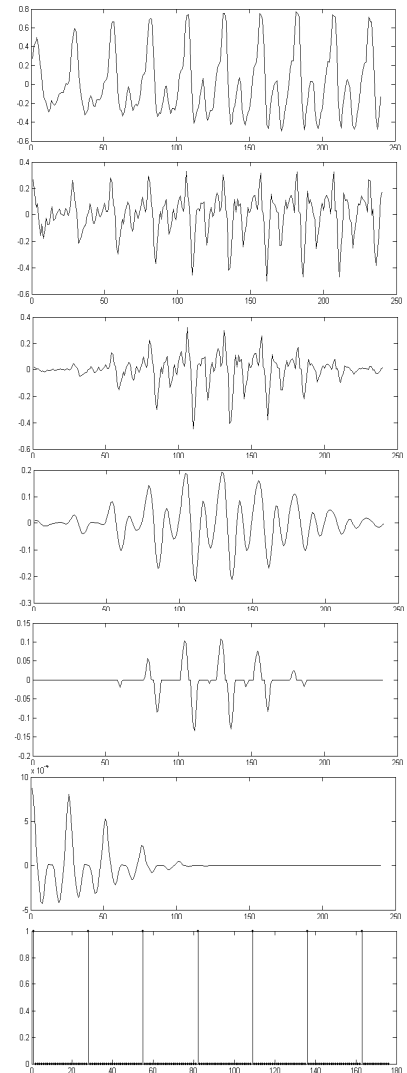
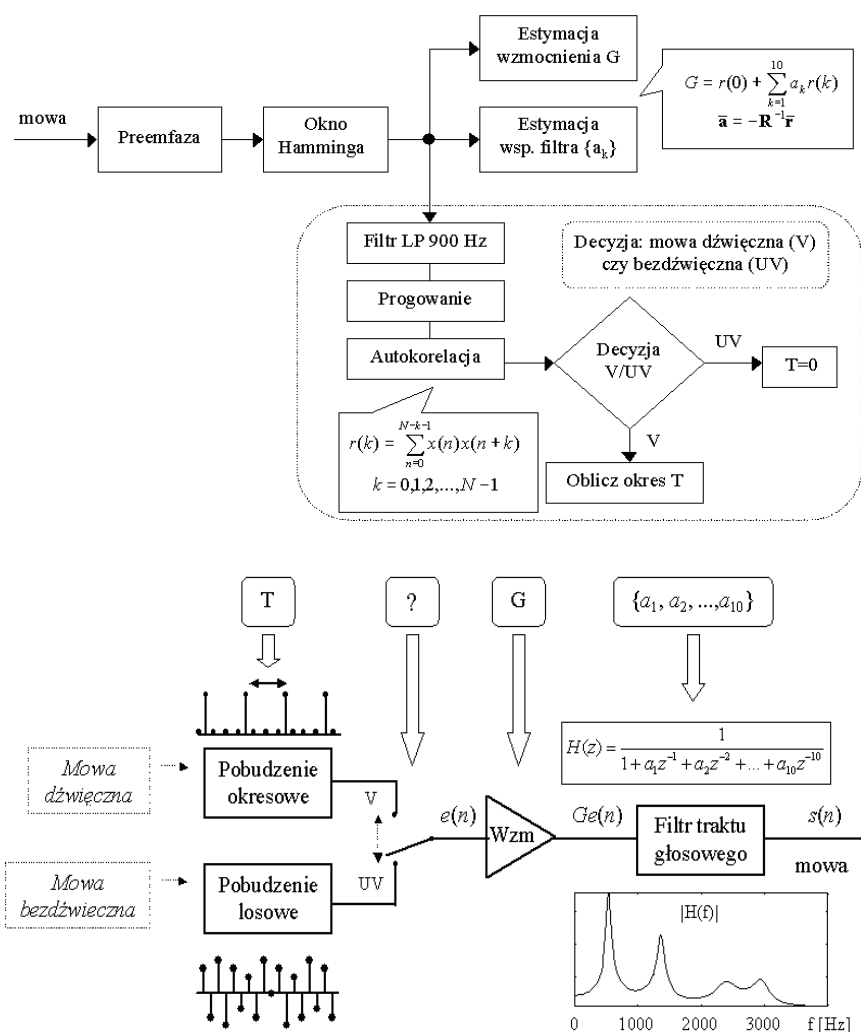


## 1. Prosty koder/dekoder (2+1 pkt)

Poniżej przedstawiono koder oraz dekodery (lewa strona rysunku) uproszczonego modelu kompresji mowy LPC-10 oraz przykładowe sygnały, dźwięcznego fragmentu mowy (prawa strona rysunku) w kolejności od góry do dołu: sygnał, preemfaza, okno Hamminga, filtr LP, progowanie, autokorelacja, pobudzenie dźwięczne. Kod realizujący ten schemat znajduje się w pliku **lpc10.m** (rozdział 19, tabela 19-4 [TZ2005]).



- Wybierz trzy fragmenty mowy reprezentujące: głoskę dźwięczną, głoskę bezdźwięczną oraz stan przejściowy, następnie wyświetl:
  - sygnał czasowy oraz widmo gęstości widmowej mocy sygnału przed oraz po preemfazie,
  - charakterystykę amplitudowo-częstotliwościową filtra  $H(z)$ ,
  - sygnał czasowy przed i po progowaniu,
  - funkcję autokorelacji sygnału z zaznaczonymi progami,
  - decyzję o „dźwięczności” głoski, jeżeli jest dźwięczna dodatkowo częstotliwość tonu podstawowego,
  - ramkę zsyntezowaną (porównaj ramkę oryginalną i zsyntezowaną w dziedzinie czasu oraz częstotliwości).
- Sprawdź jak działa program (jaka jest jakość mowy) w przypadku zmiany liczby biegunów filtra  $H(z)$  na 10, 8, 6, 4, 2.
- Wykonaj kompresję sygnałów: **mowa1.wav** (czysty dźwięk, pojedynczy mówca), **mowa2.wav** (dwóch mówców + tło), **mowa3.wav** (próbka muzyczna).

4. (**opcjonalnie**) Wykonaj kwantyzację współczynników „a” transmitancji  $H(z)$ . Kolejne pary zespolone skwantuj na następującej liczbie bitów: 8, 6, 6, 4, 4. Sprawdź dla jakich kombinacji uzyskasz największą kompresję danych (oblicz przepływność bitową kodeka przyjmując 8 bitów dla  $G$  oraz 6 bitów dla  $T$ ) przy zadowalającej jakości mowy.

## 2. Sygnał pobudzenia (2+1 pkt)

Wykonaj następujące eksperymenty z sygnałem pobudzenia. Dekoder zmodyfikuj tak aby:

- ignorował decyzję V/UV (dźwięczna/bezdźwięczna), ustaw pobudzenie zawsze na bezdźwięczne ( $T=0$ ),
- zachował decyzję V/UV ale w przypadku głoski dźwięcznej obniżył dwukrotnie częstotliwość tonu podstawowego,
- zachował decyzję V/UV ale w przypadku głoski dźwięcznej ignorował informację o częstotliwości tonu podstawowego, ustaw go na stałą wartość (np.  $T=80$ ),
- ignorował decyzję V/UV, ustaw pobudzenie na bezdźwięczne lecz zamiast szumu użyj próbek z pliku `coldvox.wav`.

W koderze z ćwiczenia 1 zastosowano impulsowe pobudzenie dla głosek dźwięcznych. Jest to mało realistyczna sytuacja. Zastosuj pobudzenie tzw. sygnałem „resztkowym”. W tym celu:

1. Wybierz dźwięczny fragment mowy o stałej amplitudzie i częstotliwości tonu podstawowego.
2. Oblicz  $H(z)$  dla tego fragmentu, następnie przefiltruj go filtrem „odwrotnym” o transmitancji  $1/H(z)$ , otrzymasz sygnał resztkowy.
3. W dekodrze, jako pobudzenie dla wszystkich głosek dźwięcznych wykorzystaj jeden okres sygnału resztkowego. **Opcjonalnie** oblicz średni sygnał resztkowy uśredniając kilka okresów tego sygnału, a następnie użyj go jako pobudzenia okresowego.

## 3. Sygnał resztkowy (1+1 pkt)

Zmodyfikuj koder i dekodser tak, aby zamiast pobudzenia okresowego wykorzystać sygnał resztkowy (ang. *residual*), którego sposób wyznaczania podano w ćwiczeniu 2. Wykonaj kodowanie z pełnym sygnałem resztkowym, tj. dla każdej ramki dźwięcznej oblicz sygnał resztkowy, wykorzystując odwrotność filtru  $H(z)$ , obliczonego w tej ramce. Nie upraszczaj sygnału do jednego okresu. Porównaj jakość tak wykonanej kompresji dla wszystkich trzech próbek dźwiękowych.

**Opcjonalnie** wykonaj uproszczenie sygnału resztkowego według następującej procedury:

- oblicz widmo sygnału resztkowego:  $w = \text{abs}(\text{fft}(x))$ , gdzie  $x$  to ramka sygnału wejściowego o długości 256 próbek,
- uprość (wygładź) widmo  $w$  poprzez zastosowanie filtru LP na pierwszych 128 współczynnikach widma,
- aproksymuj widmo za pomocą wielomianu rzędu od 5 do 10 (sprawdź jaki rząd będzie odpowiedni),
- prześlij współczynniki aproksymacji widma z koder do dekodera i w dekodrze odtwórz sygnał pobudzenia z obwiedni widma metodą odwrotnej dyskretnej transformacji Fouriera (pamiętaj o (a)symetrii widma sygnału o wartościach rzeczywistych).

W ten sposób, 256 współczynników widma sygnału rezydualnego zostało zapisanych za pomocą kilku liczb, tracąc przy okazji informację o fazie poszczególnych harmonicznych. Jest to jak najbardziej uzasadnione, ponieważ ilość informacji (w sensie entropii) w sygnale rezydualnym jest znacznie mniejsza niż 256 współczynników, a człowiek nie jest wrażliwy na fazę w sygnale monofonicznym. Podobna operacja jest wykonywana w koderze mowy HVXC standardu MPEG-4.

## 4. Inne algorytmy (0+2 pkt)

Wszystkie zadania w tym punkcie są opcjonalne.

Kluczowym algorytmem w koderze LPC-10 jest detekcja głosek dźwięcznych. Błędna detekcja powoduje zawsze znaczne pogorszenie jakości dźwięku. Spróbuj zaprojektować i wykonać lepszy algorytm U/V. W tym celu możesz wykorzystać następujące obserwacje:

- głoska dźwięczna ma zazwyczaj wyższą energię niż bezdźwięczna (trudno krzyczyć szeptem ;-)),

- częstotliwość tonu podstawowego w sąsiednich ramkach jest podobna (zmienia się w sposób ciągły i powoli),
- częstotliwość tonu podstawowego nie zmienia się skokowo z ramki na ramkę (można go śledzić adaptacyjnie),
- ton podstawowy jest silnie okresowy więc można spróbować zastosować analizę częstotliwościową do jego wyznaczania; dodatkowo zauważ, że głoski dźwięczne posiadają ton podstawowy i jego wyższe harmoniczne.

Zastąp funkcję Matlaba odwracająca macierz: `inv(R)` algorytmem Levinsona, opisanym w [TZ2005, rozdz. 20]. Ta funkcja jest wykorzystywana w telefonach komórkowych ze względu na mniejszą złożoność obliczeniową.

Spróbuj przestrajać filtr syntezy częściej niż co 160 próbek. W tym celu należy dokonać interpolacji stanów filtra (zastosowanie współczynników LSP, str. 567 w [TZ2005]).

Przedstaw filtr syntezy w postaci kratowej, przelicz współczynniki liniowej predykcji „a” na gamma i skwantuj je. Filtr syntezy w postaci kratowej został opisany w rozdz. 20 [TZ2005]. Struktura kratowa jest mniej poddana na kwantyzację, więc nie powinno zachodzić wzbudzenie filtra w dekodерze.