

Cyfrowe formaty fonii

Więcej dźwięków z komputera

W ciągu kilku ostatnich lat komputer zmienił się z procesora tekstu/grafiki w uniwersalne urządzenie odpowiednie, między innymi, do zapisywania i odtwarzania sygnałów fonii i wizji.

W szczególności, w odniesieniu do reprodukcji fonii powstało wiele różnych formatów. Niniejszy artykuł opisuje najważniejsze z nich w skrócie, a obecnie popularny MP3 w szczególności.

Dziś komputer bez dźwięku jest rzadkością. Czy będzie to wyrafinowana gra komputerowa czy proste piski ostrzegawcze towarzyszące rozmaitym komendom, dźwięk jest nieodzowny. Dźwięk stał się ważny nawet dla użytkowników Internetu. Obecnie możesz słuchać rozgłośni radiowych lub łądować muzykę za pośrednictwem sieci. W wyniku wszystkich tych możliwych zastosowań i różnorodności typów komputerów, jakie pojawiły się w ciągu minionych mniej więcej dziesięciu lat, powstała mnogość formatów plików do przechowywania sygnałów fonii.

Kwantyzacja

Dzięki wprowadzeniu płyt kompaktowych (CD) niemal 20 lat temu, wielu ludzi wie, co należy rozumieć przez "dźwięk cyfrowy". W skrócie dla tych, którzy nie wiedzą: analogowy sygnał fonii jest przetwarzany przez przetwornik analogowo-cyfrowy (ADC) na wielką liczbę słów binarnych. Proces ten nazywa się kwantyzacją lub digitalizacją. Szybkość, z jaką odbywa się przetwarzanie w przypadku CD wynosi 44100 razy na sekundę dla lewego kanału i tyleż samo dla prawego.

Ponieważ rozdzielczość typowych przetworników stosowanych w CD wynosi 16 bitów, wskutek kwantyzacji sygnału fonii powstaje strumień danych o szybkości około 1,4Mb/s. Strumień ten, wraz z dodatkowymi danymi korekcji błędów i innymi informacjami, jest rejestrowany na płycie. W trakcie odtwarzania następuje po prostu odwrócenie procesu: laser odczytuje dane z płyty, a przetwornik cyfrowo-analogowy (DAC) tłumaczy je na sygnały analogowe.

W komputerze sygnały fonii są przetwarzane w podobny sposób. Niezbędna elektronika ADC i DAC mieści się na karcie dźwiękowej, natomiast twardy dysk funkcjonuje zazwyczaj jako medium przechowywania. W komputerze strumień bitów jest zapisywany w pliku danych.

Przechowywanie słów binarnych wymaga wielkiej ilości pa-

mięci: około 1MB na każde sześć sekund dźwięku o jakości CD. Istnieje kilka sposobów ograniczania wymaganej pamięci. Jednym z nich jest obniżenie częstotliwości próbkowania z 44,1kHz do, powiedzmy, 10kHz, co zmniejsza szerokość pasma. Innym jest rejestrowanie sygnału w formacie monofonicznym zamiast stereofonicznym, co niestety także pogarsza jakość. Trzecim jest obniżenie rozdzielczości z 16 bitów do, powiedzmy, 8 bitów, co powoduje pogorszenie stosunku sygnału do szumów i zwiększenia zniekształceń. A czwartą, szeroko stosowaną metodą jest kompresja sygnału w trakcie zapisu i dekompresja w trakcie odtwarzania.

Niemal każdy system komputerowy (Atari, Unix, Intel i tak dalej) ma swój własny specyficzny format plików, co spowodowało pewne zamieszanie, jak to istniejące w przypadku formatów graficznych. Można powiedzieć, że istnieje oszałamiająca mnogość formatów, co nie pomaga użytkownikowi. W **tab. 1** zawarto przegląd najczęściej spotykanych formatów kodowania fonii wraz z ich krótkim opisem.

Kompresować albo nie kompresować

Kompresja plików fonii jest naprawdę złożona, ponieważ nie można łatwo określić pewnych podobieństw pomiędzy wielką liczbą próbek. W konsekwencji, istnieje tylko kilka bezstratnych metod kompresji sygnałów fonii. Jedną z nich jest standard ADPCM (pod Windows) dla popularnego formatu WAV.

Jest jednak możliwa kompresja sygnałów fonii, jeżeli dopuścimy pewną utratę szczegółów, tak jak przy obróbce cyfrowych obrazów. W przypadku obrazów cyfrowych metoda kompresji została opracowana przez Joint Photographic Expert Group (JPEG), podgrupę Joint Technical Committee 1 (JTC1) ISO (International Standards Organization), konsultanta Narodów Zjednoczonych i IEC (International Electrotechnical Commission). In-

Editorial items appearing on pages 13..19 are the copyright property of (C) Segment B.V., the Netherlands, 1998 which reserves all rights.

Tab. 1. Przegląd najczęściej spotykanych formatów fonii.

AIF	Format początkowo opracowany dla Macintosha. Jest dość popularny w Internecie i oferuje wiele różnych szybkości próbkowania i rozdzielczości. Przeglądarki Netscape i Microsoft pracują w nim bez trudności.
AU	Zapoczątkowany przez firmy NeXT i SUN, stał się popularny w sieci web i do dziś ciągle jest stosowany. Ma wiele wariantów, ale zazwyczaj dane są skompresowane do formatu 8-bitowego w standardzie "u-law". Większość przeglądarek może w nim pracować.
ES	Potokowy format fonii firmy EchoCast. Dostępne są odtwarzacze dla Windows i Macintosha.
IFF	Format dźwięku Amiga obsługujący tylko 8-bitowy dźwięk mono: umożliwia swobodny wybór szybkości próbkowania.
LCC	Wysoko skompresowany format (możliwe są współczynniki do 1:50), aktualnie dostępny tylko dla Windows, ale inne wersje są opracowywane.
MID	Format MIDI nie jest prawdziwym formatem fonii, ale standardem do wymiany danych sterujących pomiędzy elektronicznymi instrumentami muzycznymi. Przeglądarki mogą obsługiwać MIDI za pośrednictwem rozszerzeń.
MOD	Oryginalny format Amigi będący reminiscencją MIDI. Plik MOD zawiera bank z próbkami i instrukcjami, jak te próbki powinny być odtwarzane. Wymaga rozszerzenia MOD.
MP3	Aktualnie najbardziej popularny format fonii. Dostępne rozmaite rozszerzenia.
RA, RAM, RPM	Popularny w Internecie potokowy format fonii firmy RealAudio. Dostępne są rozszerzenia (plug-ins) dla niemal wszystkich platform.
SND	Apple, Amiga i Tandy stosowały ten przyrostek dla plików fonii. Pewne warianty są kompatybilne z formatem AU.
STR	Format dla profesjonalnej obróbki dźwięku na Macintoshu.
VDO	Inny potokowy format fonii. Wymaga rozszerzenia (plug-in) takiego, jak VDO live player.
VMD	Potokowy format fonii zwany Internet Wave. Dostępny wraz z bezpłatnym koderem/dekoderem dla Windows.
VOC	Format Voice, opracowany przez Creative Labs, twórcę dobrze znanych kart Soundblaster.
WAV	Często stosowany format stał się dobrze znany od czasu wprowadzenia Windows. Oferuje wiele różnych szybkości próbkowania, rozdzielczości i współczynników kompresji.
XDM	Format MPEG firmy StreamWorks dla potokowej fonii. Istnieje specjalny odtwarzacz dla Windows.

na podgrupa JTC1, Motion Pictures Expert Group lub MPEG, ustanowiła standard dla kompresji dźwięku. Obecnie MPEG-1 Layer 3 (MP3) jest formatem popularnym wśród użytkowników Internetu.

Fonia w czasie rzeczywistym

W pierwszych dniach istnienia fonii w komputerze cały plik foniczny musiał być załadowany do pamięci komputera, zanim mógł zostać odtworzony. Było to szczególnie przykre w trakcie skrośnego ładowania plików poprzez Internet.

Wraz z nastaniem szybkich modemów stało się możliwe rozwinięcie technik umożliwiających odtwarzanie plików fonicznych w trakcie ładowania skrośnego. Szczególnie godny uwagi udział miał w tym proces zwany „AOD” (ang. Audio-On-Demand), „real time radio” lub „streaming radio”, opracowany przez firmę RealAudio.

MP3: dobra jakość przy wysokiej kompresji

Aktualnie MP3 jest najbardziej popularnym spośród wszystkich formatów fonii. W bardzo krótkim czasie ten protokół kompresji fo-

onii zajął silną pozycję jako kompresor muzyczny. Wiele współczesnych komputerów sprzedaje się z zainstalowanym oprogramowaniem kodowania i dekodowania MP3. Również wielka liczba plików MP3 jest rozsiewana poprzez Internet i są już kompilacje płyt CD zawierające pliki MP3. Niektórzy producenci zaczęli udostępniać MP3 Walkman: półprzewodnikowy odtwarzacz CD, stosujący jako nośnik pamięć „flash”.

Opracowany dla DAB

Jako udziałowiec ogólnoeuropejskiego projektu Eureka 147 (opracowania naziemnego systemu cyfrowej radiofonii - DAB), niemiecki Fraunhofer Institut fuer Integrierte Schaltungen (Instytut Fraunhofera ds. Układów Scalonych) opracował układ kodujący (codec - coder/decoder) dla DAB.

Odpowiedni algorytm kodowania uwzględnia pewne właściwości słuchu ludzkiego i na tej podstawie określa, czy pewien aspekt dźwiękowy fragmentu muzyki ma szansę dotarcia do słuchacza, czy nie. W zależności od tego prawdopodobieństwa określa następnie, czy odpowiednie dane

powinny być włączone do strumienia danych, czy nie. Skutkuje to systemem eliminującym redundancję (nadmiarowości informacji), umożliwiając zasadniczą redukcję danych bez pogarszania dźwięku.

Algorytm ten został dalej wzbogacony przez IIS we współpracy z Uniwersytetem w Erlangen i został zaakceptowany jako ISO MPEG-1 Layer 3 (IS 11172-3 i IS 13818-3).

Warstwy

Bez redukcji danych, sygnały fonii zawierają próbki o szerokości 8 lub 16 bitów, pobierane z częstotliwością dwukrotnie większą niż częstotliwość w tych sygnałach najwyższa.

Zostało już powiedziane, że digitalizacja sygnałów fonii skutkuje przepływem strumienia bitów o wartości około 1,4Mb/s. Najdoskonalsze obecnie systemy kompresji potrafią go ograniczyć w stosunku 1:12 bez słyszalnego pogorszenia dźwięku. Możliwe są współczynniki redukcji w stosunku aż do 1:24 i ciągle jeszcze jakość dźwięku będzie lepsza od otrzymanego po obniżeniu częstotliwości próbkowania lub rozdzielczości dla otrzymania porównywalnej kompresji.

Standard MPEG-1 opisuje trzy warstwy kompresji: Layer 1, Layer 2 i Layer 3. Wszystkie one są zdolne do wytworzenia dźwięku o jakości niemal CD. W tab. 2 zawarto charakterystykę tych trzech warstw. Definicje zawarte w normie odnoszą się jedynie do stosowanego kodera i formatu danych. Informacja ta umożliwia producentom projektowanie dekodów zgodnie z ich własnymi wymaganiami.

Jeśli nie stosuje się reprodukcji stereofonicznej i, co więcej, dopuszcza się ograniczoną szerokość pasma, można zastosować nawet jeszcze wyższe współczynniki kompresji. Najwyższa z warstw, Layer 3, stosuje najniższy strumień bitów (kbit/s) i tworzy najlepszą jakość dźwięku.

Wszystkie trzy kodeki są hierarchicznie kompatybilne, co oznacza, że dekodery dla warstw Layer 3 może być również stosowany dla warstw 1 i 2. Nie jest to jednak możliwe w odwrotną stronę. Im wyższy numer warstwy, tym bardziej złożony staje

Tab. 2. Tryby fonii MPEG-1.

Współczynnik kompresji	metoda	opis
1:4	Layer 1	sygnał stereo generuje strumień bitów 384kbit/s
1:6 - 1:8	Layer 2	sygnał stereo generuje strumień bitów 256 - 192kbit/s
1:10 - 1:12	Layer 3	sygnał stereo generuje strumień bitów 128 - 112kbit/s

się koder i tym wyższe współczynniki kompresji można stosować.

W tab. 3 zawarto współczynniki kompresji możliwe do uzyskania w Layer 3 i gdzie każdy z nich można zastosować. Testy odsłuchowe przy różnej publiczności wykazują, że osiągi Layer 3 pozostają doskonałe przy współczynnikach kompresji 1:12 - przy związanym z tym strumieniem bitów 64kb/s na kanał fonii. Jeśli w pewnych zastosowaniach pasmo akustyczne zostanie ograniczone do 10kHz, jest możliwa dobra reprodukcja stereo przy współczynniku kompresji 1:24.

Jednolita struktura

Wszystkie trzy warstwy kompresji mają tę samą strukturę. Zastosowana w nich technika kodowania jest znana jako perceptualne kształtowanie szumu albo perceptualne kodowanie transformaty podpasma. Koder analizuje składowe widmowe sygnału fonii za pomocą banku filtrów (patrz rys. 1) i korzysta z modelu psychoakustycznego do określenia dostrzegalnych poziomów szumów. Z kolei informacja podlega kwantyzacji i kodowaniu w sposób zapewniający, że zostaną wzięte pod uwagę dwa ważne warunki: maksymalny strumień bitów i efekt maskowania.

Wszystkie trzy warstwy stosują ten sam bank filtrów o 32 podpasmach. Wszystkie one dopuszczają częstotliwości próbkowania 32kHz, 44,1kHz, 48kHz i są w stanie pracować ze strumieniami bitów 32kb/s lub większymi.

Tab. 3. Ułatwienia dostępne w MPEG-1 Layer 3

Jakość dźwięku	Współczynnik kompresji	Szerokość pasma (kHz)	Tryb	Strumień bitów (kb/s)
Telefoniczna	1:96	2,5	mono	8
Lepsza niż krótkofalowa	1:48	4,5	mono	16
Lepsza niż średniofalowa	1:24	7,5	mono	32
Radio FM	1:24 - 1:26	11	stereo	56 - 64
Niemal CD	1:16	15	stereo	96
CD	1:12 - 1:14	>15	stereo	112 - 128

Tło

Dla uzyskania zasadniczej redukcji szerokości niezbędnego pasma cyfrowego, w MPEG-1 Layer 3 stosowane są różne techniki i skróty. Najważniejszymi z nich są:

- dolny próg słyszalności,
- efekt maskowania,
- odkładanie bajtów,
- łączenie stereo,
- kodowanie Huffmana.

Omówimy je kolejno:

Dolny próg słyszalności

Badania wykazują, że dolny próg ludzkiego słyszenia nie jest liniowy: wznosi się pomiędzy 2kHz i 5kHz. Jego właściwości zostały opisane przez Fletchera i Munsona. Nie jest konieczne kodowanie dźwięku leżącego poniżej progu, ponieważ słuchacz i tak nie może go usłyszeć.

Efekt maskowania

Wykorzystuje się fakt, że ludzkie słuch nie postrzega słabych dźwięków, które są całkowicie lub częściowo zamaskowane przez dużo silniejsze. Badania wykazują, że skutek maskowania pewne dźwięki nie muszą być kodowane, co pozwala zaoszczędzić wcale niemało miejsca. Stąd wszystkie kodery MPEG-1 Layer 3 zawierają model psychoakustyczny, w którym jest wbudowana ta właściwość ludzkiego słuchu.

Odkładanie bajtów

Często zdarza się, że muzyczny pasaż nie może być zakodowany przy dostępnym strumieniu bitów. A więc jakość dźwięku musi być chwilowo adaptowana, by umożliwić strumieniowi bitów zmieszczenie się w pojemności kanału

cyfrowego. W takich warunkach MPEG-1 Layer 3 stosuje bufor udostępniający pewną dodatkową pojemność. Opróżnienie bufora następuje, gdy dźwięk jest kodowany z szybkością bitów niższą niż dostępna w kanale.

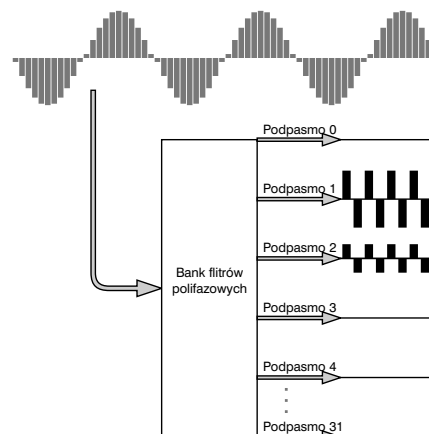
Łączenie stereo

Wiele małych systemów stereo hi-fi stosuje wspólny głośnik najniższych tonów (woofler). Pomimo tego słuchacz odnosi wrażenie, że dźwięk jest emitowany nie z tego głośnika, ale raczej z głośników-satelitów. Badania wykazują, że poniżej pewnych częstotliwości ludzki słuch nie jest w stanie rozstrzygnąć, z którego kierunku dochodzi dźwięk. Techniki kompresji mogą wykorzystać tę właściwość poprzez pomijanie informacji stereo poniżej pewnej częstotliwości granicznej. Oznacza to, że poniżej tej częstotliwości sygnał jest kodowany tylko w formie monofonicznej.

Kodowanie Huffmana

Kodowanie MPEG-1 Layer 3 stosuje klasyczną technikę: kodowanie Huffmana. Stosuje się je po przeprowadzeniu rzeczywistej kompresji danych, do zakodowania informacji cyfrowej. A więc nie jest to system kompresji, ale bardzo skuteczna technika kodowania. Algorytm Huffmana generuje kod o zmiennej długości i całkowitej liczbie bitów. Sygnały ważne transponują się na krótkie kody, te mniej znaczące na dłuższe.

Ponieważ kody Huffmana mają specyficzny nagłówek, dekodują się doskonale, pomimo swojej zmiennej długości. Dekodowanie jest bardzo szybkie, ponieważ



Rys. 1. Każda warstwa MPEG-1 wykorzystuje bank 32 filtrów. To, czy sygnał podlega maskowaniu, czy nie, określa się po jego skwantowaniu.

można korzystać z tablic. Technika ta daje oszczędność miejsca rzędu mniej więcej 20 procent.

Technika Huffmana jest idealnym uzupełnieniem kompresji zależnej od percepcji. W pasażach zawierających wiele częstotliwości jednocześnie, kodowanie zależne od percepcji zapewnia znaczną redukcję poprzez wyeliminowanie

sygnałów zamaskowanych. Ponieważ wówczas pojawia się kilka identycznych sygnałów, to kodowanie Huffmana ma niewielki skutek.

W trakcie pasażu z kilkoma różnymi dźwiękami występuje niewiele efektów maskowania. Dzieje się tak wtedy, gdy kodowanie Huffmana oszczędza znacznie wię-

cej miejsca, ponieważ jest to informacja ze znaczną redundancją (nadmiarowa). A więc takie pasáže mogą być reprezentowane przez krótkie kody.

EE

Artykuł publikujemy na podstawie umowy z redakcją miesięcznika "Elektor Electronics".