

# KODER AMR-CELP<sup>1</sup>

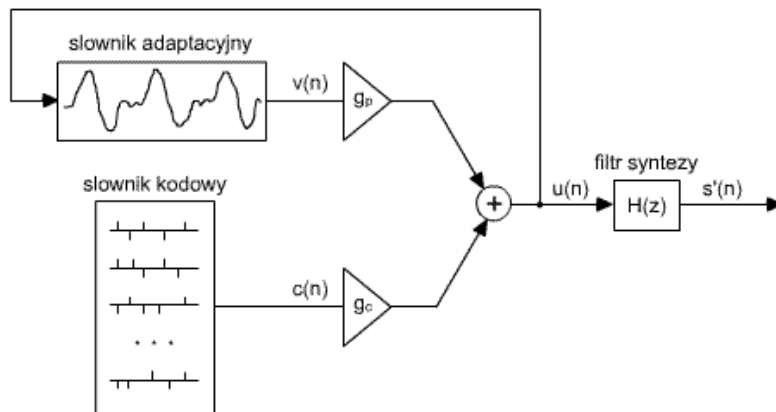
## Instrukcja laboratoryjna

### 1. Zasada działania koderów CELP

Aby zrozumieć zasadę działania kodera CELP, najlepiej przyjrzeć się modelowi syntezy mowy, który jest w nim wykorzystywany. Uproszczony schemat takiego modelu, używany przez wiele koderów typu CELP, przedstawiono na Rys. 1.1. Trzema najważniejszymi blokami funkcjonalnymi są tu filtr syntezy mowy  $H(z)$ , adaptacyjny słownik pobudzeń, oraz stały słownik pobudzeń stochastycznych. Pobudzenia te zwane są również pobudzeniami kodowymi (ang. *code excitations*). W słowniku adaptacyjnym znajdują się sygnały pobudzające wykorzystane w bezpośredniej przeszłości (korzystanie z takich pobudzeń nazywane jest niekiedy predykcją długookresową i jest bardzo korzystne dla kodowania mowy dźwięcznej, mającej charakter zbliżony do okresowego). Filtr syntezy realizuje natomiast „krótkookresową” predykcję liniową, stąd nazwa CELP – *code excited linear prediction*.

Sztuczny sygnał mowy  $s'(n)$  tworzony jest przez filtr syntezy  $H(z)$ , na wejście którego podawane jest pobudzenie  $u(n)$ . Jest ono sumą pobudzeń generowanych przez adaptacyjny słownik pobudzeń (sygnał  $v(n)$ ) oraz słownik pobudzeń kodowych (sygnał  $c(n)$ ). Oba te pobudzenia sumowane są ze wzmocnieniami  $g_p$  i  $g_c$ . W związku z powyższym pobudzenie filtra syntezy wynosi:

$$u(n) = g_p \cdot v(n) + g_c \cdot c(n). \quad (1.1)$$

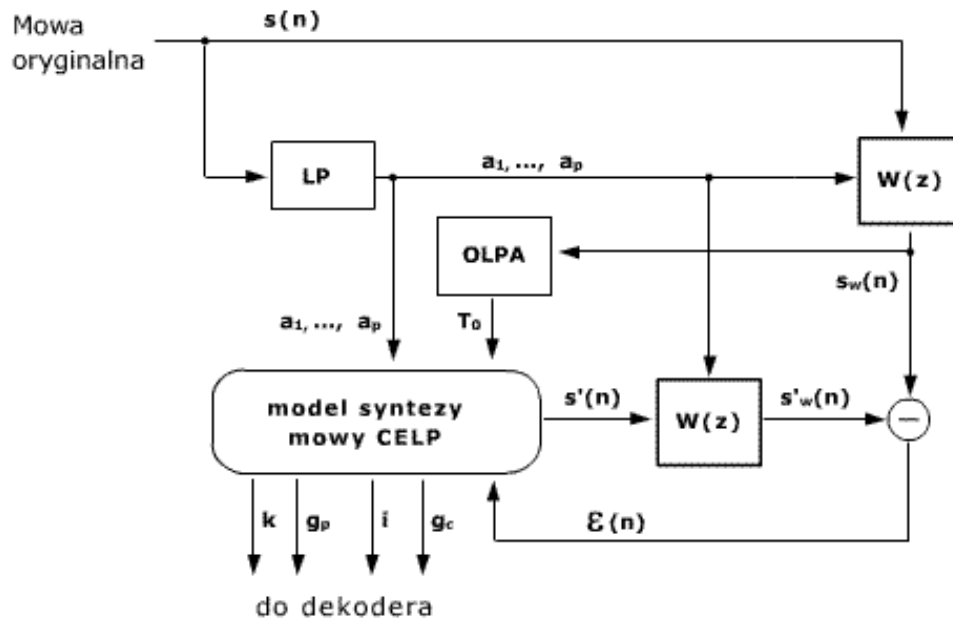


Rys. 1.1 Uproszczony schemat modelu syntezy mowy CELP

Na Rys. 1.2 przedstawiono uproszczony schemat blokowy kodera CELP. Zawiera on między innymi model syntezy mowy przedstawiony na Rys. 1.1. Zadaniem kodera jest wyznaczyć parametry sygnału  $u(n)$  – w tym wypadku  $k$ ,  $g_p$ ,  $i$  oraz  $g_c$ , a także współczynniki predykcji  $a_1, \dots, a_p$ , opisujące filtr syntezy  $H(z)$ . Charakterystyczną cechą koderów CELP jest, że ekstrakcji tych parametrów dokonuje się przy użyciu

<sup>1</sup> Na podstawie pracy dyplomowej P.J.Szelemeja, Inst. Telekomunikacji P.W., 2004

tzw. analizy przez syntezę (ang. *analysis by synthesis*). Polega to na tym, że koder zawiera w sobie dekodery lokalny, za pomocą którego dokonuje syntezy sztucznego sygnału i porównuje go z sygnałem oryginalnym, minimalizując błąd  $\varepsilon(n)$  w zamkniętej pętli zwrotnej. Jak widać na Rys. 1.2 błąd obliczany jest między sygnałami przepuszczonymi przez tzw. filtr percepcyjny o transmitancji  $W(z)$ . Filtr percepcyjny ma za zadanie takie ukształtowanie widma sygnału błędu  $\varepsilon(n)$ , aby zmniejszyć dokuczliwość tego sygnału (częściowe wykorzystanie zjawiska maskowania).



**Rys. 1.2 Schemat ideowy kodera CELP.**

- $s(n)$  – sygnał oryginalny
- $s_w(n)$  – sygnał percepcyjny
- $s'(n)$  – sygnał sztuczny
- $s'_w(n)$  – sztuczny sygnał percepcyjny
- $\varepsilon(n)$  – błąd
- $a_1, \dots, a_p$  – współczynniki predykcji
- $T_0$  – okres drgań strun głosowych (okres tonu krtaniowego)
- $k$  – opóźnienie dla pobudzenia adaptacyjnego
- $g_p$  – wzmacnienie pobudzenia ze słownika adaptacyjnego
- $i$  – indeks pobudzenia stochastycznego
- $g_c$  – wzmacnienie pobudzenia stochastycznego
- LP – obliczanie wsp. predykcji liniowej (ang. *Linear Prediction*)
- OLPA – obliczanie okresu tonu krtaniowego w pętli otwartej (ang. *Open Loop Pitch Analysis*)

## 2. Zagadnienia związane z przetwarzaniem sygnału

### 2.1 Przetwarzanie blokowe

Jak wspomniano w poprzednich punktach, w koderach CELP sygnał wejściowy opisuje się za pomocą szeregu parametrów, które przesyłane są z kodera do dekodera. Syntetyczny sygnał wyjściowy tworzony jest przy pomocy modelu aparatu mowy sterowanego tymi parametrami. W celu dokonania ekstrakcji parametrów, koder musi zgromadzić fragment sygnału o odpowiedniej długości. Zazwyczaj jest to od 20 do 30 ms sygnału (wyjątkiem jest LD CELP [G. 728]), co przy częstotliwości próbkowania  $f_s=8\text{kHz}$  wymaga buforowania od 160 do 240 próbek. Taki fragment sygnału nosi miano ramki (ang. *frame*). Zazwyczaj dzieli się ją na kilka (zwykle 2 do 4) podramek (ang. *subframes*). Parametry takie jak współczynniki predykcyjnego filtra syntezy, czy przybliżony okres tonu krtaniowego, obliczane są 1 - 2 razy na ramkę, a dobieranie parametrów dla słownika adaptacyjnego i stałego odbywa się dla każdej podramki. W związku z przetwarzaniem wektorowym (ramki, podramki) występuje opóźnienie (tzw. opóźnienie algorytmiczne) sygnału odtworzonego w odbiorniku względem sygnału oryginalnego.

### 2.2 Ocena jakości sygnału po kompresji

Jednymi z najważniejszych parametrów koderów są: stopień kompresji (przepływność binarna), jakość sygnału po kompresji (generowanego w dekodерze) oraz opóźnienie algorytmiczne. Parametry te są ze sobą ściśle powiązane i zazwyczaj poprawa jednego z nich powoduje pogorszenie pozostałych.

O ile opóźnienie i stopień kompresji są parametrami, które można łatwo obiektywnie zmierzyć, porównanie sygnału wygenerowanego sztucznie, z sygnałem oryginalnym, nie zawsze jest proste. Jeśli koder zaprojektowany jest tak, aby sygnał po kompresji naśladował przebieg czasowy sygnału oryginalnego, można użyć kryterium obiektywnego, jakim jest stosunek mocy sygnału do mocy zakłóceń (szumu kwantyzacji etc.), oznaczany jako SNR (ang. *Signal to Noise Ratio*). Dla fragmentu sygnału parametr ten, wyrażony w decybelach, oblicza się ze wzoru:

$$SNR = 10 \cdot \log_{10} \left( \frac{\sum_n s(n)^2}{\sum_n (s(n) - s'(n))^2} \right) [\text{dB}], \quad (2.1)$$

gdzie  $s(n)$  oznacza sygnał oryginalny, a  $s'(n)$  sygnał wygenerowany sztucznie. Jak widać, szumem nazywa się różnicę między tymi sygnałami.

Ze względu na niestacjonarny charakter mowy, która składa się z głosek dźwięcznych o dużej mocy w stosunku do głosek bezdźwięcznych, często stosuje się SNR segmentowy, który jest wartością średnią SNR liczonych dla fragmentów mowy o czasie trwania 10 – 30 ms.

$$SNR_{seg} = \frac{1}{M} \sum_{i=1}^M SNR(i) [\text{dB}] \quad (2.2)$$

Taka wersja SNR jest bardziej czuła na uszkodzenie cichych fragmentów sygnału.

W przypadku koderów, w których sygnał sztuczny ma brzmienie podobne do sygnału oryginalnego, ale jego przebieg czasowy jest zupełnie inny (np. w wokoderach), stosowanie SNR jest bezcelowe. W takich przypadkach pozostaje jedynie stosowanie metod subiektywnych oceny jakości sygnału, które opisane są przez wiele standardów. Do ważniejszych należy metoda ACR, gdzie przeprowadza się test słuchowy na grupie ludzi, którzy oceniają sygnał w skali bezwzględnej od 1 do 5. W metodzie CCR porównywany jest sygnał wygenerowany sztucznie, z frazą oryginalną w skali od -3 do 3, a w metodzie DCR ocenia się pogorszenie jakości w stosunku do frazy oryginalnej w skali od 1 do 5.

### 3. Standard AMR CELP

#### 3.1 Wstęp

Standard AMR CELP, ustanowiony został przez ETSI (ang. *European Telecommunications Standardization Institute*), z myślą o zastosowaniach w sieciach telefonii komórkowej trzeciej generacji – UMTS (ang. *Universal Mobile Telecommunications System*). Opisuje on koder typu ACELP, który może pracować w wielu trybach, różniących się przepływnością binarną strumienia wyjściowego i stopniem kompresji sygnału (patrz Tab. 3.1). W rzeczywistości jest to osiem oddzielnych koderów ACELP, objętych jednym wspólnym standardem. Jednym z nich jest GSM EFR (12.2 kb/s), używany w sieciach GSM (ang. *Global System for Mobile Communications*), który miał zastąpić RPE GSM (13 kb/s).

Koder sterowany jest zewnątrz i umożliwia zmianę trybu co 20 ms (co jedną ramkę sygnału), co jest szczególnie przydatne np. wtedy, gdy w kanale występują zakłócenia o charakterze chwilowym. Można wtedy obniżyć przepływność binarną strumienia generowanego przez koder, na rzecz zwiększenia liczby bitów przeznaczonych na kodowanie transmisyjne, zabezpieczające sygnał przed błędami. Także w przypadku chwilowych zmian natężenia ruchu w kanale telekomunikacyjnym, tryb kodera można zmieniać tak, aby jak najwięcej użytkowników mogło korzystać z kanału, przesyłając sygnał możliwie najlepszej jakości. Sterowanie zewnątrz oznacza, że koder nie bierze udziału w podejmowaniu decyzji odnośnie trybu działania.

Tryb	Parametry	Podramka				Suma
		1	2	3	4	
<b>12.2 kb/s</b> <b>(GSM EFR)</b>	$a_1, \dots, a_{10}$ (2 zestawy)					38
	$k$	9	6	9	6	30
	$g_p$	4	4	4	4	16
	$c_i$	35	35	35	35	140
	$g_c$	5	5	5	5	20
						<b>244</b>
<b>10.2 kb/s</b>	$a_1, \dots, a_{10}$					26
	$k$	8	5	8	5	26
	$c_i$	31	31	31	31	124
	$g_c$ i $g_p$	7	7	7	7	28
						<b>204</b>
<b>7.95 kb/s</b>	$a_1, \dots, a_{10}$					27
	$k$	8	6	8	6	28
	$g_p$	4	4	4	4	16
	$c_i$	17	17	17	17	68
	$g_c$	5	5	5	5	20
						<b>159</b>

7.40 kb/s (DAMPS EFR)	$a_1, \dots, a_{10}$					26
	$k$	8	5	8	5	26
	$c_i$	17	17	17	17	68
	$g_c$ i $g_p$	7	7	7	7	28
						148
6.70 kb/s	$a_1, \dots, a_{10}$					26
	$k$	8	4	8	4	24
	$c_i$	14	14	14	14	56
	$g_c$ i $g_p$	7	7	7	7	28
						134
5.90 kb/s	$a_1, \dots, a_{10}$					26
	$k$	8	4	8	4	24
	$c_i$	11	11	11	11	44
	$g_c$ i $g_p$	6	6	6	6	24
						118
5.15 kb/s	$a_1, \dots, a_{10}$					23
	$k$					20
	$c_i$					36
	$g_c$ i $g_p$					24
						103
4.75 kb/s	$a_1, \dots, a_{10}$					23
	$k$	8	4	4	4	20
	$c_i$	9	9	9	9	36
	$g_c$ i $g_p$	8		8		16
						95

**Tab. 3.1 Przydział bitów w poszczególnych trybach działania kodera AMR CELP**

### 3.2 Przetwarzanie sygnału

W koderze AMR CELP, we wszystkich trybach działania, długość ramki wynosi 160 próbek, co przy częstotliwości próbkowania równej 8000 Hz, odpowiada 20 ms. Każda ramka składa się z czterech podramek o długości 40 próbek (5 ms). Dla każdej ramki znajdowane są współczynniki predykcji krótkookresowej, które opisują transmitancję filtra percepcyjnego  $W(z)$  i predykcyjnego filtra syntezy  $H(z)$ . Następnie przy użyciu metod autokorelacyjnych, na podstawie sygnału percepcyjnego (sygnału oryginalnego, przepuszczonego przez filtr percepcyjny), wyznaczany jest okres tonu krtaniowego, który stanowi pierwsze przybliżenie dla adaptacyjnego słownika pobudzeń, realizującego predykcję długookresową. Słownik ten, wyznacza dokładną wartość opóźnienia, z jakim kopiowany jest fragment pobudzenia z przeszłości (z odpowiednim wzmocnieniem), w celu przybliżenia szukanego pobudzenia dla filtra syntezy w danej podramce. Następnie, w celu poprawy jakości sygnału syntetycznego, jest ono korygowane, przez dodanie pobudzenia algebraicznego ze słownika stałego. Parametry wyznaczonych pobudzeń oraz współczynniki predykcji przesyłane są do dekodera, w którym po zdekodowaniu służą do wygenerowania sygnału syntetycznego.

### 3.3 Algorytmy przeszukiwania słownika algebraicznego

#### Tryb 12.2 kb/s

Zadaniem słownika jest rozmieszczenie  $N_p = 10$  impulsów na 40 dozwolonych pozycjach, w taki sposób, aby odległość między sztucznym sygnałem percepcyjnym  $s'_w(n)$  i oryginalnym sygnałem percepcyjnym  $s_w(n)$  była jak najmniejsza. Dozwolone pozycje dzielone są na pięć ścieżek, a na każdej z nich rozmieszczane są dokładnie dwa impulsy (Tab. 3.2).

Ścieżka	Pozycje
1	0,5,10,15,20,25,30,35
2	1,6,11,16,21,26,31,36
3	2,7,12,17,22,27,32,37
4	3,8,13,18,23,28,33,38
5	4,9,14,19,24,29,34,39

**Tab. 3.2 Podział dozwolonych pozycji na ścieżki - tryb 12.2 kb/s**

Impulsy  $i_0, \dots, i_9$ , wstawiane są parami  $\{i_0, i_5\}$ ,  $\{i_1, i_6\}$ ,  $\{i_2, i_7\}$ ,  $\{i_3, i_8\}$ ,  $\{i_4, i_9\}$  na odpowiednio wybrane ścieżki. Istnieje możliwość ustawienia dwóch impulsów na jednej pozycji.

### **Tryb 10.2 kb/s**

Pozycje wektora **c**, podzielone są tu na cztery ścieżki, na tej samej zasadzie, jak przy przepływności 12.2 kb/s:

Ścieżka	Pozycje
1	0,4,8,12,16,20,24,28,32,36
2	1,5,9,13,17,21,25,29,33,37
3	2,6,10,14,18,22,26,30,34,38
4	3,7,11,15,19,23,27,31,35,39

**Tab. 3.3 Podział dozwolonych pozycji na ścieżki - tryb 10.2 kb/s**

Na każdej z nich wstawiane są dokładnie dwa impulsy.

### **Tryb 7.95, 7.40 i 6.70 kb/s**

Wektor kodowy dzielony jest na cztery ścieżki dla trybu 7.95 i 7.40 kb/s (Tab. 3.4).

Ścieżka	Pozycje
1	0,5,10,15,20,25,30,35
2	1,6,11,16,21,26,31,36
3	2,7,12,17,22,27,32,37
4	3,4,8,9,13,14,18,19,23,24,28,29,33,34,38,39

**Tab. 3.4 Podział dozwolonych pozycji na ścieżki – tryb 7.95 i 7.40 kb/s**

Dla przepływności 6.70 kb/s, pozycje dzielone są na trzy ścieżki

Ścieżka	Pozycje
1	0,5,10,15,20,25,30,35
2	1,3,6,8,11,13,16,18,21,23,26,28,31,33,36,38
3	2,4,7,9,12,14,17,19,22,24,27,29,32,34,37,39

**Tab. 3.5 Podział dozwolonych pozycji na ścieżki – tryb 6.70 kb/s**

Na każdej ścieżce umieszczany jest dokładnie jeden impuls

### **Tryb 5.90**

Pozycje wektora kodowego dzielone są na dwie ścieżki. Na każdej z nich wstawiany jest jeden impuls.

Ścieżka	Impuls	Pozycje
---------	--------	---------

1	$i_0$	1,3,6,8,11,13,16,18,21,23,26,28,31,33,36,38
2	$i_1$	0,1,2,4,5,6,7,9,10,11,12,14,15,16,17,19,20, 21,22,24,25,26,27,29,31,32,34,35,36,37,39

**Tab. 3.6 Podział dozwolonych pozycji na ścieżki – tryb 5.90 kb/s**

Taki słownik zawiera  $16 \times 32 = 512$  pobudzeń algebraicznych.

### **Tryb 5.15 i 4.75 kb/s**

Pozycje dzielone są na dwie ścieżki, w różny sposób w zależności od numeru podramki. Każda ścieżka zawiera jedynie osiem pozycji. Dla każdej podramki istnieją dwa dopuszczalne podziały na ścieżki i przeszukiwane są wszystkie dozwolone kombinacje w obydwu dopuszczalnych podzbiorach ścieżek. W związku z powyższym, objętość słownika wynosi  $2 \cdot 8^2 = 128$  kombinacji dla każdej podramki.

## **3.4 Post-processing**

Na samym końcu, w dekodерze, syntetyczny sygnał mowy  $\hat{s}(n)$  filtrowany jest przy pomocy dwóch filtrów. Pierwszy z nich podbija formanty, czyli fragmenty widma sygnału związane z rezonansami traktu głosowego, charakterystyczne dla dźwięcznych fragmentów mowy. Obszary (w dziedzinie częstotliwości) o niskiej słyszalności są natomiast przytłumiane. Dzięki temu poprawia się subiektywna ocena jakości sygnału, ponieważ część dźwięczna sygnału, o dużej energii jest wzmacniana, a część słabo słyszalna, zawierająca również szum, jest przytłumiana. Drugi filtr lekko „podbija” wyższe częstotliwości.

Należy zaznaczyć, że do celów badania jakości mowy syntetycznej, przy użyciu metod obiektywnych (patrz p. 2.2), należy stosować sygnał nieprzefiltrowany przez postfiltry, jako że z tego punktu widzenia uszkodzają one sygnał.

# **4. Obsługa programu**

## **4.1 Wstęp**

Zadaniem aplikacji AMRCELP jest zilustrowanie działania głównych modułów koder (np. ekstraktora tonu krótłaniowego, czy słowników pobudzeń), dla różnych trybów działania (przepływności binarnych strumienia wyjściowego). Implementacja nie jest w 100% zgodna ze specyfikacją [ETSI AMR], ponieważ nie zawiera elementów takich jak kontrola wzmocnień przy dużej stopie błędów w kanale transmisyjnym, kwantyzacja wzmocnienia słownika algebraicznego i konwersja współczynników predykcji krótkookresowej do dziedziny częstotliwości. Moduły te nie mają jednak znaczącego wpływu na wynik kodowania sygnału, w przypadku, gdy nie transmituje się parametrów między koderem i dekodерem. W związku z powyższym, nazwy trybów działania implementacji koder (np. 12.2 kb/s, czy 5.90 kb/s) są umowne, jako że nie ma fizycznego kodowania parametrów, przy użyciu wyspecyfikowanej liczby bitów (patrz Tab. 3.1) - bezpośrednio po wyznaczeniu wszystkich parametrów, obliczany jest sygnał syntetyczny.

Zadaniem aplikacji jest zilustrowanie wszystkich etapów kodowania sygnału, przy pomocy wykresów przebiegów czasowych sygnałów pośrednich, używanych w

koderze (np. błędu predykcji, pobudzeń, sygnałów percepcyjnych etc.), wykresów widm wybranych sygnałów i charakterystyk filtrów, w dziedzinie częstotliwości oraz wyświetlania wybranych parametrów (np. okresu tonu krtaniowego i wzmocnienia, dla każdej podramki). Poza tym obliczane są miary obiektywne (SNR), służące do oceny jakości sygnału syntetycznego. Możliwe jest również zapisanie sygnału syntetycznego do pliku, w celu umożliwienia odsłuchów i porównania sygnału wyjściowego z sygnałem wejściowym.

## **4.2 Menu główne**

Po uruchomieniu programu AMRCelp.exe, pojawia się okno menu głównego. Umożliwia ono wybór pliku wejściowego i wyjściowego, wybór trybu pracy kodera oraz przejście do okna wykresów, lub zapisanie sygnału wyjściowego do pliku. Jedynym obsługiwanym formatem plików jest „wave”, PCM, mono, 16 bitów na próbkę, 128 kb/s (częstotliwość próbkowania 8000 Hz).

### **Plik wejściowy i wyjściowy**

Plik wejściowy podać można przy pomocy okienka kontekstowego, które pojawia się po wybraniu „Plik->Otwórz...” z menu na górze okna głównego. Ścieżka do wybranego pliku pojawi się w górnym polu tekstowym, a poniżej wyświetlona zostanie nazwa pliku wyjściowego, wygenerowana automatycznie na podstawie nazwy pliku wejściowego i trybu kodera. Nazwę tę można zmienić ręcznie.

### **Wybór trybu działania kodera**

Tryb kodera można wybrać z listy znajdującej się poniżej okienek tekstowych. Możliwy jest również wybór sygnału, który zostanie zapisany do pliku wyjściowego. Może to być sygnał wyjściowy przed, lub po post-processingu (patrz p. 3.4). W zależności od dokonanego wyboru, zaproponowana nazwa pliku wyjściowego ulegnie odpowiednim zmianom. Na tym etapie można również wyłączyć słownik adaptacyjny (predyktor długookresowy).

### **Wybór trybu działania aplikacji**

Na dole okienka głównego, znajdują się przyciski: „Wykresy” i „Kodowanie”. Pierwszy z nich powoduje przejście do okna analizy sygnałów, a drugi -zapisanie sygnału wyjściowego dekodera do pliku. Format pliku wyjściowego jest taki sam, jak format pliku wejściowego (patrz wyżej). W przypadku wybrania okna analizy sygnałów, sygnał wyjściowy nie zostanie zapisany do pliku.

#### **4.2.1 Okno analizy sygnałów**

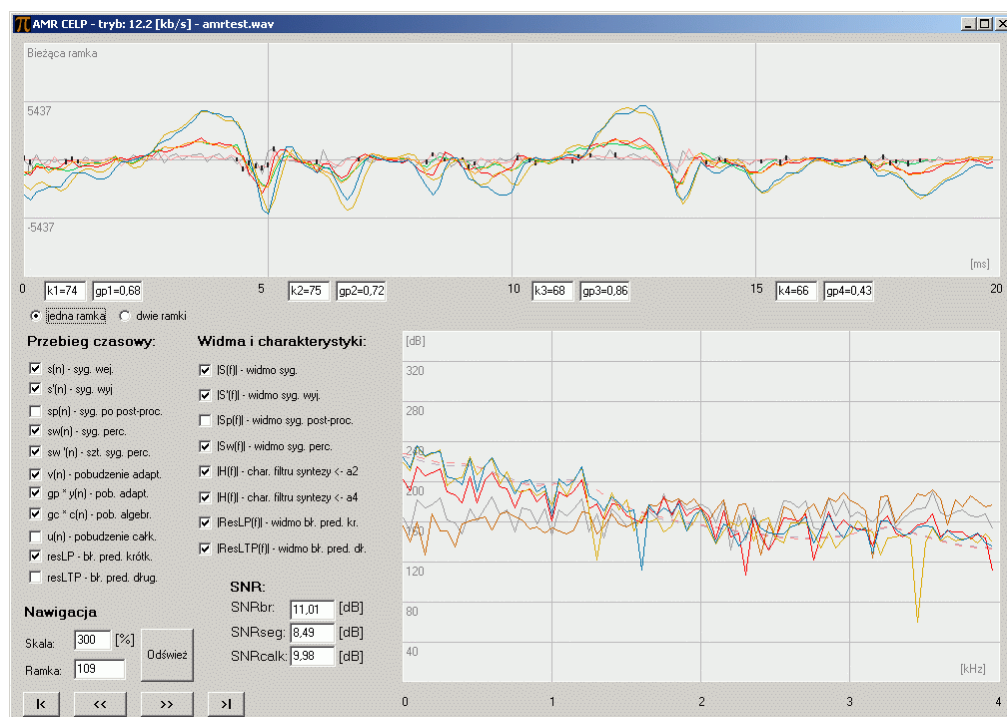
Okno analizy sygnałów przedstawione zostało na Rys. 4.1 i Rys. 4.2. Zawiera ono następujące komponenty:

- Wykresy przebiegów czasowych wybranych sygnałów
- Wykresy widm wybranych sygnałów oraz charakterystyk filtrów
- Menu pozwalającego na wybór sygnałów, które są wykresiane
- Parametry dodatkowe w okienkach tekstowych
- Menu nawigacyjne





**Rys. 4.1 Okno analizy sygnałów – wyświetlanie dwóch ramek**



**Rys. 4.2 Okno analizy sygnałów – wyświetlanie jednej ramki**

## Wykresy przebiegów czasowych

Na wykresie przebiegów czasowych, umieszczonym w górnej części okna, oglądać można jedną, lub dwie ramki sygnału. Wyboru trybu można dokonać za pomocą przycisków umieszczonych bezpośrednio pod wykresem, po lewej stronie. Gdy wykreślane są dwie ramki, lewa część wykresu odpowiada ramce poprzedniej, a część prawa – ramce bieżącej. Na wykres nakładana jest siatka pomagająca w

identyfikacji ramek i podramek sygnału oraz wartości dodatnich i ujemnych. Odstęp między dwiema sąsiednimi liniami pionowymi odpowiada jednej podramce, czyli 5 ms, lub 40 próbek. Oś wyskalowana jest w stosunku do bieżącej ramki. W związku z tym wartość -20 ms (tylko w trybie wykreślenia dwóch ramek) odpowiada początkowi poprzedniej ramki, a wartość 0 ms – początkowi bieżącej ramki sygnału. Linie poziome pozwalają na określenie przybliżonych wartości próbek sygnałów (etykiety z wartościami zmieniają się odpowiednio do skalowania wykresu).

Zmiana skalowania osi rzędnych możliwa jest przy pomocy okienka tekstowego w menu nawigacyjnym. Skalę ustawia się w procentach. Wartość 100% odpowiada zakresowi: od  $-2^{15}$  do  $2^{15}$ .

### **Wykresy widm sygnałów i charakterystyk filtrów**

W dolnej, prawej części okna, znajduje się pole zawierające wykresy wybranych widm i charakterystyk. Oś odciętych odpowiada zakresowi częstotliwości 0 - 4 kHz i jest wyskalowana liniowo. Oś rzędnych wyskalowana jest logarytmicznie. Aby ułatwić określanie dynamiki sygnałów, na wykres nałożona jest siatka wyskalowana co 1 kHz i co 20 dB.

Widma sygnałów wykreślane są na podstawie bieżącej ramki sygnału. Do ich obliczenia używana jest szybka transformata Fouriera FFT (ang. *Fast Fourier Transform*). Charakterystyka częstotliwościowa  $|H(f)|$ , filtru  $H(z)$ , „podąża” za widmem sygnału. Aby to uwidocznić, jest ona nakładana na widmo sygnału poprzez dodanie do niej energii błędu predykcji krótkookresowej, wyrażonej w decybelach.

### **Menu wyboru wykresów**

Lewa część menu wyboru wykresów, znajdującego się w lewej części okna, pozwala na wybór sygnałów, których przebiegi czasowe rysowane są na wykresie górnym. Aby wybrać dany sygnał, należy kliknąć myszką na okienku obok jego nazwy – okienko zostanie zaznaczone, a wykres odświeży się automatycznie. Nazwy podane w menu są skrótowe – po ustawieniu kursora nad daną pozycją, wyświetlana jest pełna nazwa. Menu pozwala na wybór następujących sygnałów:

- sygnał wejściowy
- sygnał wyjściowy bez post-processingu  
Sygnał ten powinien być możliwie bliski sygnałowi oryginalnemu. Używany jest do obliczenia SNR.
- sygnał wyjściowy po post-processingu  
Jest to ostateczny sygnał wyjściowy dekodera, który w normalnych warunkach (w systemie telekomunikacyjnym) podawany jest do odsłuchu.
- sygnał percepcyjny  
Sygnał oryginalny, przefiltrowany filtrem  $W(z)$ . Jest sygnałem docelowym kodera AMR CELP – używa się go w kryteriach służących do wyboru optymalnych pobudzeń.
- syntetyczny sygnał percepcyjny  
Sygnał wyjściowy dekodera (bez post-processingu), przefiltrowany filtrem  $W(z)$ . Za jego pomocą, można zweryfikować poprawność działania algorytmów, służących do konstrukcji pobudzenia filtru syntezy.
- pobudzenie adaptacyjne
- wzmocnione pobudzenie słownika adaptacyjnego
- wzmocnione pobudzenie adaptacyjne, przefiltrowane filtrem  $H(z)W(z)$

Jest to pobudzenie słownika adaptacyjnego  $v(n)$ , pomnożone przez wzmacnienie obliczone dla tego słownika  $g_p$ , przefiltrowane percepcyjnym filtrem syntezy (filtracja ciągła). Sygnał ten pozwala na obserwację fragmentów pobudzenia, które wymagają korekty pobudzeniem algebraicznym, generowanym przez słownik stały.

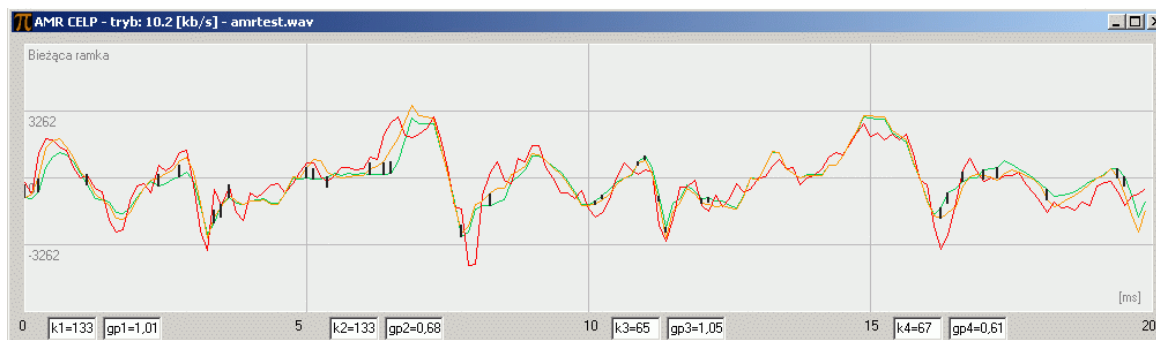
- wzmacnione pobudzenie algebraiczne

Pobudzenie algebraiczne wykreslane jest w postaci pojedynczych impulsów. Możliwe są trzy sposoby wykreslania:

- Impulsy „przyklejone są” do osi odciętych;
- Impulsy dorysowywane są do pobudzenia adaptacyjnego  $v(n)$ ;
- Impulsy dorysowywane są do przefiltrowanego pobudzenia adaptacyjnego (patrz Rys. 4.3);

To, która opcja jest aktywna, zależy od tego, które pobudzenia są wykreslane. Jeśli wykreslane jest pobudzenie  $v(n)$ , używana jest opcja b). Jeśli wykreslane jest przefiltrowane pobudzenie, a  $v(n)$  nie, używana jest opcja c). Jeśli ani  $y(n)$ , ani  $v(n)$  nie są wykreslane, używana jest opcja a). Należy podkreślić, że sumowanie przefiltrowanego pobudzenia adaptacyjnego i nieprzefiltrowanego pobudzenia algebraicznego nie ma sensu z matematycznego punktu widzenia, pozwala jednak w łatwy sposób „intuicyjnie” zweryfikować poprawność algorytmów konstruowania pobudzeń algebraicznych.

- błąd predykcji krótkookresowej
- błąd predykcji długookresowej



**Rys. 4.3 Sposób wykreslania pobudzenia algebraicznego (czarny), gdy przefiltrowane pobudzenie adaptacyjne jest wykreslane (zielony). W tle sygnał percepcyjny oryginalny (czerwony) i syntetyczny (pomarańczowy)**

Prawa część menu wyboru wykresów służy do definiowania, które widma i charakterystyki mają być wykreslane na wykresie dolnym. Możliwe opcje, to:

- widmo sygnału wejściowego
- widmo sygnału wyjściowego przed post-processingiem
- widmo sygnału wyjściowego po post-processingu
- widmo sygnału percepcyjnego
- charakterystyka filtru syntezy

Wykreślana jest charakterystyka  $|H(f)|$  filtru  $H(z)$ , obliczona na podstawie nieinterpolowanych współczynników predykcji krótkookresowej. W trybie 12.2 kb/s dostępne są dwa zestawy współczynników (obliczone dla drugiej i czwartej podramki), a więc można wykreślić dwie charakterystyki filtru syntezy. W pozostałych trybach możliwe jest wykreślenie jednej

charakterystyki, na podstawie współczynników obliczonych dla czwartej podramki (pierwsza opcja jest wtedy nieaktywna). Jak wspomniano wyżej, charakterystyki filtru syntezy nakładane są na widmo sygnału wejściowego, poprzez dodawanie do niech energii błędu predykcji krótkookresowej, wyrażonej w decybelach. Wykresy charakterystyk rysowane są za pomocą linii przerywanych.

- widmo błędu predykcji krótkookresowej
- widmo błędu predykcji długookresowej

### Parametry sygnałów

Pod wykresem przebiegów czasowych znajdują się okienka tekstowe, w których wyświetlana jest całkowita wartość okresu tonu krtaniowego  $k$  (podana w próbkach), wyznaczona w słowniku adaptacyjnym, oraz wartość wzmocnienia  $g_p$  dla pobudzenia znalezionej w tym słowniku. Wartości te wyświetlane są dla każdej podramki w bieżącej ramce sygnału, bezpośrednio pod wykresem.

W dolnej części okna analizy sygnałów, znajdują się trzy okienka tekstowe, w których wpisywane są trzy wartości SNR (patrz p. 2.2), liczone w różny sposób:

- $SNR_{br}$  - liczony dla bieżącej ramki sygnału
- $SNR_{seg}$  - uśredniona według wzoru (2.2) wartość SNR, liczona na podstawie wartości  $SNR_{br}$  wyznaczonych dla ramek o numerach  $1, \dots, N$ , gdzie  $N$  jest numerem bieżącej ramki
- $SNR_{calc}$  - wartość liczona od początku sygnału do końca bieżącej ramki

Zgodnie z powyższym, trzy wymienione wartości są sobie równe dla pierwszej ramki sygnału, a dla kolejnych ramek ich wartości zazwyczaj się różnią.

### Menu nawigacyjne

Menu to, znajdujące się w dolnej części okna wykresów, po lewej stronie, służy do przechodzenia do dowolnego fragmentu sygnału i skalowania wykresu górnego. Numer bieżącej ramki wyświetlany jest w polu tekstowym „Ramka:”. Wartość tę można zmienić ręcznie – przejście do wymaganej ramki nastąpi po kliknięciu na przycisku „Odśwież”.

Skala wykresu górnego w procentach, wyświetlana jest w okienku tekstowym „Skala:”. Można ją zmienić ręcznie – po kliknięciu na przycisku „Odśwież”, wykres zostanie przeskalowany. Wartość 100%, odpowiada zakresowi od  $-2^{15}$  do  $2^{15}$ .

Do przechodzenia między ramkami używa się następujących przycisków:

- $|<$  - Przejście do pierwszej ramki.
- $<<$  - Przejście do poprzedniej ramki. W zależności od odległości od początku sygnału, operacja ta może zająć różną ilość czasu, ze względu na fakt, że wymaga ponownego przeprowadzenia obliczeń od pierwszej ramki do poprzedniej.
- $>>$  - Przejście do kolejnej ramki.
- $>|$  - Przejście do ostatniej ramki. Wymaga różnej ilości czasu – potrzebne jest obliczenie wszystkich ramek począwszy od bieżącej, do końca.

Należy zauważyć, że w celu obliczenia wartości  $SNR_{seg}$  i  $SNR_{calc}$ , dla całego sygnału wejściowego, należy użyć przycisku  $>|$ .

## 5. Testy

### 5.1 Obserwacja sygnałów w koderze AMR\_CELP

Uruchom symulator AMRCELP z predykcją długookresową (tzn. ze słownikiem adaptacyjnym). Dla wybranej ramki mowy dźwięcznej:

- Porównaj widmo sygnału wejściowego  $S(f)$  i charakterystykę częstotliwościową filtra syntezy  $H(f)$ . Jaki jest związek  $H(f)$  i  $S(f)$ ?
- Zaobserwuj przebieg czasowy i widmo sygnału percepcyjnego  $sw(n)$ . Czym różnią się one od przebiegu czasowego i widma sygnału wejściowego  $s(n)$ ?
- Zaobserwuj sygnał pobudzający  $u(n)$  i jego składniki: pobudzenie adaptacyjne  $v(n)$  i algebraiczne  $c(n)$ . Jaka jest rola pobudzenia adaptacyjnego (wytwarzanego w predyktorze długookresowym) w kodowaniu mowy dźwięcznej?

### 5.2 Badanie jakości sygnału mowy w funkcji przepływności kodera

Dla wybranej frazy mowy przeprowadź symulacje kodera AMR CELP (z predykcją długookresową, z post-processing'iem) wykorzystując dostępne przepływności binarne. Notuj  $SNR$  i  $SNR_{seg}$  [dB] oraz odsłuchuj frazy poddane kompresji. Wykreśl  $SNR_{seg}$  w funkcji przepływności binarnej. Powyżej jakiej przepływności jakość mowy można uznać za wystarczającą dla rozmowy telefonicznej?

### 5.3 Badanie wpływu predykcji długookresowej na jakość sygnału

Powtórz symulacje z poprzedniego punktu, ale bez użycia słownika adaptacyjnego (predykcji długookresowej). Wykreśl  $SNR_{seg}$  w funkcji przepływności binarnej w tym samym układzie współrzędnych. Pamiętaj, aby zredukować przepływność w stosunku do wybranej z menu, gdyż opóźnienie ( $k$ ) i wzmacnienie ( $g_p$ ) sygnału ze słownika adaptacyjnego nie są obecnie transmitowane. Należy tu wykorzystać Tab. 3.1. Np. przepływność 12.2 kbit/s (244 bity/ramkę razy 50 ramek/s) należy zmniejszyć o  $(30+16) \cdot 50 = 2300$  bit/s; będzie ona wynosiła 9.9 kbit/s.

Czy opłaca się stosować predykcję długookresową (słownik adaptacyjny) czy należy z niej zrezygnować, przeznaczając zaoszczędzone bity na zakodowanie większej liczby impulsów sygnału pobudzającego?

## Bibliografia

- [Atal 82] Atal, B. S. and J. R. Remde (1992), *A New Method of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates*, IEEE ICASSP
- [bell] <http://www.bell-labs.com/org/1133/Heritage/Vocoder/>
- [Drozdek 99] A. Drozdek (1999), *Wprowadzenie do kompresji danych*, Wydawnictwo Naukowo – Techniczne, Warszawa
- [Dymarski 02] P. Dymarski (2002), *Predykcyjne i wektorowe metody kompresji sygnału mowy*, Oficyna Wydawnicza PW, Warszawa
- [ETSI AMR] Norma ETSI EN 301 704 V7.2.1 (2000-04),  
*Digital cellular telecommunications system (Phase 2+)*  
*Adaptive Multi-Rate (AMR) speech transcoding*  
*(GSM 06.90 version 7.2.1 Release 1998)*
- [ETSI RPE] Norma ETSI (1992), *Recommendation GSM 6.10 Full-Rate Speech Transcoding*
- [G. 728] Norma ITU-T G.728: *Coding of speech at 16 kbit/s using low-delay code excited linear prediction*
- [Haykin] S. Haykin (1988), *Digital Communications*, John Wiley & Sons
- [Hanzo] L. Hanzo (2001), *Voice Compression and Communications*, John Wiley & Sons
- [Jayant 90] N. Jayant, P. Noll (1990), *Digital coding of waveforms*, Prentice Hall
- [Kondoz 95] A. M. Kondoz (1995), *Digital Speech*, Wiley & Sons
- [Kula 02] S. Kula (2002), *Obiektywne i subiektywne metody badania jakości mowy przesyłanej w sieciach telekomunikacyjnych, oraz Pomiary w telekomunikacji*
- [Marm] Paul Marmelstein and Yasheng Qian (2000), *Analysis by Synthesis Speech Coding with Generalized Speech Prediction*, University of Quebec, Canada
- [NCS 92] National Communications Systems (1992), *Details to Assist in Implementation of Federal Standard 1016 CELP*, Arlington, VA
- [Rabiner 75] L. R. Rabiner, B. Sold (1975), *Theory and application of digital signal processing*, Prentice – Hall
- [Wai 03] Wai C. Chu (2003), *Speech coding algorithms: Foundation and evolution of standardized coders*, Wiley – Interscience