The Basics of High Fidelity
Part 5: Audio Compression

So we have defined transparency (Part 1) and the resulting incompatible ideals for sound reproduction “here and now” versus “there and then” (Part 2) (augmented reality versus virtual reality). We have constructed the ideal “there and then” loudspeaker reproduction on the basis of reflection optimization (Part 3 & 4). In part 5 we will deal with audio compression for which we can return to the simple transparency ideal using direct digital representation comparison.

A first, mathematical principle, that can be exploited in audio compression is redundancy. When an audio signal contains repetitions we can simply send the signal part that is repeated and send a description of how to repeat it. Although there are some algorithmic difficulties to overcome, especially when compressing “live” signals, such a system can be build and a compression factor of about 2 can be reached with advanced signal processing.

A second, perceptual principle, that can be exploited is irrelevancy. When a signal component is inaudible you don’t need to code it. In its simplest form it is already exploited by limiting the maximum frequency that is coded in the signal. It is no use to code a 25 kHz audio signal if nobody is able to detect it. In a more advanced implementation you construct a time-frequency representation and make an analysis of each time-frequency component to see if it is masked by another time-frequency component that makes it inaudible.

Everybody knows from experience that masking is a powerful concept, when we are perceiving a loud sound any soft sound that occurs simultaneously becomes inaudible. Figure 1 illustrates this for a pure sinusoidal tone using a time (1A) and frequency domain representation (1B). If we listen to such a loud tone a second, softer, tone will be inaudible and all simultaneous sinusoids within the blue area are inaudible. These soft tones thus need not to be coded into the audio representation.

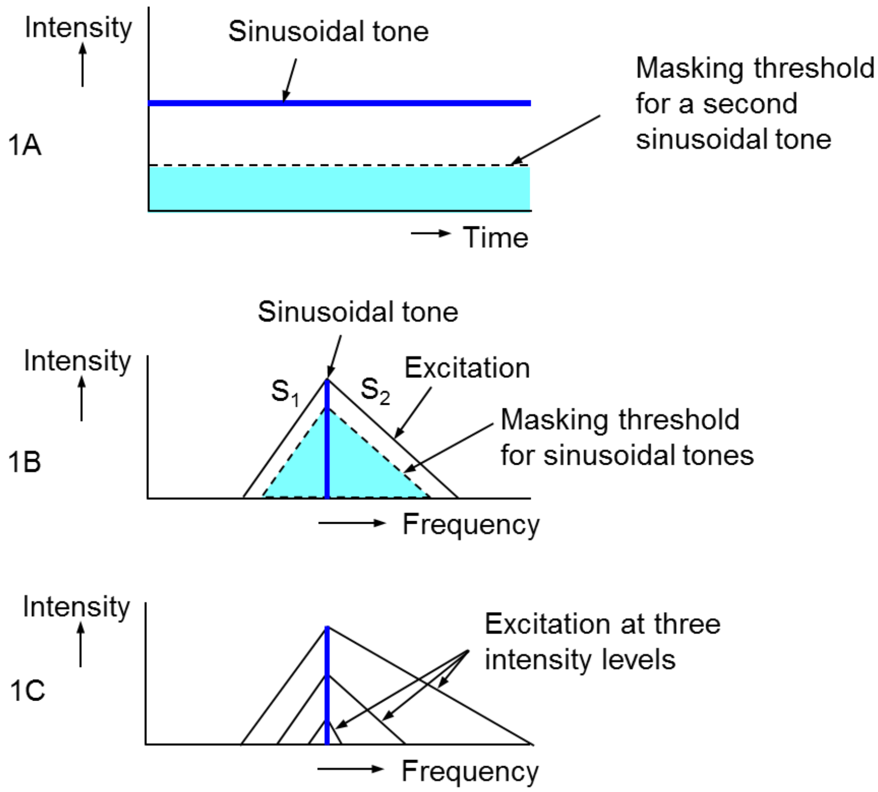

Figure 1. Masking of a soft sinusoidal tone by a second loud one in time- (1A) and frequency domain (1B). All tones within the blue triangle in 1B are inaudible. For loud sinusoidal tones the right slope (S_2) is steeper for low loudness values. Interference that leads to audible beats are not taken into account.

For a short pulse, which by definition is composed of an infinite number of sinusoids, a similar picture can be made. A second softer pulse will be inaudible, even if it does not occur simultaneously. The final masking picture looks the same as for pure sinusoids, only the time and frequency axis have been interchanged (see Figure 2).

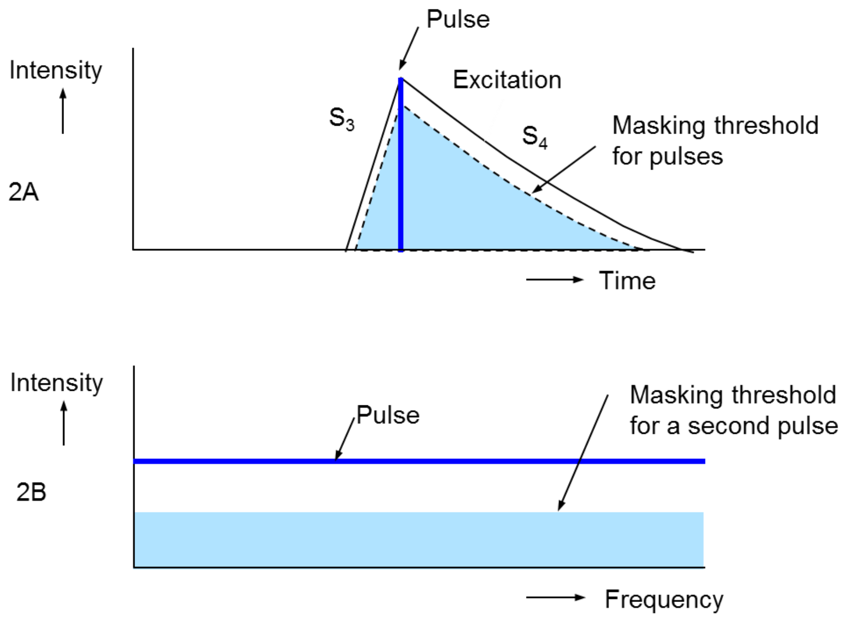

Figure 2. Masking of a soft pulse by a loud pulse in time- (2A) and frequency domain (2B). All pulses within the blue triangle in 2A are inaudible.

Music is more than just a few sinusoids or pulses but as a first approximation it can be constructed from the addition of a large set of short duration sinusoids. For such a signal the masking threshold is drawn in Figure 3A and B in both the time- and frequency domain. When drawn in a time-frequency representation we get a “tent” shaped figure as given in Figure 3C. If we design an audio codec in such a manner that all masked components are not coded and all distortions that are introduced by the coded fall within the blue “tent” shaped figure, the codec will be perceptually transparent while at the same time we need significantly less bits in the coding of the audio signal.

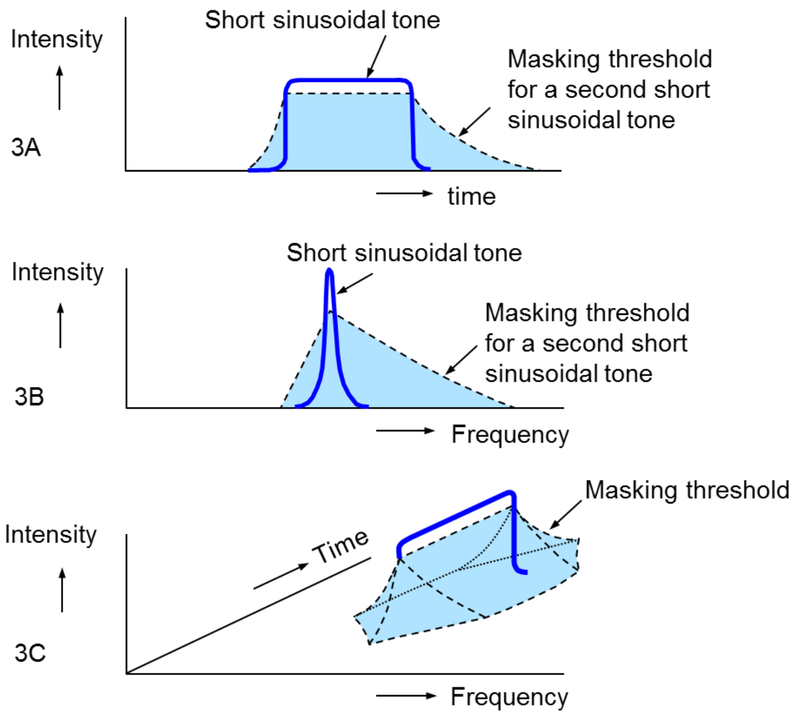

Figure 3. Masking of soft short sinusoidal tones by a loud short duration sinusoid in time (3A), frequency (3B) and time-frequency domain representation. All tones with the blue “tent” shaped figure are inaudible.

The amount of compression that can be reached is strongly dependent on the number of audio channels that have to be coded, a single mono channel, two strongly correlated stereo channels, or a larger set of highly correlated surround channels (5.1, 7.1, …) that can be effectively coded by using “binaural cue coding”. For standard stereo CD audio modern coding schemes like MP3, AAC, WMA, OGG, … allow for compression factors up to 10 with only marginal loss of audio quality. A problem that can arise in stereo/surround coding is that our auditory system is extremely sensitive to binaural timing differences, making perceptual testing difficult. Also the reproduction of compressed audio over loudspeakers may give rise to strong unmasking effects caused by interference between the direct sound and strong first reflections. And finally, some types of postprocessing, e.g. the one proposed in Part 7 where the left minus right signal is used to create a diffuse field that allows for an optimal immersion experience, can lead to disturbing degradations that were not noticeable in a standard headphone test. So we may conclude that audio compression has made HiFi a complex field of engineering.

Go to Part 6: Subjective Testing.

John G. Beerends
Published in Hifi Video Test 2/2008 (in Dutch), translated and updated over the period 2012-2022.