



**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**

**WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI
i ELEKTRONIKI
KATEDRA TELEKOMUNIKACJI**

**Przegląd wybranych standardów kodowania
dźwięku i obrazu**

KKRR'98

Piotr Pacyna



Poznan, 1998 r.

1. Wstęp

Ostatnie dziesięciolecie to okres w którym następował dynamiczny rozwój sieci lokalnych LAN, ewoluowały systemy centralowe sieci PSTN, upowszechniały się sieci z integracją usług N-ISDN a potem B-ISDN. Sieci te otworzyły nowe możliwości w zakresie komunikacji interpersonalnej i umożliwiły realizację nowych usług telekomunikacyjnych integrujących różne formy przekazu informacji w postaci cyfrowej: przekaz foniczny, przekaz wizyjny oraz przekaz danych numerycznych.

Oferując nowe możliwości stwarzały również nowe problemy, które musiały być rozwiązane. Rozwiązania dotyczące zasad funkcjonowania systemów audiowizualnych w tych środowiskach sieciowych i zasad współpracy tych systemów osiągnięto na drodze standaryzacji. Wymagało to osiągnięcia kompromisów pomiędzy interesami wielu sił rynkowych i uwzględnienia aspektów techno-ekonomicznych proponowanych rozwiązań.

Producenci systemów telekomunikacyjnych oraz ich użytkownicy oczekiwali, by wprowadzane standardy umożliwiały rozwój rynku w sposób uporządkowany, oraz nie stanowiły czynnika hamującego. Dlatego standardy są ogólnikowe w tych kwestiach, gdzie duży stopień dowolności jest dopuszczalny. Na przykład standardy nie narzucają sposobu implementacji systemów. Wskazówki dotyczące implementacji mają na ogół charakter ilustracyjny a nie normatywny. Z drugiej jednak strony standardy są bardzo szczegółowe w kwestiach zasadniczych oraz tych, w których swoboda implementacyjna lub swoboda interpretacyjna mogłaby być przyczyną niezgodności sprzętu. Przykładem tego jest zawsze bardzo precyzyjnie określona składnia syntaktyczna i semantyka protokołów komunikacyjnych, zarówno przenoszących informacje użytkowe jak sterujących pracą systemu.

Równie ważne oczekiwania związane były z ochroną inwestycji. Ich cechą charakterystyczną są zawsze wysokie nakłady i długi czas zwrotu kosztów sięgający kilkunastu lat. Aby umożliwić migrację systemów w kierunku rozwiązań nowocześniejszych konieczne było zapewnienie współpracy z systemami starszymi, zbudowanymi według historycznie wcześniejszych standardów. Dlatego kolejne standardy promują podejście ewolucyjne. Nowe rozwiązania, jeśli się dało, wprowadzano jako rozszerzenia standardów obowiązujących. Zachowanie wstecznej zgodności kolejnych wersji standardu było możliwe dzięki elastycznej strukturze składniowej protokołów (*extension fields*). Tam jednak gdzie postęp technologiczny i wiedza otworzyły całkiem nowe możliwości wprowadzano nowe standardy, które wypełniały powstającą lukę. Takie podejście zawsze powoduje, że nowsze systemy są łatwiej akceptowane i łatwiej osiągnięta jest na rynku pewna masa krytyczna, po przekroczeniu której nowe technologie zostają uznane za ugruntowane.

Decydujący głos w dziedzinie standaryzacji systemów audiowizualnych na przestrzeni ostatnich lat miała ITU-T (*International Telecommunication Union - Telecommunication Standardisation Sector*, dawniej CCITT), choć niektóre ważne standardy zostały opracowane przez inne komitety i organizacje jak *International Organization for Standardisation* (ISO) i *ATM Forum* a potem, po pewnych korektach merytorycznych i związanych z terminologią przyjęte przez ITU.

2. Standardy kodowania obrazu nieruchomego

W ostatnich latach opracowano lub udoskonalono szereg algorytmów kompresji dostosowanych do obrazów statycznych (kolorowych i czarno-białych), obrazów silnie kontrastowych i o płynnych przejściach pomiędzy barwami. Poniżej omówiono te najważniejsze.

2.1 Standard JPEG

JPEG jest standardem kompresji obrazów nieruchomych kolorowych i czarno-białych, opracowanym przez *Joint Photographic Experts Group*. Dobrze spełnia swe zadania w przypadku obrazów fotograficznych o płynnych przejściach między barwami. Osiąga natomiast gorsze wyniki w przypadku kodowania prostych rysunków o ostrych konturach, takich jak na przykład kadry pochodzące z filmów rysunkowych.

JPEG implementuje zarówno techniki stratne jak i bezstratne, choć o jego sile stanowią głównie te pierwsze. Podczas kodowania wykorzystuje naturalne ograniczenia zmysłu wzroku człowieka w zakresie postrzegania, a szczególnie to, że zmiany koloru są dostrzegane słabiej niż zmiany jasności obrazu. JPEG przechowuje informacje o pełnym kolorze co oznacza, że kolor każdego elementu obrazu (pixel'a) jest przed zakodowaniem reprezentowany na 24 bitach. Kompresja JPEG jest kilkuetapowa. W algorytmie stosowane są: dyskretna transformata kosinusowa, kwantyzacja współczynników transformaty, kodowanie długości ciągów (*run-length encoding - RLE*) i kodowanie Huffmana.

Standard definiuje cztery tryby pracy. W trybie **kodowania sekwencyjnego** składniki obrazu (luminancja i obie składowe chrominancje) kodowane są za pomocą prostych skaningu od lewej do prawej i z góry na dół.

Podczas **kodowania progresywnego** obraz jest kodowany w kilku przebiegach tak, że kolejne przebiegi wnoszą poprawę rozdzielczości obrazu wskutek dodatkowych bitów na których reprezentowany jest każdy element obrazu. Tryb ten jest przeznaczony dla aplikacji prowadzących transmisje po łączach o małej przepustowości.

W trybie **kodowania bezstratnego** obraz jest kodowany tak, aby po zdekodowaniu możliwe było wierne odtworzenie oryginału. Pixeli kodowane są jako suma lub różnica lub kombinacja sum i różnic wartości innych (sasiadujących z nim) pixeli.

Podczas **kodowania hierarchicznego** obraz jest kodowany w wielu rozdzielczościach tak, że niższa rozdzielczość może być dostępna bez potrzeby dekompresji w pełnej rozdzielczości. Osiąga się to poprzez zastosowanie różnych tablic kwantyzacji do zakodowania tego samego obrazu.

JPEG typowo osiąga w kompresji stratnej współczynnik kompresji pomiędzy 10:1 a 20:1 przy którym stratność jest praktycznie niedostrzegalna. Kompresja 30:1 do 50:1 jest możliwa z małą lub średnią degradacją jakości. Wada algorytmów JPEG w trybie kompresji stratnej jest to, że przy kilkukrotnej kompresji i dekompresji tego samego obrazu, za każdym razem następuje pogorszenie jakości.

3. Standardy kodowania obrazu ruchomego

3.1 Standard kodowania H.261

H.261 jest najbardziej rozpowszechnionym i ugruntowanym standardem kodowania sygnału wideo zatwierdzonym przez ITU w 1990. Ponieważ był projektowany z myślą o

sieciach N-ISDN definiuje kodowanie obrazu ruchomego dla predkosci bedacymi wielkrotnosciami 64 Kbit/s. Dlatego okresla sie go terminem 'px64', gdzie p jest parametrem przyjmujacym wartosci od 1 do 30. Odpowiada to predkosciom od 64 do 1920 Kbit/s.

Koder obsluguje wytlaczenie obrazy bez przeplotu o czestotliwosci powtarzania od 10 do 30 obrazów na sekunde. Moga byc stosowane dwa formaty obrazu. Dla potrzeb wideotelefonii ($p=1,2$) opracowano format QCIF (*Quarter Common Intermediate Format*). Format CIF (*Common Intermediate Format*) jest uzywany dla wideokonferencji i wymaga wiekszych szybkosci transmisji danych ($p \geq 6$). Obsluga formatu QCIF jest obowiazkowa, natomiast formatu CIF jest opcjonalna - patrz Tabela 1.

P*64 jest koderem hybrydowym bazujacym na dyskretnej transformacie kosinusowej, kwantyzacji obrazu i kodowaniu RLE oraz metoda Huffmana. Wyodrebnia z obrazu skladowa luminancji oraz dwie skladowe chrominancji (Y, Cb, Cr). Skladowe różnicowe koloru sa reprezentowane z rozdzielczoscia zmniejszona dwukrotnie w kierunku poziomym i dwukrotnie w kierunku pionowym. Współczynnik kwantyzacji jest zmienny, tak by zachowac stala szybkość bitowa na wyjściu kodera.

Koder moze pracowac w trybie INTRA, w którym kazdy blok obrazu (8x8 pixeli) kodowany jest niezaleznie, lub INTER gdzie ramki moga byc kodowane różnicowo wzgledem poprzedniej ramki. W tym drugim przypadku blad predykcji liczony jest pomiedzy odpowiadajacymi sobie makroblokami biezacej i poprzedniej ramki, chyba ze stosowana jest opcjonalna kompensacja ruchu. Gdy jest stosowana, obszar poszukiwania najlepszego dopasowania wynosi $-15 \div +15$ w poziomie i tyle samo w pionie.

Podczas gdy w MPEG-1 opóźnienie kodowania i dekodowania wynosi od 0.5 do 1 sekundy, tutaj wynosi 150 milisekund. Jest to takie opóźnienie, przy którym rozmowy za posrednictwem wideotelefonu i wideokonferencji nie traca jeszcze dialogowego charakteru.

Stabilnosc standardu, przeplywnosci bitowe i rozdzielczosc przestrzenna obrazu powoduja, ze jest powszechnie stosowany w systemach wideokonferencyjnych.

3.2 H.263

Standard H.263 jest rezultatem osiagnietego w ostatnim czasie postepu w zakresie kodowania i kompresji sygnalów wizyjnych [2]. Podstawowa konfiguracja kodera H.263 jest oparta na standardzie H.261. H.263 pierwotnie ukierunkowany byl na aplikacje wideofoniczne dla bardzo niskich przeplywnosci bitowych, za które uwaza sie przeplywnosci ponizej 64 Kbit/s. W koncowej postaci obejmuje znacznie wiekszy zakres przeplywnosci, umozliwiajac kodowanie obrazów w pieciu trybach (sub-QCIF, QCIF, CIF, 4CIF, 16CIF) z rozdzielczosciami w kanale luminancji od 128x96 pixeli (sub-QuarterCIF) do 1408x1152 pixeli (4CIF) - Tabela 1.

Format	Lumin. pixeli	Lumin. linii	H.261	H.263
SQCIF	128	96	Nie	Tak
QCIF	176	144	Tak	Tak
CIF	352	288	Opcja	Opcja
4CIF	704	576	Nie	Opcja
16CIF	1408	1152	Nie	Opcja

Tabela 1. Tryby pracy koderów H.261 i H.263.

Oprócz podstawowego trybu kodowania dostępne są opcje nieograniczonych wektorów przemieszczenia (*unrestricted motion vectors*), kodowanie arytmetyczne zorientowane składniowo (*syntax-based arithmetic coding*), zaawansowana predykcja ruchu (*advanced prediction*), oraz interpolacja dwukierunkowa (*PB mode*). Następnym tych nowości jest większa odporność na błędy. Sposób w jaki prowadzona jest kompensacja ruchu (*motion compensation*) powoduje wzrost wydajności kodowania.

3.3 MPEG-1

Rodzina standardów MPEG to aktualnie dwa standardy: MPEG1 i MPEG2. Kryterium podziału stanowią przepływności bitowe, a więc zastosowania, dla których zostały one opracowane. Struktura dokumentów opisujących oba standardy jest wspólna. Tworzą one cztery wzajemnie uzupełniające się części [6]:

część 1: *System* – określa synchronizację i multipleksację składowych audio i wideo,

część 2: *Video* – definiuje kompresję sygnału wideo,

część 3: *Audio* – definiuje kompresję sygnału audio,

część 4: *Compliance testing* – opisuje procedury badania strumienia MPEG, proces dekodowania i testowania zgodności strumienia z prawidłami zawartymi w częściach 1-3.

MPEG-1 (ISO 11172-2) jest algorytmem hybrydowym opartym na metodach stosowanych przy kodowaniu obrazów statycznych (JPEG), kompresji poprzez szacowanie ruchu oraz interpolacji dwukierunkowej. MPEG-1 różni się trzy rodzaje ramek obrazu (tzw. ramek): ramki I (*intraframes*), ramki P (*predicted frames*) i ramki B (*bidirectional frames*). Ramki I kompresowane są całkowicie niezależnie od innych, dzięki czemu możliwe jest odtworzenie zawartego w nich obrazu nawet w przypadku uszkodzenia ramek sąsiednich. Kosztem tego jest stosunkowo duży ich rozmiar i duży udział w całkowitej objętości strumienia wyjściowego. Obrazy predykcyjne mogą mieć mniejszy rozmiar dzięki temu, że zamiast kodowania zawartości ramki, kodowane są referencje do identycznych lub podobnych elementów obrazu zawartych w poprzednich lub kolejnych ramkach. Ramki typu B, to takie, które są zależne zarówno od ramek poprzednich jak i kolejnych. Tutaj kodowany jest obraz różnicowy powstały z odjęcia obrazu danej ramki oraz obrazu powstałego jako interpolacja obrazów ramki poprzedzającej i następującej po aktualnie kodowanej.

MPEG-1 jest zoptymalizowany dla strumienia rzędu 1,5 Mbit/s na który składa się: 1,15 Mbit/s video + 256 Kbit/s audio + informacja synchronizująca.

3.4 MPEG-2

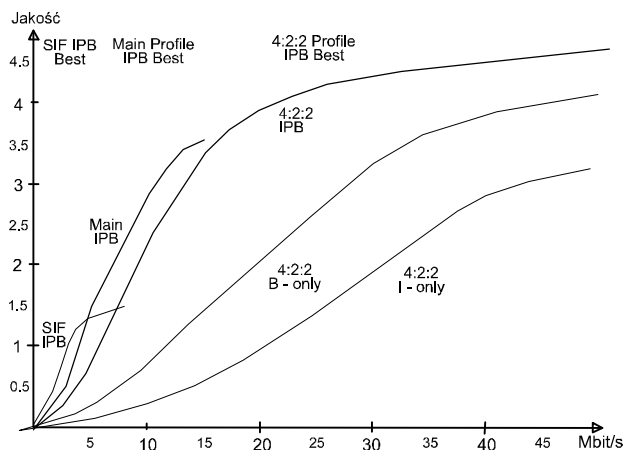
Standard MPEG-2 był opracowywany dla kompresji sygnału audio (ISO-13818-3, H.222) i wideo (ISO 13818-2; H.262) w zastosowaniach profesjonalnych i półprofesjonalnych. Kodek MPEG-2 charakteryzuje się następującymi właściwościami:

- ?? Jest stosowany do kompresji obrazów o różnych proporcjach, takich jak 4x3, 16x9 i inne
- ?? Dopuszcza ramki o rozmiarze 16 K x 16 K.
- ?? Posiada możliwość rejestracji 5 kanałów dźwiękowych i kilku ścieżek zawierających napisy w różnych wersjach językowych.
- ?? Obsługuje sygnał zarejestrowany w trybie progresywnym lub z przeplotem.
- ?? Rejestruje szczegółowo charakterystyki kompresowanego obrazu.
- ?? Definiuje 2 formaty strumieni dla mediów transmisyjnych o różnych właściwościach: *Transport Streams* i *Program Streams*.

		Profil				
		Simple	Main	SNR spatial	High	profil 4:2:2
Level	Typ ramki	I, P	I, P, B	I, P, B	I, P, B	I, P, B
	Rozdzielczość chrominancji	4:2:0	4:2:0	4:2:0	4:2:2	4:2:2
High	Max. strumień		80 Mb/s		100 Mb/s	
	kolumn/wierszy		1920/1152		1920/1152	
	ramek/sek.		60		60	
High1440	Max. strumień		60 Mbit/s	60 Mbit/s	80	
	kolumn/wierszy		1440/1152	1440/1152	1440/1152	
	ramek/sek.		60	60	60	
Main	Max. strumień	15 Mbit/s	15 Mbit/s	15 Mbit/s	20 Mbit/s	50 Mbit/s
	kolumn/wierszy	720/576	720/576	720/576	720/576	720/608
	ramek/sek.	30	30	30	30	30
Low	Max. strumień		4 Mbit/s	4 Mbit/s		
	kolumn/wierszy		352/288	352/288		
	ramek/sek.		30	30		

Tabela 2. Profile i poziomy w standardzie MPEG2 oraz odpowiadające im parametry kodowania sygnału wizyjnego.

Aby ułatwić stosowanie standardu wprowadzono pięć tak zwanych profili (*profile*), które w przybliżeniu wyznaczają obszary zastosowań - Tabela 2 [6,8,9]. Każdy profil wykorzystuje podzbiór pełnej składni strumienia bitowego MPEG2, co umożliwia budowę koderów i dekoderów z przeznaczeniem do konkretnych zastosowań, bez potrzeby implementowania całego standardu. Dla poszczególnych profili wprowadzono tak zwane poziomy (*level*). Określają one dopuszczalne szybkości bitowe strumienia MPEG2, nakładając tym samym wymagania na moc obliczeniową koodera. Pozwala to różnicować produkty pod względem zaawansowania technologicznego a więc i ceny.



Rys. 1. Osiągana jakość wideo w standardzie MPEG2 w funkcji wielkości strumienia bitowego.

W zależności od stosowanego profilu i poziomu, uzyskuje się strumienie o różnej jakości i różnych szybkościach bitowych. Związek pomiędzy szybkością bitową strumienia a postrzegana jakość sygnału wizyjnego, wyrażony według skali *mean opinion score* przedstawiony jest na Rys. 1 [8].

3.5 MPEG4

Przystąpienie do standaryzacji nowych metod przetwarzania cyfrowej informacji akustycznej i wizyjnej motywowane było pojawieniem się całkowicie nowych i bardzo obiecujących technik.

Istota i zarazem wspólna cecha wcześniejszych standardów (MPEG1, MPEG2, H.261 oraz H.263) jest rastrowa reprezentacja obrazu. Jej następstwem jest kilkietapowe przetwarzanie obejmujące separacje składowych chrominancji i luminancji, transformacje do dziedziny częstotliwości (*Discrete Cosine Transform* - DCT), decymacje składowych przy wykorzystaniu odpowiednich tablic kwantyzacji i wydajne kodowanie. Działania te prowadzi do ograniczenia przestrzennej nadmiarowości informacyjnej. Drugim charakterystycznym elementem, wykorzystywanym w różnym stopniu przez różne standardy jest kompensacja ruchu. Kompensacja ogranicza nadmiarowość informacyjną na skali czasu (*temporal redundancy*), narzucając przy tym blokową interpretację obrazu.

Standard MPEG4 podobnie jak wcześniej wymienione standardy, określa zasady reprezentacji dźwięku, obrazu ruchomego i strumieni danych, oraz rozszerza je o nowe elementy jak np. grafiki rastrowe i wektorowe [3, 4]. Jego odmiennosc wynika z obiektowego podejścia do kodowanego materiału. Elementy składowe przekazu reprezentowane są w postaci tzw. obiektów audiowizualnych (*primitive audio-visual objects* - AVO). Obiekty mogą być pochodzenia naturalnego tzn. być zarejestrowane przy pomocy kamery wideo lub mikrofonu; syntetyzowane lub mogą pochodzić z przetworzenia innych obiektów przy pomocy dostępnych technik obróbki sygnałów wizyjnych i fonicznych.

MPEG4 standaryzuje zasady reprezentacji całej grupy obiektów o różnym pochodzeniu i różnych właściwościach oraz metody ich dekodowania. Do standardowych obiektów zalicza się na przykład mówiące i poruszające się postacie, dwu- i trójwymiarowe reprezentacje twarzy mówiącej osoby, głos, tło, tekst, grafiki a także obiekty statyczne (np. dwu- lub trójwymiarowe biurko) i inne [4]. Każdy obiekt jest kodowany niezależnie od otaczającego go kontekstu, co umożliwia jego wyizolowanie i powtórne wykorzystanie. Obiekty elementarne mogą być łączone w celu wytworzenia obiektów złożonych, które po odpowiednim skomponowaniu tworzą sceny audiowizualne. Budowanie scen jest możliwe

ponieważ standard zapewnia metody synchronizacji, dzięki czemu zachowane są ich wzajemne relacje przestrzenne i czasowe obiektów. MPEG4 umożliwia skomponowanie sceny poprzez:

- 1) rozmieszczenie pewnej liczby obiektów audiowizualnych w trójwymiarowej przestrzeni wirtualnej stanowiącej scenę,
- 2) ustalenie porządku prezentacji obiektów i ich wzajemnej synchronizacji,
- 3) projekcje sceny na dwuwymiarową 'płaszczyznę widzenia' odpowiednio do perspektywy odbiorcy, a także zmianę perspektywy (lub punktu odsłuchu) w trakcie dekodowania i prezentacji informacji.

Efekt ten uzyskuje się poprzez wykorzystanie języka MSDL (*MPEG4 Syntactic Description Language*). Jest to język opisu scen audiowizualnych oparty na algorytmach dekodowania obiektów oraz algorytmach komponowania obiektów w sceny.

Istotną, unikalną cechą tego standardu jest interaktywność, która otwiera możliwość wpływania odbiorcy na przebieg prezentacji. Interaktywność standardu zapewnia architektura *Delivery Multimedia Integration Framework (DMIF)*, która rozwiązuje zagadnienia współpracy interaktywnych aplikacji multimedialnych z różnymi mediami. Architektura gwarantuje, że aplikacje, które z niej korzystają, są ekranowane od specyficznych właściwości medium, którym może być sieć podkładowa, trwałe nośniki informacji, eter lub inne. Ze względu na różnorodność informacji przekazywanej w strumieniu MPEG4 architektura DMIF jest zorientowana na zapewnianie różnych parametrów jakościowych (*quality of service*), stosownie do wymagań wynikających z rodzaju informacji.

3.5.1 Reprezentacja obiektów wizualnych

Składowanie, transmisja i manipulacja obiektami audiowizualnymi, statycznymi i ruchomymi realizowana jest przez obszerną kolekcję narzędzi. Manipulowanie obiektami AV wykracza daleko poza wkomponowywanie obiektów w sceny, synchronizację lub interaktywność. Obejmuje również reprezentację obiektów dwu- i trójwymiarowych w postaci siatek, reprezentację tekstury tych obiektów a także zmieniającej się w czasie geometrii, skalowalność przestrzenną/czasową i jakościową obiektów i tekstur oraz dostęp swobodny do tych obiektów, tj. odtwarzanie wprzód/wstecz, normalne lub przyspieszone, z podglądem lub bez.

Wobec dużych na ogół wymagań obrazu wideo w zakresie przepustowości skoncentrowano się na parametrycznym opisie obiektów wideo, w celu późniejszego syntetyzowania ich w terminalu odbiorczym. Najbardziej zaawansowaną metodą dotyczy parametryzacji wyglądu i mimiki ludzkiej twarzy oraz kształtu i dynamiki postaci.

W przypadku twarzy kształt, powierzchnia twarzy, jej kolor, odcień i mimika są wyrażane za pomocą komend języka *Facial Definition Parameter/Facial Animation Parameter*. Komendy te tworzą strumień informacji przekazywany do terminala odbiorczego, w którym po uwzględnieniu odpowiednich współczynników wzmocnienia są wykorzystywane do przekształcenia „standardowej” twarzy w twarz odpowiadającą wyglądkowi mówcy. Zestaw komend *Lip Shape Patterns* umożliwia nadanie odpowiedniego kształtu ustom w zależności od wypowiedzianych głosek i wersji językowej, którą posługuje się mówca. Komendy te mogą też być stosowane by dostosować kształt ust do wersji językowej, jeżeli realizowany jest dubbing.

Animacja ciała polega na takim przetworzeniu wirtualnego, trójwymiarowej „standardowej” postaci, by odpowiadała fizjonomia postaci rzeczywistej. Do opisu postaci ludzkiej wykorzystuje się komendy *Body Parameter Set (BPS)* oraz *Body Animation Parameter (BAP)* [4].

4. Kodowanie sygnału audio

W ogólności algorytmy kompresji audio tworzą dwie grupy. Pierwsza grupa to tzw. kodery falowe (*waveform coders*), druga to kodery źródłowe (*source coders*) [11].

Kodery falowe podejmują próbę transmisji oryginalnego sygnału akustycznego z możliwie małym zniekształceniem, zależnie od dostępnego pasma przepustowego. Dostarczają sygnał o dobrej jakości i o małych opóźnieniach. Metody takie jak PCM lub ADPCM należą do podstawowych. Kodery źródła analizują sygnał mowy dla określenia ich charakterystycznych parametrów: proporcji mowy i ciszy, barwy głosu (*pitch*), głośności, i innych, a następnie transmitują parametry do dekodera, gdzie są wykorzystywane dla cyfrowego odtworzenia sygnału. Ich zaleta, w porównaniu do koderów falowych są małe wymagania wobec pasma przepustowego.

Najczęściej używane algorytmy kompresji głosu wymienione są poniżej.

?? PCM (*Pulse-Code Modulation*) - modulacja podczas której sygnał analogowy jest próbkowany z częstotliwością 8 KHz tworząc cyfrowy strumień danych, w którym każda próbka reprezentowana jest za pomocą 8 bitów. Daje to strumień o wielkości 64 Kbit/s. W podstawowej metodzie kodowania PCM, każda próbka reprezentowana jest przez słowo kodowe z równomiernie rozmieszczonymi poziomami kwantyzacji. Poprzez zastosowanie odpowiedniej transformacji stopień kwantyzatora (*quantizer step*) może być logarytmiczny, co otwiera możliwość rejestrowania większego zakresu wartości przy tej samej liczbie bitów słowa kodowego. Istnieją zasadniczo dwie metody transformacji: μ -law oraz A-law. Obie stosują 8 bitów na próbkę dla reprezentacji takiego przedziału wartości, który byłby osiągalny przy 14 bitach na próbkę w podstawowym PCM. Odpowiada to kompresji 1.75:1, zakładając przepustowość 64 Kbit/s na wyjściu kodera A-law lub μ -law. Ze względu na logarytmiczną naturę transformacji, próbki sygnału o małej amplitudzie sygnału są kodowane z większą precyzją niż próbki o dużej amplitudzie. Standard μ -law stosowany jest powszechnie w telefonii cyfrowej ISDN w Stanach Zjednoczonych i Japonii. Standard A-law PCM stosowany jest w sieciach ISDN w Europie.

?

?? ADPCM (*Adaptive Differential PCM*). W podstawowej metodzie kodowania PCM próbka dźwięku rejestrowana jest niezależnie od sąsiednich próbek. Ponieważ próbki te mają zwykle podobne wartości, ich wartości można przewidzieć z pewną dokładnością na podstawie wartości próbki poprzedniej. Metoda ADPCM wykonuje taką predykcję i koduje jedynie różnice wartości rzeczywistej zarejestrowanej przez koder i wartości przewidzianej. Dla kodowania różnicy zwykle przeznaczają się cztery bity. Kodery ADPCM potrafią adaptować się do charakterystyki sygnału poprzez zmianę współczynnika kwantyzacji lub parametrów bloku prognozującego wartość próbki. Przewagę tego algorytmu w stosunku do algorytmów μ -law lub A-law wyraża współczynnik kompresji, który w typowych przypadkach osiąga wartość 2:1.

ITU-T wystandardyzowało wiele metod ADPCM. Na przykład standard G.722 definiuje subpasmowe adaptacyjne kodowanie różnicowe (*Sub-Band ADPCM*, SB-ADPCM),

które polega na podziale pasma na dwa subpasma: górne i dolne. W każdym subpasmie sygnały kodowane są z wykorzystaniem ADPCM. G.722 pracuje w trzech trybach: 48, 56 i 64 Kbit/s.

Standardy ADPCM są powszechnie stosowane w systemach wideokonferencyjnych.

?? ATC (*Adaptive Transform Coding*). Algorytm ATC jest kombinacją TDHS (*Time Domain Harmonic Scaling*), LPC (*Linear Predictive Coding*) i VQ (*Vector Quantization*). Idea VQ polega na tym, że jedno słowo kodowe reprezentuje kilka próbek. Jest to więc kompresja stratna, stosowana w tych strumieniach w których kolejne próbki są ze sobą skorelowane. Cechy ATC: mała złożoność i zmienna szybkość przetwarzania. Wymagane pasmo: 8-16 Kbit/s

Oprócz uniwersalnych standardów kodowania audio istnieje pewna grupa standardów opracowanych wyłącznie dla kompresji sygnału mowy. Wykorzystując modele symulujące trakt wokalny człowieka potrafią osiągać szczególnie dobre rezultaty.

?? LPC (*Linear Predictive Coding*) dopasowuje sygnał mowy do prostego, analitycznego modelu traktu głosowego. Parametry najlepszego dopasowania są transmitowane i wykorzystywane przez dekodery dla wygenerowania syntetyzowanego sygnału mowy, zbliżonego do oryginału. Standardem, który w Stanach Zjednoczonych wykorzystuje proste kodowanie LPC jest U.S. Federal Standard 1015, który wymaga pasma 2,4 Kbit/s. Również GSM wykorzystuje odmianę LPC, nazywaną RPE-LPC (*Regular Pulse Excited - Linear Predictive Coder with a Long Term Predictor Loop*). Zakładając próbkowanie 8KHz wymagane jest pasmo 13 Kbit/s.

?

?? CELP (*Code Excited Linear Prediction*) przeprowadza podobne modelowanie sygnału mowy jak kodery LPC. Ponadto oblicza różnice pomiędzy wartościami zarejestrowanymi a modelowymi i transmituje je wraz z parametrami modelu. Transmitowane są nie tyle wartości różnic, lecz indeksy do tablicy różnic znanej zarówno koderowi jak dekodery (*common code book*). Dodatkowe informacje pozwalają dalej ograniczyć wielkość strumienia na wyjściu z kodera. W Stanach Zjednoczonych algorytm CELP jest stosowany w zaleceniu U.S. Federal Standard 1016 i wymaga 4,8 Kbit/s. Standard ITU-T G.728, jest implementacją odmiany algorytmu CELP nazywanej LD-CELP (*Low Delay CELP*). G.728 charakteryzuje się dużą złożonością obliczeniową i wymaga pasma 16 Kbit/s. Zatwierdzony w 1996 roku standard G.729 wykorzystuje zmodyfikowaną wersję algorytmu CELP (CS-ACELP) i określa kodowanie sygnału mowy przy przepływności 8 Kbit/s. Trwają prace nad rozszerzeniami tego algorytmu dla przepływności 6,4 i 12 Kbit/s.

Szczegółowe porównanie charakterystyk koderów jest trudne, ze względu na zasadniczo różne podejścia do sposobu kodowania jak również ich specyficzne funkcje (tzw. opcje), które wzbogacają funkcjonalność kodera i przynoszą dalszą poprawę jego parametrów. Dlatego w (Tabela 6, źródło: [14]) podano jedynie ich podstawowe cechy. Z tabeli tej wynika, że wspólny cel wszystkich metod kodowania jakim jest zapewnienie zadowalającej jakości głosu przy minimalnych kosztach osiągany jest głównie poprzez implementacje układów kodujących o coraz większej mocy obliczeniowej, które potrafią sprostać

algorytmom takim jak ACELP. Implementacje tego algorytmu dowodza, ze glos moze byc kompresowany do tak niskiej przeplywnosci jak 6 Kbit/s i zachowywac dosyc dobra jakosc.

Koder	PCM	MP-MLQ /ACELP	AD-PCM	LPC	RPE-LTP	LD-CELP	CS-ACELP
Standard	G.711	G.723.1	G.726	DOD*FS 1015	ETSI-GSM	G.728	G.729
Typ	analiza czasowa	analiza-synteza	analiza czasowa	parametryczny	analiza-synteza	analiza-synteza	analiza-synteza
Pasma	64 Kbit/s	5.3/6.3 Kbit/s	32 Kbit/s	2,4 Kbit/s	13 Kbit/s	16 Kbit/s	8 Kbit/s
Opóźnienie	0,125 ms	100 ms	0,3 ms	50 ms	50 ms	3 ms	30-40 ms
Złożoność obliczeniowa	0,1 MIPS	18/14 MIPS	12 MIPS	7 MIPS	2,5 MIPS	33 MIPS	20 MIPS
MOS	4,2-4,4	4	4,0-4,1	2,3	3,6	4	4

Tabela 3. Porównanie koderów audio.

4.1.1 Metody oceny jakości dźwięku

Miara jakości dźwięku jest Jednostka Zniekształcenia Kwantyzacji (*Quantising Distortion Unit*, QDU) zdefiniowana w zaleceniu G.113 [11]. Zgodnie z nim, Jednostka Zniekształcenia to ilość zniekształcenia lub utraty jakości równoważna tej, która jest wprowadzana do traktu głosowego przez pojedynczą konwersję analogowo/cyfrową PCM oraz przez konwersję odwrotną. Jeżeli kilka traktów głosowych jest łączone w sposób szeregowy dla uzyskania pojedynczego połączenia, wówczas całkowite zniekształcenie jest sumą zniekształceń wprowadzanych przez kolejne konwersje, na wszystkich odcinkach łącza. Taki sposób oceny zniekształceń nosi nazwę *hopping-u*. Jeżeli na całym odcinku transmisja odbywa się w postaci cyfrowej, wówczas *hopping* nie ma zastosowania, gdyż konwersja dokonywana jest wyłącznie w krańcowych punktach łącza. W istocie na jakość przekazu wpływa znacznie więcej czynników zależnych od rodzaju sprzętu rejestrującego, systemu transmisyjnego i medium. Do nich zalicza się głośność, tłumienie, przesłuch międzykanałowy i jakość łącza. Gdy transmisja odbywa się w trybie cyfrowym, znaczenie niektórych spośród nich maleje, natomiast pojawiają się inne: opóźnienia transmisyjne wprowadzane przez kodek i szum kwantyzacji [11].

Wartość MOS	Jakość
4.0?5.0	wysoka
3.0?4.0	średnia
< 3.0	niska

Tabela 4. Ocena jakości według metody Mean Opinion Score.

ACELP jest według skali MOS oceniany na 4-4.2.

Pomimo iż odsłuch dźwięku jest sprawą bardzo indywidualną, szeroko akceptowaną miarą jakości głosu jest subiektywna ocena MOS (*Mean Opinion Score*) - patrz Tabela 4. Stosując tę skalę ocenia się, że algorytmy PCM i ADPCM oferują wysoką jakość transmisji. MOS wynosi około 4.3 dla PCM i 4.0 dla ADPCM. MOS dla ATC wynosi od 2.0 do 3.8, zależnie od szybkości przetwarzania sygnału głosowego. Dla algorytmu ACELP testy wykazują, że uzyskiwana jakość jest zbliżona do ADPCM 32 Kbit/s.

4.2 MPEG1 audio

4.2.1 Koncepcja

Koncepcja koderza MPEG Audio opiera się na ograniczaniu wielkości strumienia audio poprzez usunięcie z niego tej części sygnału, która jest nieistotna z punktu widzenia słuchacza. W tym celu wykorzystuje się naturalne niedoskonałości narządu słuchu a zwłaszcza tzw. efekt maskowania, który ogranicza zdolność usłyszenia szumu wprowadzanego przez proces kwantyzacji oraz tzw. efekt ograniczonej selektywności częstotliwościowej. Efekt maskowania to zjawisko 'zagłuszania' słabego sygnału (szumu) przez silny sygnał występujący w jego sąsiedztwie. 'Sąsiedztwo' może być dwójakiego rodzaju: sąsiedztwo na skali czasu (*temporal masking*) lub sąsiedztwo na skali częstotliwości (*concurrent masking*) [10].

4.2.2 Cechy charakterystyczne koderza audio

Próbkowanie sygnału audio w koderze MPEG1 odbywa się z częstotliwościami 32 KHz, 44.1 KHz, 48 KHz. Strumień zawierający skompresowany sygnał może mieć jedną z predefiniowanych wielkości, obejmujących przedział od 32 – 224 Kbit/s na każdy kanał. Osiągane współczynniki kompresji wahają się od 2.7 do 24.

4.2.3 Tryby pracy

Strumień MPEG-1 Audio może przenosić 1 lub 2 kanały dźwiękowe, kodowane w jednym z czterech trybów [11]. Są to:

?? **Tryb monofoniczny** z jednym kanałem dźwiękowym.

?? **Tryb podwójny-monofoniczny**, który tworzy dwa niezależne kanały audio, funkcjonalnie równoważne sygnałowi stereo.

?? **Tryb stereo**, w którym bity słowa kodowego są współdzielone przez oba kanały dla lepszego ich wykorzystania.

?? **Tryb joint-stereo**, który wykorzystuje korelację sygnałów słyszanych w obu kanałach.

4.2.4 Ogólna charakterystyka warstw

Koder zdefiniowany w standardzie określa trzy warstwy (poziomy) kompresji.

Layer-1: 32 kbps – 448 kbps, zalecana dla strumieni powyżej 192 Kb/s na kanał.

Layer-2: 32 kbps – 384 kbps, optymalna dla strumieni ok. 128 Kb/s na kanał

Layer-3: 32 kbps – 320 kbps, zaprojektowana dla strumienia 64 Kb/s na kanał

Warstwa	Opóźnienie
Layer I	19 ms (< 50)
Layer II	35 ms (100)
Layer III	59 ms (150)

Tabela 5. Teoretyczne i faktyczne opóźnienia kodowania audio dla różnych warstw.

Podstawowe różnice pomiędzy warstwami I–III są następujące: wzrost złożoności koderza, wzrost całkowitego opóźnienia, wzrost wydajności rozumianej jako jakość dźwięku odniesiona do wielkości strumienia wyjściowego. Różnice te znajdują odbicie w opóźnieniach kodowania - Tabela 5

4.3 MPEG2 audio

Cechy charakterystyczne standardu:

?? Standard obsługujący wiele kanałów: 5 kanałów wysokiej jakości + 1 kanał niskiej częstotliwości (5.1) .

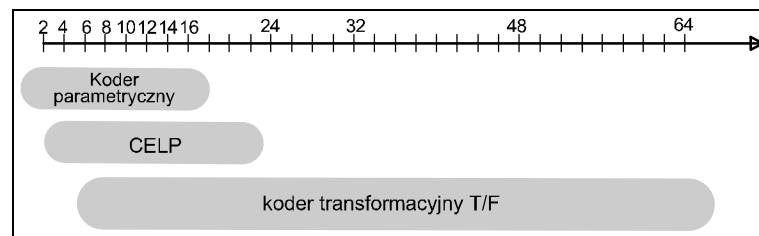
- ?? Wielojęzykowość: max. 7 tzw. *commentary channels*.
- ?? Nowe, niższe wielkości strumienia wyjściowego.
- ?? Nowe, niższe częstotliwości próbkowania. W MPEG1: 32 KHz, 44,1 KHz oraz 48 KHz. W MPEG2 dodatkowo: 16 KHz, 22,05 KHz, 24 KHz.
- ?? Wsteczna zgodność z MPEG1.
- ?

4.4 MPEG4 audio

4.4.1 Dźwięk pochodzenia naturalnego

MPEG4 standaryzuje kodowanie audio z przepływnościami od 2 Kbit/s do 64 Kbit/s. Dla tych przepływności określona jest składnia strumienia bitowego oraz proces dekodowania. Wysoka jakość kodowania jest wynikiem zastosowania trzech różnych koderów. W przedziale od 2 do 4 Kbit/s kodowanie ukierunkowane jest na sygnał mowy. Tu próbkowanie odbywa się z częstotliwością 8 KHz. W przedziale 4-16 Kbit/s próbkowanie odbywa się z częstotliwością 8/16 KHz. W obu tych zakresach wykorzystywane są metody kodowania parametrycznego. Kodowanie mowy w zakresie 6-24 Kbit/s wykorzystuje algorytm CELP. Częstotliwości poniżej 16 Kbit/s wykorzystują metody transformacyjne czasowo-częstotliwościowe (Twin-VQ oraz AAC) - Rys. 2

Ponadto istnieje możliwość stosowania szeregu usprawnień polepszających wydajność kodowania [4]. Standard dopuszcza również stosowanie serii G.7xx oraz MPEG-audio, które mogą być swobodnie wybierane i które są rozpoznawane przez dekodery MPEG4.



Rys. 2. Zakresy pracy koderów audio w standardzie MPEG4.

4.4.2 Syntetyzowana mowa

Możliwości standardu MPEG4 w odniesieniu do sygnałów akustycznych naturalnych i syntetyzowanych określa się terminem *Synthetic-Natural Hybrid Coding audio delivery* (SNHC). SNHC to zestaw narzędzi, które umożliwiają interpretację strumienia strukturalizowanego tekstu (*text-to-speech decoder* - TTS). Strumień taki oprócz treści przekazu zapisanej w postaci tekstowej może zawierać informacje o specyficznych cechach mówcy, w tym wysokości i barwie głosu, wieku, płci, sposobie wymawiania głosek, a także kształtu ust w zależności od wymawianych głosek, jeżeli odtwarzanie obiektu audio ma być połączone z odtwarzaniem obiektu wideo reprezentującego twarz mówcy.

4.4.3 Syntetyzowana muzyka

Dla syntetyzowania muzyki wykorzystywany jest język opisu składu instrumentalnego oraz specyficznych cech poszczególnych instrumentów (*Structured Audio Orchestra Language* - SAOL). MPEG4 nie wskazuje jednej, konkretnej metody syntezy dźwięku, lecz jedynie język służący do definiowania/określania stosowanej metody, np. MIDI. SAOL służy

również do definiowania dźwiękowych efektów specjalnych (efektu przestrzeni, echa, filtrów) oraz wielu innych, w tym efektów mieszanych. Sterowanie synteza realizowane jest przy pomocy odpowiednich „procedur” (*score*) przekazywanych w strumieniu, których podstawowa funkcja jest aktywowanie/wyciszanie instrumentów w trakcie ‘odtworzenia’ muzyki.

5. Środowiska sieciowe

Do roku 1996 w ITU-T opracowano lub zatwierdzono najbardziej kluczowe uregulowania dotyczące budowy systemów komunikacji audiowizualnej, zasad ich implementacji i współpracy. Rekomendacje te stanowią tzw. serie H, która wciąż jest rozszerzana o nowe dokumenty. W rekomendacjach tej serii tych zarysowuje się podział na cztery grupy odpowiadające czterem obszarom zastosowań. Są to publiczne sieci telefoniczne (GSTN), sieci lokalne (LAN), sieci wąskopasmowe (N-ISDN) oraz sieci szerokopasmowe (B-ISDN) - Tabela 6 [1]. Prace standaryzacyjne dotyczące systemu audiowizualnego dla środowiska sieci mobilnych są w toku.

Siec	GSTN	N-ISDN	siec LAN z QOS	siec LAN bez QOS	B-ISDN (ATM)
Szybkość bitowa kanału (max.)	28.8 Kbit/s	1920 Kbit/s	do 6 lub 16 Mbit/s	10 / 100 Mbit/s	600 Mbit/s
Definicja systemu	H.324	H.320	H.322	H.323	H.310
Kodowanie audio	G.723.1	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.723.1 G.728	G.711 G.722 G.728 ISO 11172-3
Kodowanie wideo	H.261 H.263	H.261	H.261	H.261 H.263	H.261 H.262
Transmisja danych	T.120 i in.	T.120 i in.	T.120 i in.	T.120 i in.	T.120 i in.
Kontrola systemowa	H.245	H.242	H.242	H.245	H.242 dla trybu H.321 H.245 dla trybu H.310
Mux. i synchr. Strumieni składowych	H.223	H.221	H.221	H.225.0	H.222.0, H.222.1
Sygnalizacja dla zestaw. Połączenia	standardy narodowe	Q.931	Q.931	Q.931 H.225.0	Q.2931

Tabela 6. Systemy audiowizualne w środowiskach sieciowych GSTN, N-ISDN, LAN i B-ISDN.

Rdzeniem systemu w każdym z wymienionych środowisk jest rekomendacja oznaczona jako ‘total system’. Dla sieci GSTN jest to H.324 natomiast dla N-ISDN H.320. System dla sieci lokalnych bez gwarancji jakościowych (sieci z usługami typu *best effort*) jest zdefiniowany w zaleceniu H.323, a dla sieci LAN z gwarancjami jakościowymi (ISLAN16-T) w H.322. Standardy te definiują architekturę systemów audiowizualnych w tych środowiskach sieciowych oraz wskazują na inne rekomendacje specyfikujące poszczególne elementy składowe systemu. Elementami składowymi są standardy określające procedury nadzoru i zarządzania systemem oraz definiujące zasady reprezentacji różnych mediów (audio, wideo, danych i in.), zasady multipleksacji i synchronizacji tych mediów, i inne.



Rys. 3 Wspólny stos protokołów rekomendacji serii H.

Wspólna cecha systemów audiowizualnych zdefiniowanych w serii H jest architektura systemu z wyraźnie zarysowanymi granicami warstw, w tym warstwy kodowania, multipleksacji i synchronizacji elementarnych strumieni informacyjnych, warstwy adaptacji do sieci, warstwy interfejsu sieciowego oraz warstwy sterowania i zarządzania systemem (Rys. 3).

5.1 Standard H.324

H.324 definiuje terminal multimedialny dla usług wideotelefonicznych o niskiej przepływności bitowej dla sieci GSTN (*Global Standard Telephone Network*). Jest to grupa zaleceń dla kodowania i transmisji sygnałów audio, wideo i strumieni danych przekazywanych poprzez modemy V.34. W standardzie tym zakłada się przepustowość 28,8 Kbit/s. Sygnał audio kodowany jest zgodnie z G.723.1, tzn. z szybkością 5.3 Kbit/s i 6.4 Kbit/s, w związku z czym pozostaje 23,5 lub 22,4 Kbit/s dla wideo i innych sygnałów, których kodowanie zgodne jest z H.261, H.263, T.120 i in.

5.2 Standard H.320

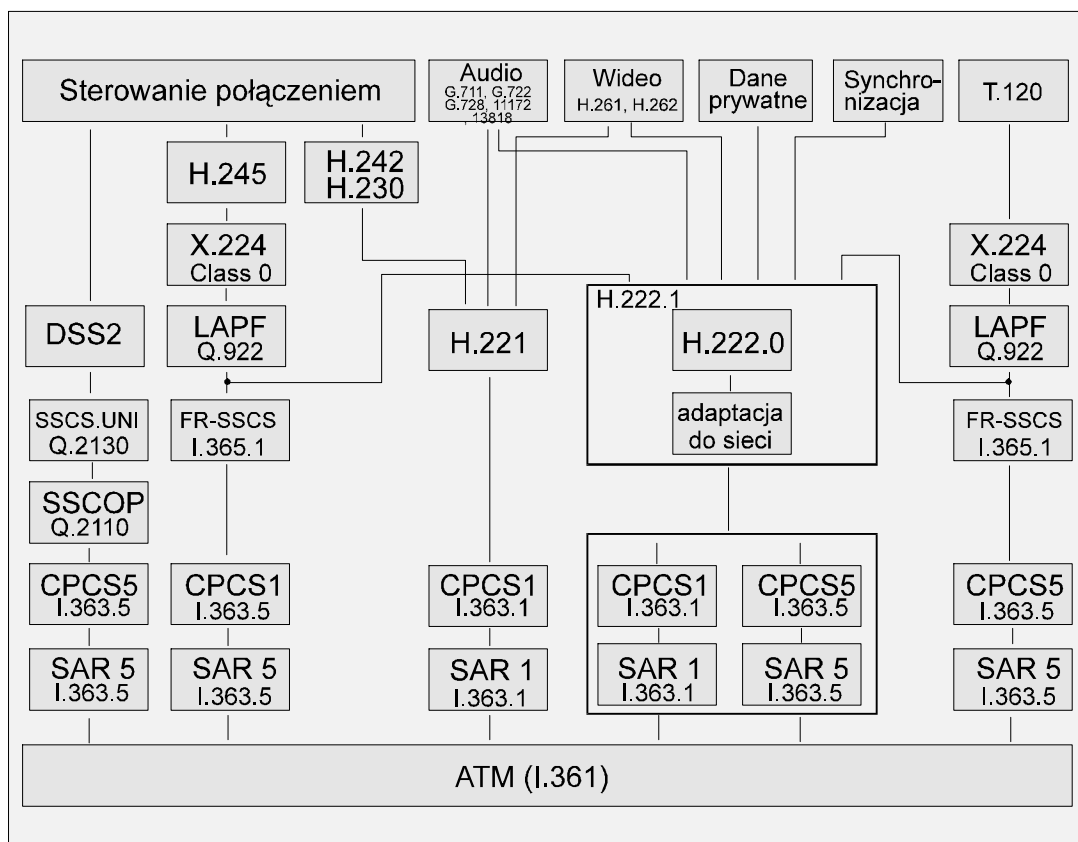
Standard H.320 został opracowany dla budowy systemów audiowizualnych w sieciach ISDN i obejmuje szybkości bitowe od 64 Kbit/s do 1920 Kbit/s. W istocie H.320 jest zestawem rekomendacji. H.261 definiuje strumień wideo jako wielokrotność 64 Kbit/s. G.711, G.722, G.728 określają kodowanie audio.

Inne standardy: H.221, H.230 i H.242 specyfikują negocjacje połączeń, multipleksowanie i ramkowanie danych.

5.3 Systemy audiowizualne w sieciach szerokopasmowych - H.310

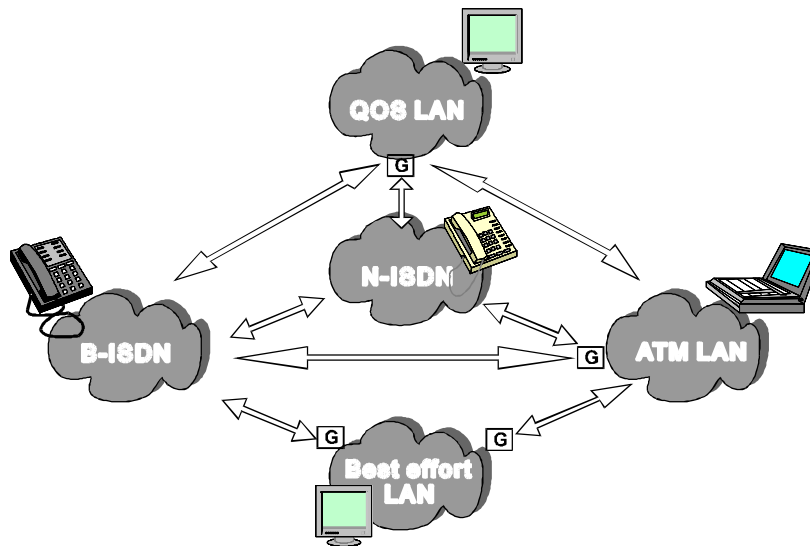
Najnowszym, nie wystandaryzowanym jeszcze w pełni etapem rozwoju sieci teletransmisyjnych zdolnych przenosić sygnały multimedialne są sieci B-ISDN. Ich cechą podstawową jest szerokie pasmo przepustowe. Cecha ta, jakkolwiek niekonieczna do realizacji niektórych usług audiowizualnych, oznacza zwykle możliwość uzyskiwania wyższej jakości usług lub możliwość prowadzenia transmisji z mniejszymi opóźnieniami.

Niewatpliwa zaleta technologii szerokopasmowych jest zdolność integracji usług, która oznacza dostęp do całej gamy usług za pośrednictwem jednego, standardowego terminala poprzez zunifikowane przyłącze. Technologia szerokopasmowa oznacza także możliwość różnicowania wymagań jakościowych QOS (*quality of service*) poprzez wyodrębnienie kilku klas usług, a również takich wymagań przez poszczególne usługi. Jakkolwiek spodziewane korzyści z zastosowania szerokopasmowej infrastruktury sieciowej są bardzo duże, to jednak ich osiągnięcie nie jest jeszcze w pełni możliwe ze względu na obszerną listę nie rozwiązanych problemów. Niemniej kluczowe ustalenia są już poczynione. W szczególności znane są kluczowe protokoły wykorzystywane w systemie H.310 i ich wzajemne powiązania - Rys. 4 [1].



Rys. 4 Współpraca protokołów w systemie H.310.

Istotnym krokiem w kierunku integracji wystandardyzowanych wcześniej systemów dla środowisk PSTN, N-ISDN oraz LAN z technologią B-ISDN było wyznaczenie zasad współpracy tych systemów [1]. Odbyło się to poprzez wskazanie wspólnych elementów i wprowadzenie dodatkowych, umożliwiających budowę systemów w heterogenicznym środowisku sieciowym -(Rys. 5).



Rys. 5 Współpraca systemów audiowizualnych w różnym heterogenicznym środowisku sieciowym.

Jako przykład zapewniania współpracy systemów może posłużyć adaptacja systemu audiowizualnego dla sieci N-ISDN (H.320) do systemu stosowanego w sieciach B-ISDN (H.310), zawarta w rekomendacji H.321. Konsekwencją tej adaptacji jest dostępność dwóch trybów pracy szerokopasmowego terminala audiowizualnego: tzw. trybu *native ATM mode* (tryb podstawowy) albo w trybie *H.321 mode* - trybu zgodności z N-ISDN.

W trybie podstawowym H.310 zdefiniowane jest dziewięć różnych wariantów terminala. W zależności od realizowanych funkcji mogą być budowane i eksploatowane terminale nadawcze (*send only terminal - SOT*), terminale odbiorcze (*receive only terminal - ROT*), lub terminale nadawczo-odbiorcze (*receive and send terminals - RAST*).

Tryb współpracy z H.320 zdefiniowany w H.321 rozwiązuje wiele problemów wynikających z różnic systemów wąsko- i szerokopasmowych, w tym np. dopuszczalne szybkości bitowe. Terminale H.320 (N-ISDN) dopuszczają szybkości bitowe będące wielokrotnością 64 Kbit/s, podczas gdy w sieciach szerokopasmowych dozwolone są praktycznie dowolne. Dlatego terminal szerokopasmowy (H.310) definiuje dwie szybkości podstawowe: 6144 Kbit/s (=96x64 Kbit/s) oraz 9216 Kbit/s (144x64 Kbit/s). Wskazanie takich właśnie szybkości jest nieprzypadkowe. Odpowiadają one strumieniom generowanym przez koder wideo H.262 (MPEG2) pracujący w trybie Main Profile@Main Level (MP@ML), czyli w usługach integrujących sygnał wideo o średniej, lub wysokiej jakości. Inne szybkości bitowe są dla terminala szerokopasmowego opcjonalne. Negocjacja szybkości bitowej pomiędzy komunikującymi się terminalami dokonywana jest za pośrednictwem protokołu H.245, który jest podstawowym protokołem sterowania (*system control protocol*) wspólnym dla H.310/H.323/H.324 i nadającym się do wykorzystania w innych systemach.

5.4 Standard H.245

Standard H.245 definiuje komunikaty i procedury wymiany informacji sterującej pomiędzy terminalami. Tworzą one system sygnalizacji międzyterminalowej dzięki której po zestawieniu połączenia urządzenia końcowe mogą informować się nawzajem i uzgadniać

możliwości kodowania i dekodowania informacji multimedialnej, oraz zarządzać zasobami własnymi i odległego terminala przez cały czas trwania połączenia.

Rekomendacja H.245 składa się z trzech części które definiują składnię i semantykę wiadomości sygnalizacyjnych oraz procedury wymiany.

Bezpośrednio po nawiązaniu połączenia następuje wymiana informacji o możliwościach przetwarzania informacji multimedialnej (*capability exchange*). Terminale wymieniają informacje o rozpoznawanych (akceptowanych) standardach kodowania wideo, co odbywa się poprzez wskazanie obsługiwanego standardu (H.261, H.262 lub H.263). Podobnie wskazywane są możliwości odbioru strumieni audio - sygnalizowane są obsługiwane standardy serii G. Terminal może sygnalizować różne kombinacje swoich możliwości, funkcje dostępne warunkowo lub ewentualne ograniczenia. Terminale mogą także wymieniać informacje o preferencjach dotyczących sposobu nadawania, zestawiać i kasować kanały logiczne służące do przesyłania strumieni audio, wideo i danych.

Kolejna faza - faza uzgadniania protokołów - obejmuje sygnalizację w kanale logicznym, która umożliwi przekazanie do terminala odbiorczego informacji o zasobach niezbędnych do odbioru i dekodowania strumieni. Terminal nadawczy, który zada przydzielenie zasobów w terminalu odbiorczym może oczekiwać na pozytywne potwierdzenie przydziału, odmowę takiej alokacji, a także zadanie zaprzestania transmisji.

H.245 nie jest jedynym protokołem sterującym implementowanym w urządzeniach. Zbliżone funkcje pełni standard DSM-CC (*Digital storage Media - Command and Control*) wprowadzony przez DAVIC (*Digital Audio Video Council*) i wykorzystywany m. in. w platformie DVB (*Digital Video Broadcasting*). Oba protokoły mają wiele cech wspólnych. Istnieje więc duże prawdopodobieństwo, że urządzenia implementujące albo H.245 albo DSM-CC będą ze sobą współpracować [1]. Niemniej jednak przyszły system multimedialnej komunikacji mobilnej (H.324/M) będzie wykorzystywał H.245.

6. Inne standardy

6.1 Standard T.120

Standard T.120 definiuje transport danych w multimedialnych połączeniach wielopunktowych w sieciach POTS, ISDN, LAN, B-ISDN. W szczególności definiuje współdzielenie deski kreslarskiej (*whiteboard*) lub plików.

W istocie T.120 jest zestawem rekomendacji: zalecenia T.122 i T.125 definiują komunikację wielopunktową. T.124 dostarcza mechanizm ustanawiania i zarządzania konferencją. T.127 definiuje wymianę plików, natomiast T.126 przeglądanie i interakcje za pośrednictwem współdzielonych obrazów.

7. PODSUMOWANIE

Prace standaryzacyjne dotyczące metod kodowania i kompresji sygnałów audio i wideo postępują w miarę jak rozwijane są i doskonalone nowe metody przetwarzania tych sygnałów. I tak na przykład wiadomo już dzisiaj, że formalne zatwierdzenie standardu MPEG4 w styczniu 1999 roku nie oznacza jego zamknięcia. Już teraz dyskutowany jest *Proposal Description Package*, który ma w przyszłości udoskonalić i uzupełnić jego możliwości.

Dalszy postęp w zakresie transmisji sygnałów multimedialnych będzie niewątpliwie warunkowany opracowaniem skuteczniejszych metod ograniczania pasma wymaganego

dla sygnału wideo. Do najbardziej obiecujących technik, które mogą ograniczyć pasmo zalicza się:

- ?? kodery skalowalne z hierarchiczną kompensacją ruchu i zwiększona ochrona przed błędami (*multiscale motion compensation, unequal forward error correction*),
- ?? modelowanie przestrzenne z dużą wiernością szczegółów i techniki animacji (*model based video coding*),
- ?? techniki kompensacji ruchu w oparciu o obszary o arbitralnym kształcie (*region-based motion compensation*),
- ?? rozpoznawanie wzajemnego położenia przestrzennych obiektów na podstawie analizy obrazu dwuwymiarowego oraz ruchu kamery rejestrującej obraz (*structure from motion techniques*).

Niezależnie od tego, jakiego rodzaju będą nowe standardy i kiedy się pojawią, należy się spodziewać, że wskutek ugruntowania się tych, które zostały dotychczas zatwierdzone, najbliższe lata przyniosą duże zmiany w zakresie i profilu świadczonych usług telekomunikacyjnych.

Może się okazać, że wizja społeczeństwa korzystającego z integracji telekomunikacji, informatyki, telewizji interaktywnej i filmu stanie się bliższa niż się obecnie sądzi. Okazuje się bowiem, że wymiana pewnych elementów dotychczas właściwych dla tych dziedzin powoduje eksplozję nowych możliwości. Standardy, choć powstają z pewnym opóźnieniem, wskazują drogę realizacji tych możliwości. Prowadzi to do rozwoju nowych usług i do zwrócenia się ku usługom wcześniej znanym, które dzięki nowym środkom technicznym nabierają nowych walorów użytkowych. Podstawowa idea standaryzacji, by móc realizować usługi w heterogenicznych środowiskach sieciowych, zbudowanych w oparciu o współpracujące urządzenia pochodzące od różnych producentów, pozostaje wciąż ta sama. Natomiast w miarę implementacji nowych standardów jej realizacja staje się bliższa.

Literatura

1. Okubo S., Dunstan S., Morrison G., Nilsson M., Radha H., Skran D. L., Thom G., *ITU-T Standardization of Audiovisual Communication Systems in ATM and LAN Environments*, IEEE JSAC, August 1997.
2. Girod B., Younes K.B, Bernstein R., Eisert P., Farber N., Hartung F., Horn U. Steinbach E., Stuhlmuller K., Wiegand T., *Recent Advances in Video Compression*, IEEE 1996.
3. Pereira F., *MPEG4: a New Challenge for the Representation of Audio-Visual Information*, Picture Coding Symposium 1996.
4. ISO, *Overview of the MPEG-4 Version 1 Standard*, Fribourg 1997, dokument ISO/IEC JTC1/SC29/WG11 N 1909.
5. Guyader A. Le, *Adoption par l'UIT-T de deux nouvelles normes de codage de la parole a debit reduit: G.729 et G.723.1*, l'echo des Recherches, no. 168-169 1997, CNET France Telecom.
6. Pacyna P., *Dystrybucja strumieni MPEG w sieciach szerokopasmowych*, sprawozdanie z grantu KBN, grudzien 1997.
7. Pacyna P, Papier Z., *Dystrybucja Strumieni MPEG2 z wykorzystaniem technologii ATM*, Krajowe Sympozjum Telekomunikacji KST'97, Bydgoszcz, 1997.
8. *MPEG-2 Digital Video Technology and Testing*, materialy firmy Hewlett Packard, 1997.
9. Rodriguez A., Fibush D. K., Bilow S. C, *MPEG-2 Fundamentals for Broadcast and Post-Production Engineers*, materialy firmy Tektronix, 1996.
10. Pan, *A Tutorial on MPEG/Audio Compression*, IEEE 1995.
11. Ochmanek G., Wieteska P., *Sieci Frame-Relay w zastosowaniach multimedialnych*, raport wew. KT AGH, Kraków 1996.
12. Chiang, D. Anastassiou, *Hierarchical Coding of Digital Television*, IEEE Telecommunications Magazine, May 1994.
13. Karwowska-Lamparska, *Telewizyjne Systemy Cyfrowe*, Wydawnictwa Komunikacji i Laczności, Warszawa 1994.
14. *r&d eclairer l'avenir*, miesiecznik CNET, France Telecom. Styczen 1998.
15. standardy serii G, H, oraz T ITU-T omawiane w tekście.