

Jak działa dźwięk w komputerze?

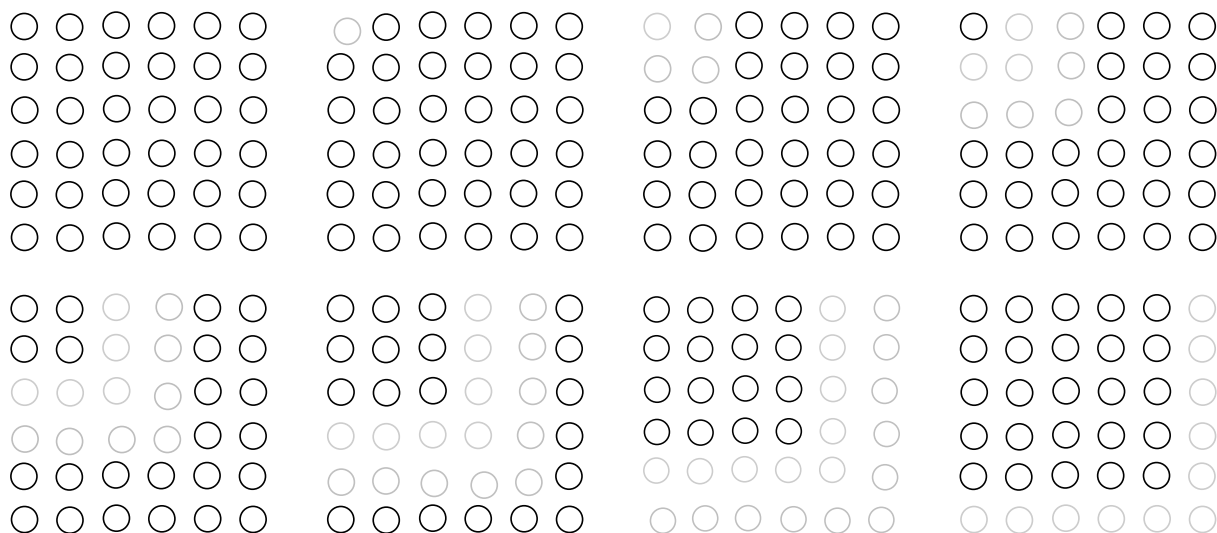
Jakub Juszczakiewicz

2025-12-25

Żyjemy w czasach w których do posłuchania ulubionej piosenki wystarczy komputer lub urządzenie elektroniczne posiadające część jego właściwości. Korzystanie z elektronicznych odtwarzaczy jest powszechne i codzienne. Można słuchać praktycznie wszędzie, wykorzystując do tego słuchawki i smartphone czy siedząc w zaciszu domowych i delektując się dźwiękami z radia cyfrowego radia internetowego. Ale jak to się dzieje, że w komputerze można z taką „łatwością“ przechowywać dźwięki? To nie jest aż tak skomplikowane, jak może się wydawać. Zacznijmy jednak od fizyki - czym jest dźwięk - żeby łatwiej było potem zrozumieć resztę cyfrowego świata muzyki.

1 Fizyka

Fizycy mówią, że dźwięk to podłużna fala mechaniczna rozchodząca się w ośrodku gazowym. Magia, ale to niewiele mówi tak na prawdę. Zacznijmy od tego, że dźwięk nie rozchodzi się jak wiatr. Powietrze w którym rozchodzi się dźwięk może całkowicie stać w miejscu. Jeśli weźmiemy jakieś źródło dźwięku, np. głośnik, to on pobudza najbliższe cząsteczki powietrza do niewielkiego ruchu przez „wywarcie na nie ciśnienia”. Te cząsteczki zachowują się troszkę podobnie do wahadełek - są jakby popychane (bez dotykania się) i potem wracają z powrotem na swoje miejsce. Każda cząstka która zbliża się do kolejnej przekazuje tę zmianę (energię) kolejnej i wraca.



Powyżej jest uproszczony przykład jak może wyglądać takie przekazywanie dźwięku z lewego górnego rogu do prawego dolnego.

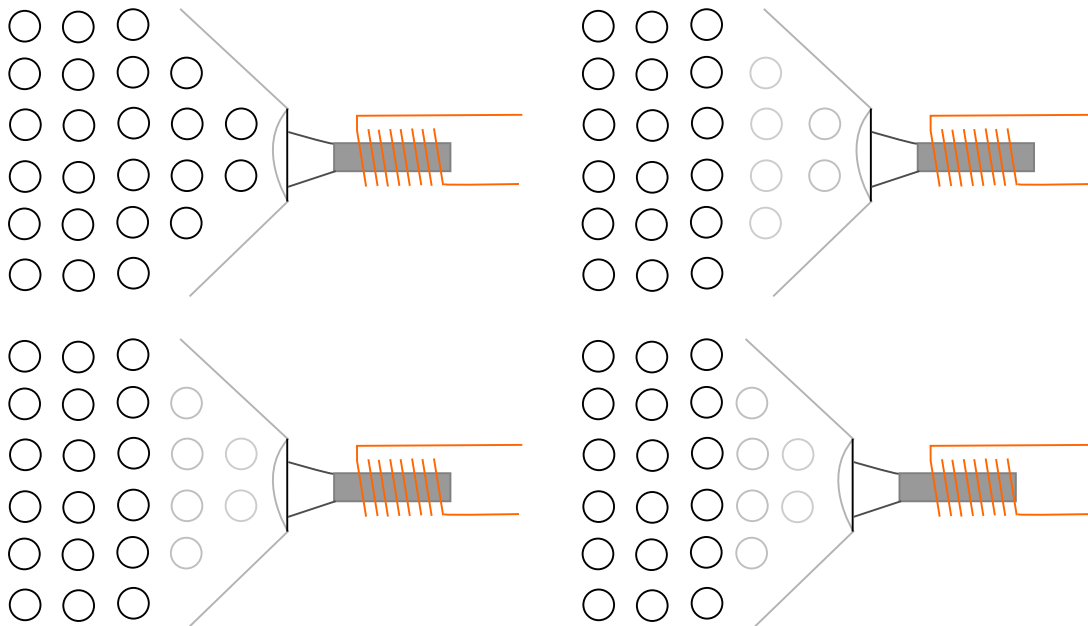
A jak się dzieje, że mamy wysokie i niskie dźwięki? Jeśli mamy nutę a - 440Hz, to cząstki będą przesuwały się z taką samą częstotliwością - czyli 440 razy na sekundę będą np. w jednej z 2 skrajnych pozycji - będą miały 440 cykle ruchu na sekundę. Wyższe tony będą wymagały szybszych ruchów, a niższe wolniejszych. Głośność z kolei będzie determinowana przez to, jak bardzo się oddalą od miejsca początkowego.

Ponieważ komputery są urządzeniami elektronicznymi, to do komunikacji ze światem potrzebują coś, co zamieni prąd elektryczny w postać odbieralną przez zmysły i w drugą stronę. W przypadku dźwięków służą do tego głośniki i mikrofony.

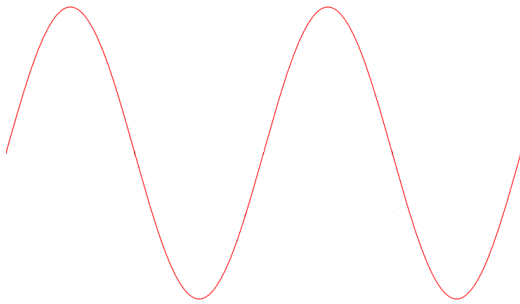
1.1 Elektromagnesy, mikrofony i głośniki

Mamy już część bazy fizycznej, więc czas na kolejny krok. Jak zamienić dźwięk na prąd elektryczny. Żeby to zrozumieć, trzeba mieć podstawy działania elektromagnesu, bo to jest jedno z serc zarówno głośnika jak i mikrofonu. Elektromagnes to drut (zwykle miedziany) pokryty cienką warstwą izolatora (żeby nie było zwarcia; zwykle przezroczystego) owinięty wokół kawałka metalu zwanego rdzeniem (np. z ferrytu). Zasada jest taka: jak przez (dowolny) drut płynie prąd, to wokół tego drutu tworzy się pole magnetyczne. Drut wtedy przez to, że płynie przez niego prąd staje się takim słabym magnesem którego pole magnetyczne jest zawinięte wokół przewodu - kręci się równoległe do niego. Elektromagnes wykorzystuje to zjawisko i potęguje wspomnianym wcześniej rdzeniem. Czyli jak do elektromagnesu podłączymy prąd - zaczyna przyciągać w zależności od napięcia, jakie podamy. Ale może to też działać w drugą stronę - jak będziemy blisko elektromagnesu ruszać (najlepiej szybko) jakimś zwykłym magnesem - w drucie pojawi się napięcie. To ta, jak z silnikami elektrycznymi - silnik napędzają elektromagnesy. A silnik (prądu stałego) jest też prądnicą jeśli odwrócimy proces.

Mikrofon składa się z elastycznej membrany która jest na sztywno podłączona do zwykłego magnesu który porusza się wewnątrz elektromagnesu. Jeśli dźwięk dociera do tej membrany, to cząsteczki powietrza tak, jak między sobą - wybierają ciśnienie na membranę, która się przesuwa - przesuując zwykły magnes. Magnes poruszając się wewnątrz elektromagnesu (w postaci np. cewki, czyli bez rdzenia - sam drut) indukuje napięcie w drucie. Takie napięcie potem trafia przez wzmacniacz do jakiejś karty dźwiękowej.



A co z głośnikiem? Głośnik działa dokładnie odwrotnie jak mikrofon - prąd zmienny porusza magnesem który popycha i ciągnie membranę. Jedną z ważniejszych kwestii jest to, że napięcie które jest wynikiem działania mikrofonu odzwierciedla te chwilowe zmiany ciśnienia (położenia cząstek). Jeśli do mikrofonu dotrze idealny ton (np. nuta A - 440Hz z kamertonu) to wykres (przebiegu) napięcia będzie idealnym sinusem. Np. gdy podłączymy oscyloskop do mikrofonu i przyłożymy do tego mikrofonu wzbudzony kamerton, to oscyloskop pokaże ładną falkę sinusa.



(Co)sinus dla przypomnienia.

2 Dźwięk cyfrowy

Wiemy już czym jest dźwięk i jak go zamienić na prąd. Ale to nadal jest dźwięk analogowy. Żeby można było go trzymać i przetwarzać w komputerze - potrzeba wersji cyfrowej. Ponieważ pamięć komputerów nie jest nieskończona, trzeba ograniczyć ilość informacji o dźwięku. Odpowiadają za to konwertery analogowo-cyfrowe, które robią dwie rzeczy: próbują i kwantują dźwięk z mikrofonu. Potem taki dźwięk może być trzymany w pamięci komputera, więc i przetwarzany. Jeśli potrzebujemy go odtworzyć, to proces się odwraca przez przetworniki cyfrowo-analogowe, ale po kolei...

2.1 Próbkiwanie

Mamy np. z mikrofonu to z kamertonu o częstotliwości 440Hz. Sinus napięcia w przewodzie z mikrofonu też ma 440Hz. Przetwornik analogowo-cyfrowy z większą częstotliwością sprawdza, w którym miejscu tego sinusa teraz jest. Sprawdza, czyli próbuje. Np. robi to z częstotliwością 44100Hz. Czyli 44100 razy na sekundę sprawdza, jakie jest aktualne wychylenie membrany w mikrofonie. To jest pierwsze ograniczenie ilości danych do przetwarzania. Popularne standardy próbkowania to np. 22050Hz, 44100Hz (płyta CD audio), 48000Hz (płyta DVD audio), 96000Hz czy 192000Hz.

Każdy wynik takiego sprawdzenia nazywamy próbką - stąd próbkowanie.

2.2 Kwantyzacja

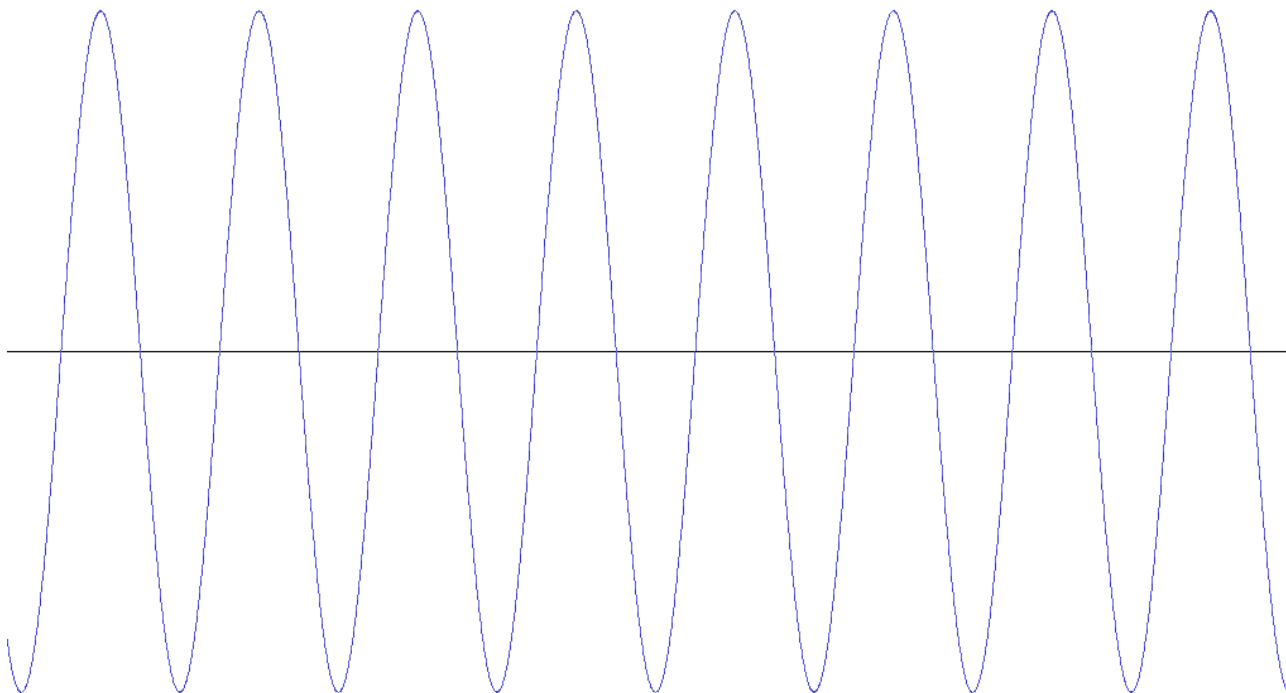
Drugim aspektem zamiany sygnały analogowego w cyfrowy jest to, jak dokładnie zapisać każdą próbkę. Znow - pamięci nie mamy nieskończenie wiele, więc musimy ograniczyć ilość danych. Można przyjąć, że każdą próbkę zapiszemy np. w 1 czy 2 bajtach. Bajt ma 8 bitów. Bit to cyfra w zapisie dwójkowych. My używamy zapisu dziesiętnego - mamy 10 cyfr (od 0 do 9). Na 3 cyfrach możemy zapisać liczby od 000 do 999, czyli tysiąc różnych. Mając do dyspozycji tylko dwie cyfry (0 i 1) przy 8 pozycjach (bitach) możemy zapisać liczby od 00000000 do 11111111, co po przeliczeniu na nasze to od 0 do 255. A jak ktoś jest „sprytny“, to można zapisać liczby¹ od -128 do 127. Przy 16bitach (2 bajty) mamy zakresy od 0 do 65535, czy ze minusami od -32768 do 32767. I teraz tak. Założmy, że membrana ma zakres wychylenia od 0% do 100%. Jeśli każdą próbkę zapiszemy w jednym bajcie, to 0% to będzie 0, a 100% to będzie

¹Wykorzystuje się do tego tzw. kod dopełnienia do dwóch.

255. Czyli najmniejsza różnica jaką możemy zapisać to ok. 0,4%. Jeśli sinus dźwięku będzie powoli narastał, od 0% do 100%, to będziemy zapisywać w komputerze kolejne próbki na zasadzie:

wychylenie:	0.0%	0.1%	0.2%	0.3%	0.4%	0.5%	0.6%	0.7%	0.8%	0.9%
wartość próbki:	0	0	1	1	1	1	2	2	2	2
odzwierciedlenie:	0.0%	0.0%	0.4%	0.4%	0.4%	0.4%	0.8%	0.8%	0.8%	0.8%

Jeśli zrobimy to samo, ale na 2 bajtach, a nie na 1, to dokładność „w pionie” będzie ok. 0.0015% a więc 256 razy dokładniej.



2.3 Jak dźwięk jest zapisywany

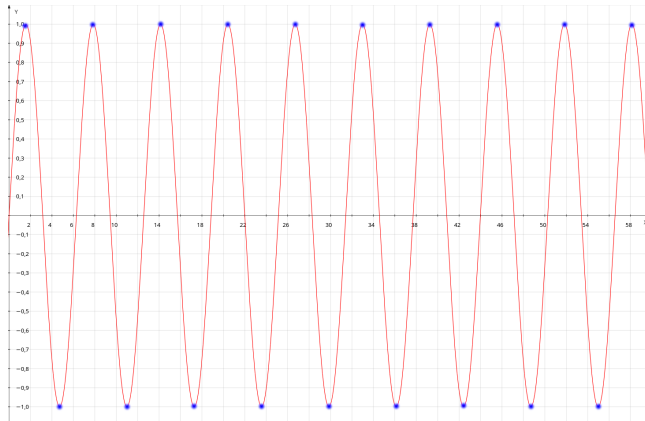
Dźwięk np. w plikach *.wav, jak również na płytach audio jest zapisywany jako ciąg próbek dla każdego kanału osobno. Na płycie CD audio próbkowanie jest 44100Hz, 2 kanały (lewy i prawy) oraz 2 bajty na próbkę. Daje to razem ok. 10MiB na minutę dźwięku, bo:
 $2 * 2 * 44100 * 60 = 10584000$.

2.4 Odtwarzanie

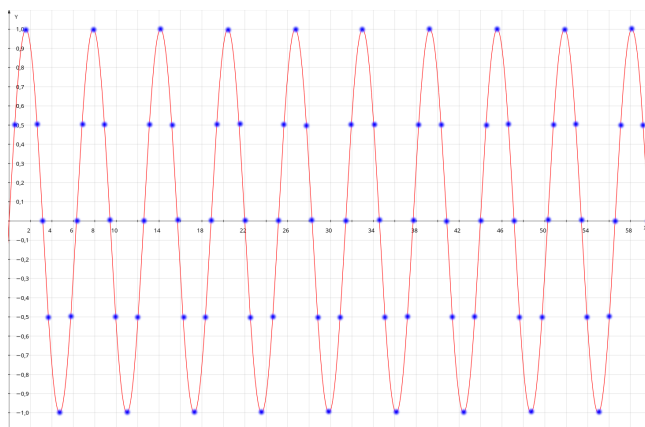
To została jeszcze kwestia odtwarzania. Tutaj nie powinno być zaskoczenia: podczas odtwarzania karta dźwiękowa za pośrednictwem przetwornika cyfrowo-analogowego bierze każdą próbkę po kolei i określonym taktie (wg częstotliwości próbkowania) podaje na wyjściu napięcie zgodne z wartością próbki. Taki dźwięk jest w przybliżeniu zgodny z analogowym, ale im częstotliwość tonu odtwarzanego dźwięku jest bliższa częstotliwości próbkowania - tym bardziej ten sinus będzie „kanciasty“, przy czym nie należy tego traktować dosłownie. Bezwładność membrany w głośniku czy kwestie fizyczne elektroniki takie jak pojemność i indukcyjność spowodują wygładzenie fali tonu do sinusa. Przy odtwarzaniu takiego sygnału nie ma możliwości „szarpania” dźwięku.

3 Pasma przenoszenia dźwięku

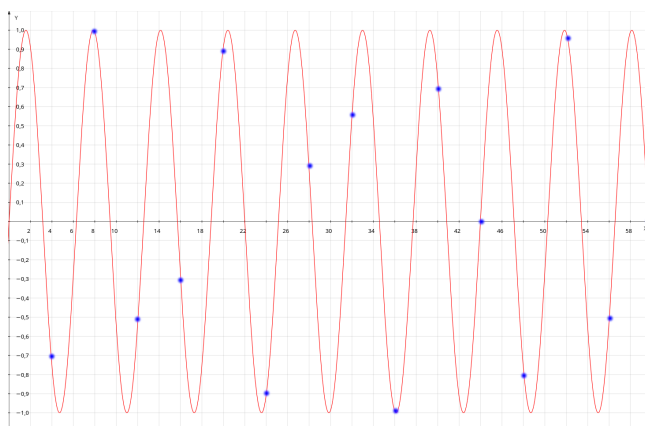
W teoriach związanych z przetwarzaniem dźwięku funkcjonuje takie twierdzenie jak twierdzenie o próbkowaniu zwane również twierdzeniem Nyquista–Shannona. Odnosi się ono do maksymalnej częstotliwości jaką można zapisać przy określonym próbkowaniu. O co chodzi? Spójrzmy na poniższy rysunek.



Niebieskie kropki oznaczają pobranie próbki a czerwony przebieg - pojedynczy ton. Jest to sytuacja kiedy częstotliwość próbkowania jest 2 razy większa niż częstotliwością tonu. Tutaj wszystko gra. W tym miejscu warto nadmienić, że jest to sytuacja idealna, gdzie próbki akurat trafiają w miejsce maksymalnego wychylenia membrany. Jakby były przesunięte, to próbkowany sygnał byłby zapisany jako cichszy, ale sama idea nadal działa. Teraz drugi rysunek:



Tutaj sytuacja jest jeszcze lepsza - częstotliwość próbkowania jest wielokrotnie większa niż częstotliwość tonu. A co jeśli sytuacja będzie odwrotna? Jest do tego jeszcze jeden rysunek:



Widać na nim, że jakbyśmy nie kombinowali, to takiego tonu poprawnie nie zapiszemy. I teraz wracając do wspomnianego twierdzenia. Mówi ono o tym, że jak mamy jakiś sygnał (np. ton) to potrzebujemy co najmniej 2 razy większą częstotliwość próbkowania, żeby taki sygnał poprawnie zapisać. Albo w drugą stronę, nie jesteśmy w stanie zapisać próbkując sygnału o częstotliwości większej niż połowa częstotliwości próbkowania. Co to nam daje? Jeśli ludzie słyszą maksymalnie 20kHz, to jeśli dźwięk będziemy zapisywać z częstotliwością próbkowania powyżej 40kHz, to zapiszemy wszystko co da się usłyszeć. Tylko jest jeszcze jeden mały szkopuł. Co, jeśli mikrofon zbiera wyższe częstotliwości? W kartach dźwiękowych stosuje się specjalne filtry - filtry dolnoprzepustowe. Działają one tak, że mają jakąś częstotliwość znamionową (np. właśnie 20kHz) i powyżej tej częstotliwości wyciszają wszystkie tony na zasadzie im wyższy - tym bardziej go wycisz. To wyciszanie postępuje szybko ze wzrostem częstotliwości i dość szybko całkowicie wygłusza wyższe tony, ale nie jest to przejście punktowe - jest pewien zakres częstotliwości które są wyciszone a nie odcięte. Dlatego na płytach CD przyjęto częstotliwość próbkowania 44100Hz - żeby zapisać tony do 20kHz i żeby filtry miały możliwość wygłuszenia wyższych tonów.

4 Stratna kompresja dźwięku

Ponieważ dźwięk w komputerze zajmuje dużo miejsca, pojawiła się potrzeba zmniejszenia jego rozmiarów. Ze względu na chaotyczną strukturę dźwięku (muzyka to nie pojedynczy ton, tylko zestawienie wielu), trudno jest go skutecznie kompresować bez utraty jakości. Ale z pomocą przychodzi zbiór opisów, który nazywamy „psychoakustycznym modelem słuchu”. Jest to zestaw który opisuje np. że jeśli dwa tony w muzyce będą odpowiednio blisko siebie, to cichszego człowiek nie usłyszy. Więc powstał pomysł, żeby powycinać z dźwięku, to czego i tak nie usłyszymy. W dużym uproszczeniu robi się to tak, że dźwięk zamienia się z próbek na zestaw składowych tonów. Można to skojarzyć z wizualizacją equalizera, gdzie w rytm muzyki wykres słupkowy pokazuje, jak głośne są niskie tony, jak głośne wysokie (zwykle od lewej niskie - do prawej wysokie) W takiej postaci (tylko dokładniejszej niż na takim wykresie) wycina się „ile się da“, np. żeby zostało tylko tyle danych, żeby w każdej sekundzie było ich 128 kilobitów. I tak pocięty dźwięk się zapisuje. Często się słucha plików MP3 o „tempie bitowym” 128kbit/s... to właśnie to. Zostaje wtedy z oryginalnego dźwięku mniej niż 10%. Odtwarzanie takiego formatu polega na zamianie tej postaci „jak z equalizera“ z powrotem na próbki, ale... to co się wycięło, to tego już nie ma. Ponieważ dźwięk taki ma mniej tonów, mówimy o stratnej kompresji. Algorytmem który umożliwia szybką zamianę „próbek na tony” czyli formalnie zmianę dziedziny dźwięku z dziedziny czasu na dziedzinę częstotliwości jest szybka transformata Fouriera. Tutaj warto jeszcze dodać, że są formaty kompresji bezstratnej jakie jak FLAC (Free Losses Audio Codec) który zmniejsza pliki z dźwiękiem o ok. 50% bez gubienia dźwięków.