrum of the sweep response is then divided by the spectrum of the excitation signal. An IFFT yields the desired IR in which the second half, corresponding to negative arrival times, can be chopped off.

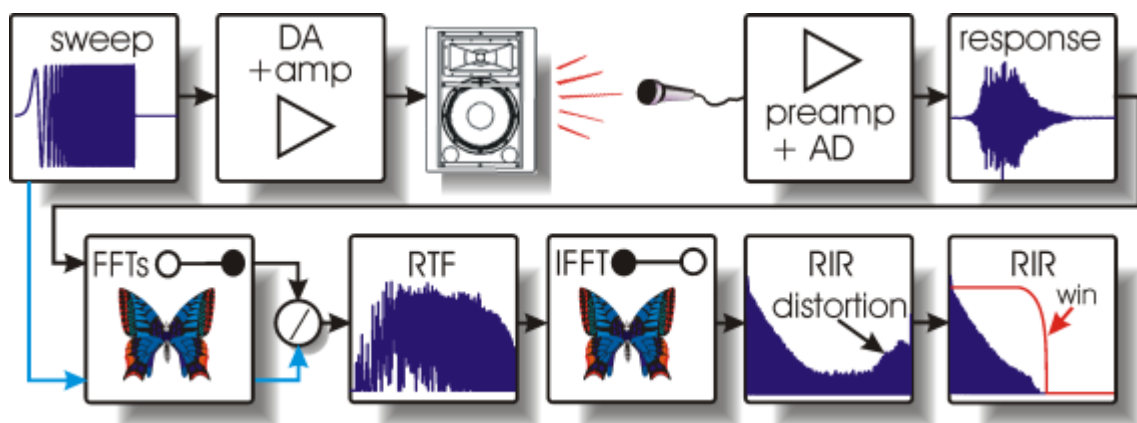


Fig. 6e. Alternatively to the linear deconvolution, a circular deconvolution using an FFT size equal to the acquisition time may be employed. In this case, however, the distortion products could smear into the decay of the IR. This means that the length of the excitation signal has to be chosen sufficiently longer than the decay time. The distortion products will then appear in the noise floor where they can be safely discarded by windowing without affecting the reverberant tail.

In order to actually place all distortion products to negative times relative to the direct sound in the acquired IR, a linear deconvolution suited for non-periodic signals would be necessary. Instead of this, it is also possible to maintain the normal FFT operation and reference multiplication as used in measurements with periodic stimuli, provided that either the excitation sweep is considerably longer than the DUT's IR or zeros are inserted to stuff the DUT's sweep response to double the length (Fig. 6b). In both cases, the distortion products will appear at the end of the recovered IR where they can be interpreted as bearer of negative arrival times. They can be chopped off without corrupting the actual IR, as in the first case, the latter has already decayed into the noise floor, and in the second case no causal information cannot reside in this region.

However, both cases mean that an FFT block length longer than required for the final spectral resolution must be employed as a first signal-processing step. For the goal to capture high-quality RIRs advocated in this paper, it is even advisable to use sweeps that are considerably longer than the IR. This is the best means of increasing the signal-to-noise ratio and decreasing the influence of time variance.

There is an important difference concerning the noise floor in the IRs obtained by linear and circular deconvolution. Using a circular deconvolution results in a noise floor which is basically constant in both amplitude and frequency distribution, up to the point where the first distortion products appear. The linear deconvolution, however, yields a decaying noise tail which is increasingly low-pass filtered towards its end. This stems from the fact that this last part of the deconvolution result originates from steady noise convoluted with a sweep in reverse order (i.e. from high to low frequencies). The user should be aware of this affect and not confound the decreasing noise floor with the reverberant tail of the room.

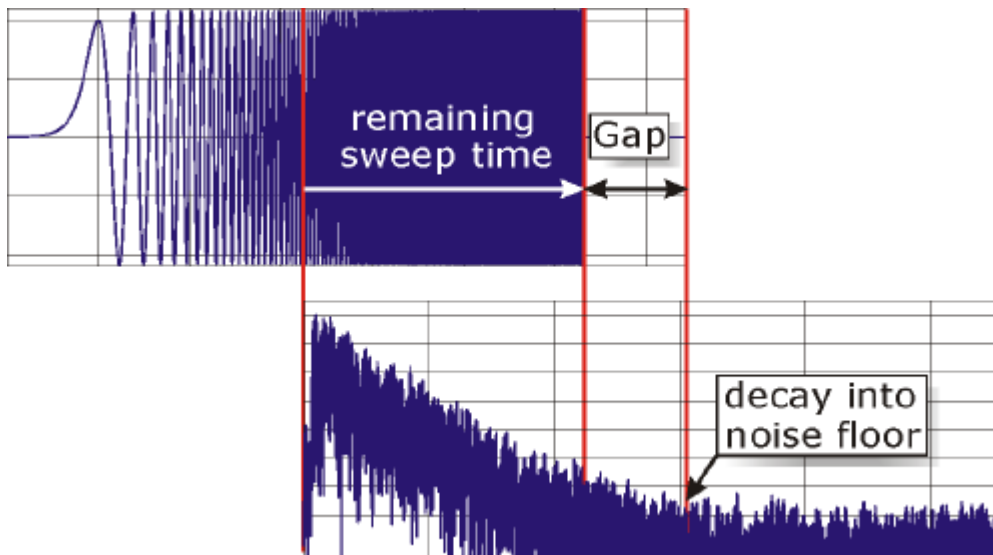


Fig 6f. When the sweep is shorter than the reverberation time, the minimum gap length has to be calculated according to the following rule: For any frequency, the gap length must not be smaller than the time until the reverberation for this frequency decays into the noise floor, minus the remaining sweep time at that frequency.

The data-acquiring period in non-periodic measurements must be made sufficiently large so as to capture all delayed components. This means that the sweep always must be somewhat shorter than the capturing period and the subsequent FFT length. In room

acoustic measurements, it is very beneficial that the reverberation times at the highest frequencies are usually much shorter than the ones encountered at low frequencies. Thus, the sweep must be shortened only by a time correspondent to the reverberation at high frequencies, provided that the entire sweep is long enough to avoid that low frequency reverberation is stumbling behind the high frequency components.

2 A COMPARISON BETWEEN SWEEP AND MLS-BASED MEASUREMENTS

2.1 Measurement Duration

To avoid errors, the AD capture period must to be at least as long as the IR itself (in practice, the time until the response decays below the noise level) in any measurement. This is obvious for the measurement with a non-periodic pulse. All of its energy is emitted at the very beginning, and the AD converter simply must collect samples until the IR has decayed. In case of a non-periodic sweep being used as excitation signal, the capture period must be a little bit longer, but in general not much. This is due to the sweep's nice property of starting at the low frequencies. With normal DUTs such as loudspeakers, the largest signal delays will occur precisely at low frequencies. Thus, while sweeping through the high frequencies, there should be sufficient time to catch the delayed low-frequency components. For loudspeaker measurements, the decay for the highest frequencies is usually so short that AD capturing can be stopped almost immediately after the excitation signal swept through (provided that the sweep is considerably longer than the delay at low frequencies, and of course taking into account the propagation time between loudspeaker and microphone). In room acoustic measurements, the gap of silence following the emission of the sweep usually must be as long as the reverberation at the highest frequencies.

In the case of periodic excitation, the period length and AD capture time do not necessarily have to be longer than the IR length. In contrast, using a shorter length would lead to time-aliasing, i.e., "folding back" the tail of the IR that crosses the end of the period and adding it to the beginning of the IR. Depending on the amount of energy folded back, this creates more or less tolerable errors. Compared to non-periodic pulse or sweep-measurements, the need to emit the excitation signal twice means longer measurement durations than required physically. Moreover, only half of the total energy fed to the DUT is used for the analysis.

2.2 Crest factor

The crest factor is the relation of peak to RMS voltage of a signal, here expressed in dB. If either the measurement system or the DUT is limited by a distinct voltage level, the peak value of any considered excitation signal must be normalized to this value to extract the maximum possible energy in a measurement. The RMS level will then be lower according to the crest factor. Thus, the crest factor indicates how much energy is lost when employing a certain excitation signal, compared to the ideal case of a stimulus whose RMS voltage equals the peak value (crest = 0dB). For this reason, it has almost become a kind of sport among signal theorists to devise excitation signals with the lowest possible crest factor.

The assumption that a certain voltage level defines the upper limit in a measurement is mostly true for purely electrical measurements (for instance, audio gear such as EQs, mixers, etc). In acoustic measurements, it is only valid when the driving amplifier is the weak link in the chain. Even then, most amplifiers can reproduce surge peaks with 2 or 3 dB higher level than continuous power.

If the danger of overheating loudspeaker voice coils is the primary restriction, the total energy fed to the loudspeaker is more important than the crest factor. However, very high crest factors should always be avoided, as single high-level peaks could cause distortion.

At first glance, a bipolar MLS produced by a first-order hold output seems to be the ideal excitation signal in the sense of extracting maximal energy. The peak value equals its RMS value. However, the resulting ideal crest factor of 0dB cannot be exploited in practice. As soon as an MLS goes out to the real world and passes through a filter, the rectangular waveform can change considerably. In particular, the steep anti-aliasing filters used in the oversampling stages of audio DA-converters cause dramatic overshoot. In order to avoid drastic distortion caused by clipping filter stages, MLSs must therefore be fed to the DA converter with a level at least 5 to 8 dB below full scale, depending on the anti-aliasing filter characteristics. This means that MLS cannot be ejected distortion-free with the same energy as a (swept) sine that features a crest factor of merely 3.01 dB.

But even if the MLS is produced by a hardware generator, it will not retain its favorable crest factor for a long time. Power amplifiers are always equipped with an input low-pass filter to reject radio interference and to avoid transient intermodulation induced by high slew rates of the audio signal - precisely what prevails in an unfiltered MLS. A typical input filter would be a second-order Butterworth low-pass with a cutoff frequency of maybe 40 kHz. The overshoot produced by such a filter is much more moderate than that of a steep anti-aliasing filter (Fig. 7), but still merits consideration.

There are cases in which the restriction on the driving level is not the voltage at the measurement system output, but at some internal nodes of the DUT. For example, if a resonance with high gain is encountered in a chain of equalizer stages, the excitation signal will most likely first be clipped at the output of that stage. In these cases, a sweep must be fed with a level that is lower by the amount of gain at the resonance frequency of that specific filter. In contrast, filtered MLSs tend to assume a Gaussian amplitude distribution, with 1% of the amplitudes reaching a level more than 11 dB above the RMS value [13]. In most loudspeaker and room acoustical measurements, even with the presence of moderate resonances, a sweep can still be fed with a higher RMS level than an MLS. In practice, distortion already occurs gradually with MLSs long before the clip level of the driving amplifier is reached, as section 2.5 will examine more thoroughly.

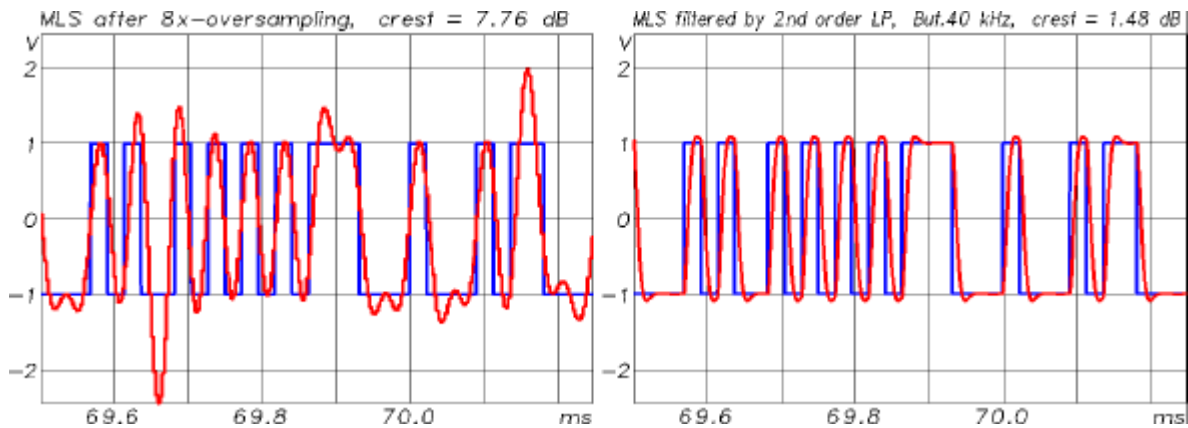


Fig. 7. MLS passed through anti-aliasing low-pass of 8x oversampler (left) and 2nd order Butterworth low-pass with corner frequency of 40 kHz (right).

2.3 Noise Rejection

Any measurement principle using excitation signals with equal length, spectral distribution and total energy will lead to exactly the same amount of noise rejection, if the entire period of the unwrapped IR is considered. For every frequency, the SNR solely depends on the energy ratio of the DUT's response to the extraneous noise captured in the measurement period. The difference between the various measurement methods lies merely in the way that the noise is distributed over the period of the recovered IR.

Clearly, using the same spectral distribution in an excitation signal requires the same inverted coloration in the deconvolution process. That is why the magnitude of an interfering noise source will not vary when changing the stimulus type. The phases, however, will turn out to be very different. Still, some kinds of noise sources will appear similarly in all measurements, as their general character is not altered by manipulating the phases. Monofrequency noise, such as hum, is an example. Likewise, uncorrelated noise (for instance, air conditioning) will still appear as noise.

Any other disturbance, however, will be reproduced quite differently, depending on the type of stimulus. Short, impulsive noise sources, such as clicks and pops, will be transformed into noise when using noise as a stimulus. In contrast, they will become audible as time-inverted sweeps in a sweep measurement. Usually, steady noise is considered to be more unobtrusive than other error signals. However, the time-inverted sweeps that give the IR tail a bizarre melodic touch do not sound too disturbing as long as their level is low. Generally, if any loud noise suddenly happens to appear, the measurement should simply be repeated, or, when averaging, the specific period should be discarded from the synchronous averaging process.

2.4 Time Variance

Time variances tend to haunt measurements whenever they are performed over long distances outdoors, when synchronous averaging is performed over a long period, or when the DUT itself is not reasonably time invariant. The first case often holds for measurements in stadiums or in open-air sites under windy weather conditions. The second is an issue when very low SNRs force the use of several hundred or even thousands of synchronous averages. In such long periods, a slight temperature drift or

movement of the air can thwart the averaging process. Finally, any kind of analogue recording device is an example of a DUT which is inherently time-variant itself.

It is well known that periodic noise sequences in general and MLS in particular are extremely vulnerable to even slight time variances. A considerable amount of theoretical work has already been performed to explain these effects in detail [27-29]. While the complicated equation framework in these publications looks threatening, it is neither easy to compare the effects of time variance in practice, as these tend to have an erratic and unpredictable nature. It is likely that two outdoor measurements performed in series are affected quite differently from wind gusts. Only a simulation allows circumventing this “time variance of the time variance”. Fig. 8 shows a small example.

A noise sequence and a sweep, both with white flat spectrum, have been submitted to a slight sinusoidal time variance of ± 0.5 samples. To simulate this jitter, the signals have first been oversampled by a factor 256 (without filtering, simply inserting 255 consecutive zeros after each sample). Then the exact arrival times of the samples have been shifted in the range of ± 128 according to the sinusoidal jitter curve. The resulting disturbance of the base band spectra is negligible at low frequencies, but then increases dramatically for the jittered noise spectrum. In contrast, the jittered sweep spectrum only displays a minor corrugation at the high end that could easily be removed by applying gentle smoothing.

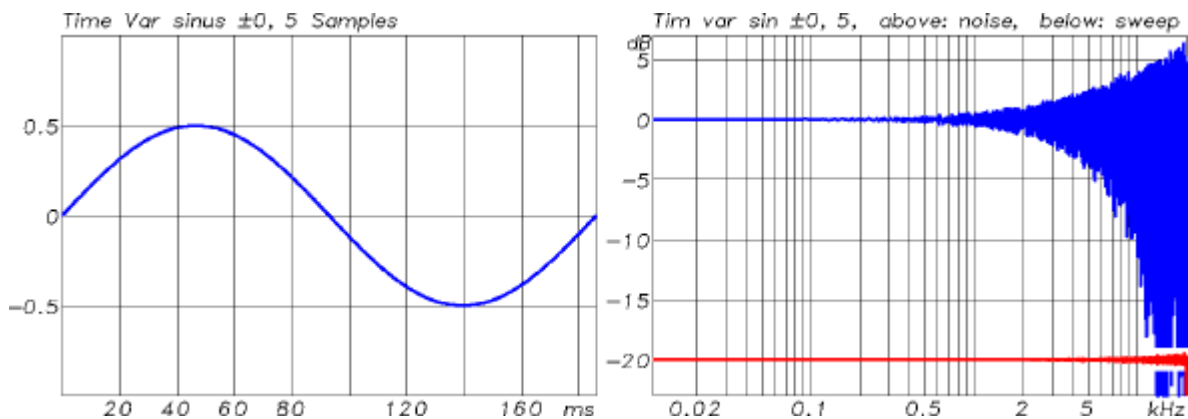


Fig. 8. Artificial sinusoidal time variance of ± 0.5 samples imposed on noise signal (above) and sweep (below) and the resulting spectra.

Another demonstrative test is the frequency response measurement of an analogue tape deck, a kind of machine that always suffers from “wow and flutter” to some degree. As the magnetic tape material tends to saturate much earlier at higher frequencies, an emphasis of 24 dB at low frequencies has been applied to both the MLS and the sweep used in this experiment. In addition, the sweep’s envelope has been tailored to decrease by 12 dB above 5 kHz while maintaining the initial coloration. Section 4.4 will reveal how this is accomplished. Both stimuli were normalized to contain identical energy and were recorded with the same input level. As the results in Fig. 9 show, the measured frequency response using MLS as excitation becomes pretty noisy above 500 Hz. The one using a sweep as stimulus also contains some time-variance induced contamination, but to a much lesser extent. Besides the deleterious effects of time variance, distortion certainly also plays an important role in polluting the recovered frequency responses of an analogue recorder. However, to minimize intermodulation products, the recording level has been adjusted to about 20 dB below the tape’s saturation limit in this trial.

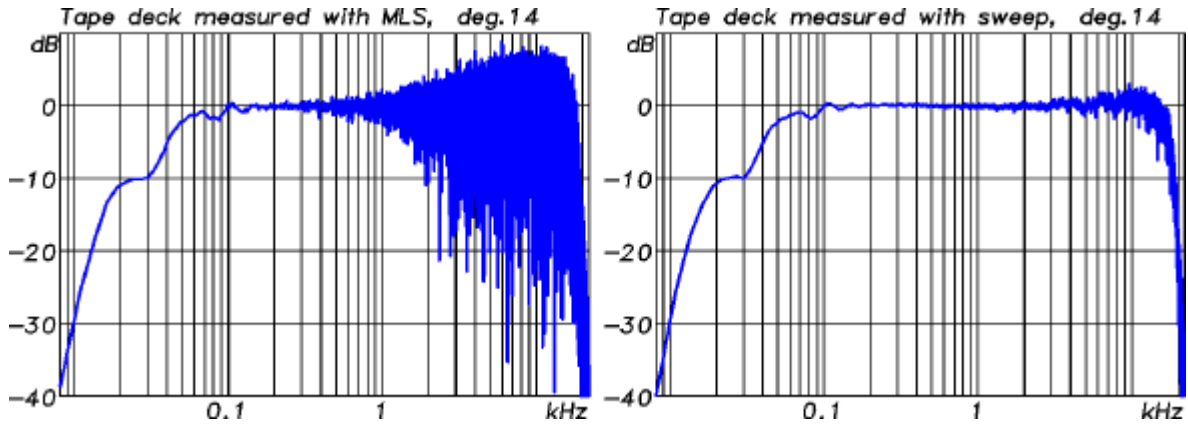


Fig. 9. Transfer function of analogue tape deck measured with MLS (left) and sweep (right), both with identical coloration and energy.

Even when measuring systems that are virtually free of time variance themselves, using pseudo-noise as an excitation is disadvantageous when adjustments (volume, EQ or other) are made within the measurement period. In this case, gross errors occur in the displayed frequency response. Sweep and impulse measurements do not display this unfavorable reaction to time variance and are thus more pleasant for fine-tuning sound systems with continually repeated measurements.

2.5 Distortion

Hardware audio engineers greatly endeavor to optimize the dynamic range of their circuits, which ideally should match that of our auditory system, that is, encompassing up to 130 dB. Yet in the past the SNR in acoustical measurements seems to have been fairly neglected by acousticians.

At relatively calm sites, it is less the background noise that limits the quality of the acquired IRs, but primarily the distortion produced by the loudspeaker employed. In any measurement using noise as excitation signal, these distortion products will be distributed as noise over the entire period of the IR. The reason is that the distortion products of a stimulus with (pseudo-) random phases also have more or less random phases, and the deconvolution process again involves random phases that will eventually produce an error spectrum with random phases, corresponding to a randomly distributed noise signal. As this error signal is correlated with the excitation signal, synchronous averaging does not improve the situation.

While it is true that the noise level diminishes relative to the IR peak value when a longer sequence is chosen, it can hardly be reduced to an acceptable level. Room acoustic measurements already involve lengthy sequences. For example, increasing the length of an MLS of degree 18 by a factor of 128 would theoretically reduce the noise level from a typical value of -65 dB to -86 dB. While it would even be feasible to process an MLS of degree 25 (length: 12 minutes, 41 seconds at 44.1 kHz sampling rate) on a computer generously equipped with memory, it would not work out to reduce the noise level in this manner, because when using such long sequences, the vulnerability to time variance becomes predominant.

Instead of reducing the influence of distortion products by spreading them out over an ever increasing measurement period, it is far more beneficial to simply exclude them

totally from the recovered IR. This can be accomplished easily by using sweeps as the excitation signal, as explained in section 1.8. The great improvement that can be realized in room acoustical measurements by replacing MLSs with sweeps is demonstrated in Fig. 10.

The RIR of a reverberant chamber has been measured here, using an MLS and a sweep of degree 20, both with exactly the same energy and the same pre-emphasis of 20 dB at low frequencies. In favor for the MLS, the volume has been adjusted carefully so as to yield minimum contamination of the IR. Indeed, this optimization of the level by trial and error is crucial to MLS measurements [34], as too much power leads to excessive distortion, shown by an increasing noise floor with lumpy structure, whereas a low level leads to more background noise which impairs the measurement.

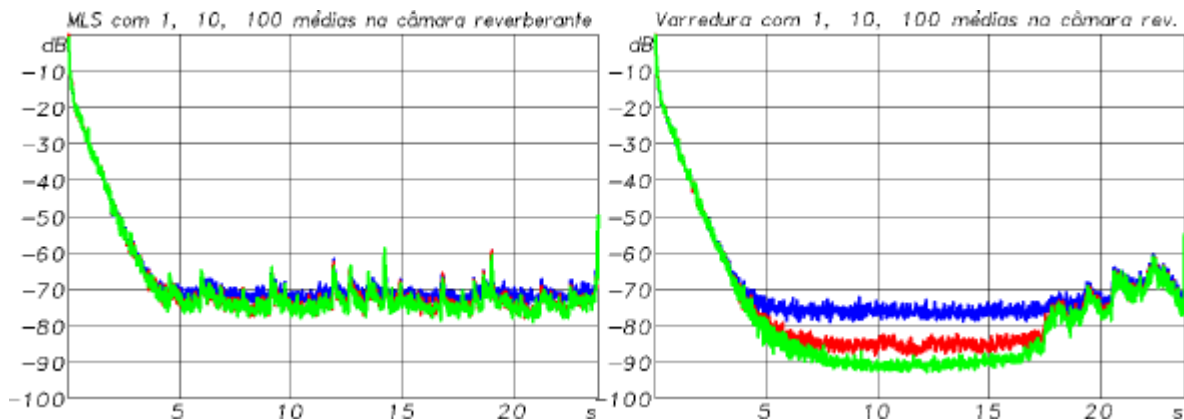


Fig. 10. Measurement of RIR with 12" coaxial PA-speaker in a reverberant chamber. 1, 10 and 100 synchronous averages were performed. Left: with MLS, right: with sweep of identical coloration and energy. The curves are compressed to 1303 values, each of them representing the maximum of 805 consecutive samples.

After this adjustment, 1, 10 and 100 synchronous averages have been performed with both excitation signals. As shown clearly in Fig. 10, even when using no averages, the IR decays into a noise level that is already lower by 5 dB in the case of the sweep measurement. The accumulated distortion products reside at the end of the measurement period. As the length of the excitation signal has been chosen to be almost 8 times longer than the reverberation time, the distortion products can easily be separated from the actual IR. Executing 10 synchronous averages reduces the noise floor by the expected 10 dB when using a sweep. In contrast, only a minor decrease of the noise floor is noticeable when an MLS serves as stimulus, showing that the correlated intermodulation products now prevail in the recovered IR. Performing 100 synchronous averages does not result in any noteworthy improvement in the MLS measurement, whereas a further decrease is recognizable in the sweep measurement, albeit somewhat less than the expected 10 dB. Time variance by a small temperature drift might come into play here, as the 100 averages require almost 40 minutes to complete.

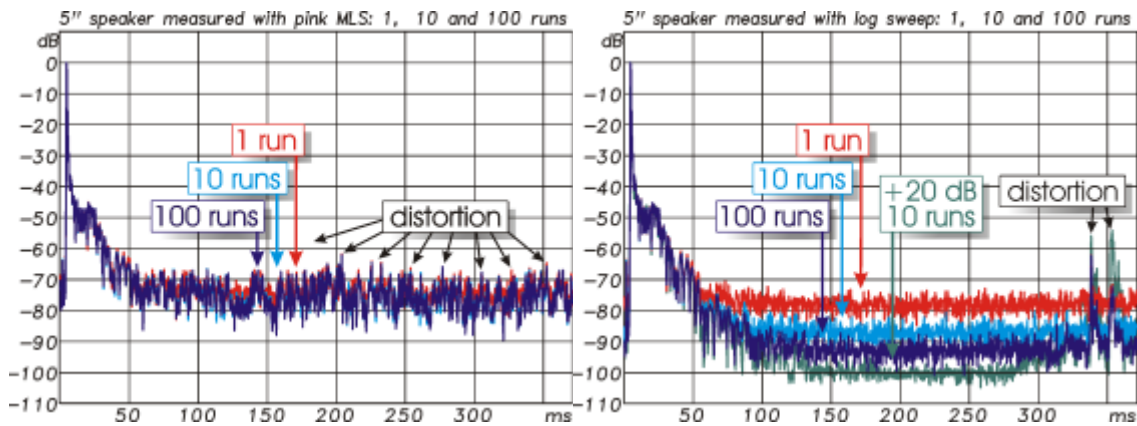


Fig. 10b) Similar experiment as in fig.10 with a small 5" speaker in an anechoic chamber. The level again has been adjusted so to yield minimum pollution of the IR when measuring with MLS. The sweep could be ejected with 20 dB more power, which leads to 100 dB of SNR with just 10 averages (dark green curve on the right side).

In this experiment, the output level has been optimized for use with the MLS. The level of the emitted sweep could have been raised by 15 dB without causing amplifier clipping, and indeed the noise floor dropped exactly by this value when doing so. Thus, an SNR of 100 dB could have been reached with just 10 averages of the sweep. In contrast, it is completely impossible to achieve this high SNR with an MLS measurement, regardless of the level and the number of averages.

The loudspeaker used in this setup was a coaxial 12" PA woofer/tweeter combination optimized for high-efficiency. This kind of speaker certainly produces more distortion than high-fidelity types with soft suspension and long voice coils, but even with the latter types, an SNR of 90 dB or better proves unfeasible in MLS-measurements.

This is shown with a similar experiment using a small soft-suspension 5" coaxial speaker measured in an anechoic chamber with much shorter excitation signals (Fig. 10b). Again, the level was adjusted so as to achieve optimal SNR with the pink MLS. Fed with the same energy, the pink sweep already yields a better SNR with just one run. Its level adjusted in favor for the MLS was so low that it was possible to increase it by 20 dB without causing clipping of the employed 20-Watt amplifier. Doing so, the distortion products at the end of the period rose considerably, but the SNR still increased almost by the same amount as the additional amplifier gain. With just 10 averages (total measurement length: 3.5 seconds), the 100 dB goal was almost reached.

However, driving the transducer considerably into its non-linear range is of course only acceptable in a comparative measurement, in which the transducer itself is not the object of investigation. If the transducer's frequency response itself shall be estimated, operating in its considerably non-linear range will obviously lead to an apparent loss of sensitivity.

Apart from acoustical measurements, there are more measurement situations in which the required linearity for MLS measurements is not fulfilled. This holds especially for signal paths including psycho-acoustic coders. An obvious example are cellular phones which use very high compression to achieve low bit rates. Preliminary experiments showed that short excitation signals produce unpredictable and erratic results, regardless of whether MLS or sweeps were used. Extending the length to degree 18 at 44.1 kHz

sampling rate delivered more or less reliable results with a sweep (linear between 100 Hz to 7kHz), whereas the measurement with an MLS of same length and coloration produced a very rugged curve that hardly allows recognition of the transfer function (Fig. 11). However, it must be admitted that these results only hold when the IR is transformed to the spectral domain without prior processing. Applying a narrow window with a width of only 80 ms around the main peak relieved the situation to a great extent (Fig. 12). Nonetheless, surprising differences appear above 2 kHz between the measurements with MLS and sweeps. The encoder seems to produce different results for broad-band noise and signals that appear almost sinusoidal in a short term analysis.

Another example that certainly appeals more to the audio community than low-fidelity telephone-quality encoding is the popular MPEG 1/layer 3 compression. Fig. 13 shows the transfer function of a coder for the common rate of 128 kbit/s, as measured with MLS and sweep. The advantage of using a sweep becomes overwhelming in this application. In fact, the broad-band noise that the coder must deal with when an MLS is the excitation signal presents the worst case for psycho-acoustic coding. All frequency bands contain energy, so none falls below the masking level that would allow omitting it. Consequently, all bands must be subjected to a fairly coarse quantization to achieve the required bit rate, resulting in distortion that disturbs the MLS measurement significantly. In contrast, the sweep glides through only a couple of bands per analysis interval, allowing one to quantize these with high resolution while discarding the others that contain no signal energy.

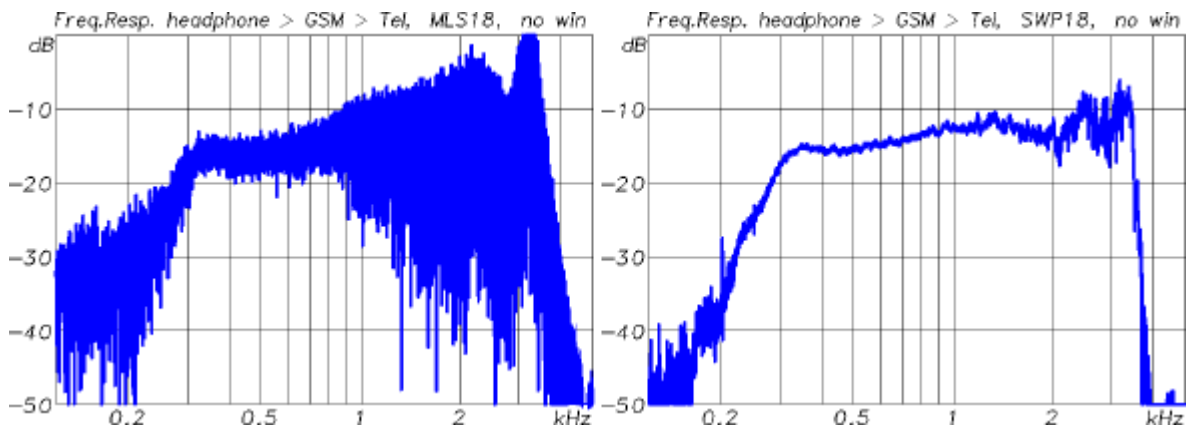


Fig. 11. Transfer function of a GSM cellular (fed acoustically by a headphone), received by a fixed phone. Left: Measurement with linear MLS of degree 18. Right: Measurement with linear sweep of same length.

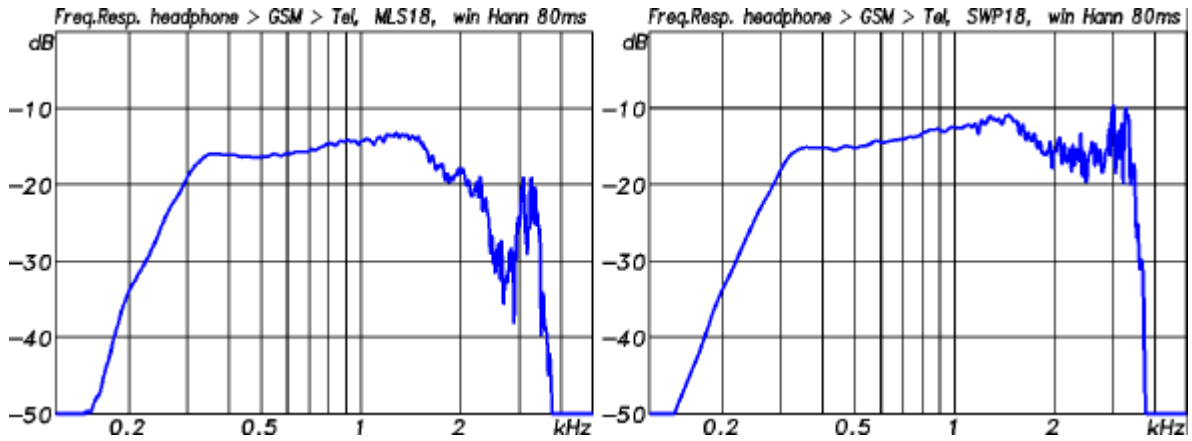


Fig. 12. Same measurement as in Fig. 11, but with a window (Hann, width 80 ms) applied to the recovered IRs.

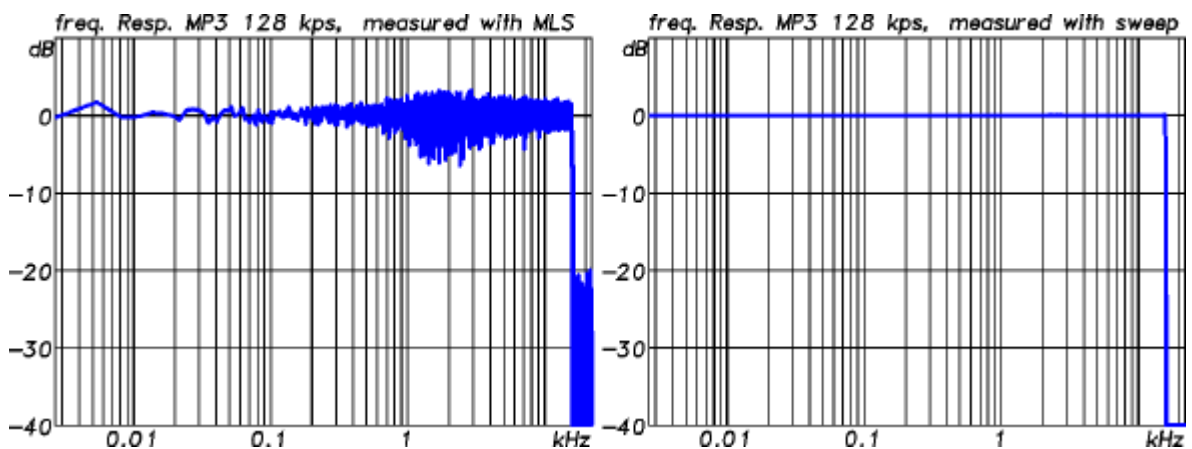


Fig. 13. Transfer function of an MPEG3 Coder at 128 Kbit/s, measured with linear MLS of degree 14 (left) and linear sweep of same length (right).

To summarize, when measuring data-compressing coders, sweeps bear the advantage of considerably reduced signal complexity compared to noise sequences. This makes coding them an easy task. While in “natural” measurements using sweeps, the distortion products can be isolated and rejected by windowing, the advantage in measuring signal paths involving coders lies more in the fact that the generation of distortion is simply prevented.

3 PREEMPHASIS

In almost any acoustical measurement, it is not advisable to use an excitation signal with white spectral contents. When measuring a loudspeaker in an anechoic chamber, two effects account for a substantial loss of SNR at low frequencies: the loss of sensitivity (with 12 or 24 dB per octave) below the (lowest) resonance of the bass cabinet, and the increase of ambient noise in this frequency region due to the wall’s-decreasing insulation against outside noise. So in order to track the speaker’s low-frequency roll-off (if even possible, see [30, 31]) without the corrupting effects of low SNR, a strong emphasis of more than, say 20 dB is required to establish reasonable measurement certainty.

This also results in a better contribution of the power fed to a multi-way loudspeaker. The woofer typically endures much higher power than the tweeter. But when using a white excitation signal, the tweeter must bear the brunt of the excitation signal's energy. A dome tweeter can be overheated and damaged by as little as a few watts, and this limit can easily be exceeded by any power amplifier. Hence, a substantial emphasis of lower frequencies is highly desirable due to this consideration as well.



Fig.14a: Typical background noise from an air condition unit (curve obtained from 30 energetic averages) and various pre-emphasis curves to overcome lacking SNR at low frequencies, compared to a white excitation signal of same total energy. To limit subsonic energy, the pink and red pre-emphasis curves were chosen to not rise further below 30 Hz.

A third reason worth mentioning is less of a physical but rather of a social nature: When measuring “on site” (such as in a concert hall or stadium), an excitation signal with bass boost will sound warmer and more pleasant than white noise, and hence is more acceptable by other people present. Moreover, a strong increase of power in the low band will not correspond to the impression of much increased loudness due to the decreased sensitivity of our hearing sense in this frequency region. While this may not appear rigorously proven, experience from hundreds of measurement sessions in public has shown that the maximum applicable volume is dictated by the persons congregating in the venue, not by the loudspeakers or the amplifiers (see also [1]). When setting up a PA-system for large-scale sound reinforcement, dozens of technicians not related to the audio discipline (lighting etc.) are also present. For instance, riggers climbing on trusses near a loudspeaker cluster would surely benefit from less obtrusive excitation signals. In these cases, using white noise as stimulus even bears a potential health risk and security problem. An MLS accidentally fed with full level to a powerful sound system may cause hearing damage or even lead to accidents of startled personnel. This holds especially true in the vicinity of horn-loaded drivers. Using excitation signals with decent low-frequency emphasis relaxes the situation to some extent. In particular, long sweeps can be interrupted before reaching the excruciating mid frequencies when it becomes clear that their level is too high.

3.1 Equalizing loudspeakers for room acoustical measurements

Measuring RIRs is one of the most common tasks in room and building acoustics. All typical parameters that describe the acoustical properties of a room (or, to be more precise, the acoustical transmission path between two points in the room, using a source and a receiver with distinct directivity) such as reverberation, clarity, definition, center time, STI and many others, can be derived from it. A close study of the IR can help to identify acoustical problems such as unwanted reflections or an undesirable ratio between direct sound and reverberation. Examining the associated room transfer function (obtained by FFT) can reveal disturbing room modes or, of course, tonal misbalance of a sound reinforcement system.

Another very interesting application is the creation of “virtual reality” by convolution of anechoic audio material with binaural RIRs measured with a dummy head. We will later see that only sweeps are capable of fulfilling this task with sufficient dynamic range.

Even today, capturing RIRs for reverberation time measurements is occasionally done using non-electroacoustic impulsive sources. Firing a pistol (a delicate action especially in churches) or popping balloons are common means. While achieving high levels in some frequency bands, these methods have very poor repeatability and produce unpredictable spectra. The low frequency energy content is usually scant, especially for pistols because of their small dimension. Even the omnidirectionality is not at all guaranteed [1]. The only way to avoid these severe drawbacks is the use of an electroacoustical system, which thus brings a loudspeaker into play.

Obviously, when using a loudspeaker without any further precautions, the acquired room transfer functions will be colored by the loudspeaker’s frequency response. This is particularly a problem when the RIR is to be used for auralization. To worsen things, the frequency response is direction dependent. For room acoustical measurements, the ISO 3382 prescribes that the loudspeaker to be used be “as omnidirectional as possible” (a condition that in practice hardly can be satisfied over 2 kHz).

No loudspeaker will be able to produce a frequency-independent acoustical output. This is not a dramatic problem with reverberation time measurements in octaves or third-octaves, as long as the deviation within these bands does not become too high. However, coloration of the room transfer function (RTF) by the loudspeaker’s frequency response is highly undesirable for auralization purposes. In these cases, it is necessary to use a pre-emphasis to remove this coloration. Of course, this equalization could also be done by post-processing the RTFs with the inverse of the speaker’s response, but this would not improve the poor S/N ratio at frequency regions where the acoustical output of the loudspeaker is low. That is why it is especially advantageous to pre-filter the excitation signal in order to allow for a frequency-independent power output.

Loudspeaker equalization is but one component of pre-emphasis that should be applied in room acoustical measurements. Additionally, the measurement can be enhanced to a great extent by adapting the emitted power spectrum to the ambient noise spectrum. And in most of the cases, this background noise tends to be much higher in the lower frequency regions. Thus, in order to achieve an S/N-ratio that is almost constant with frequency, there should be an extra pre-emphasis that reflects the background noise spectrum.

While a frequency independent S/N-ratio is certainly desirable for room acoustical analysis (especially for the extraction of reverberation times in filtered bands), it may be argued that in order to minimize audible noise, RIRs acquired for auralization purposes should have a noise floor that matches our hearing's sensitivity at low levels. Noise shaping to reduce the perceived noise in recordings with a fixed quantization (for instance, the 16 bits of a CD) attempts the same goal. A very expert introduction to this area is given in [32]. To achieve a noise level that is particularly low in the spectral regions of high ear sensitivity, an emphasis equal to such a sensitivity curve (such as the *E* curve) would have to be applied to the excitation signal. On the other hand, giving an extra boost to the mid-frequency region of highest hearing sensitivity will lead to particularly annoying excitation signals. In any case, the question of which emphasis is most suitable is a multifaceted problem and may be answered differently for every measurement scenario.

One conundrum is how to handle measurements for acoustic power equalization of measurement loudspeakers. Usually, the acoustical power response of the speaker is obtained by magnitude averaging many transfer functions measured in the diffuse field of a reverberant chamber. A correction with the inverse of the reverberation times, $10 \log \{1/T(f)\}$, converts the diffuse-field sound-pressure spectrum into a spectrum proportional to the acoustical power.

After this, a smoothing over 1/6 or 1/3 octave is indispensable to obtain a non-corrugated curve. But this smoothing will not be sufficient for low frequencies, as can be seen in the top right curve in Fig. 14. In this range, the chamber's transfer function consists of only a few single high-Q modes. It is then a good idea to replace these peaks in the response by a smoothed, sloping curve with the theoretical decrease of the speaker's sensitivity (12 dB/octave for closed box or 24 dB/octave for vented systems).

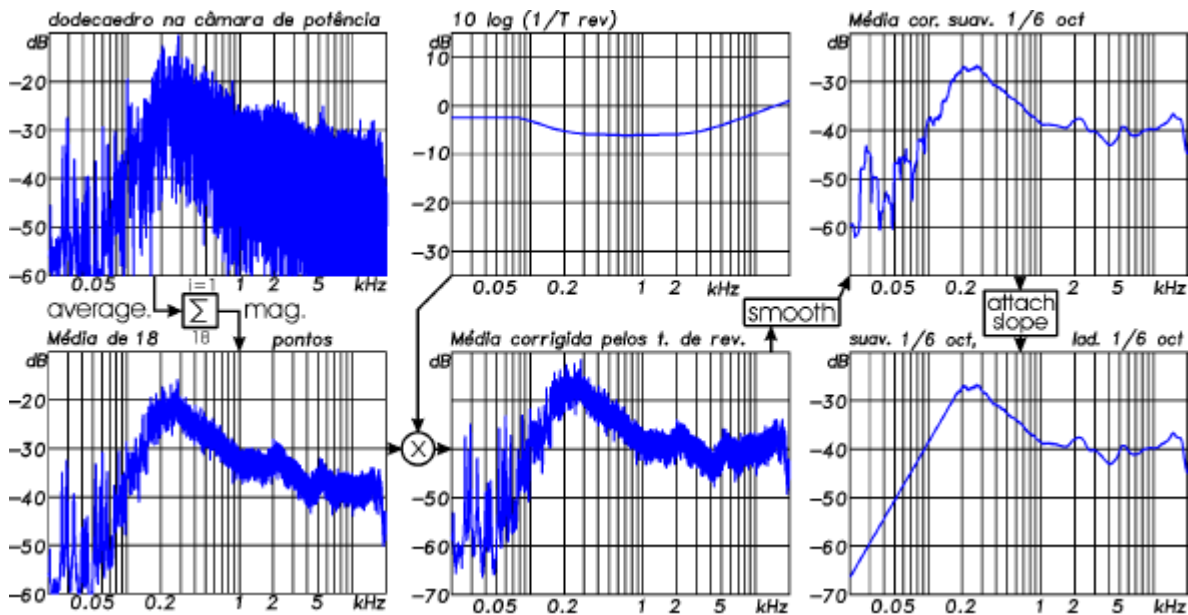


Fig. 14. Steps to construct the acoustical power spectrum of a loudspeaker.

The problem now is how to deal with the phase and the associated group delay. Of course, the phase of a reverberant chamber's transfer function cannot be used. A possible compromise is to combine the acoustic power magnitude response (as obtained in the reverberant chamber) with a free-field phase response (as obtained by measuring

the speaker's sensitivity on-axis). Obviously, merging the amplitude of one measurement with the phase of another leads to an artificial spectrum that will also correspond to a synthetic IR with some artifacts. Nevertheless, this seems to be a viable way to minimize amplitude and phase distortion in room acoustical measurements.

It is interesting to note here that if a sweep is to be created, only the target's magnitude response will influence the excitation signal. Having no influence on the creation of the sweep, the measurement loudspeaker's phase can thus only be equalized by post-processing (that is, applying the inverse of this phase to the reference file). Thus, for general room acoustical measurements, the signal processing will always consist of a combination of pre- and post-processing.

4 SWEEP SYNTHESIS

Sweeps can be created either directly in the time domain or indirectly in the frequency domain. In the latter case, their magnitude and group delay are synthesized and the sweep is obtained via IFFT of this artificial spectrum. Some of the formulas given here are written differently from the typical mathematical standards, but their form is suitable for direct implementation in software.

The two most commonly known types are the linear and the logarithmic sweep. The linear sweep has a white spectrum and increases with fixed rate [Hz] per time unit:

$$\frac{f_2 - f_1}{T_2 - T_1} = \text{const} \quad (1a)$$

Japanese scientists [35, 36] have been using linear sweeps for good reasons long before the MLS technology got popular in acoustics. They refer to the linear sweep as “time stretched pulse” (TSP), but of course any broad-band excitation signal, be it a sweep or a noise signal, could be considered a pulse whose energy has been spread out in time.

In the last years, the originally white “TSP” has been improved for use in acoustics by giving it pre-emphasis at lower frequencies [37]. Also known are bandpass-filtered sweeps for specific purposes [38].

Logarithmic sweeps have a pink spectrum, meaning their amplitude decreases with 3 dB/octave. This also means that every octave contains the same energy. The frequency increases with a fixed fraction of an octave per time unit:

$$\frac{\log(f_2/f_1)}{T_2 - T_1} = \text{const} \quad (1b)$$

4.1 Construction in the time domain

Sweeps can be synthesized easily in the time domain by increasing the phase step that is added to the argument of a sine expression after each calculation of an output sample. A linear sweep as used in TDS has a fixed value added to the phase increment:

$$x(t) = A \sin(\varphi) \quad (1)$$

$$\varphi = \varphi + \Delta\varphi$$

$$\Delta\varphi = \Delta\varphi + Inc\varphi$$

In contrast, a logarithmic sweep is generated by multiplying the phase increment by a fixed factor after each new output sample calculation. So the last line in equation (1) simply changes to:

$$\Delta\varphi = \Delta\varphi \cdot Mul\varphi \quad (2)$$

The value of φ for the first sample is 0 while the start value of $\Delta\varphi$ corresponds to the desired start frequency of the sweep:

$$\Delta\varphi_{START} = 2\pi \cdot f_{START} / f_S \quad (3)$$

The factor $Inc\varphi$ used for the generation of a linear sweep depends on the start- and stop-frequency, the sampling rate f_S and the number of samples N to be generated:

$$Inc\varphi = 2\pi \cdot \frac{f_{STOP} - f_{START}}{f_S \cdot N} \quad (4)$$

In contrast, the factor $Mul\varphi$ necessary to create a logarithmic sweep is calculated by:

$$Mul\varphi = 2^{\frac{\log_2(f_{STOP} - f_{START})}{N}} \quad (5)$$

where $\log_2$ is the logarithm dualis (logarithm with base 2).

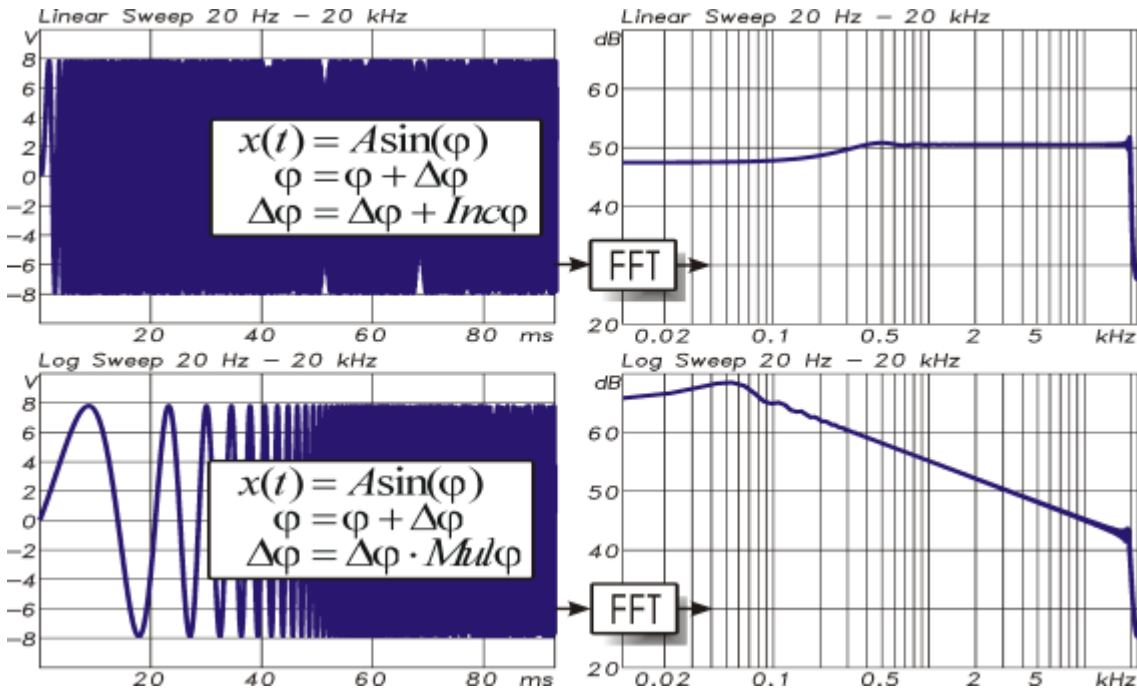


Fig. 15. Sweeps created in the time domain and their spectral properties.

While sweeps generated in the time domain have a perfect envelope and thus the same ideal crest factor as a sine wave (3,02 dB), their spectrum is not exactly what is expected. The sudden switch-on at the beginning and switch-off at the end are responsible for unwanted ripple at the extremities of the excitation spectrum, as can be seen in Fig. 15. Half-windows can be used to mitigate the impact of switching, but do not entirely suppress it. Normally, these irregularities will have no effect on the recovered frequency response when correcting them with a reference spectrum derived by inversion of the excitation spectrum (as obtained by a reference measurement with output connected to input). If, however, the deconvolution is simply done with the time-inverted and amplitude-shaped stimulus, as proposed in [2], or if the correction is not feasible, as with TDS, then errors (such as considerable pre-ringing in the calculated IR) can be expected due to the imperfections near the start and end frequency of the sweep.

4.2 Construction in the frequency domain

Constructing the sweep in the spectral domain avoids these problems. The synthesis can be done by defining the magnitude and the group delay of an FFT-spectrum, calculating real and imaginary parts from them, and finally transforming the artificial sweep spectrum into the time domain by IFFT. The group delay, while sometimes not easily interpretable for complex signals, is a well-defined function for swept sines, describing exactly at which time each instantaneous frequency occurs. For sweeps, the group delay display looks like a time-frequency distribution with the vertical and horizontal axis interchanged (albeit lacking the third dimension, which contains the magnitude information).

If a constant temporal envelope of the sweep is desired (guaranteed naturally by the construction in the time domain), the magnitude and the group delay must have a certain relationship to each other. In the case of a linear sweep, the magnitude spectrum must be white. In the case of a logarithmic sweep, the magnitude spectrum must be pink, that is, with a slope of -3 dB per octave. The associated group delay for the linear sweep can then simply be set by independently from the magnitude spectrum:

$$\tau_G(f) = \tau_G(0) + [\tau_G(f_S/2) - \tau_G(0)] \cdot \frac{f}{f_S/2} \quad (6)$$

with $\tau_G(0)$ and $\tau_G(f_S/2)$ being the desired group delays at DC and the Nyquist frequency ($f_S/2$).

The group delay of a logarithmic sweep is slightly more complicated:

$$\tau_G(f) = A + B \cdot \log_2(f) \quad (7)$$

with

$$B = \frac{\tau_G(f_{END}) - \tau_G(f_{START})}{\log_2(f_{END} / f_{START})} \quad A = \tau_G(f_{START}) - B \cdot \log_2(f_{START}) \quad (8)$$

Normally, f_{START} will be the first frequency bin in the discrete FFT-spectrum, while f_{END} is equal to $f_S/2$. Of course, $\tau_G(f_{START})$ and $\tau_G(f_{END})$ must be restricted to values that

fit into the time interval obtained after IFFT. The phase is calculated from the group delay by integration:

$$\varphi(f) = \varphi(f - df) - 2\pi \cdot df \cdot \tau_G(f) \quad \text{with} \quad df = f_s / 2^N \quad (8a)$$

Phase and magnitude can then be converted to real and imaginary parts by the usual sin/cosine expressions.

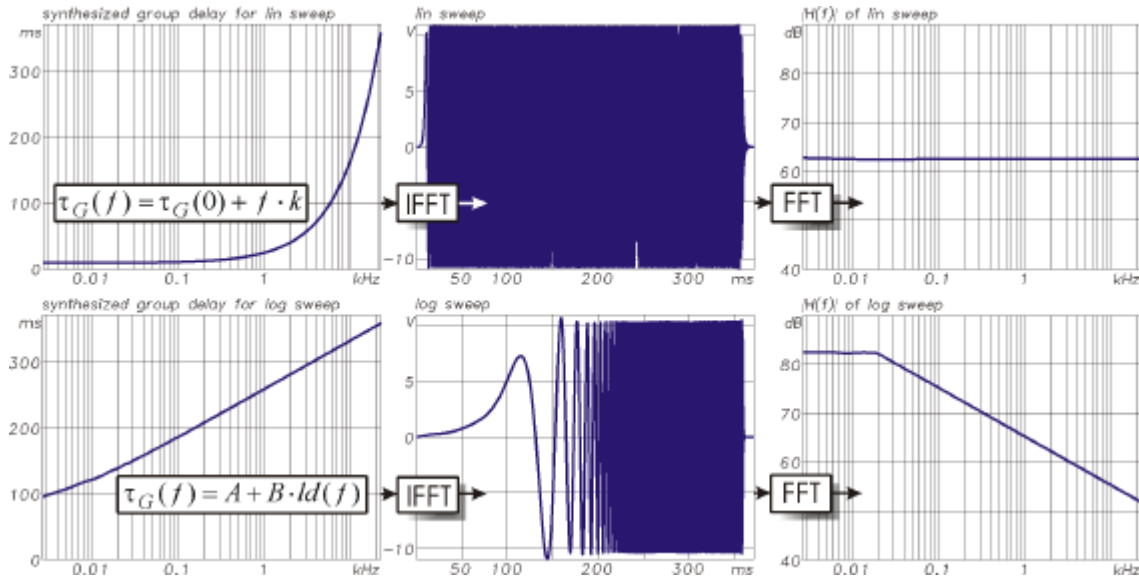
As stated before, creating a sweep in the time domain will cause some contaminating effects in the spectral domain. Likewise, synthesizing a sweep in the spectral domain will cause some oddities in the resulting time signal. First, it is important that the phase resulting from the integration of the constructed group delay reaches exactly 0° or 180° at $f_s/2$. This condition generally must be fulfilled for every spectrum of a real time signal. It can be achieved easily by subtracting values from the phase spectrum that decrease linearly with frequency until exactly offsetting the former end phase:

$$\varphi_{NEU}(f) = \varphi_{ALT}(f) - \frac{f}{f_s/2} \cdot \varphi_{END} \quad (9)$$

This is equivalent to adding a minor constant group delay in the range of $\pm 0,5$ samples over the whole frequency range.

Even satisfying this condition, the sweep will not be confined exactly to the values given by $\tau_G(f_{START})$ and $\tau_G(f_{END})$, but spread out further in both directions. This is a direct consequence of the desired magnitude spectrum in which the oscillations that would occur with abrupt sweep start and stop points are precisely *not* present. Because of the broadening, the group delay for the lowest frequency bin should not be set to zero, but instead be a little higher. In this way, the sweep's first half-wave has more time to evolve. However, it will always start with a value greater than zero, while the remaining part left of the starting point folds back to "negative times" at the end of the period. There it can "smear" into the high-frequency tail of the sweep if the group delay chosen for $f_s/2$ is too close to the length of the FFT time interval. A safe way to avoid contaminating the late decay of the tail by low-frequency components is to simply choose an FFT block length that is at least double the desired sweep length.

To force the sweep's desired start and end point to zero, fade operations of the first half-wave and the tail are indispensable to avoid switching noise. By doing this, it is clear that a deviation from the desired magnitude spectrum occurs. But it can be kept insignificant by choosing sufficiently narrow parts at the very beginning and end of the sweep. The resulting spectra and time signals for both the linear and logarithmic sweeps are shown in Fig. 16. The horizontal platform below 20 Hz for the logarithmic sweep has been introduced deliberately to avoid too much subsonic signal energy. Both sweeps cover the full frequency range from DC (included) to $f_s/2$.



**Fig. 16. Sweeps created in the spectral domain by formulating group delay.
Control of the fade-in/fade-out impact on the sweeps by FFT.**

The ripple introduced by the fading operations (half-cosine windows are most suitable) can easily be kept under 0.1 dB and should not cause any concern, as its impact on measurement results is canceled by performing and applying the reference measurement.

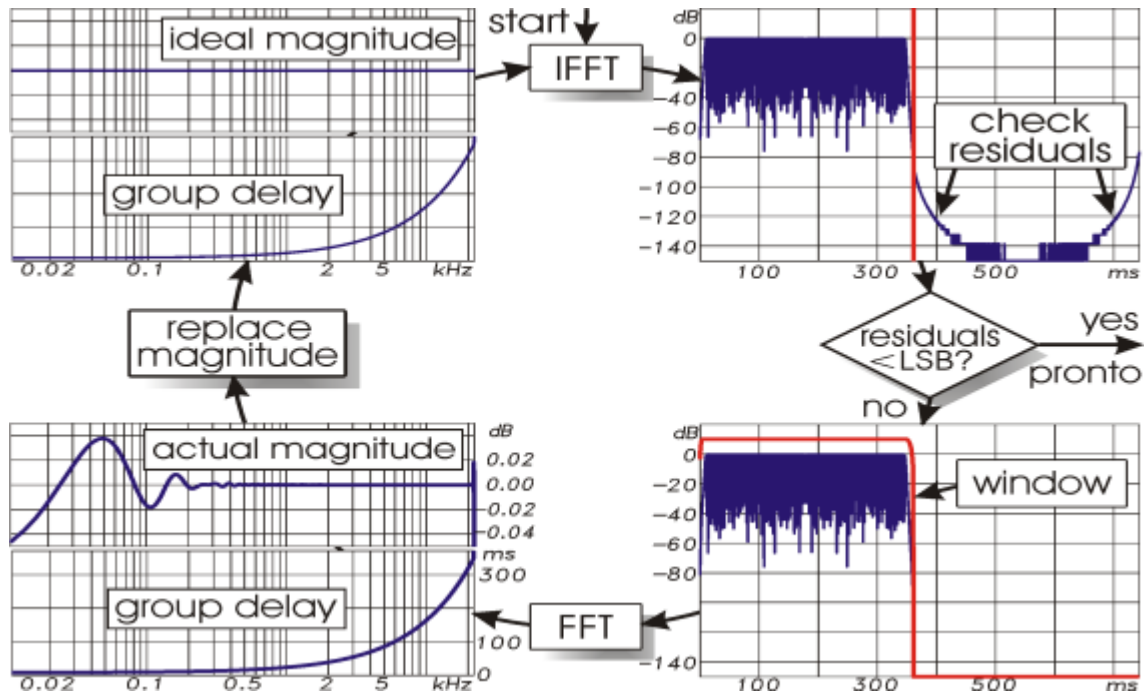


Fig. 17. Iteration to construct broad-band sweeps with perfect magnitude response.

However, in cases when an exact amplitude compensation is not feasible (for example, in a TDS device) or not intended, the iterative method depicted in Fig. 17 permits to reject the ripple completely, establishing exactly the desired magnitude spectrum while maintaining the sweep confined to the desired length. The rather primitive iteration consists of consecutively performing the fade-in/out-operation, transforming the time

signal to the spectral domain, replacing the slightly corrugated magnitude spectrum with the target magnitude while maintaining the phases, and eventually back-transforming the manipulated spectrum into the time domain. Before imposing the fade-in/out windows another time, the residuals outside the sweep interval are examined. If their peak value is below the LSB of the sweep's intended final quantization, the windowing is omitted and the iteration ends. Usually, the process converges rapidly and even a sweep quantized with 24 bits is available after only approximately 15 iterations. However, the perfect magnitude response is traded off by a very light distortion of the phase spectrum. Hence, the derivative group delay will be slightly warped, but the effect is almost unperceivable and restricted to the very narrow frequency strips around DC and $f_s/2$. The iteration works best with broadband sweeps covering the full frequency range between 0 Hz and $f_s/2$.

4.3 Sweeps with arbitrary magnitude spectrum and constant temporal envelope

So far we have restricted discussion to linear and logarithmic sweeps. If their magnitude spectra is altered from the dictated white or pink energy contribution to something different, it is clear that their temporal envelope would not be constant any more. This would entail an increasing crest factor and hence an energy drop, given a fixed peak value as maximum amplitude. Now it would be very attractive to use sweeps with just an arbitrary energy contribution without losing the advantage of a low crest factor. This can be achieved easily by making the group delay grow proportionally to the power of the desired excitation spectrum. In general, the energy of a sweep in a particular frequency region can be controlled by either the amplitude or the sweep rate at that frequency. A steeply increasing group delay means that the corresponding frequency region is stretched out substantially in time, with the instantaneous frequency rising only slowly. This way, much energy is packed into the respective spectral section.

The group delay for the arbitrary-magnitude, constant-envelope sweep can be constructed starting with $\tau_G(f_{START})$ for the first frequency and then increasing bin by bin:

$$\tau_G(f) = \tau_G(f - df) + C \cdot |H(f)|^2 \quad (10)$$

with C being the sweep length divided by the excitation spectrum's energy:

$$C = \frac{\tau_G(f_{END}) - \tau_G(f_{START})}{\sum_{f=0}^{f_s/2} |H(f)|^2} \quad (11)$$

The process is illustrated by an example in Fig.18. The sweep under construction shall serve to equalize a loudspeaker and feature an additional low-frequency boost. The desired excitation signal spectrum in the upper left originates from the inverted loudspeaker response. It is further emphasized with a first order low-shelf filter with 10 dB gain and then high-pass filtered at 30 Hz to avoid an excess of infrasonic energy being fed to the loudspeaker by the new equalizing excitation signal to be generated.

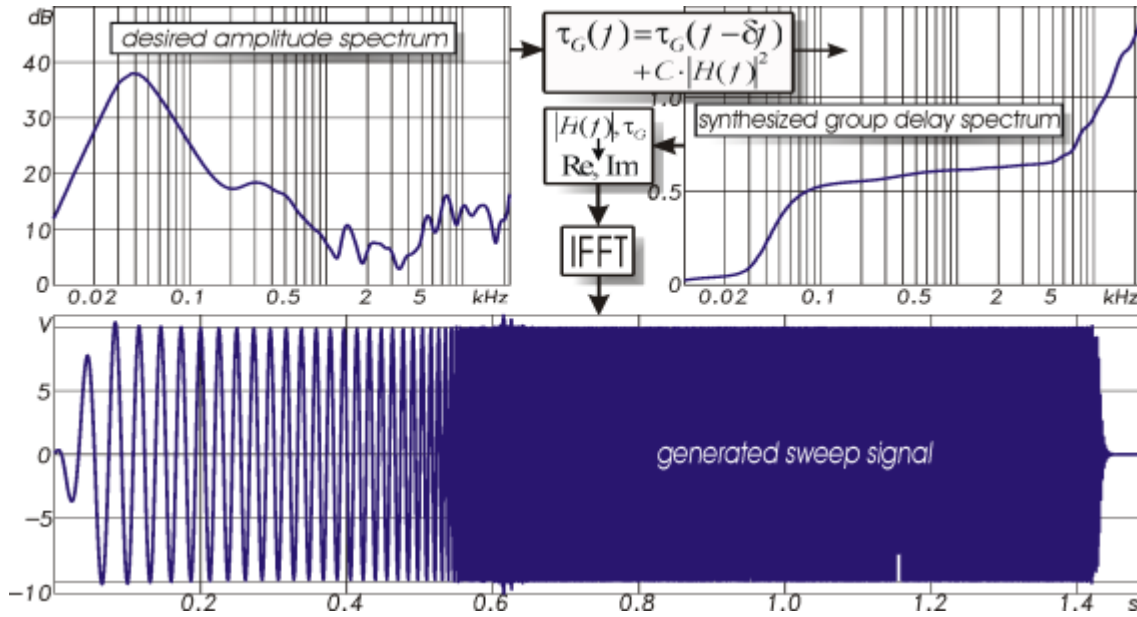


Fig. 18. Generation of a sweep with nearly constant envelope from an arbitrary magnitude spectrum.

The constructed group delay shows a relative steep inclination up to 0.5 second, and the resulting time signal reveals that the frequency in this range increases only gradually from the start value. Thus, the sweep will contain a lot of energy in that frequency region. From 80 Hz to 6 kHz, the group delay increases by merely 200 ms, so the whole midrange is swept through in this short time. Above 6 kHz, the group delay again inclines due to the increased desired magnitude and the frequency rises more moderately, thereby extending the HF part of the sweep.

It is noticeable that the sweep's amplitude is not entirely constant, as would be desirable to achieve the ideal crest factor. At the beginning and at approximately 650 ms, a slight overshoot cannot be avoided. This is another imperfection caused by the sweep's synthetic formulation in the frequency domain. To keep these disturbances small, a slight smoothing of the magnitude spectrum helps. Doing so, the sweep's crest factor can normally be kept below 4 dB, leading to an energy loss of less than 1 dB. Of course, the iterative method described in 4.2 can be combined with this algorithm to reduce deviations from the desired magnitude response near the band limits. Doing so, the crest factor even decreases slightly.

4.4 Sweeps with arbitrary magnitude spectrum and prescribed temporal envelope

Is the constant-envelope sweep with freely definable magnitude spectrum the optimum excitation signal for room acoustical measurements? Well, if the amplifier power should be the restricting limit of the measurement equipment, one would say, yes. The almost constant envelope of the excitation signal allows drawing the amplifier's maximum power throughout the whole measurement, thereby pumping the maximum possible energy into the RUT (room under test) in the distinct time interval. However, lacking amplifier power is rarely a point of concern today (except for battery powered portable equipment). It is much more likely that the power handling capabilities of the deployed loudspeakers have to be considered carefully to avoid damage. In the case of a multi-

way system, for example, a dodecahedron composed of coaxial speakers, perhaps supported by a powerful subwoofer, each way has its own power limit. A constant-envelope sweep must have its power adjusted to the weakest link of the equipment, mostly the tweeters. This would leave the power handling capability of the woofer (which exceeds that of the tweeters often by a factor of 10 or more) for the most part idle. Additionally, an extra boost is often desirable precisely in the low-frequency region to overcome the increasing ambient noise floor.

It becomes clear that in the case of loudspeaker impasses, the instantaneous sweep power should be controllable according to the frequency just being swept through. This can be accomplished by controlling the amplitude of the sweep in a frequency-dependent fashion. To do so, only a minor modification of the sweep creation process revealed in section 4.3 is necessary. As depicted in Fig. 19, the trick is to first *divide* the target spectrum by the “desired-envelope” spectrum. The resulting spectrum is the base for the group delay synthesis, using the same formula as in the “constant-envelope” case. The synthesized spectrum indeed corresponds to a sweep with constant envelope, but with reduced energy at lower frequencies, according to the inverse “desired envelope” spectrum. After creating the real/imaginary pair, the sweep spectrum will now be *multiplied* with the desired-envelope spectrum, reestablishing the desired magnitude response. The IFFT will now produce a sweep that no longer has a constant envelope, but faithfully features the frequency-dependent amplitude imposed by the “desired-envelope” spectrum.

This is the most general form to create an appropriate sweep signal. Two crucial degrees of freedom are offered here: Any user-defined spectral distribution along with any user-defined definition of the frequency-dependent envelope (instantaneous power) will be transformed into a swept sine wave suitably warped in amplitude and time.

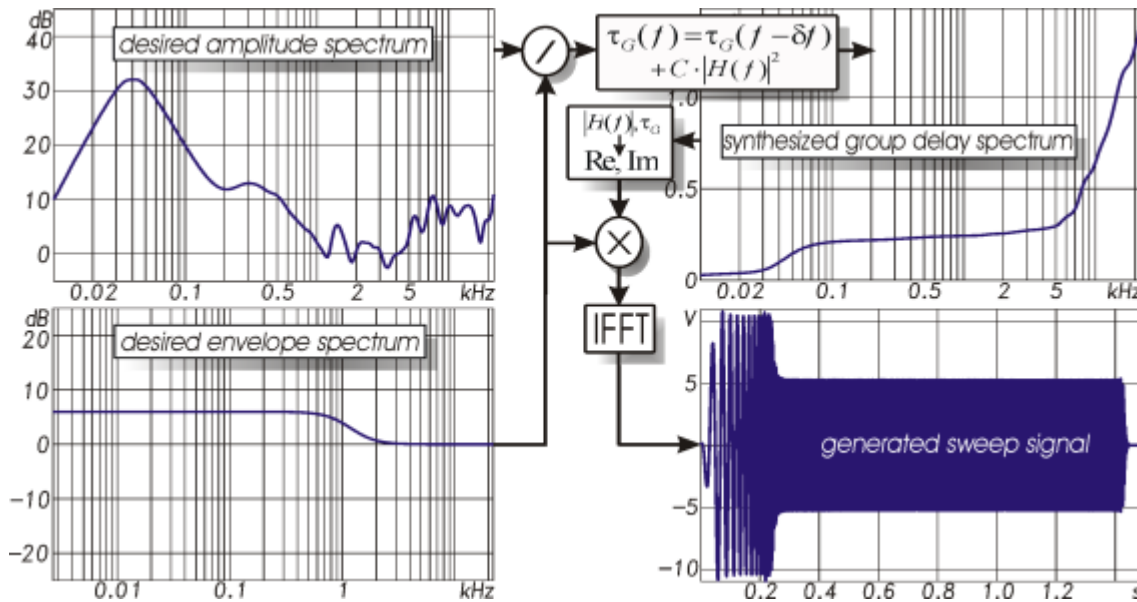


Fig. 19. Sweep creation with desired envelope and arbitrary magnitude spectrum.

It must be admitted that this approach does not entirely remedy the danger of overheating sensitive tweeters. Compared to a “constant-envelope” sweep with identical energy contents, a “desired-envelope” sweep only leads to stretching out the same energy fed to the tweeter in time, resulting in the same heat-up if the sweep length is

smaller than the time constant of the voice coil's thermal capacity. Room acoustical measurements with long sweeps (many seconds), however, benefit from this expanding. The "desired-envelope" sweeps are also advantageous if the tweeter is threatened not only by overheating, but also by mechanically caused damage (for example, by excessive forces in compression drivers). The reduced distortion due to the lesser level may also be motivating, although it is the major asset of sweep measurements that the distortion products can be separated so well from the actual IR.

Another application of sweeps with controlled decrease of the envelope at higher frequencies is the measurement of analogue tape recorders. The envelope can be adapted to the frequency-dependent saturation curve of the tape, thus making optimum use of the tape's dynamic range at every frequency. This way, an apparent drop in response at frequencies where the level gets close to the saturation limit is obviated. In compact cassette tape decks with their narrow and very slowly running tapes, the necessary decrease of the envelope can reach 20 dB or more, depending on the tape material. The saturation curve itself can be determined by deliberately feeding a sweep with excessive level that causes overload at all frequencies. When the distortion products are removed from the IR, the remaining spectrum of the main IR is a good estimate of the frequency-dependent maximal input level for recordings and measurements.

4.5 Dual channel sweeps with speaker equalization and crossover functionality

In many acoustical measurements, multi-way loudspeakers are employed to cover as much of the audio range as possible. For example, an omnidirectional dodecahedron will usually exhibit poor response below approximately 200 Hz, depending on its size. Thus, it is advantageous to support it with a subwoofer to circumvent the need of excessive pre-emphasis. A normal closed or vented box design with just one chassis will still be sufficiently omnidirectional in this frequency range. As all sound cards and most measurement systems are equipped at least with a stereo DA converter, it is advantageous to make use of the two channels to include an active crossover functionality into the excitation signal. This can be achieved with some extensions of the sweep generation methods described in the sections 4.3 and 4.4.

First, as in any equalization task, it is necessary to measure both speakers at the same position to establish valid magnitude, phase and delay relationships. Bearing in mind the speaker's power handling capabilities and according to their magnitude responses, an appropriate crossover point can then be selected. At this frequency, the phase and group delay difference between both spectra is read out and stored. They will be needed later. Now, after an optional smoothing, both spectra are inverted and treated with a band-pass to confine them to the desired frequency range to be swept through (see Fig. 20). At this point, a desired additional emphasis is also applied to the stereo spectrum. Should the excitation signal feature a frequency-controlled envelope, a division by the desired envelope spectrum has to be executed now. After these preparatory steps, the crossover function can be brought into play by multiplying the first channel with a low-pass-filter and the second with the corresponding high-pass filter.

An optional smoothing effect (with constant width on a linear frequency scale) may be obtained by windowing the IRs of the two spectra. For this purpose, both are

transformed to the time domain after deleting their phase. The desired window is applied and the confined IRs transformed back to the frequency domain (see Fig. 21). The window should not be too narrow to avoid too much blurring of the low frequency details.

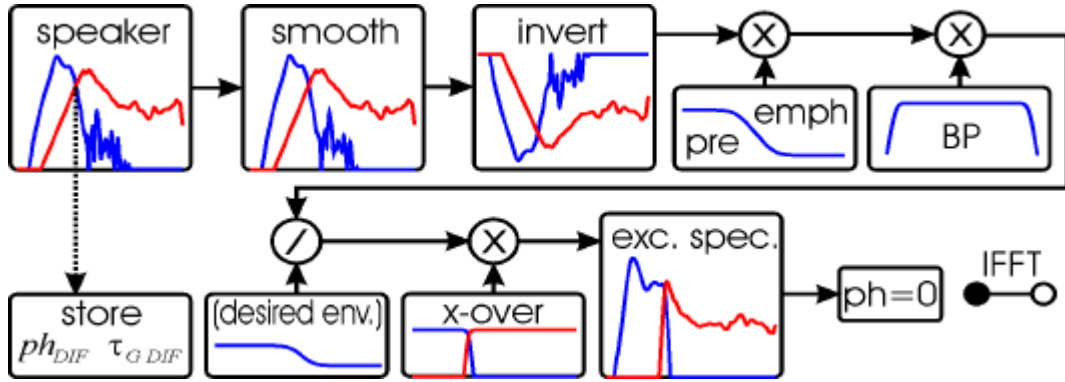


Fig. 20. Preprocessing for dual-channel sweep with 2-way speaker equalization and active-crossover functionality.

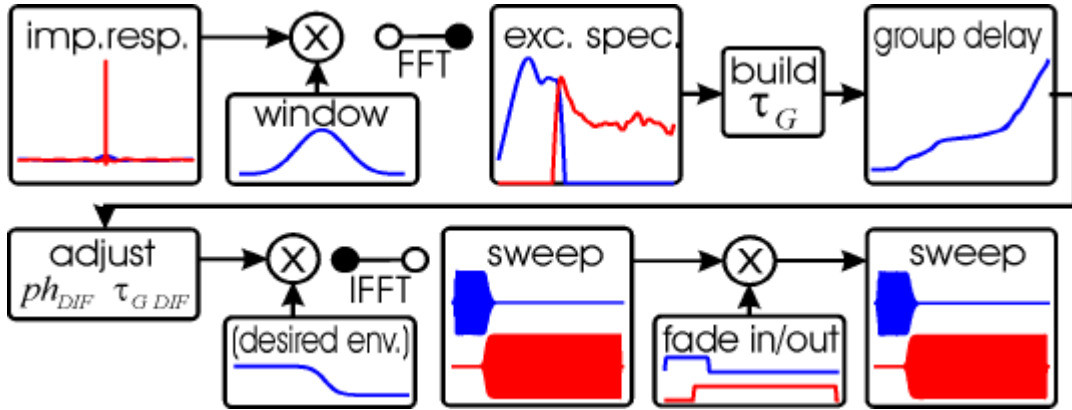


Fig. 21. Further processing of dual channel sweep: Windowing of IR, group delay synthesizing, inter-channel phase and delay adjust, fade in/out of sweep.

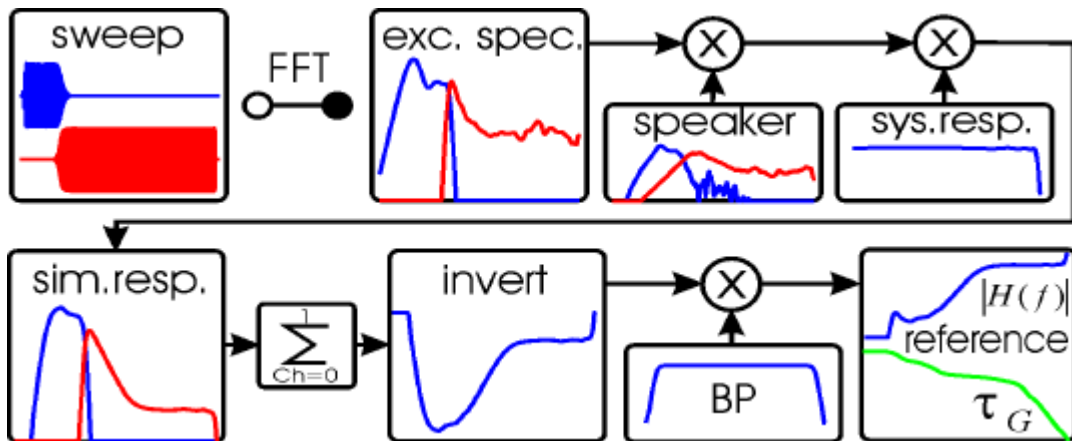


Fig. 22. Creation of reference for deconvolution of dual-channel sweep emitted by speaker.

Now the dual-channel sweep can be created by formulation of the already known relationship between the squared magnitude and group delay increase. But instead of

using just one channel, the squared magnitude of both channels must be summed here to yield the group delay growth value:

$$\tau_G(f) = \tau_G(f - df) + C \cdot \sum_{Ch=1}^2 |H(f)|^2 \quad (12)$$

Likewise, C is calculated using the sum of both channel's total energy:

$$C = \frac{\tau_G(f_{END}) - \tau_G(f_{START})}{\sum_{Ch=1}^2 \sum_{f=0}^{f_s/2} |H(f)|^2} \quad (13)$$

As the instantaneous frequency shall be the same for both channels, the group delay synthesis must be performed only once and the resulting phase can be copied into the second channel. If the excitation signal shall feature a frequency-controlled envelope, its dual-channel spectrum must be multiplied by the envelope spectrum after converting magnitude and synthesized group delay to the normal real/imaginary part representation.

The IFFT will turn out a dual-channel sweep that first glides through the low frequencies in one channel and then through the remaining frequencies in the other channels. The sum of both channels will have the desired envelope, while the relation of their amplitude at each instantaneous frequency corresponds to the relation established in the spectral domain.

In the crossover region that cannot be made indefinitely narrow due to the non-repetitive nature of the sweep, both channels interfere. While they are in phase in the synthesized dual-channel excitation signal, they usually would not arrive in phase at the microphone when emitted over the two loudspeakers. So a delay and phase correction are necessary to avoid drops in sound pressure level at the crossover frequency. That is why the phase and group-delay relationship of the speaker responses should be stored previously. This information can now be used to shift the signal for the loudspeaker with the smaller group delay to the right by the difference in arrival times. This can be accomplished best while still in the spectral domain by adding the appropriate group delay. The received phases at the crossover point can be brought into accordance by adding or subtracting a further small group delay to one of the channels.

It can be argued that dual-channel noise signals or sweeps with “two voices” covering independently both frequency ranges at the same time would be more advantageous, as they allow pumping energy to both speakers over the whole measurement period. However, only a “single-voiced” sweep permits excluding the distortion products from the recovered IR. Being able to do so, the stereo signal can be fed with a much higher level, which offsets the disadvantage of restricted emission time in the single ways.

Obviously, the reference spectrum necessary to deconvolve the received excitation signal cannot be created by performing the usual electrical reference measurement. By doing so, the carefully introduced loudspeaker equalization would disappear in the final results. The only viable way is to construct the reference spectrum by simulation, as depicted in Fig. 22. First, the previously generated dual-channel excitation signal is

transformed to the spectral domain. There, it is multiplied by the loudspeaker response and the self-response of the measurement system. The latter should include the response of the entire electrical signal path (converters, power amplifier, and microphone pre-amp) and the microphone response, should it not be sufficiently flat. After these operations, both channels are summed, yielding a simulation of the received sound pressure spectrum at the microphone position (colored by the response of the receiving path). This spectrum is now inverted to negate the group delay and to neutralize the chosen pre-emphasis, if any was applied. As the inverted spectrum would lead to excessive boost of the frequencies outside the selected transmission range, a multiplication with a band-pass of roughly the same corner frequencies as used in the pre-processing stages (top right of Fig. 20) is indispensable. To mute the out-of-band noise effectively, the order of this band-pass should be somewhat higher than the one used in the pre-processing. The application of this inevitable band-pass is a critical step, as it means that the acoustical measurement results are convoluted with its IR. The resulting effects can be quite disturbing. For example, using a linear-phase band-pass filter will obviously produce pre-ringing of the recovered RIRs. This is undesirable for auralization purposes. In this case, it is more advisable to use minimum-phase IIR-type filters. On the other hand, a linear phase band-pass filter might induce fewer errors in the classical room acoustical parameter evaluations. In any case, the filter order should be as moderate as possible to keep the filter's IR sufficiently narrow.

In general, these considerations apply to any broadband room acoustical measurement. At the last, the active pre-equalization technique presented here allows acquiring RIRs that are free of coloration by the measurement loudspeaker and feature a high, frequency-independent SNR.

5 DISTORTION MEASUREMENT

So far, it has been shown that the harmonic distortion artifacts can be removed entirely from acquired IRs when measuring with sweeps. In room acoustical measurements, they are usually simply discarded as the loudspeaker is not the object of investigation. In loudspeaker measurements, however, it is very interesting to relate them to the fundamental to evaluate the frequency-dependent distortion percentage. Indeed, this can be done separately for every single harmonic, as has already been proposed by Farina [2]. To illustrate the technique, Fig. 23 shows the group delay of a logarithmic sweep and its first four harmonics and Fig. 24 shows the time-frequency energy distribution from a signal with similar distortion.

Obviously, for a given instantaneous frequency of the fundamental, all its corresponding harmonics have the same group delay. In the example, the sweep fundamental reaches 400 Hz after 2 seconds. Consequently, the second-order harmonic trace intersects the horizontal second line at 800 Hz, the third at 1.2 kHz and so on. Now focusing on just one specific frequency in the diagram, the harmonic traces intersecting this vertical line have a lesser group delay than the sweep fundamental, since they belong to fundamentals with lower frequencies that have been swept through earlier. Multiplying these curves by the reference spectrum (that is, subtracting the group delay of the fundamental) leads to displacing them into the negative range while the fundamental will reside at $t=0s$, as desired.

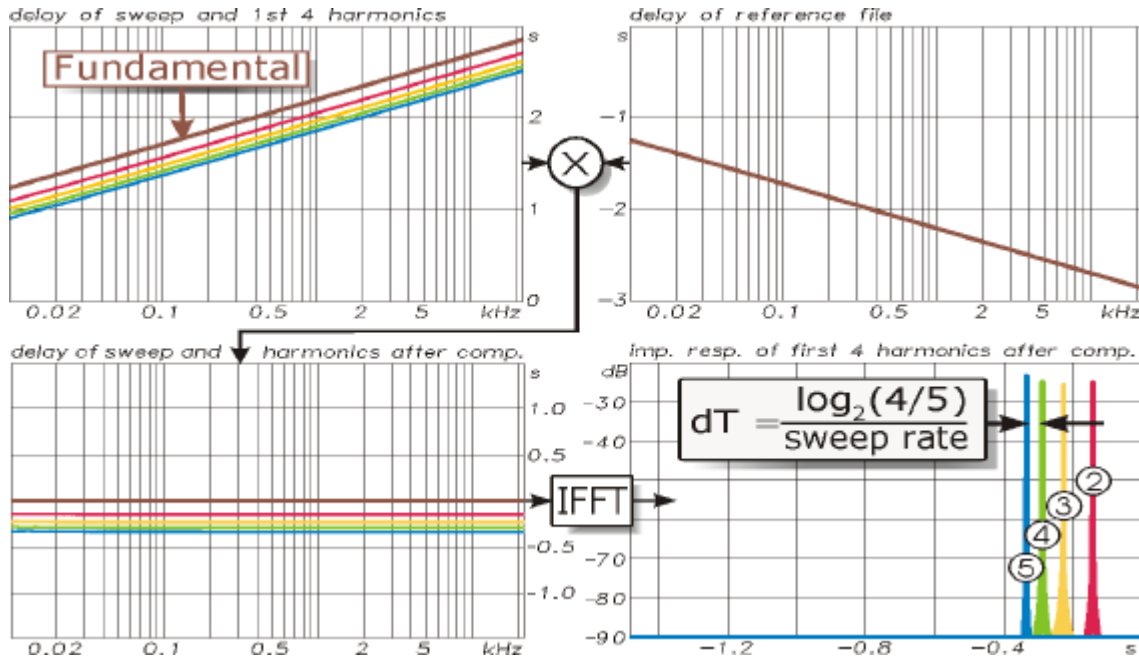


Fig. 23. Group delay of fundamental and first four harmonics (upper left), reference file (upper right), deconvolved sweep with harmonics (lower left) and IR positions of harmonics (lower right).

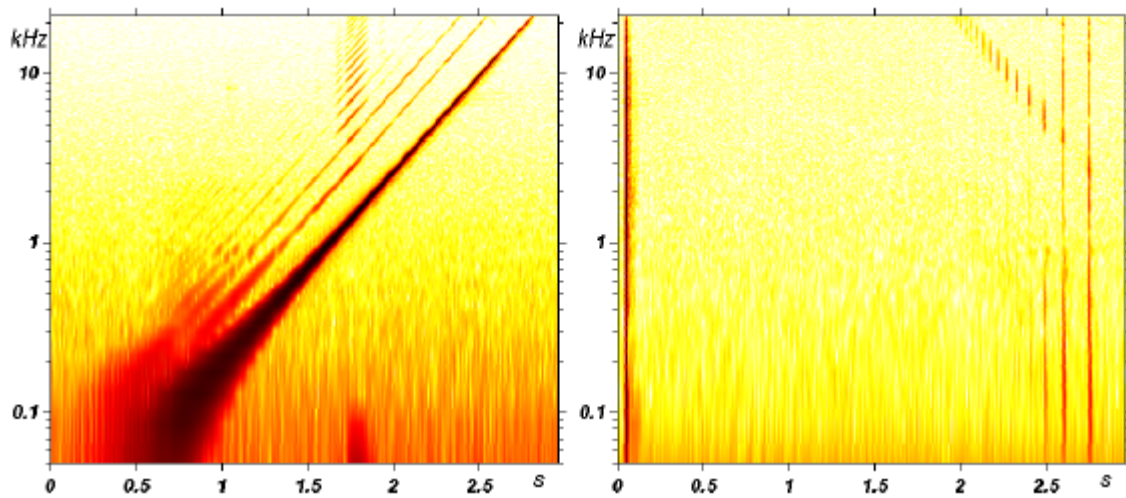


Fig.24: Time-Frequency diagrams of logarithmic sweep and harmonics (left) and IR (right).

Only in the case of a logarithmic sweep, the harmonics will all feature a frequency-independent constant group delay after the application of the reference spectrum. Other group delay courses of the sweep could be used, but would require the use of a separate reference spectrum for each harmonic to warp them to straight lines. Moreover, the distance between the components of the individual harmonics would not be frequency-independent. Thus, a logarithmic sweep is the preferable excitation signal, similar to the one that has already been used for such a long time in the venerable level recorder (1.1).

An IFFT of the deconvolved spectrum yields the actual IR at the left border and a couple of “harmonic impulse responses” (HIRs) at negative times near the right border,

with the second-order HIR situated rightmost and the upper order HIRs following from right to left. Their distance between each other can be calculated by:

$$Dist_{HIR\ AB} = \frac{\log_2(ord_A/ord_B)}{sweep\ rate\ [oct/s]} \quad (14)$$

To evaluate the frequency-dependent distortion fraction for every harmonic, the time signal is separated into the fundamental's IR and the single HIRs by windowing. Each of the isolated HIRs is submitted to a separate FFT. The FFT block length used therefore, can be much shorter than the one used for the initial deconvolution of the sweep response, thereby speeding up the whole process. To relate the frequency contents of one HIR spectrum to the fundamental, a spectral shift operation according to the order of the specific harmonic must be performed. For example, the spectral components of the fifth-order HIR are shifted horizontally to one-fifth of their original frequency. After this operation, the shifted spectrum can be divided by the fundamental spectrum, yielding the frequency-dependent distortion fraction (see Fig. 25).

It would seem that a correction of $-10 \log \{\text{order of the harmonic}\}$ must be applied to the results to compensate for the emphasis of higher frequencies imposed by the reference spectrum. Surprisingly, as exposed by many trials, this operation must be omitted to yield the correct results.

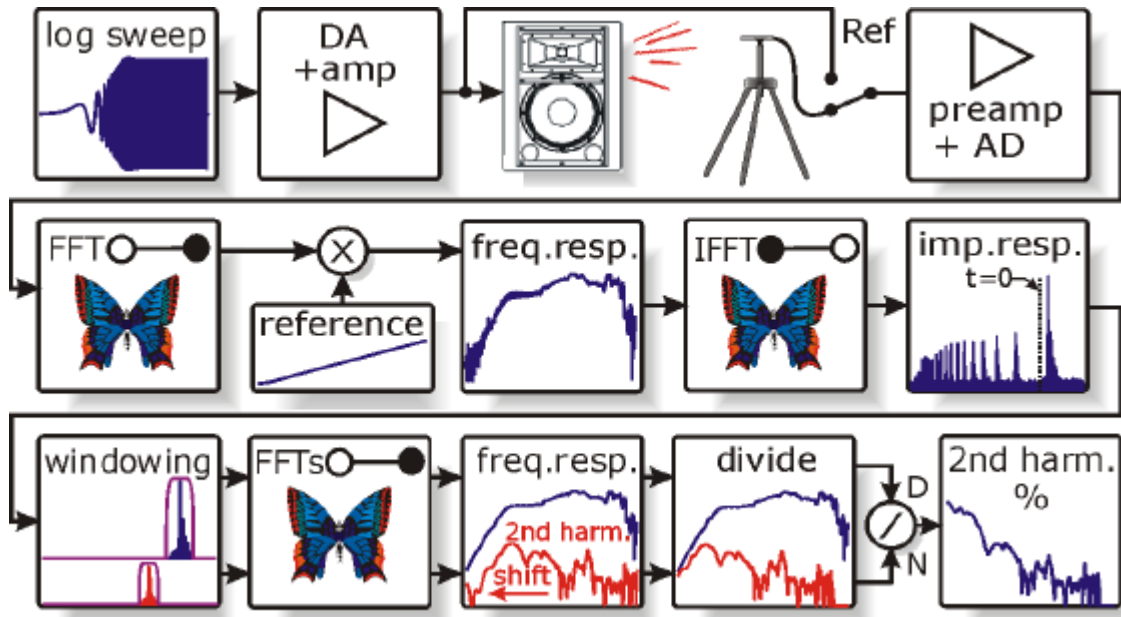


Fig.25: Signal processing stages for evaluation of transfer function and 2nd-order harmonic with logarithmic sweep.

Compared to the standard single-tone excitation and analysis with fixed frequency increments, this method is many times faster and usually establishes a much higher frequency resolution, at least in the mid and high-frequency ranges. However, some disadvantages should not be overlooked. First of all, the measurement is restricted to anechoic conditions, unless very long sweeps are used. Too much reverberation at the measurement site would lead to smearing of the distinct HIRs into each other by delayed components, thus thwarting their separation. This certainly is a significant shortcoming, as the strength of IR-based measurement is precisely the ability to reject

reverberation by windowing, provided that the time gap between direct sound and first reflection is sufficiently large.

Other problems are related to the mandatory use of windows to separate the individual HIRs from one other. Of course, all the usual problems associated with windowing [33] apply here. In particular, the choice of the window type constitutes a tradeoff between main-lobe width (equivalent to the spectral resolution) and side-lobe suppression. To avoid an energy loss and subsequent underestimate of distortion components that are not exactly situated under the window's top, a Tukey-style window [33] should be used. However, this kind of window features the rectangular window's poor side-lobe suppression of just 21 dB. This can lead to filling up with artifacts those frequency regions in which the distortion plummets to very low values. The spectral smoothing caused by any window is constant on a linear frequency scale. On the usual logarithmic display, this means that the spectral resolution becomes extremely high at high frequencies, while perhaps lacking details at the low end of the distortion spectrum. If a higher resolution is desired, the sweep must be made longer to space the HIRs further apart, thus permitting the use of wider windows.

The window width must decrease according to equation Equ. (14) to separate the higher order HIRs. This entails a reduced resolution of the corresponding distortion spectrum. However, when compressing it to the right to relate it with the fundamental spectrum, the resolution becomes higher than that of lower order HIRs. In practice, the desired resolution of the second-order harmonic at the low end of the loudspeaker's transmission range dictates the sweep rate and hence the excitation signal length.

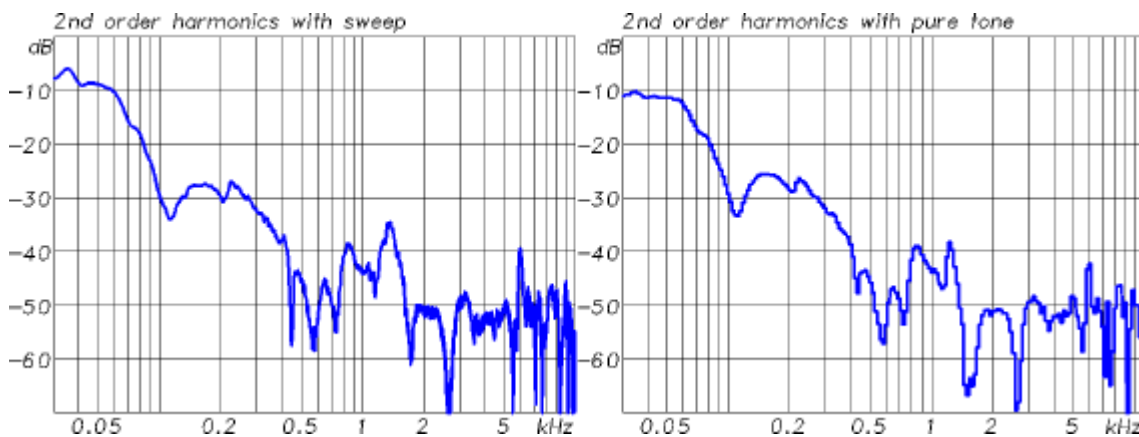


Fig.26: Comparison of 2nd-order harmonic acquired with sweep (left) and traditional pure tone testing in 1/24-octave increments (right).

Another problem of the fast distortion analysis is lacking SNR, especially for the higher-order harmonics which usually have fairly low levels. While pure-tone testing results in much energy being packed in the steady fundamental and its harmonics, the sweep technique distributes each harmonic's energy continuously over the whole frequency range. This is why the distortion spectra for the higher-order harmonics tend to become rather noisy, especially at higher frequencies. To alleviate this problem, it is a good practice to extend the sweep to an even longer length than necessary to achieve the desired spectral resolution. The windows can then be made narrower than necessary to isolate the single HIRs, thus rejecting the noise floor that resides between them.

Finally, even when all precautions have been taken to guarantee a high-precision measurement, it cannot be denied that sometimes, unexplainable differences between the steady tone testing and the sweep method occur in some frequency regions. Fig. 26 displays an example of such a discrepancy. Between 1.3 and 1.7 kHz, the second-order harmonic trace acquired with sweep and steady sine testing look quite different. The reasons for these occasional divergences are not obvious, although perhaps different voice coil temperatures have some effect.

In spite of these uncertainties, the sweep-based distortion testing is very attractive, as it is so much faster than the conventional pure-tone testing. In production testing, it does not only allow occasional spot checks, but it enables checking 100% of the manufacture, even if the merchandise is of inexpensive mass production.

6 CONCLUSIONS

FFT techniques using sweeps as excitation signals are the most advantageous choice for almost every transfer-function measurement situation. They allow feeding the device under test with high power at little more than 3dB crest factor and are relatively tolerant of time variance and totally immune against harmonic distortion. Choosing an adequate sweep length allows complete rejection of the harmonic distortion products. Moreover, these can be classified into individual frequency-dependent harmonics, allowing a complete and ultrafast distortion analysis over the entire frequency range, together with the evaluation of the transfer function.

When it comes to capturing RIRs for auralization purposes, there is no alternative to sweep measurements: The high dynamic range, in excess of 90 dB, desirable for this purpose is unattainable with MLS or noise measurements. Furthermore, even for standard RT measurements that do not require such a high dynamic range, sweeps are attractive because of the ease of increasing the dynamic range up to 15 dB compared to MLS-based measurements, using the same amplifier, loudspeaker, and measurement duration. Thus, there is little point in using MLS. Even MLS-related advantages of saving memory and processing time have almost completely lost their relevance with today's computer technology.

From a programmer's point of view, not having to include the MLS generation and Hadamard transform may save precious development time when writing a measurement program.

While acquiring transfer functions with MLS may be mathematically elegant, the authors are of the opinion that using sweeps to do so is elegant from a system theory and signal-purity point of view. Measuring with sweeps is also more natural. After all, bats do not emit MLSs to do their acoustic profiling.

7 ACKNOWLEDGEMENT

The work has been supported by the Brazilian National Council for Scientific and Technological Development (CNPq). It sponsored a six month's journey in the laboratory of acoustics of INMETRO of one of the authors (in 1999). While the scholarship initially was not exactly intended to develop measurement technology,

many of the ideas regarding sweep measurements have evolved and been tested in this period, thanks to an intense exchange of ideas by the two authors.

8 REFERENCES

- [1] David Griesinger, “Beyond MLS – Occupied Hall Measurement with FFT Techniques”, 101st AES convention, Los Angeles, November 1996, preprint 4403. Available on the web.
- [2] Angelo Farina, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-sine technique”, *J.AES*, vol. 48, p. 350, 108th AES Convention, Paris 2000, Preprint 5093. Available on the web.
- [3] Richard C. Heyser, “Acoustical Measurements by Time Delay Spectrometry”, *J.AES*, vol. 15, 1967, pp. 370-382
- [4] Richard C. Heyser, “Loudspeaker Phase Characteristics and Time Delay Distortion”, *J.AES*, vol. 17, 1969, pp. 30-41
- [5] Richard C. Heyser, “Determination of Loudspeaker Signal Arrival Times, Parts I, II & III”, *J.AES*, 1971, pp. 734-743, pp. 829-834, pp. 902, or *AES Loudspeakers Anthology*, vol. 1-25, pp. 225
- [6] AES / Richard C. Heyser, “Time Delay Spectrometry - An Anthology of the Works of Richard C. Heyser”, *AES*, New York 1988
- [7] John Vanderkooy, “Another Approach to Time-Delay Spectrometry”, *J.AES*, vol. 34, July 1986, pp. 523-538
- [8] Richard Greiner, Jamsheed Wania, Gerardo Noejovich, “A Digital Approach to Time Delay Spectrometry”, *J.AES*, vol. 37, July 1989, pp. 593-602
- [9] Peter d’Antonio, John Konnert, “Complex Time Response Measurements Using Time-Delay Spectrometry (Dedicated to the late Richard C. Heyser)”, *J.AES*, vol. Vol. 37, September 1989, pp. 674-690
- [10] Henrik Biering, Ole Z. Pederson, “Comments on ‘Another Approach to Time-Delay Spectrometry’ and author’s reply”, *J.AES*, vol. 35, March 1987, pp. 145-146
- [11] Henrik V. Sorensen, Douglas L. Jones, Michael T. Heideman, Sidney Burrus, “Real-valued Fast Fourier Transform Algorithms”, *IEEE Trans. Acoustics, Speech, Signal Proc.*, June 1987, p. 849
- [12] Johan Shoukens, Rik Pintelon, “Measurement of Frequency Response Functions in Noise Environments”, *IEEE Trans. Instrumentation and Measurement*, December 1990
- [13] Douglas D. Rife & John Vanderkooy, “Transfer Function Measurement with Maximum-Length Sequences”, *J.AES*, vol. 37, June 1989, pp. 419-444
- [14] Johan Shoukens, Rik Pintelon, Yves Rolain, “Broadband versus Stepped Sine FRF Measurements”, *IEEE Trans. Instrumentation and Measurement*, April 2000, No.2

- [15] J.M. Berman, L.R. Fincham
The Application of Digital Techniques to the Measurement of Loudspeakers
J.AES, vol. 25, 1977, pp. 370-384, or *AES Loudspeakers Anthology*, vol. 1-25, p. 436
- [16] Chris Dunn, Malcom Omar Hawksford, "Distortion Immunity of MLS-Derived Response Measurements", *J.AES* May 1993, pp. 314. Available on the web.
- [17] L.R. Fincham, "Refinements in the Impulse Testing of Loudspeakers", *J.AES*, vol. 41, March 1985, pp. 133-140
- [18] F.J. MacWilliams, N.J.A. Sloane, "Pseudo-Random Sequences and Arrays", *Proc. IEEE*, 1976, vol. 84, p. 1715
- [19] Thomas W. Parks, James H. McClellan, Lawrence R. Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters", *IEEE Trans. Audio Electroacoustics*, 1973, p. 506
- [20] E.D. Nelson, M.L. Fredman, "Hadamard Spectroscopy", *J. Opt. Soc. Am.*, 1970, pp. 1664
- [21] H. Alrutz & Manfred R. Schroeder, "A Fast Hadamard Transform Method for the Evaluation of Measurements using Pseudorandom Test Signals", *Proc. 11th ICA*, Paris 1983, pp. 235
- [22] Jeffrey Borish & J. Angell, "An Efficient Algorithm for Measuring the Impulse Response Using Pseudo-Random Noise", *J.AES*, vol. 33, 1983, pp. 478-488
- [23] Jeffrey Borish, "Self-Contained Crosscorrelation Program for Maximum Length Sequences", *J.AES*, vol. 33, 1985, pp. 888-891
- [24] Manfred R. Schroeder, "Integrated -Impulse Method for Measuring Sound Decay without using Impulses", *J.ASA*, 1979, p. 497
- [25] Eckard Mommertz, Swen Müller, "Measuring Impulse Responses with Preemphasized Pseudo Random Noise derived from Maximum Length Sequences.", *Applied Acoustics*, 1995, vol. 44, p. 195
- [26] Jeffrey Borish, "An Efficient Algorithm for Generating Colored Noise Using a Pseudorandom Sequence", *J.AES*, vol. 33, March 1985, pp. 141-144
- [27] Michael Vorländer, Heinrich Bietz, „Der Einfluss von Zeitvarianzen bei Maximalfolgenmessungen“, *DAGA*, 1995, p. 675
- [28] Michael Vorländer, Malte Kob, "Practical Aspects of MLS Measurements in Building Acoustics", *Applied Acoustics*, vol. 52, p. 239
- [29] Peter Svensson, Johan L. Nielsen, "Errors in MLS Measurements Caused by Time Variance in Acoustic Systems", *J.AES*, vol. 47, November 1999, pp. 907
- [30] C. P. Jane and A. J. M. Kaizer, "Time-Frequency Distributions of Loudspeakers: The Application of the Wigner Distribution", *J.AES*, vol. 31, April 1983, pp. 198-223

[31] D. B. Keele Jr., „Low-Frequency Loudspeaker Assessment by Nearfield Sound-Pressure-Measurements”, *J.AES*, vol. 22, 1974, pp. 154-162, or *AES Loudspeakers Anthology*, vol. 1-25, p. 344

[32] John Vanderkooy, Stanley Lipshitz, Robert Wannamaker, “Minimally Audible Noise Shaping”, *J.AES*, vol. 39, 1991, pp. 836-852

[33] Fredric J.Harris, “On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform”, *Proc. IEEE*, January 1978, p. 51

[34] John S. Bradly, “Optimizing the Decay Range in Room Acoustics Measurements using Maximum-Length-Sequence Techniques”, *J.AES*, vol. 44, pp. 266-273, April 1996.

[35] Nobuharo Aoshima, “Computer-generated pulse signal applied for sound measurement”, *J.ASA*, May 1981, p. 1484

[36] Yôiti Suzuki, Futoshi Asano, Hack-Yoon Kim, Toshio Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses”, *J.ASA*, February 1995, p. 1119

[37] Masanori Morise, Toshio Irino, Hideki Banno, Hideki Kawahara, “A Test Signal Robust to Background Noise in the Measurement of Acoustic Impulse Responses: Warped TSP” , 34th Internoise, August 2005, Rio de Janeiro.

[38] John C. Burgess, “Chirp Design for Acoustical System Identification”, *J.ASA*, 1992, p.1525-1530

[39] Eckard Mommertz, Swen Müller, “Applying the Inverse Fast Hadamard Transform to Improve MLS Measurements”, Proceedings ICA 1995.

[40] H. Herlufsen, “Dual Channel FFT Analysis (Part I, II)”, Brüel & Kjær Technical Review No. 1-1984. Available on the web.

[41] Eric Benjamin, "Extending Quasi-Anechoic Electroacoustic Measurements to Low Frequencies," 117th AES Convention, San Francisco, November 2004, preprint 6218.

[42] Scott Norcross, John Vanderkooy, „A Survey of the Effects of Nonlinearity on Various Types of Transfer-Function Measurements“, 99th AES Convention, preprint 4137.

9 BIBLIOGRAPHY

[50] A.J. Berkhout, D. de Vries, M. M. Boone, “A New Method to Acquire Impulse Responses in Concert Halls”, *J.ASA*, 1980, p. 179

[51] A.J. Berkhout, M.M. Boone, C. Kesselman, “Acoustic Impulse Response Measurement: A New Technique”, *J.AES*, vol. 32, October 1984, pp. 740-746

[52] John D. Bunton, Richard H. Small, “Cumulative Spectra, Tone Bursts, and Apodization”, *J.AES*, vol. 30, June 1982, pp. 386-395, or *AES Loudspeakers Anthology*, vol. 26-31, p. 322

- [53] Richard C. Cabot, "Audio Measurements", *J.AES*, vol. 35, June 1987, pp. 477-500
- [54] Angelo Farina, "Non-Linear Convolution: A New Approach for the Auralization of Distorting Systems", presented at the *110th Convention of the AES*, Amsterdam, 2001 May.
- [55] Panagiotis D. Hatziantoniou, John N. Mourjopoulos, „Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses“, *JAES*, vol. 48, pp. 259-280
- [56] Paul S. Kovitz, "On the Repeatability of TDS Energy-Time Curve and MLS Impulse Response Measurements", Sabine Centennial Symposium, *127th ASA Convention*, June 1994, p. 129
- [57] Stanley P. Lipshitz, Tony C. Scott, John Vanderkooy, "Increasing the Audio Measurement Capability of FFT Analyzers by Microcomputer Postprocessing", *J.AES*, vol. 33, 1985, pp. 626-648
- [58] Anders Lundeby, Tor Erik Vigran, Heinrich Bietz, Michael Vorländer, "Uncertainties of Measurements in Room Acoustics", *Acustica*, 1995, p. 344
- [59] Eckard Mommertz, „Mobiles Meßsystem zur Aufnahme von Raumimpulsantworten“, *DAGA*, 1990, p. 843
- [60] Johan L. Nielsen, "Improvement of Signal-to-Noise Ratio in Long-Term Measurements with High-Level Nonstationary Disturbances", *J.AES*, 1997, vol. 45, pp. 1063-1066
- [61] E.P. Palmer, R.D. Price, S.J. Burton, "Impulse-response and transfer-function measurements in rooms by m-sequence cross correlation", *J.ASA*, 1986, Vol. 80, p. 56
- [62] M. Poletti – "Linearly swept frequency measurements, time-delay spectrometry, and the Wigner distribution" – *J.AES*, vol. 36, 1988 June, pp. 457-468
- [63] Douglas Preis, Voula Chris Georgopoulos, "Wigner Distribution Representation and Analysis of Audio Signals: An Illustrated Tutorial Review", *J.AES*, December 1999, pp. 1043-1053
- [64] Allan Rosenheck, Kurt Heutschi, "Tone Bursts for the Objective and Subjective Evaluation of Loudspeaker Frequency Response in Ordinary Rooms", *J. AES*, vol. 47, April 1999, pp. 252-255
- [65] Johan Schoukens, Rik Pintelon, Edwin van der Ouderna, Jean Renneboog, "Survey of Excitation Signals for FFT based Signal Analyzers", *IEEE Trans. Instrumentation and Measurement*, September 1988, p. 342
- [66] Johan Shoukens, Yves Rolain, Rik Pintelon, "Improved Frequency Response Measurements for Random Noise Excitations", *IEEE Trans. Instrumentation and Measurement*, February 1998, No.1, p. 332
- [67] Manfred R. Schroeder, "Number Theory in Science and Communication *Springer Verlag Berlin Heidelberg New York Tokyo*, 1986

- [68] Manfred R. Schroeder, "Synthesis of Low-Peak-Factor Signals and Binary Sequences with Low Autocorrelation", *IEEE Trans. Info. Theory*, 1970, p. 85
- [69] Christopher J. Struck, Steve F. Temme, "Simulated Free Field Measurements", *J.AES*, vol. 42, 1994, pp. 467-482
- [70] Michael Vorländer, "Application of Maximum Length Sequences in Acoustics", *17^o Encontro da Sociedade Brasileira de Acústica (SOBRAC)*, December 1996, p. 35
- [71] Xiang-Gen Xia, "System Identification Using Chirp Signals and Time-Variant Filters in the Joint Time-Frequency Domain", *IEEE Trans. Signal Proc.*, August 1997, p. 2072
- [72] Edgar Villchur, "A Method for Testing Loudspeakers with Random Noise Input", *AES Loudspeakers Anthology*, vol. 1-25, p. 96
- [73] John Vanderkooy, "Aspects of MLS measuring systems", *JAES*, vol. 42, April 1994, p. 219
- [74] Douglas D. Rife, "Comments on "Distortion Immunity of MLS D. Rifed Impulse Response Measurements" and author's reply", *J.AES*, June 1994, p. 490
- [75] Won-Jin Kim and Youn-Sik Park, "Non-Linearity Identification and Quantification Using an Inverse Fourier Transform", *Mechanical Systems and Signal Processing* 1993 7(3) pp. 239-255
- [76] G.R. Tomlinson, "Developments in the Use of the Hilbert Transform for Detecting and Quantifying Non-Linearity Associated with Frequency Response Functions,, *Mechanical Systems and Signal Processing* 1987 1(2), pp. 151-171
- [77] Guy-Bart Stan, Jean Jaques Embrechts and Dominique Archambeau, "Comparison of Different Impulse Response Measurement Techniques", *J.AES*, April 2002, pp. 249-262
- [78] Gottfried K. Behler, Swen Müller, "Technique for the Derivation of Wide Band Room Impulse Responses", EAA Symposium on architectural acoustics, Madrid, October 2000. Available on the web.
- [79] Timo Peltonen, "Response Measurements: Overview and Practical Aspects". Available on the web.
- [80] Juha Merimaa, Timo Peltonen and Tapio Lokki , "Concert Hall Impulse Responses Pori, Finland: Reference", May 6, 2005. Available on the web.
- [81] Patrizio Fausti and Angelo Farina, "Acoustic Measurements in Opera Houses: Comparison between Different Techniques and Equipment", *Journal of Sound and Vibration*, 2000, p.213-229. Available on the web.
- [82] Stefan Tenbohlen, Simon A Ryder, "Making Frequency Response Analysis Measurements: A Comparison of the Swept Frequency and Low Voltage Impulse Methods", 18th International symposium on high voltage engineering, Netherlands, 2003. Available on the web.

[83] Michael Vorländer, Eckard Mommertz, “Guidelines for the Application of the MLS technique on Building Acoustics and in Outdoor Measurements”, Proceedings Internoise 1997.

[84] Swen Müller, Paulo Massarani, “Medições da Resposta em Frequência de Sistemas de Sonorização, 1. encontro do SemEA, UFMG, Belo Horizonte, June 2002. Available on the web.

[85] Swen Müller, Paulo Massarani, “Criação de Varreduras com Ênfase Arbitrária”, 1. encontro do SemEA, UFMG, Belo Horizonte, June 2002. Available on the web.

[86] Fumiaki Satoh, Jin Hirano, Shinichi Sakamoto, Hideki Tachibana, “Sound Propagation Measurement Using Swept-sine Signal”, 34th Internoise, Rio de Janeiro, August 2005.

[87] Igor Nikolic, Ole-Herman Bjor, Svein Arne Nordby, “Swept-Sine Method Improves Signal-to-Noise Ratio in Building Acoustics Applications”, 34th Internoise, Rio de Janeiro, August 2005.

[88] Bruno Sanches Masiero, Fernando Iazzetta, “Estudo e Implementação de Métodos de Medição de Resposta Impulsiva”, 1. Seminário de Música, Ciência e Tecnologia da Acústica Musical, USP, São Paulo, November 2004. Available on the web.