

METODY KODOWANIA SYGNAŁU MOWY DO ZASTOSOWAŃ W TELEKOMUNIKACJI

Maciej Kulesza pok. 726
Katedra Systemów Multimedialnych

Plan wykładu

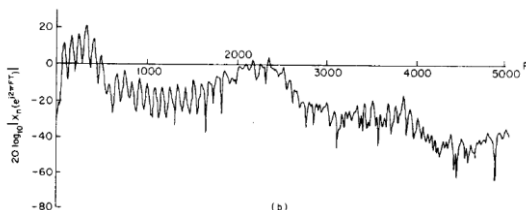
- ✧ Właściwości (charakterystyka) sygnału mowy
- ✧ Właściwości kodeków mowy
- ✧ Metody kodowania typu *waveform*
- ✧ Parametryczne metody kodowania
- ✧ Porównanie jakości kodowania uzyskiwanej dla poszczególnych algorytmów
- ✧ Szerokopasmowe kodowanie sygnału mowy
- ✧ Przyszłość kodeków dla zastosowań w telekomunikacji
- ✧ Podsumowanie

Plan wykładu

- ✦ Właściwości (charakterystyka) sygnału mowy
- ✦ Właściwości kodeków mowy
- ✦ Metody kodowania typu *waveform*
- ✦ Parametryczne metody kodowania
- ✦ Porównanie jakości kodowania uzyskiwanej dla poszczególnych algorytmów
- ✦ Szerokopasmowe kodowanie sygnału mowy
- ✦ Przyszłość kodeków dla zastosowań w telekomunikacji
- ✦ Podsumowanie



Właściwości sygnału mowy

- ✦ Widmo sygnału mowy posiada strukturę formantową



- ✦ Jakie pasmo posiada sygnał mowy?
 - ✦ W paśmie do **300-3400 Hz** występują **3-4 formanty** sygnału mowy, które posiadają znaczący wpływ na zrozumiałość
 - ✦ Pełne pasmo sygnału mowy obejmuje częstotliwości od około **50 Hz do 14 kHz**

Właściwości sygnału mowy

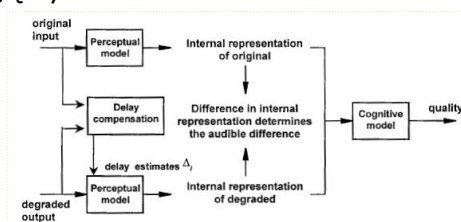
- ✦ W sygnale mowy możemy wyróżnić zasadniczo trzy części składowe:
 - Składowe dźwięczne (np. samogłoski) 
 - Składowe bezdźwięczne (np. „sz”, „s”, itp.) 
 - Stany transjentowe (stanowią sumę składowych dźwięcznych i bezdźwięcznych)
- ✦ Składowe dźwięczne sygnału mowy mają strukturę harmoniczną
 - Możliwe jest zatem wyznaczenie częstotliwości podstawowej sygnału harmonicznego

Właściwości kodeków mowy

- ✦ Parametry określające właściwości kodeków mowy
 - Wymagana przepływność strumienia bitowego
 - Opóźnienie wprowadzane przez algorytm i jego złożoność (MIPS, WMOPS)
 - Pasma kodowanej mowy
 - Zrozumiałość -> badana z wykorzystaniem list logatomowych (zrównoważone fonetycznie sekwencje pozbawione znaczenia semantycznego)
 - Jakość sygnału mowy wyrażona w skali MOS (*Mean Opinion Score*) -> określane w wyniku przeprowadzenia testów subiektywnych (oceny 1-5)

Właściwości kodeków mowy

- Przeprowadzanie testów subiektywnych mających na celu określenie jakości kodowania sygnału mowy przez poszczególne algorytmy jest uciążliwe
- ITU-T P. 862 - automatyczna metoda oceny jakości sygnału mowy PESQ (*Perceptual Evaluation of Speech Quality*)
- Wyniki uzyskiwane w skali PESQ są skorelowane z wynikami uzyskiwanymi w wyniku testów subiektywnych (MOS) i dlatego mogą być stosowane zamiennie



Plan wykładu

- Właściwości (charakterystyka) sygnału mowy
- Właściwości kodeków mowy
- Metody kodowania typu *waveform*
- Parametryczne metody kodowania
- Porównanie jakości kodowania uzyskiwanej dla poszczególnych algorytmów
- Szerokopasmowe kodowanie sygnału mowy
- Przyszłość kodeków dla zastosowań w telekomunikacji
- Podsumowanie

Metody kodowania typu *waveform*

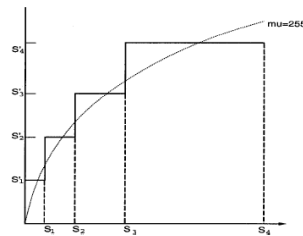
• Kodowaniu podlegają wartości próbek sygnału

– PCM (*Pulse Code Modulation*)

- $8000 \cdot 16 = 128$ kbps

– u-Law/A-Law (PCM) -> G.711

- Sygnał o rozdzielczości 14-bitowej poddawany jest kompresji w koderze i kodowany na 8 bitach (nieliniowa kwantyzacja) -> $8000 \cdot 8 = 64$ kbps

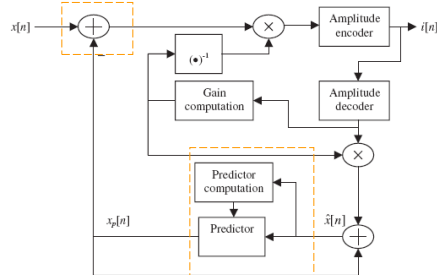


Metody kodowania typu *waveform*

– ADPCM (*Adaptive Differential PCM*) -> G.726

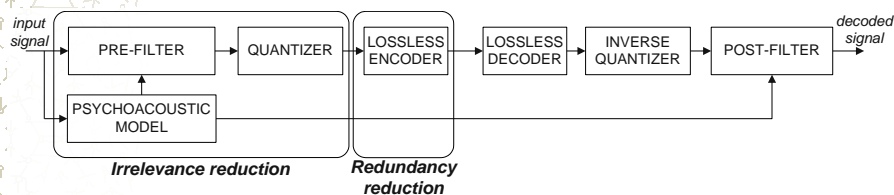
- Wartości próbek sygnału mowy są ze sobą skorelowane
- Możliwe jest zatem dokonywanie predykcji wartości próbki następnej na podstawie próbek poprzedzających
- Kodowaniu podlega sygnał błędu obliczony jako różnica pomiędzy sygnałem wejściowym a predykowanym

- Dostępne przepływności: 40, 32, 24 kbps
- Największej degradacji podlegają transjenty



Perceptualny ADPCM

- ✦ Sygnał filtrowany wstępnie zgodnie z charakterystyką odpowiadającą estymowanemu progowi słyszenia



Metody kodowania typu *waveform*

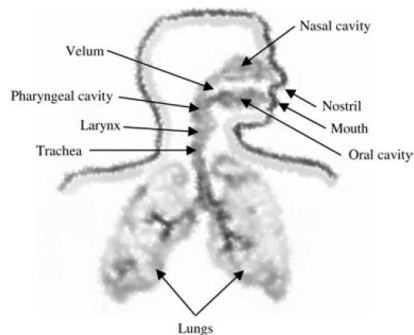
- ✦ Metody w których kodowaniu podlegają wartości próbek sygnału:
 - Pozwalają na uzyskanie **wysokiej jakości** sygnału mowy
 - Wymagają łączy zapewniających stosunkowo **szerokie pasmo**
 - Wprowadzają **marginalne opóźnienie** w procesie kodowania (rzędu pojedynczych milisekund)
 - Nie wymagają dużej **mocy obliczeniowej**

Plan wykładu

- ✦ Właściwości (charakterystyka) sygnału mowy
- ✦ Właściwości kodeków mowy
- ✦ Metody kodowania typu *waveform*
- ✦ Parametryczne metody kodowania
- ✦ Porównanie jakości kodowania uzyskiwanej dla poszczególnych algorytmów
- ✦ Szerokopasmowe kodowanie sygnału mowy
- ✦ Przyszłość kodeków dla zastosowań w telekomunikacji
- ✦ Podsumowanie

Parametryczne metody kodowania

- ✦ Parametryczne metody kodowania bazują na odpowiednim modelu sygnału mowy
 - Model uwzględnia sygnał pobudzenia oraz filtr o zmiennej w czasie charakterystyce częstotliwościowej

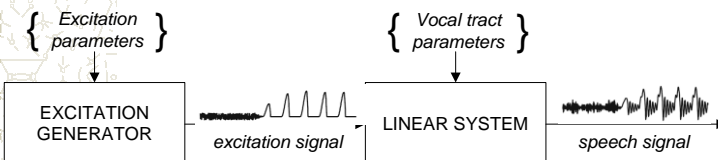


Przepływność

Zrozumiałość/Jakość

Parametryczne metody kodowania

- W dekoderyze sygnał mowy jest syntetyzowany na podstawie grupy przesłanych parametrów



- Możemy wyróżnić kilka głównych grup metod parametrycznych

- LPC10 (Linear Predictive Codec)
- MELP (Mixed-Excited Linear Prediction)
- MP/RPE-LP (Multi-Pulse/Regular Pulse Excited LP)
- CELP (Code Excited Linear Prediction)

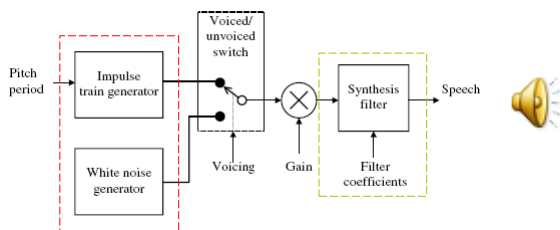
Przepliwność

Zrozumiałość/Jakość

Parametryczne metody kodowania (LPC-10)

- Każda ramka sygnału jest klasyfikowana jako dźwięczna bądź bezdźwięczna

- Dla ramek dźwięcznych sygnałem modelującym jest ciąg impulsów generowanych na podstawie informacji o częstotliwości podstawowej tonu krtaniowego
- Ramki bezdźwięczne syntetyzowane są z wykorzystaniem generatora szumu białego

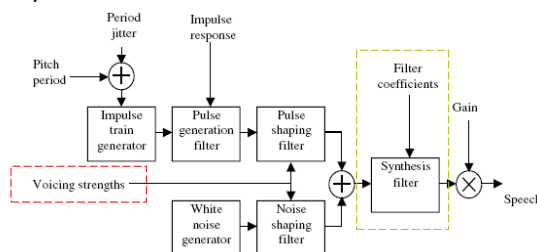


Przepliwność

Zrozumiałość/Jakość

Parametryczne metody kodowania (MELP - Mixed Excited Linear Prediction)

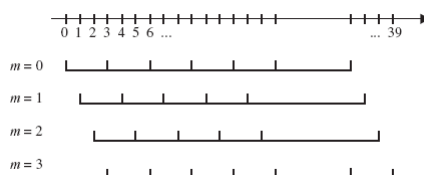
- ✦ Sygnał dzielony jest na kilka podpasm (zwykle 5)
- ✦ Dla każdego z podpasm sygnał pobudzenia określany jest jako suma sygnału generowanego na podstawie częstotliwości tonu krtaniowego oraz szumu
- ✦ Struktura harmoniczna może być celowo zaburzana w celu wierniejszego odzwierciedlenia stanów transjentowych



Zrozumiałość/Jakość
Przepływność

Parametryczne metody kodowania (MP/RPE-LP -> Multi-Pulse/Regular Pulse Excited)

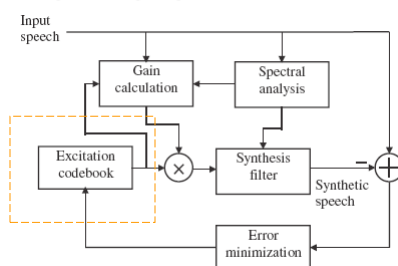
- ✦ **MP**– sygnał pobudzenia modelowany jako sekwencja impulsów, których położenie na osi czasu nie jest ograniczone (10 do 12 impulsów dla 10 ms sygnału)
- ✦ **RPE**– sygnał pobudzenia modelowany jest jako sekwencja impulsów, przy czym odległości pomiędzy nimi są stałe (10 do 12 impulsów dla 5 ms sygnału)
- ✦ Położenie na osi czasu impulsów dobierane jest zwykle na zasadzie *analysis-by-synthesis*



Zrozumiałość/Jakość
Przepływność

Parametryczne metody kodowania (CELP - Code Excited Linear Prediction)

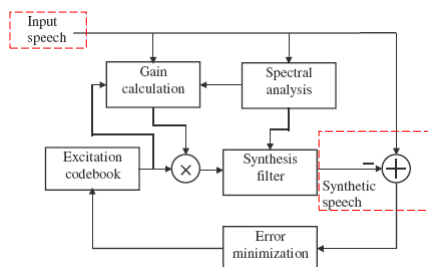
- Zamiast modelowania sygnału pobudzenia wykorzystywana jest książka kodowa, w której zapisane są określone sekwencje -> kodowanie wektorowe
- Wybór odpowiedniej sekwencji następuje na zasadzie *analysis-by-synthesis*



Zrozumiałość/Jakość
Przepływność

Parametryczne metody kodowania (CELP - Code Excited Linear Prediction)

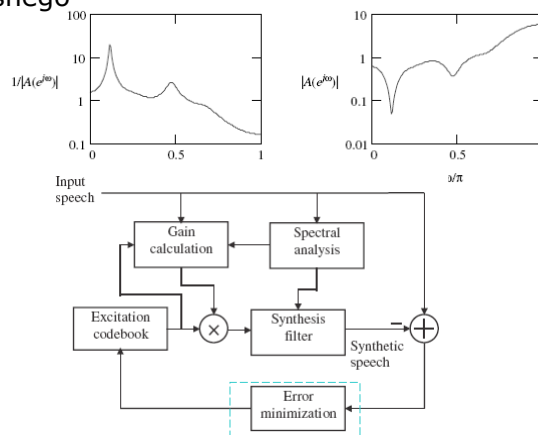
- Wybierana jest ta sekwencja dla której sygnał błędu wyznaczany jako różnica pomiędzy ramką sygnału oryginalnego i syntetyzowanego jest najmniejszy w sensie średniokwadratowym
- Indeks wybranej sekwencji z książki kodowej transmitowany jest do dekodera



Zrozumiałość/Jakość
Przepływność

Parametryczne metody kodowania (CELP - Code Excited Linear Prediction)

- Wyznaczony sygnał błędu jest perceptualnie ważony w celu wykorzystania zjawiska maskowania jednoczesnego



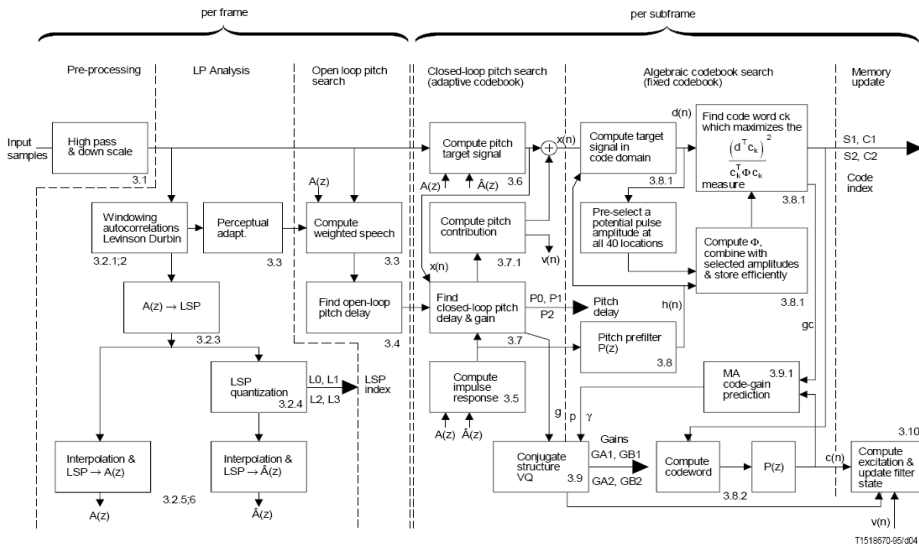
Zrozumiałość/Jakość
Przeżywność

Parametryczne metody kodowania (CELP - Code Excited Linear Prediction)

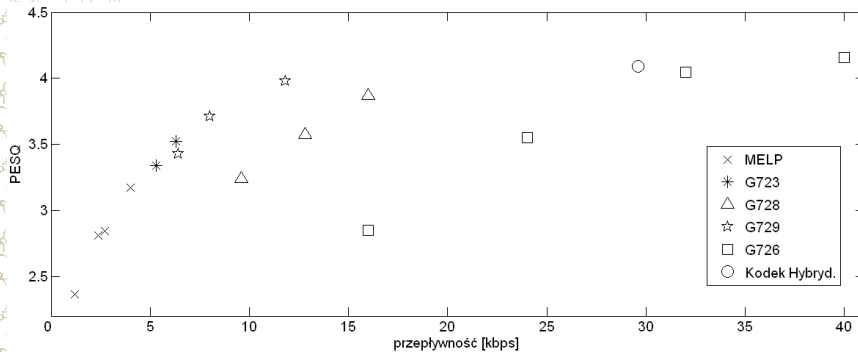
- Metody typu CELP należą do grupy najpopularniejszych kodeków wykorzystywanych w telekomunikacji
 - G.723.1 MP-MLQ/ACELP (Multi Pulse/Algebraic CELP) -> 5.3/6.3 kbps
 - G.728 LD-CELP (Low Delay-CELP) -> 9.6, 12.8, 16 kbps
 - G.729 CS-ACELP (Conjugate Structure Algebraic CELP) -> 6.4, 8, 11.8 kbps
 - GSM (AMR-NB) -> 4.75 - 12.2 kbps
 - GSM FR (Full Rate) -> 12.2 kbps

Zrozumiałość/Jakość
Przeżywność

Struktura koder G729



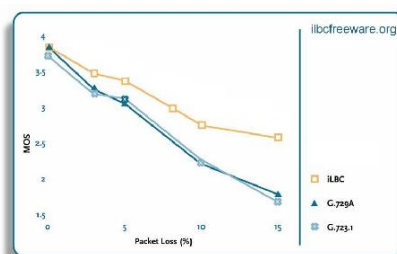
Porównanie jakości kodowania



Zestawienie próbek zakodowanych algorytmami CELP/MELP jest dostępne poprzez -> <http://www.signalogic.com>

Porównanie jakości kodowania

- ✦ W celu dokładnego zbadania właściwości kodeka konieczne jest przeprowadzenie zestawu kompleksowych testów
 - Kodowanie mowy w szumie (np. różne poziomy szumu „samochodowego”)
 - Określenie wpływu kodowanego języka na jakość (nie powinno być)
 - Wpływ zjawisk związanych z transmisją w kanale telekomunikacyjnym na jakość (np. utrata pakietów)

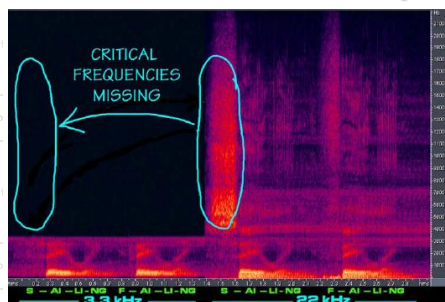


Plan wykładu

- ✦ Właściwości (charakterystyka) sygnału mowy
- ✦ Właściwości kodeków mowy
- ✦ Metody kodowania typu *waveform*
- ✦ Parametryczne metody kodowania
- ✦ Porównanie jakości kodowania uzyskiwanej dla poszczególnych algorytmów
- ✦ Szerokopasmowe kodowanie sygnału mowy
- ✦ Przyszłość kodeków mowy
- ✦ Podsumowanie

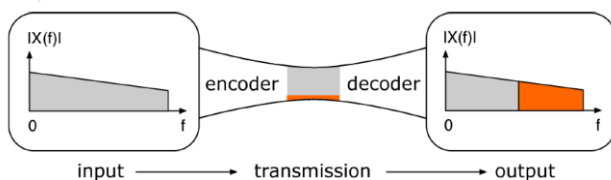
Kodowanie szerokopasmowe

- ✦ Pasma >4 kHz ma również znaczący wpływ na zrozumiałość (nie jest możliwe odróżnienie „f” od „s” dla sygnału wąskopasmowego, można pomylić głos matki i córki, trudno jest dokładnie zanotować nieznaną frazę)
- ✦ Możliwa jest prowadzenie swobodnej konwersacji



BWE – Bandwidth Extension

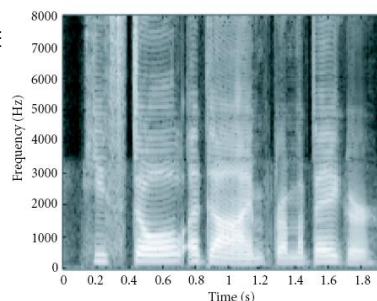
- ✦ Pomiędzy pasmem podstawowym a pasmem >4000 Hz występuje korelacja – możliwe jest „ślepa regeneracja” pasma >4000 Hz -> słaba jakość
- ✦ Możliwe jest przesyłanie niewielkiej ilości dodatkowych danych, na podstawie których dekodery dokonuje syntezy pasma >4000 Hz



- ✦ Słuch jest bardziej czuły na wierne odwzorowanie obwiedni widma niż pobudzenia w paśmie >4000 Hz

BWE – Bandwidth Extension

- ✦ Dla sygnału mowy można przyjąć prosty model pobudzenia w paśmie >4000 Hz w postaci szumu białego
- ✦ SBR (Spectral Band Replication) -> 'pobudzenie' z pasma podstawowego wykorzystywane jest również w paśmie wyższym
 - Metoda stosowana w kodeku AAC HE
- ✦ Metody nieliniowe
- ✦ inne



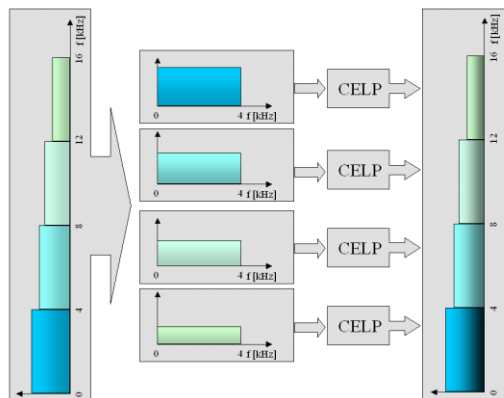
Kodowanie szerokopasmowe

- ✦ Zwykle przyjmuje się następujący podział kodeków ze względu na pasmo kodowanej mowy:
 - *narrow band* (do 4 kHz)
 - *wideband* (do 8kHz)
 - *ultra wideband* (do 16 kHz)
- ✦ W jaki sposób możliwe jest kodowanie sygnału mowy w szerokim paśmie?



Kodowanie szerokopasmowe

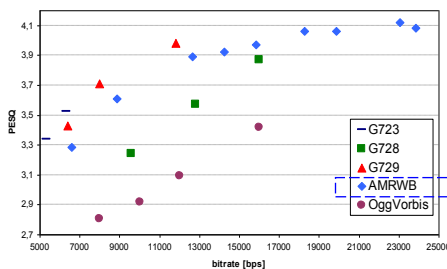
- Wielo-pasmowe kodowanie sygnału mowy z wykorzystaniem modułów CELP -> Speex (www.speex.org)



Zrozumiałość/Jakość
Przepływność

Kodowanie szerokopasmowe (AMR-WB)

- G.722.2 AMR-WB (*Adaptive Multi Rate-Wide Band*)
 - Algorytm oparty na architekturze ACELP (sygnał pobudzenia dobierany dla szerszego pasma)
 - Koduje sygnał w paśmie od 50 do 7000 Hz
 - Przepływność od 6.6 do 23.85 kbps (dla niższych przepływności kodeka więcej pasma przeznaczają się na kodowania kanałowe)
 - Standard dla telefonii stacjonarnej i bezprzewodowej



AMRWB

VoiceAge
NOKIA
Connecting People

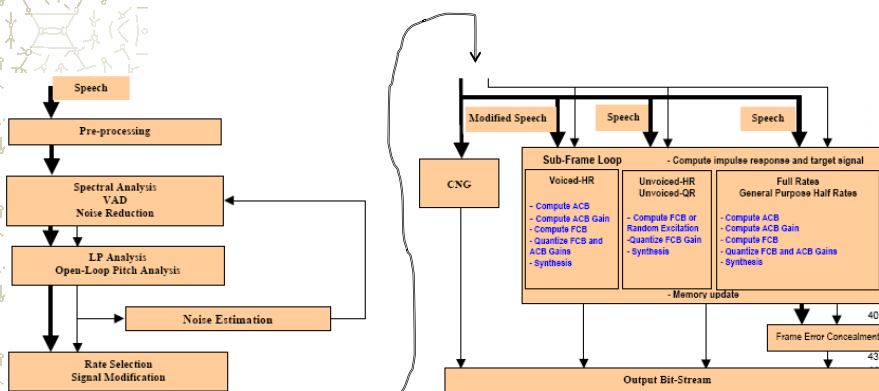
Kodowanie szerokopasmowe (VMR-WB)

VMR-WB (*Variable Rate Multi-Mode Wideband*)

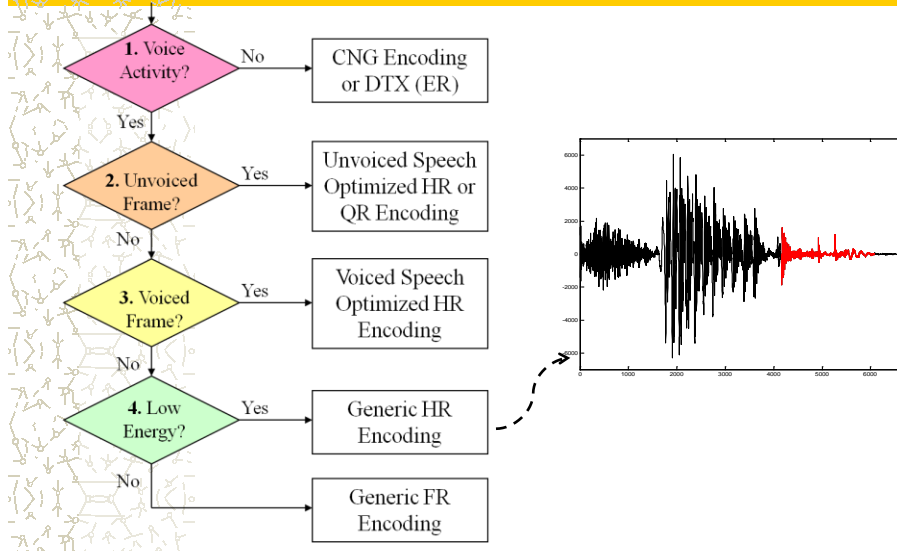
- By zapewnić optymalną jakość przy zadanej średniej przepływności, chwilowa przepływność dostosowywana jest do charakteru kodowanego sygnału (0.8-13.3 kbps)
- Koder klasyfikuje ramki na
 - „puste” (nie zawierające sygnału mowy, a jedynie szum tła)
 - dźwięczne
 - bezdźwięczne
 - transjentowe
- Jeden z trybów pracy kodeka jest zgodny ze standardem AMR-WB (minimalna utraty jakości VMR-WB->AMR-WB)
- Standard telefonii CDMA2000®



VMR-WB



VMR-WB



Przyszłość kodeków mowy (AMR-WB+)

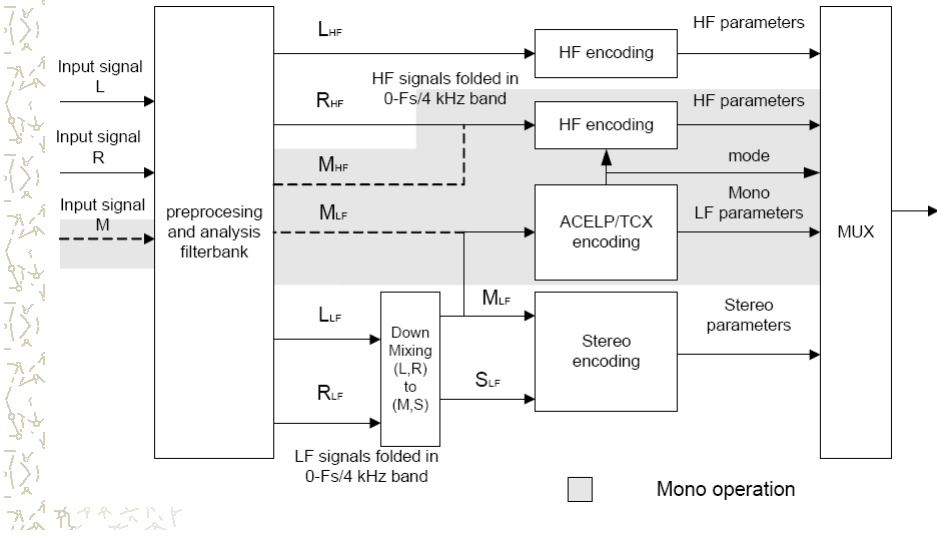
- ✦ Dotychczas stosowane kodeki mowy (CELP) bardzo słabo kodują sygnały muzyczne
- ✦ Nowoczesne kodeki mowy nie tylko kodują sygnał mowy w szerokim paśmie, ale również pozwalają na kodowanie z wysoką jakością sygnałów muzycznych



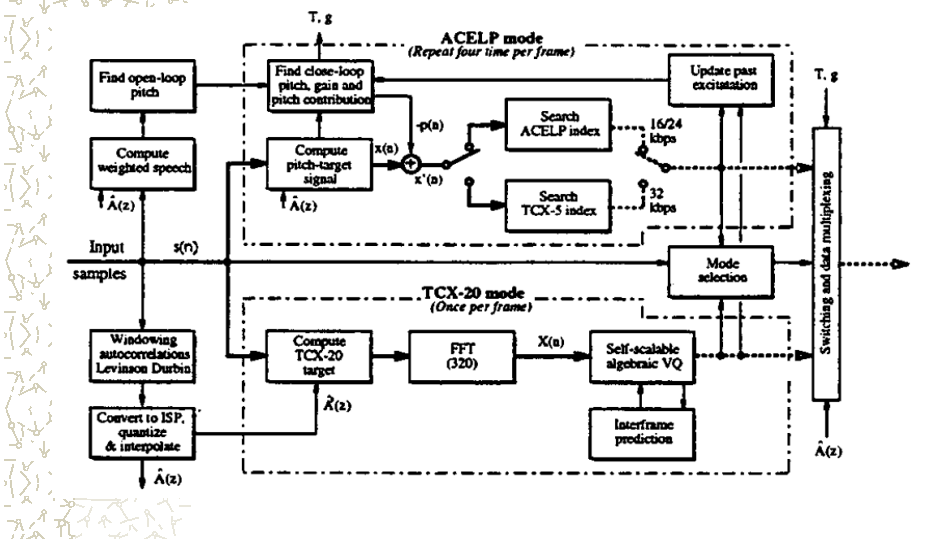
- ✦ Kodek wykorzystuje hybrydową architekturę ACELP/TCX
- ✦ <http://www.voiceage.com>



Kodek AMR-WB+



Kodek AMR-WB+

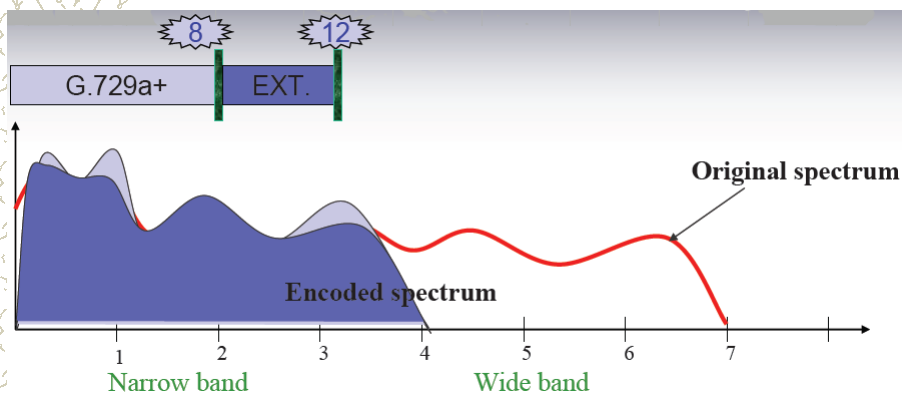


Kodek G729.1 (G.729EV)

- Zbudowany w oparciu o wąskopasmowy kodek G.729
- Adaptacyjny algorytm kodowania szerokopasmowego (50-7000 Hz, 8-32 kbps)
- Kodek hybrydowy – CELP dla niskich przepływności, CELP wraz z kodowaniem widmowym dla wyższych przepływności
- Pozwala na kodowanie sygnału mowy oraz muzycznych (dla wyższych przepływności)

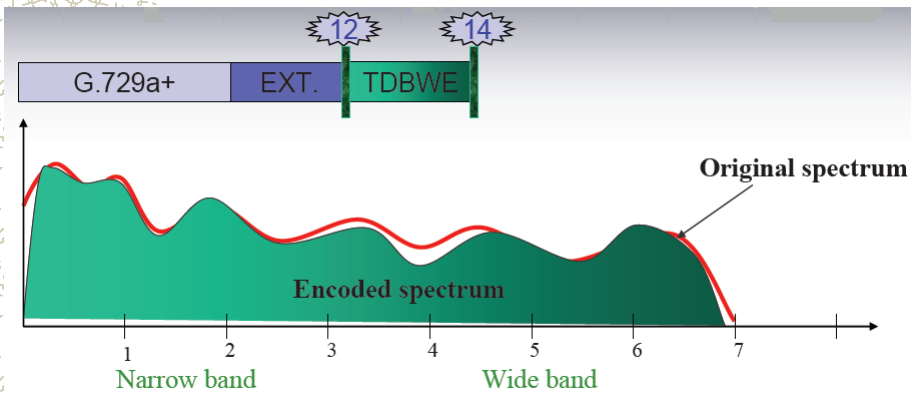


G729.1 (8, 12 kbps)



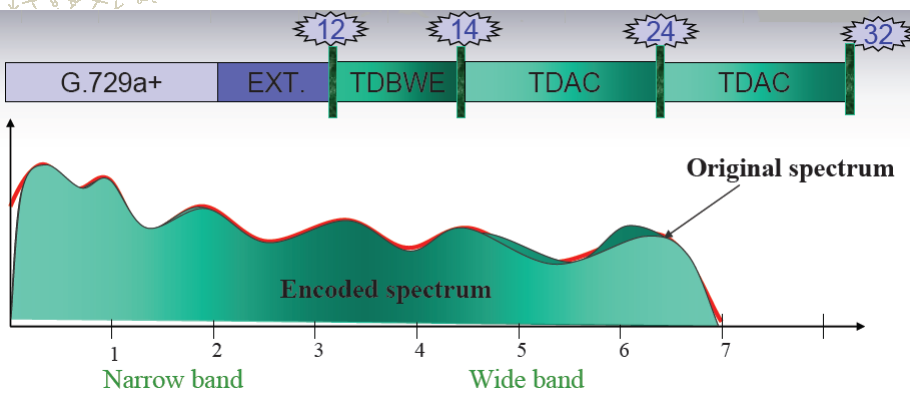
Dla niskich przepływności kodek zachowuje się jak zwykły kodek mowy G.729

G729.1 (14 kbps)



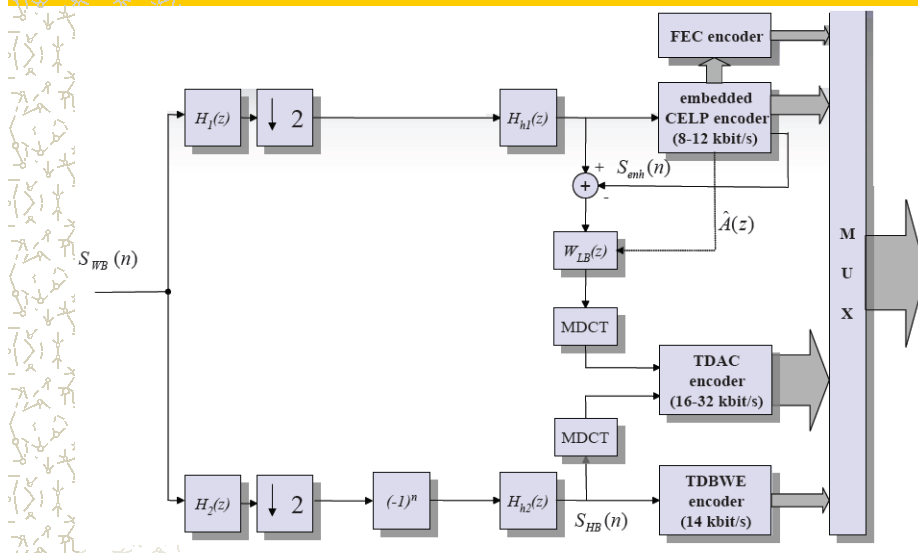
- Zastosowano metodę poszerzania widma sygnału z 4000 do 7000 Hz - TDBWE

G729.1 (24-32 kbps)



- Predykcyjne kodowanie widmowe – TDAC
 - Sygnału rezydualnego kodeka CELP (50-4000 Hz)
 - Sygnału w paśmie >4000 Hz

G729.1 (koder)



Kodek G722.1C

- ✦ Kodek stanowi rozwinięcie algorytmu G729.1
- ✦ Sygnał audio próbkowany z częstotliwości 32 kHz – kodowane pasmo zdublowane w stosunku do G729.1 (14 kHz zamiast 7 kHz)
- ✦ Celem było opracowanie algorytmu, który:
 - Jest możliwe prosty obliczeniowo (wymaga niewielkiej mocy obliczeniowej procesora i pamięci)
 - wprowadza niewielkie opóźnienie kodowania (algorytm do zastosowania w komunikacji głosowej)

Kodek G722.1C – wprowadzane opóźnienie

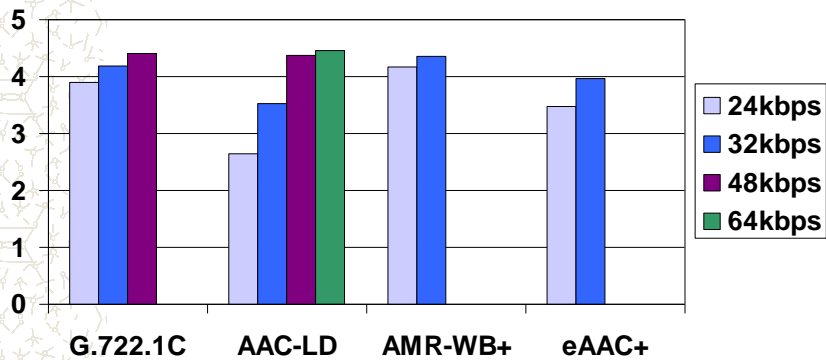
G.722.1C (ms)	eAAC+ (ms)	AMR-WB+ (ms)
40.0	129.9 ^[1]	113.8 ^[2]

Kodek G722.1C – złożoność obliczeniowa

Bit rate (kbit/s)	G.722.1C (WMOPS)	eAAC+ (WMOPS)	AMR-WB+ (WMOPS)
24	9.7	40.8	80.1
32	10.3	42.6	86.7

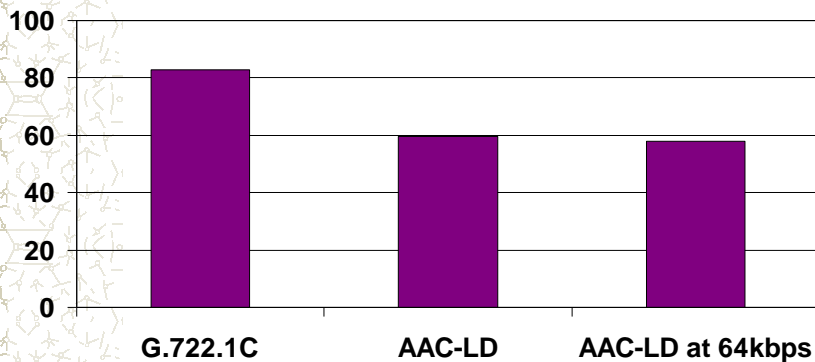
Kodek G722.1C – testy MUSHRA

Reverberant Speech with Office Noise



Kodek G722.1C – testy MUSHRA

Music and Mixed Content (48 kbps)



Kodek MPEG RM0

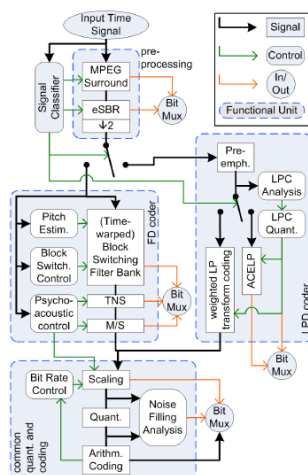


Fig. 1: Overview over the RM0 encoder

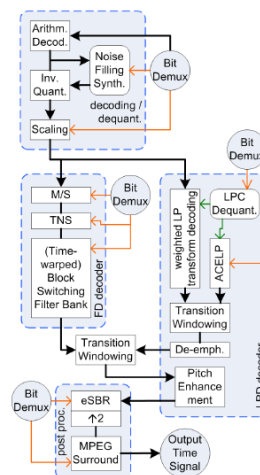


Fig. 2: Overview over the RM0 decoder

Podsumowanie (co warto zapamiętać/wiedzieć)

- ✦ Zasadę działania kodeków opartych na algorytmie predykcji liniowej (model produkcji mowy: pobudzenie+filtr)
- ✦ Różnicę w architekturze kodeków LPC-10, MELP, CELP
- ✦ Czynniki, jakie należy wziąć pod uwagę przy wyborze kodeka oraz sposoby oceny jakości kodowania
- ✦ Wytlumaczyć zasadność kodowania sygnału mowy w szerokim paśmie
- ✦ Standardy kodowania szerokopasmowego i metody doboru przepływności w kodeku VMR-WB, AMR-WB+