

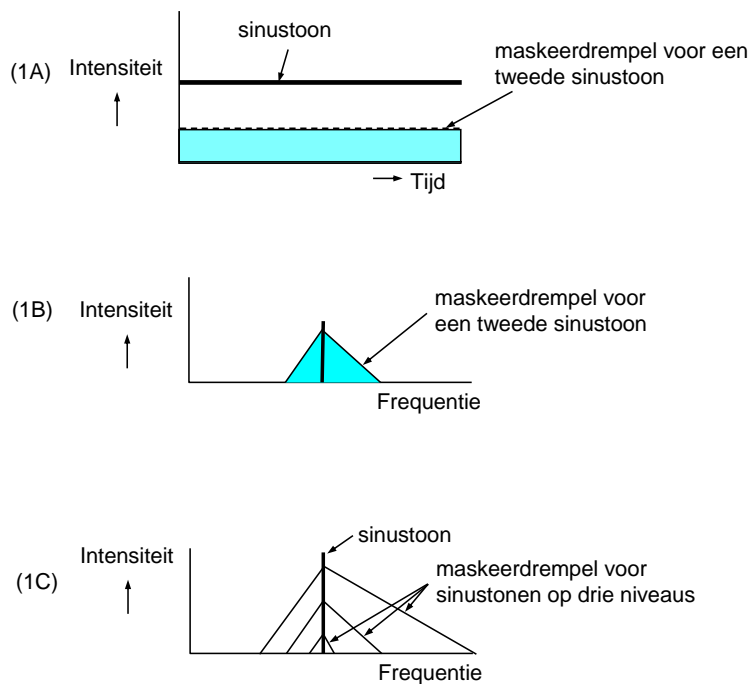
De betekenis van het begrip HiFi (updated english version [here](#))

Deel 5, Audio Compressie

We hebben met behulp van het begrip transparantie de twee onverenigbare idealen “hier en nu” versus “daar en toen” kunnen opstellen (deel 2) wat in deel 3 leidde tot een revolutionair luidsprekerontwerp voor de “daar en toen” illusie. De rol hierbij van reflecties en resonanties is in deel 4 uitgediept. In dit deel behandelen we audio compressie en gaan we terug naar de fundamentele transparantievraag waarvan in deel 1 de basis filosofie is gegeven.

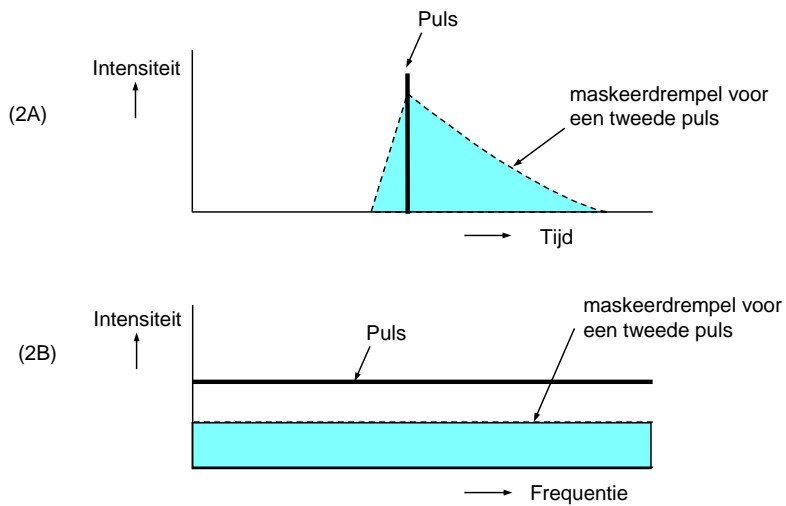
We hebben in deel 1 uitgebreid stilgestaan bij het begrip transparantie, veel HiFi apparatuur moet niets aan het geluid toevoegen of weggelaten. We kunnen dit simpelweg testen door het audio signaal wat we erin stoppen af te trekken van wat eruit komt, als de uitkomst nul is, eventueel na amplitude schaling, dan is het HiFi apparaat perfect. Nu bevat een audiosignaal vaak veel redundantie, er worden stukjes geluid herhaald. Dus is het voor het transparant opslaan van audio niet noodzakelijk om alle bits op te slaan, je kunt de data comprimeren. Bekende compressie programma's zoals zip gebruiken de redundantie om een data file kleiner te maken zonder dat ze enige kennis nodig hebben van wat de data representeert; tekst, audio, video het kan allemaal. Als je gaat kijken naar de compressie factor voor een stukje muziek dan levert deze compressie techniek niet veel op, voor muziek signalen ligt die factor ergens tussen de 1.1 en 1.3. Met compressie technieken die speciaal gebruik maken van de eigenschappen van audio kunnen we ongeveer een factor 2 halen, nog steeds niet de hoge compressie die je tegenwoordig ziet in MP3 spelers die gebruik maken van geavanceerde codecs (een apparaat dat codeert, met gebruikmaking van compressie, en decodeert).

Om tot hoge compressie factoren te komen moeten we naast de redundantie ook de irrelevantie in het audio signaal uitbuiten. Of eenvoudig gezegd, we moeten alles wat we niet kunnen horen niet coderen. Om uit te vinden wat we wel en niet kunnen horen gaan we te raden bij de psycho-akoestiek, het vakgebied wat de relatie legt tussen de meetbare eigenschappen van een signaal en de waarneming daarvan. Uit die psycho-akoestiek kennen we het verschijnsel maskering, een hard geluid maakt de waarneming van een zacht geluid moeilijk of zelfs onmogelijk. Nauwkeuriger gezegd, een hard geluid maakt de waarneming van een ongeveer gelijktijdig zacht geluid met ongeveer hetzelfde spectrum moeilijk of zelfs onmogelijk. In Figuur 1 is daar een voorbeeld van gegeven voor een zuivere sinustoon, in zowel het tijddomein (1A) als het frequentiedomein (1B). Een tweede zachtere toon is onhoorbaar als de frequentie en amplitude binnen de blauwe driehoek uit Figuur 1B liggen. Deze onhoorbare tonen hoeven we dus niet te coderen en eventuele fouten bij het coderen van de sinustoon zelf mogen dus signalen opleveren die binnen deze driehoek vallen. Bij nauwkeurig meten is de helling van de rechterkant van deze driehoek afhankelijk van de intensiteit van de sinus (zie 1C).



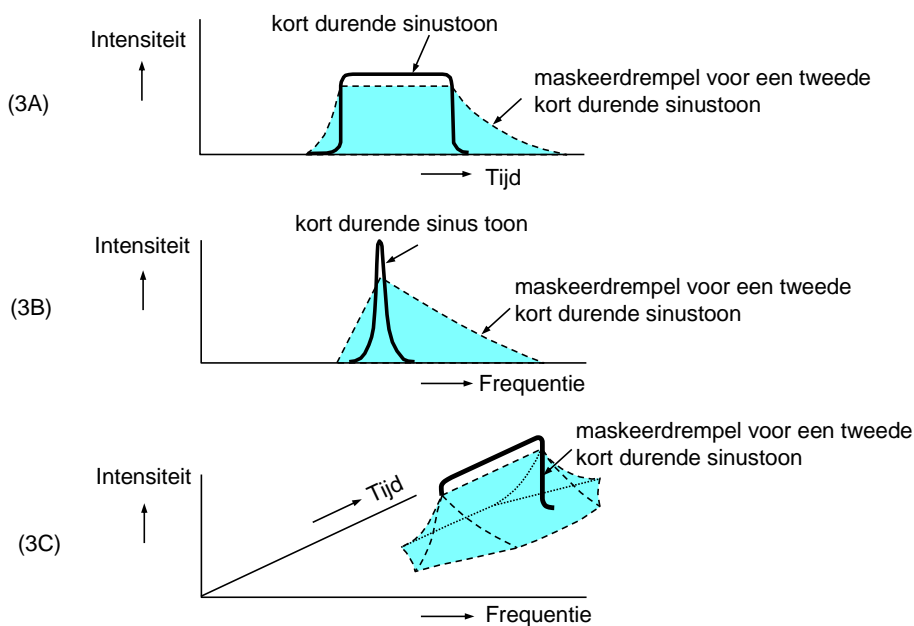
Figuur 1. Maskering van een zachte sinustoon door een harde sinustoon in tijd- (1A) en frequentiedomein (1B). Alle tonen binnen de blauwe driehoek in 1B zijn onhoorbaar. Voor luide sinustonen loopt de rechter helling vlakker dan voor zachte sinustonen (bekend als “upward spread of masking”). Er is geen rekening gehouden met interferentie die tonen met bijna dezelfde frequentie hoorbaar kunnen maken.

Voor een korte puls, die per definitie is opgebouwd uit oneindig veel sinussen, geldt iets dergelijks, een tweede, zachtere puls is onhoorbaar, zelfs als deze niet gelijktijdig klinkt. Dit levert vrijwel hetzelfde plaatje op maar dan met verwisseling van de tijd en frequentie-as (zie Figuur 2).



Figuur 2. Maskering van een zachte puls door een harde, in tijd- (2A) en frequentiedomein (2B). Alle pulsen binnen de blauwe driehoek in 1A zijn onhoorbaar.

Muziek bestaat uit meer dan alleen zuivere tonen en zuivere pulsen maar laat zich in eerste instantie wel goed benaderen door een opbouw van kortdurende sinustonen. Voor zo'n signaal ziet de maskeerdrempel eruit als in Figuur 3, waarbij ook een simultane representatie van tijd en frequentie domein is getekend. Als we een codec maken die alles weglaat wat blauw is in deze figuur is hij perceptief transparant.



Figuur 3. Maskering van een zachte kort durende sinustoon door een harde kort durende sinustoon, in tijd- (3A), frequentie- (3B) en tijd-frequentiedomein (3C). Alle tonen binnen de blauwe gebieden zijn onhoorbaar.

Als je in een codec zowel de redundantie als irrelevantie optimaal uitbuit kun je hoge compressiefactoren halen. De schattingen tot hoever je audio kunt comprimeren voordat de verschillen hoorbaar worden, liggen nogal uiteen maar een factor 10 voor een normale stereo opname lijkt haalbaar. Die factor hangt ook erg af van wat je codeert, een mono opname, een stereo opname of een meerkanaals surround opname. Bij het gebruik van “binaural cue coding” is de bit rate toename per kanaal marginaal terwijl die bij normale PCM codering linear mee omhoog gaat. Bij een 7-kanaals surround opname zal een factor 20 dan ook geen probleem zijn. Wat wel een probleem wordt is hoe je kunt testen of het systeem echt perceptief transparant is. De simpele test van het aftrekken van in en uitgang werkt niet meer, het verschil kan heel groot zijn, maar wel onhoorbaar.

En het tegenstrijdige van het hele compressie verhaal is dat er aan de andere kant van het kwaliteitsspectrum sampling rates worden verdubbeld en per sample 24 in plaats van 16 bits worden gebruikt om te komen tot een superieure geluidskwaliteit. Wie heeft er gelijk?, tijd om een test opzet te bedenken waarmee we deze vraag kunnen beantwoorden (in deel 6).

Dr. John Beerends

Gepubliceerd in Hifi Video Test 2/2008, herzien 1/2012