ON ROOM CORRECTION AND EQUALIZATION OF SOUND SYSTEMS

MATHIAS JOHANSSON, DIRAC RESEARCH AB

In this note I discuss some issues in filter design for equalization of sound systems. The emphasis is on rationale, not on experiments, and I will focus on a few common misunderstandings. I will briefly describe the basic concepts used in sound equalization, such as FIR and IIR filters, minimum and linear phase and present basic mathematical facts as well as give a background to the philosophy behind the Dirac Live approach. To limit the length of the text, I assume some basic familiarity with the topics covered. I will however refrain from the popular trend in some engineering periodicals of hiding bad ideas behind complex-looking equations. As already mentioned, the emphasis is on the logic of different approaches to equalization rather than experiments. Logic can be checked by the reader, whereas experiments carried out by others always leave room for doubt concerning experiment conditions. Furthermore, as probability theory teaches us, one lucky experiment shows nothing about the underlying rationale, but the underlying rationale will be much more indicative of future experimental outcomes.

Note: Parts of this material were presented by the author at the 123rd AES convention in NYC on the workshop "FIR or IIR? That is the question!".

FIR AND IIR FILTERS

The output of an FIR filter is simply a weighted average of the N most recent input samples. The filter order, or length, thus indicates the memory of the filter. An IIR filter does the same thing as an FIR filter, but it adds a weighted average of old output samples to the first average. Of course, the weights can be set arbitrarily, but if the weights corresponding to the old outputs are too large it is easily seen that this may lead to the output increasing exponentially; the filter is then said to be unstable. In economics, the more vivid term 'explosive' is often used. An IIR filter can become unstable due to limited precision in fixed-point processors. There are different ways of implementing a high-order IIR filter to reduce the risk of such problems. Most commonly, a high-order IIR filter is rewritten as a cascade of several second-order filters, called biquads. By this means, the effect of numerical errors can be alleviated. FIR filters cannot become unstable as they are non-recursive by their nature, and are thus easier to implement.

Due to the recursion inherent in an IIR filter, the impulse response of such a filter becomes infinite, hence the term Infinite Impulse Response filter. In contrast, an FIR filter (of order less than infinity) has a finite impulse response. It seems then that the IIR filter is more flexible. This is true in certain senses, but there are limitations of IIR filters. First, it is actually impossible to construct an IIR filter with linear phase (see below for more on the concept of phase responses); the reason is that a linear phase system must have a symmetrical impulse response, i.e. the left and the right tails (with the main pulse as the reference point) must be exact mirror versions of each other. (To be exact, it may also be anti-symmetrical, i.e. a negative mirror version.) This is obviously precluded with a causal IIR filter which has an infinite tail to the right (it remembers all past inputs) but is cut off to the left (it does not see into the future). In practice however, if we want a linear phase filter we can still get a very good approximation even with an IIR filter.

The next problem is of a different character. When designing an optimal IIR filter in the sense of minimizing the expected squared error between the filter response and a desired response, the

equations turn ugly. It is a non-linear problem, and it becomes mathematically messy to find the right response. However, just because it is difficult doesn't mean it is impossible, so this is more of the lazy mathematician's excuse for staying away from IIR filters. (But this excuse may still be relevant when designing a filter in a processor of limited capacity.)

So, why is it that most simple audio equalizers still only use IIR filters? One reason is that the filters are typically set manually, and there are closed-form formulas for constructing several simple filter types such as peak/notch filter, lowpass/highpass, shelving, etc. You can easily adjust the Q in real-time and listen to the difference. IIR filters have another strength: less coefficients are needed for a given slope of e.g. a lowpass filter as compared to an FIR. The reason is the recursion. It is like getting several filters working in succession for the price of one.

FIR filters are easier to implement (and can run on fixed-point architectures), are more flexible to work with mathematically, can have arbitrary phase responses, but require more coefficients for a given steepness of the filter. This last point is important when it comes to resolution in the low-frequency area. Intuitively, the length of the FIR filter has a relation to the impact it can have at a certain wavelength of the signal. With a sufficiently long FIR filter however, much speaks in favor of using them instead of the numerically more cumbersome IIR cousin.

In summary, within a very good approximation we can do the same thing with IIR filters as with FIR filters. The choice of either structure is more of an implementation issue and also to some extent one of application. It turns out that in order to do a proper impulse response correction of a set of acoustic transfer functions, a rather long filter is needed and little seems to be gained by using IIR filters. On the other hand, if our processor budget only allows a low-order filter (maybe 10 biquads or 20-40 FIR taps) then impulse response correction is not possible, and then it is wiser to focus on the magnitude response and use minimum-phase biquad filters.

Finally, it should be noted that impulse response correction requires a non-causal part, which cannot be modeled well by an IIR model. An IIR filter can however model the causal part of a well-design equalizer filter, which means that one of the most efficient mixed-phase equalizer implementations consists of a combination of IIR and FIR filters.

ON THE CONCEPT OF PHASE

The phase response of a filter shows how large a fraction of a period an input in the form of a pure sinusoid is shifted at the output. It is thus a representation of frequency-variable delays.

A DIGRESSION ON GROUP DELAY

The derivative of the phase response with respect to frequency is called the group delay. It is sometimes mistaken for the delay at a given frequency. In fact, the group delay only measures the delay of the *envelope* (i.e. the shape) of a narrowband signal centered at that frequency. It is nonsense to interpret this as a physical delay. To get the exact delay at a certain frequency we only rescale the phase response to turn degrees into seconds (or meters if we prefer it that way).

Example: A phase shift of 36 degrees means a delay of 1/10 period. If the frequency in question is, say, 1000 Hz, a period is 1 msec, and a phase shift of 1/10 the period means 0.1 msec.

BACK TO PHASE

A system with a linear phase response has a phase response that is a straight line when drawn as a function of frequency (on a linear frequency scale). This also means that the system introduces

a constant delay at all frequencies. However, as mentioned above, a linear-phase filter always has a symmetrical (or anti-symmetrical) impulse response. This means that if such a filter has a non-flat magnitude response it will also introduce pre- and post-ringing in the time domain. The danger of this is sometimes exaggerated among workers in the field. In practice, if the magnitude response of a linear-phase filter is smooth (which an equalization filter should be) the pre-ringing will be negligible.

At this time we must also discuss the concept of a minimum phase system. This is a very peculiar system that has the least energy delay of all systems with a given magnitude response. (In the complex plane a minimum-phase system has all zeros and poles inside the unit circle.) It is very important to emphasize that this does not in any way mean that the impulse response is the *shortest* of all systems with that magnitude response. It means that it is *earliest*; the difference is profound. That it is early is a useful feature in applications where latency needs to be kept to a minimum. For sound quality, there is no evidence that a minimum-phase equalizer sounds better than say a linear-phase one. In fact, a linear-phase filter typically has a shorter impulse response than a minimum-phase one. The term minimum-phase is misleading and much would be gained if the audio community would instead adopt the term minimum energy delay. The exact definition is as follows: Consider a minimum-phase impulse response and an arbitrary other linear system with identical magnitude response. Compare the accumulated energy at the outputs of the two systems at any time t. The accumulated energy from time zero until t will always be larger for the minimum phase system than the other system. This implies that the response must start early, for otherwise we can always construct another system that at time t close to zero has transferred more energy than the minimum-phase filter. Note that a minimum-phase system does not imply zero latency, only lower latency than other filters.

A well-known theorem from complex analysis states that any linear time-invariant system can be factored into two systems: a minimum-phase system and an all-pass system. If we design an equalizer that perfectly inverts the minimum-phase system we have thus made the magnitude response flat (this might not be our ultimate goal, but let's assume so for a moment. It does not make any difference in this context.), but what has happened to the impulse response? On occasion, I have heard competent people say that we have at least not made it worse, since we took care of the minimum-phase part (the inverse of a minimum-phase system is itself a minimum-phase system and since there is only one for any magnitude response we have automatically taken care of it). The fallacy of this statement is obvious: the all-pass system may have a much worse impulse response than the complete system. If you take away the minimum-phase factor you may have taken away the features that removed severe ringing in the all-pass factor. The factorization we chose is completely arbitrary and has no physical interpretation at all. We could just as well have chosen a different factorization.

What we need to do is look at the total system response. If we factor it, we'd better make sure to use a factorization that is meaningful also from a physical or psycho-acoustical standpoint and evaluate the final result in terms of the full response. A commonly employed model that has a well-founded physical motivation is to model transfer functions at different listening positions within a room as having a set of common poles and a position-variable transfer function consisting only of zeros. So far so good. The next step is to say: ok, we want to make a robust correction and therefore we will only invert the poles. Here the error. Even though they are indeed common to all positions, it makes no sense to just look at them in isolation from the position-dependent zeros! If you remove poles in the low-frequency region close to the unit circle, it does not mean that you have improved the over-all response. First, the poles affect not just the region of the unit circle which they happened to be closest to. They also help to keep *down* the gain at frequencies far away from that region! Similarly, the position dependent zeros

close to the high-frequency region of the unit circle will now *boost* the energy in the low-frequency region.

The lesson from both these examples is that even though a certain factorization makes sense physically or for ease of mathematical consideration, it is dangerous folly to look at each factor as an isolated system and believe that if you improve that factor the whole system will fare better. However, as elementary logic teaches us: a false premise implies any proposition, false or true. So you can produce examples which will sound great on occasion even with a faulty rationale. But that does not mean that with a better rationale you will not be able to do even better.

THE FOURIER TRANSFORM AND THE CONCEPT OF FREQUENCY

As Dennis Gabor noted in 1945, the Fourier representation of a signal shows significant departures from the human sensation of frequencies. The Fourier transform integrates a time series over an infinite time window. This has some interesting consequences. For instance, a changing frequency is an oxymoron, something that is per definition impossible as time has been literally swept out of the equations. So a siren will have a very funny-looking Fourier spectrum, which does not easily betray its "true nature". If we instead analyze the time series in terms of Gabor frames, or short-time Fourier transforms, or some other time-frequency basis, we will easily distinguish a single frequency at a time, continuously changing pitch. This time-frequency representation is also a better description of the sensation of a human listener. For a human, frequency is time-dependent. Here a distinct difference between the Fourier representation and the perceptual impression.

This has some important implications for sound equalization. When we read our magnitude response estimates, we use a very simple estimate of the perceived spectrum, completely suppressing the concept of time. For example, take a minimum-phase impulse response and reverse it. The former starts at time zero and decays with some time constant until it dies out; the latter has instead a substantial pre-ringing but no post-ringing. Now all research that has been carried out on human perception of transients say that pre-echoes and post-echoes sound completely different to us. Yet, the magnitude responses are identical. Of course, this is also manifested if we say something and then play it back in reverse. Both samples have the same Fourier magnitude responses. This illustrates that phase responses, or impulse responses, do affect the perception of sound, even for a monaural source. There is obviously some threshold of how phase sensitive we are, but the literature on this matter (which is extensive, starting from the 1930's) has concluded that this threshold, or integration time-constant, is adaptive and varies with what we listen to. What we can say for sure is that we do hear absolute phase but that the higher the frequencies the less sensitive we are (obviously, as the wavelengths become very short and the physics of wave propagation dictates that very little relevant information can be transmitted acoustically at high frequencies due to the chaotic behavior of high-frequency acoustic transfer functions). This implies that a good equalizer needs to consider also phase, not just magnitude.

An example will show this point more clearly. Consider a loudspeaker standing in a room. Mr A measures impulse responses in a certain listening volume and finds to his dismay that the magnitude response has a substantial broad dip at some rather low frequency, say 300 Hz, in all positions. He calibrates a peak filter and fills up the hole in the magnitude response, which is then confirmed by measurements. Enter Mr B. Mr B is a musician and he listens to the equalized system. "It sounds horrible! What have you done to the system!? It sounds all swollen and strange!" Mr A becomes nervous, as Mr B is an important customer, and calls his trusted friend Mr C. Mr C answers: "Ah, yes of course. The dip was really due to reflections. You should never

boost any dip, because they are typically due to reflections." So Mr A removes his equalizer filter and lets Mr B listen again. Mr B, however, is still not happy. "It is better, but it's not good. There is something hollow about the sound." At this time Mrs D enters the conversation. She's been listening, sitting quietly in a corner of the room, and says: "Mr A was wrong because he forgot about the time domain. Looking only at the magnitude of the Fourier transform and interpreting it as strongly related to our concept of frequencies, he thought that he could boost that region and obtain better sound. The problem is that he uses minimum-phase filters and consequently adds energy at that frequency early in time. But if we only look at the direct wave there is no hole to be filled in the frequency response. The hole never exists if we look at a short window at any time." Mr B frowns: "So Mr C was right to say that we cannot do anything about it. But if that's the case, why do I still hear a strange sounding oboe on my recording?" Mrs D looks sternly at him: "Mr C was wrong too. The problem is due to the time domain properties; the reflection causes the problem and it can only be corrected for by a time-domain approach. If we design a filter that reduces the reflection, you will end up with the interesting result that the hole will be gone and the oboe will sound more natural." "But," Mrs D adds, "don't take this example as evidence that you can always correct dips this way! In this case it was possible, because all positions experienced the same problem."
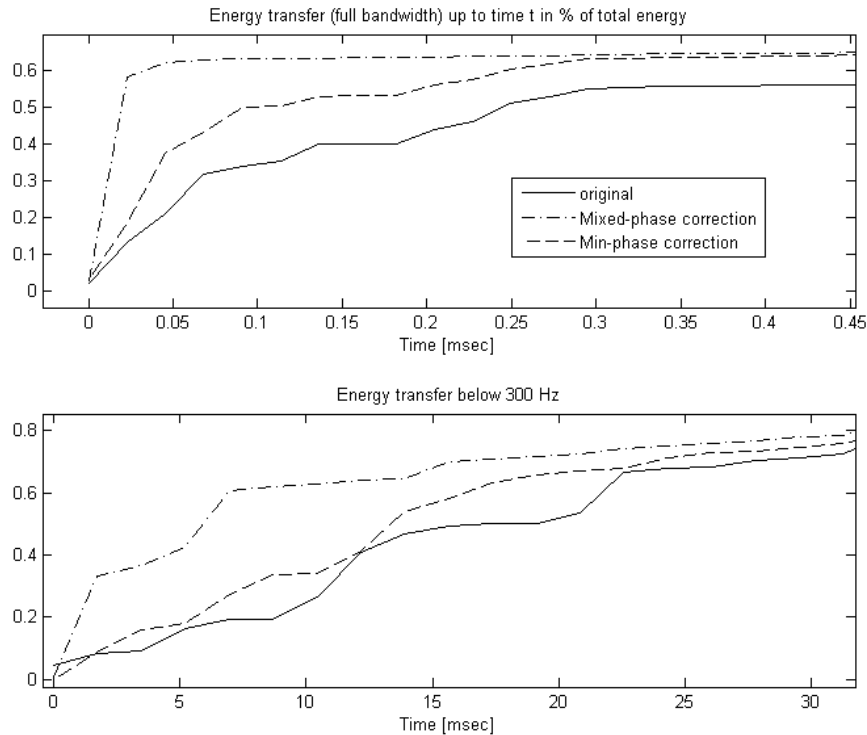
The fallacy to uncritically view the magnitude of Fourier transforms as an accurate description of what we perceive (and thereby ignore time) is evident in debates on what a "good" acoustical magnitude response should be like. The answer depends on the time-domain character of the room that is being measured. A magnitude response that sounds good in a large hall may sound terrible inside a car, as the reverberation ratios are completely different.

The basic problem with boosting dips is that it is typically done using a minimum-phase equalizer. Thus, it injects energy at the wrong time. The same is obviously also true for a linear phase filter or any other filter that does not consider what the impulse response of the total system becomes. The lesson is: Don't mix up Fourier transforms with perceived frequency responses. Perceived frequency responses are time-dependent. A joint time and frequency analysis is required in order to design a good equalizer. In addition to this, the problem of spatial variations must be considered carefully. A zero at one measurement point may be just inside the unit circle but when moving the microphone it may end up on the outside. In such a scenario it would be disastrous to look at the mean behavior (a zero on the unit circle) and optimize that response. A correction which is good for the mean response is very different from a correction which on average is good for any one of the measured responses. Zeros that are located outside the unit circle but at approximately the same spot regardless of microphone position are robustly invertible, others are not. (And again, this does not mean that we consider one zero isolated from the rest of the response.) The following zoom-in on a zero plot shows these properties. Measurements were taken on different positions in a good listening room using good (Genelec) speakers. The zeros have been drawn using different sizes for different measurement positions, in order to see how they move with respect to spatial movements.
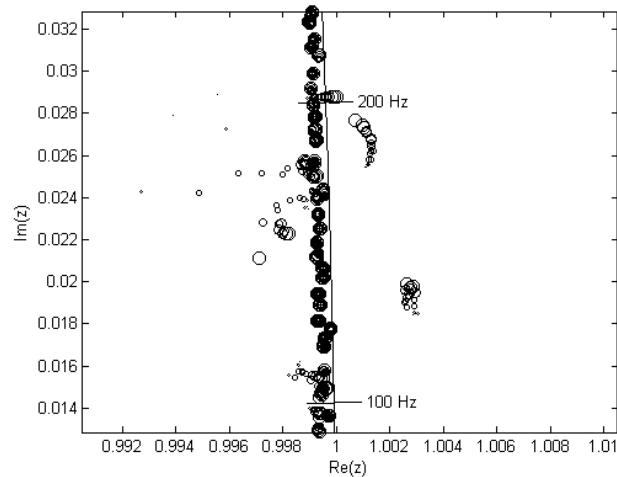
Just above 200 Hz we see a zero moving from minimum-phase to mixed-phase behavior. The average lies just inside the unit circle so a careless minimum-phase inversion would be able to cause a lot of serious time-domain ringing here. Between 150 and 200 Hz we see a zero outside the unit circle that moves a lot with position. As it is close to the unit circle and moves, a robust correction would not try to alter this behavior. The zero outside the unit circle between 100 and 150 Hz is however possible to get rid of without causing any problem at any position (the distance to the unit circle is rather large and the variation with position is small).

Two different equalizers were designed based on these measurements. The following figure shows how much of the total accumulated energy has been transferred at any time for the impulse response of the original system, the system compensated with a minimum-phase correction, and the system compensated with a mixed-phase correction. The plot was generated taking new measurements at different listening positions than the original measurements used for designing the filters. It should be emphasized that the minimum-phase correction uses basically the same robust methodology as the mixed-phase inverse except that it does not consider the zeros outside the unit circle. The upper plot shows the full bandwidth, the lower one the responses up to 300 Hz. These are basically energy step responses, a better system will thus have a faster rise time and ideally touch 1 immediately. It is clear that even in a good listening room with good speakers a substantial improvement is possible using a careful mixed-phase design. The minimum-phase filter is clearly doing a good job as well; not nearly as good as the mixed-phase design it nevertheless improves the time-domain behavior. In a large well-designed listening room the impulse response would preferably be nearly minimum-phase and that is also the case in this room. Therefore we can cause improvements by just using plain minimum-phase filters. This unfortunately does not carry over to trickier environments such as car cabins. It should finally be mentioned that the pre-ringing using the mixed-phase filter was at its highest level 60 dB below the peak in the impulse response. (Again, at other listening positions using real measurements.) We can therefore safely conclude that mixed-phase inversion is useful and can improve sound system performance without loss of robustness.

Energy transfer (full bandwidth) up to time t in % of total energy

Legend:
- original
- Mixed-phase correction
- Min-phase correction

Energy transfer below 300 Hz

We have shown that mixed-phase equalizers can improve an already good situation while increasing robustness as compared to minimum-phase filters. Even if the perceived difference in this specific case would not be huge for an average listener, the importance of this example lies in the fact that the equalizer is more robust than the minimum-phase one (as the minimum-phase filter does not consider the full impulse response) while at the same time it shows that decay times of impulse responses can be reduced by turning away from minimum-phase responses. The only disadvantage is non-zero latency, but if this is no problem, I would recommend the mixed-phase inversion as it is simply a safer method. (As a side remark, in the method used in the examples above a parameter controls the amount of pre-ringing vis-à-vis the amount of post-ringing and thus we can reach an even better system response if listening tests reveal that people are insensitive to pre-ringing of up to say -30 dB.)

FAITHFUL STEREO REPRODUCTION

Stereo reproduction is based on the fact that humans are sensitive to correlation of sounds from different sources. Even if there are two physical sound sources in a room, we can under certain circumstances perceive a single source placed somewhere in between the two physical sources. If the two sources play identical sounds, and are equally far away from us, one to the left and one to the right, we will perceive a phantom source in the middle between the two sources. When the sources are not perfectly correlated, the spatial impression is altered. Also, if one source is delayed enough we will not even hear that source; all sound will be perceived as emanating from the earlier source. This is the well-documented precedence effect. An interesting aspect of the precedence effect is that the perception depends not just on the delay but also on the spatial separation between the two sources. For instance, if the two sources are located on a straight line extending from the listener so that the listening angle is the same to both sources, the sound will be perceived as colored. If the same delay settings are used but the speakers are separated horizontally, there is no perception of coloration. We simply hear the first source. If we put an omni-directional microphone in the place of the listener, the two experimental set-ups however

yield two identical recordings (assuming the room is well-damped). Both sound awful. The second source interferes with the first one. Taking the magnitude of the Fourier transform of both recordings reveals that they are identical and have the shape of a comb; deep notches permeate the spectrum. Again, we find that the Fourier transform is a poor representation of our hearing sensation. This time the main reason is that a basic Fourier transform does not distinguish between different angles of incidence.

Let us now consider a traditional stereo loudspeaker set-up in a listening room. If the two sources have different transfer functions then they will change the correlation properties of the Left and Right signals. The signals at the ears are no longer correlated in the same manner as they were on the recording. If the difference is big, the L and R signals may have been completely decorrelated. In that case, even when L=R at the sources we hear two distinct sources instead of the expected phantom in the middle. The sound stage becomes diffuse. If we now do a proper impulse response correction of the individual transfer functions from the left speaker and the right speaker then we can reconstruct the original correlation properties from the recording and the sound stage will become distinct and coherent again. Unfortunately, the reverse situation is sometimes true in alleged room correction systems employing high-order filters. Instead of improving the correlation of the L and R signals, they unwittingly decorrelate the sound field.

The reason is that a high-order filter has a long impulse response that needs to be properly controlled. If we build an inverse filter based on a simplistic average model of the original transfer function measured in different listening positions, we might end up correcting for issues late in the tail of that model that are not in accord with perception. Consider the following plot taken from a rather poor, but large, listening room, using a very good speaker, measured in 12 different positions corresponding to two different seats in a sofa. The plot is a Fourier magnitude response based on a simple delay-and-average response, i.e. the impulse responses were aligned in time and then averaged (i.e. beamforming), before taking the Fourier transform.
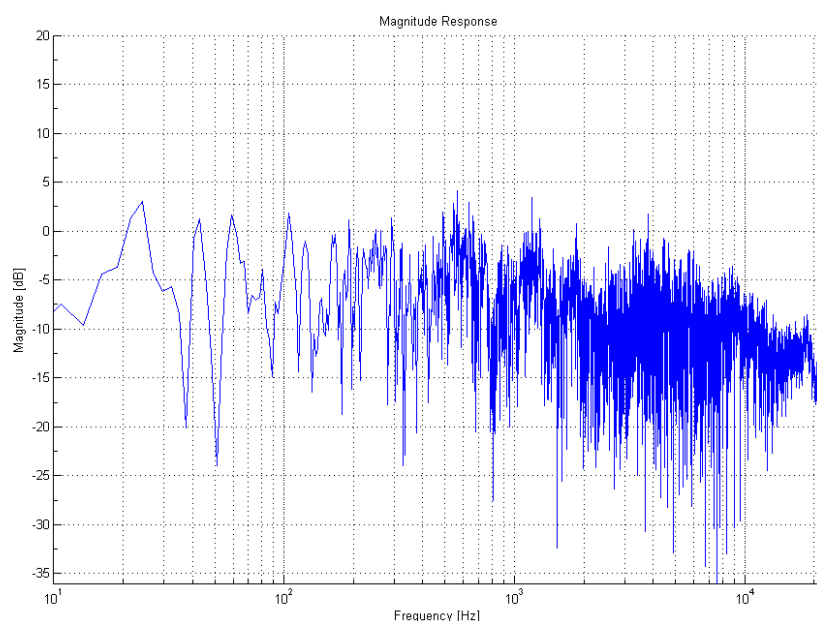

Note the amount of ripple, in particular in the high-frequency region. (If we plot this data using fractional octave smoothing, we will not *see* the ripple, but it is still there and will affect the final filter unless we are careful.) The problem with this ripple is that it is not a property of the frequency response that we perceive, only of the Fourier frequency response which does not take spatial angle nor time into account. The ripple is inaudible, because it is only a property of

the late reflections in the room (what can be regarded as late is of course frequency-dependent). However, let us now assume we build an equalizer that inverts this Fourier spectrum. Regardless of whether it is minimum-phase or not, it will still have similar variability in the high-frequency region which manifests itself in the time domain as additional high-frequency garbage in the impulse response (spread out over time). If we now use different equalizers on the left and on the right speaker, we end up with two widely different transfer functions that can have any arbitrary spatial effect, typically in the form of a weak decorrelation. The naïve averaging led us to compensate for issues that are (a) nonexistent in any single listening position and (b) psycho-acoustically irrelevant. It is well worth noting that correcting for something inaudible may lead to something audible.

Based on the same twelve measurements we used to produce the plot above, we can form estimates of more psycho-acoustically relevant representations by using different types of time-frequency analysis. The two following plots show two alternative representations, both in the frequency domain, but psycho-acoustically relevant for different purposes. The first one has been subjected to a processing which tries to resemble the human perception of transients in a reverberant room. As we can see, the high-frequency ripple is gone but still some rather powerful dips are present below 1 kHz. In somewhat simplified terms, we have removed late reflections that are not from the same direction as the direct wave.
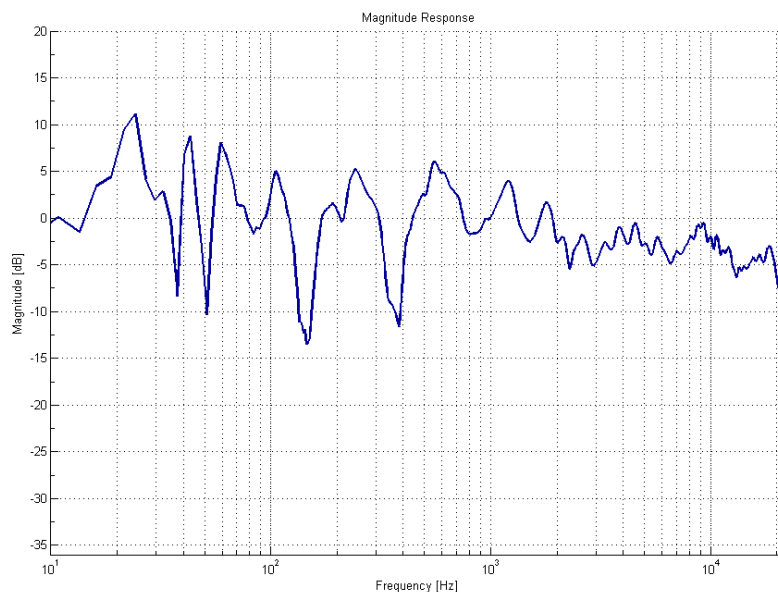
However, this picture does not give an adequate picture of what we perceive when listening for instance to a string quartet. The late reflections will still add coloration that affects our listening experience on more stationary harmonic signals. For that reason, a plot which gives a reasonable estimate of our appreciation of stationary frequency-domain coloration looks something like the following.
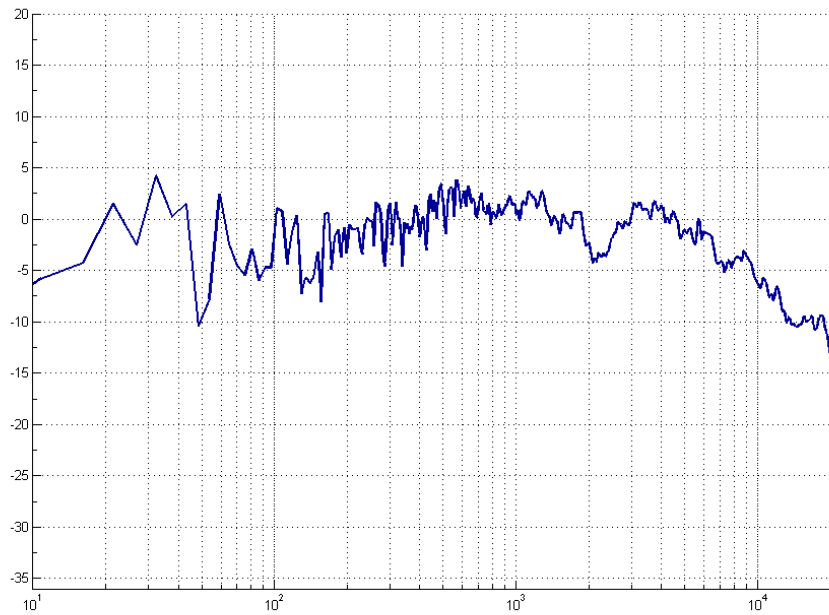
My claim is that both these plots give a meaningful picture of what we actually perceive in the listening room. The former plot for how we perceive transients. The latter for how we perceive stationary harmonic signals. That is to say, we should not base a magnitude response correction on the former, and we should not base an impulse response correction on the latter.

The key improvements that we obtain when using a mixed-phase correction lie in stereo and multi-channel reproduction. Spatial information gets distorted when the left and right transfer functions are different. Since room responses are non-minimum-phase, we need a mixed-phase inverse in order to reestablish the correct spatial image. The difference between a proper mixed-phase inverse and a minimum-phase one becomes bigger in a poor listening space, such as in a car, in particular where the left and right transfer functions differ appreciably. It should be noted that the difference in sound quality is typically not subtle. Although there are certainly cases where differences are small, that is when the major room response variations can actually be corrected for by a minimum-phase filter, the typical difference is substantial. However, as previously warned, it is difficult to do a proper impulse response correction since one very easily ends up with a decorrelation filter (and/or pre-echoes). That can however be detected: in listening tests the spatial image becomes diffuse (as opposed to distinct) and the sound may even be perceived as coming outside the region between the two speakers which is a clear sign of phase correction gone bad; by plotting the frequency response of the filter in high resolution, the filter should not have any fast variations in the high-frequency region, this is a sure sign of trying to correct for properties of the late tail in the impulse response. Fast variations in the magnitude response are indicative of similar variations in the phase response, and it is highly implausible that phase correction of room responses should be useful at high frequencies (since what would constitute a correction in one physical position is sure to cause a degradation just a few centimeters away according to the laws of wave propagation).

The end advice is that mixed-phase correction is absolutely useful and is in my mind the correct route forward to improved sound reproduction, but that it is much more difficult than a reading of the last twenty years of research on impulse response correction would lead one to think.

SOME REFLECTIONS ARE GOOD, SOME BAD

There seems to be consensus in the field that some early reflections actually help make speech more intelligible. However, it is also well documented that reflections within 5-10 ms of the main pulse in typical listening rooms are above the level where the primary source shifts or spreads (even when just listening to a single primary source). Reflections from the front and the rear (within ±40º) are perceived as detrimental to sound quality, whereas side reflections (within reasonable levels) often improve the perceived sound quality.

This can be understood from an information-theoretic perspective. Reflections that come from the front are typically very hard to distinguish from the primary source due to the placement of our ears. The transfer function of an individual front reflection is nearly the same at the left and the right ear, just like the response of the primary source. Indeed, this is the reason why a filter can correct for these reflections robustly. They are constant in a relatively large listening volume! Contrast this with reflections from the side. Such reflections vary much more with position due to the angle of incidence. There is always a big difference at the left and the right ear, and consequently side reflections give a diversity gain. Diversity is an information-theoretic concept that quantifies the number of distinguishable communication channels. The higher the diversity, the higher the Shannon capacity of the information transfer. This is utilized in mobile wireless communication systems where reflections that are independent of the direct path are used to actually increase the bit rate of the system. There is also an interesting corollary that reflects the detrimental effects on human auditory perception by front and rear reflections. Since these channels cannot be distinguished from the primary source (both are nearly constant), these channels simply cause self-interference. In the capacity formulas, this translates directly to lower signal-to-noise ratios, and lower capacity of the information transfer.

Since the effects mentioned here that reduce sound quality are approximately constant with respect to position it means that these are also effects that a non-causal mixed-phase filter can correct. And similarly, the information-bearing reflections are not possible to improve, simply because the reason that they bear information is their variation with position. If they were the same at both ears at all times, nothing could guide us to separate the source from the reflection, and consequently fidelity would be reduced. It is indeed an interesting fact, one that can be expected on account of evolution, that the human auditory system is capable of taking advantage of some of the additional information that is being provided to us from the room. At the same time, we all know intuitively that this reasoning hinges on the assumption that the reflection actually carries distinguishable and additional information. That is only a rather special case, and it is clear that most reflections will simply diminish sound quality.

SUMMARY AND RECOMMENDATIONS

FIR and IIR filters have more or less the same possibilities. In many cases, a minimum-phase EQ will do a decent job if tuned correctly. However, in order to improve spatial distinctness and to reach a sound quality which is sure to be good, care has to be taken to evaluate the whole system response. For best performance, mixed-phase EQ is required. Striking differences can be achieved in difficult listening environments such as cars by proper mixed-phase inversion. To do a mixed-phase inversion, however, latency is introduced and filter lengths typically need to be rather large. In applications where latency must be minimized or where filter order is limited, say below 100 FIR taps at 44.1 kHz sampling, my recommendation is to use a properly designed minimum-phase IIR EQ. Even then, one has to take care to achieve a design which is robust with respect to position.

If your DSP budget allows you to use high-order filters my recommendation is to use a careful mixed-phase design. A well-designed mixed-phase filter gives a faster system response than can be achieved by minimum-phase or linear-phase filters. But again, the word 'careful' is important. Some algorithms for designing high-order filters (be it minimum-phase or mixed-phase) will have a hard time staying away from adding spurious high-frequency ripple in the tail of the impulse response. This affects stereo perception in a negative way (spreading of virtual sources). Pre-echoes can also be introduced if the algorithm tries to do 'too much'. In short, mixed-phase designs have a better potential both performance-wise and robustness-wise, but that potential is considerably more difficult to achieve. (For a subwoofer channel, however, minimum-phase inversion will typically be sufficient as the combined room and subwoofer transfer function actually is well modeled by a minimum-phase system at such low frequencies.)

A mixed-phase design can be realized as an FIR filter or as an IIR filter, but typically the most efficient implementation is a mix of the two.

As described above, a proper mixed-phase filter design is similar to removing reflecting surfaces near the loudspeaker. In addition, it minimizes linear imperfections in the loudspeaker design itself. Digital mixed-phase filters therefore make a powerful and cost-effective complement to ordinary electroacoustic considerations in the design of loudspeaker systems and rooms.

ACKNOWLEDGEMENTS