AKUSTYKA MOWY

METODY POPRAWY ZROZUMIAŁOŚCI MOWY

PLAN PREZENTACJI

- Szumy i zakłócenia
- Zniekształcenia
- Metody redukcji zakłóceń
- Metody redukcji zniekształceń
- Ocena zrozumiałości mowy
- Analiza mowy w procesach sądowych

SZUMY I ZAKŁÓCENIA

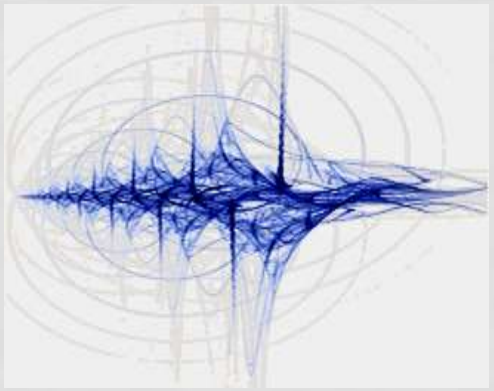
- Zarejestrowane sygnały utrudniające, bądź uniemożliwiające prawidłową postrzeganie sygnału użytecznego (np. mowy)
  - Szum klimatyzacji
  - Odgłosy z ulicy
  - Hałas przemysłowy
  - Przydźwięk z sieci
  - Inne

SZUMY I ZAKŁÓCENIA

- Zakłócenia:
  - Ze względu na zajmowane pasmo:
    - Wąskopasmowe – np. przydźwięk sieciowy
    - Szerokopasmowe – np. szum biały, różowy, brązowy
  - Ze względu na charakter procesu:
    - Stacjonarne – np. szum biały, różowy, brązowy
    - Niestacjonarne – np. odgłosy ruchu drogowego

ZNIEKSZTAŁCENIA

- Przykładowe źródła zniekształceń:
  - Przesterowanie sygnału – przekroczenie dostępnego zakresu
  - Nierówna charakterystyka przenoszenia – np. mikrofonu
  - Wynikające z charakterystyki kanału – podbicia/tłumienie określonych pasm częstotliwości
  - Jako rezultat przetwarzania dźwięku
  - Jako wynik wadliwych mechanizmów – np. głowicy odtwarzacza analogowego
  - Inne



REDUKCJA ZAKŁÓCEŃ I ZNIEKSZTAŁCEŃ

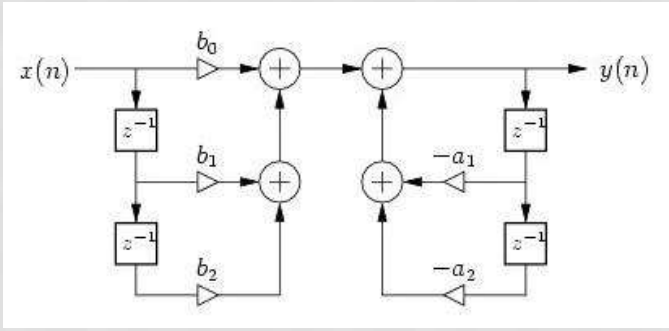
FILTRACJA

- Charakterystyka filtru:

$$H(z)=\frac{B(z)}{A(z)}=\frac{b_0+b_1z^{-1}+b_2z^{-2}+\ldots+b_Nz^{-N}}{1+a_1z^{-1}+a_2z^{-2}+\ldots+a_Mz^{-M}}$$

- Filtracja sygnału:

$$y_n=\sum_{k=0}^{n-1}h_kx_{n-k}$$



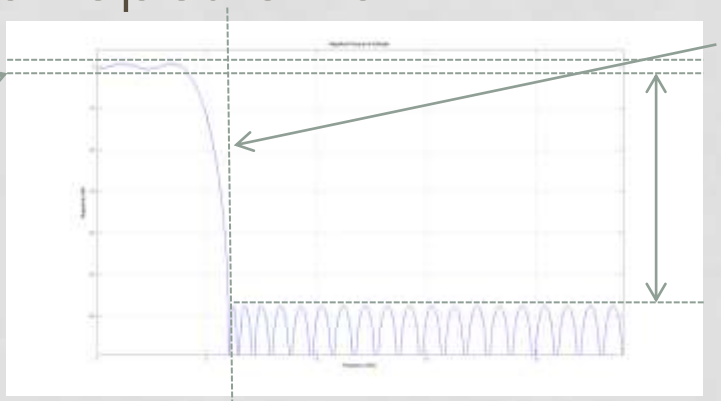
gdzie h_k to k-ty współczynnik filtru;
x jest sygnałem wejściowym, a y sygnałem po zastosowaniu filtracji

- Sygnał oryginalny:

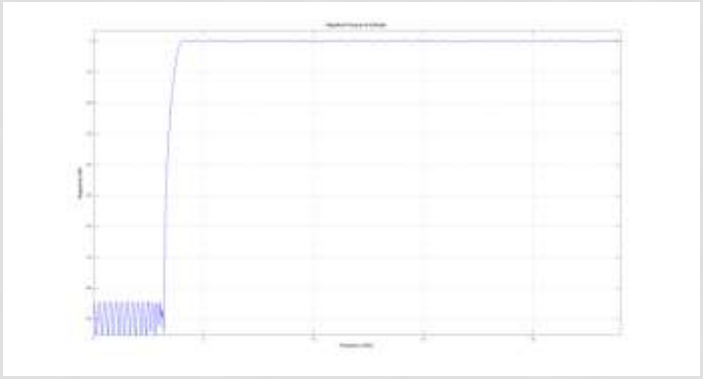
FILTRACJA

- Filtracja dolnoprzepustowa

Zafalowanie charakterystyki w pasmie przepustowym

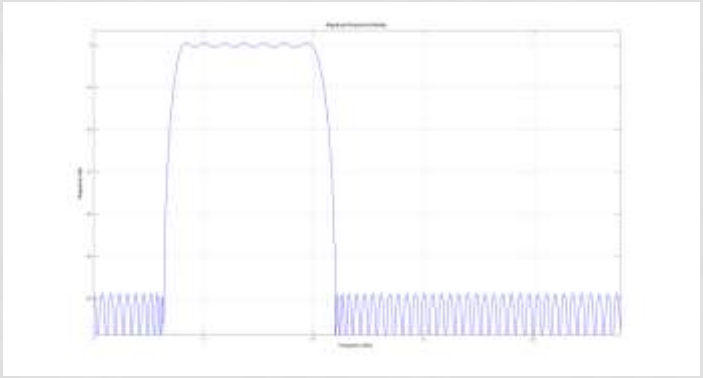


- Filtracja górnoprzepustowa

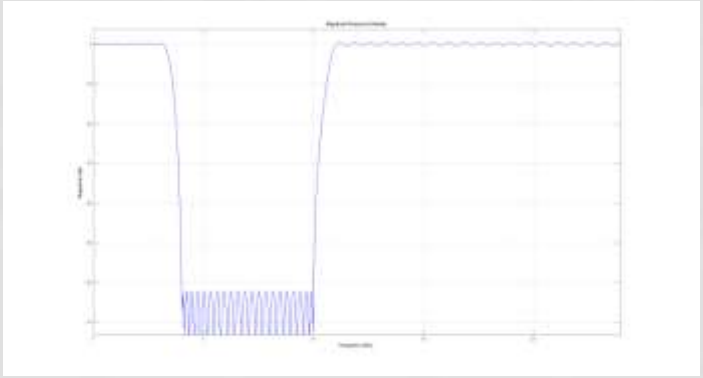


FILTRACJA

- Filtracja pasmowo-przepustowa



- Filtracja pasmowo-zaporowa



FILTRY ADAPTACYJNE

- Wymagają dodatkowego sygnału referencyjnego
- Zmienna charakterystyka filtru – konieczność wyboru algorytmu adaptacji współczynników opisujących filtr

- Algorytmy adaptacji – podział ze względu na dziedzinę przetwarzania:
  - Dziedzina czasu: LMS, NLMS, DLMS, RLS, …
  - Dziedzina częstotliwości: FDAF, TDAFDFT, PDFDAF, …

FILTRY ADAPTACYJNE

- Algorytm NLMS (Normalized Least Mean Squares)
- Minimalizacja chwilowej wartości błędu średniokwadratowego

$$w_i(k+1) = w_i(k) + \frac{2\beta}{\sigma + \sum_{i=0}^{L} x_0{}^2(k-i)} \cdot e(k) x_0(k-i)$$

gdzie w_i oznacza i-ty współczynnik filtru, β to krok adaptacji, x_0 określa filtrowany sygnał, a e – sygnał błędu

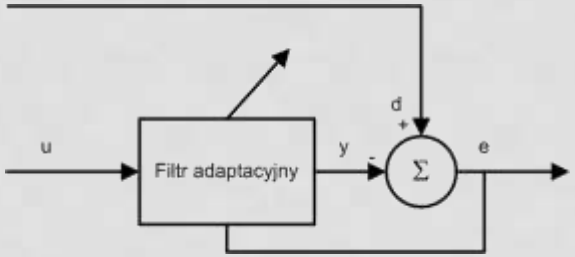
- Wady/zalety: niska złożoność, względnie słaba zbieżność algorytmu (aczkolwiek lepsza niż w przypadku LMS)

FILTRY ADAPTACYJNE

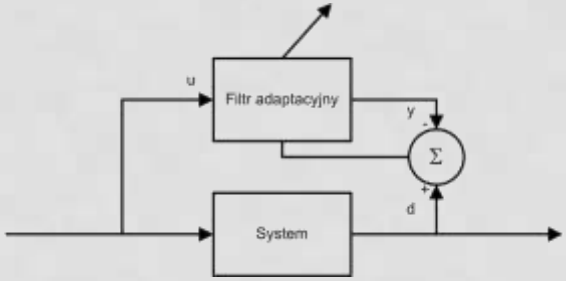
- Istotne cechy filtrów adaptacyjnych:
  - Zbieżność – czas/ilość iteracji koniecznych do ustalenia charakterystyki filtru
  - Błąd średniokwadratowy – określa stopień dopasowania filtru do modelowanego procesu
  - Złożoność obliczeniowa – związana pośrednio z rzędem filtru oraz wykorzystanym algorytmem adaptacyjnym
  - Rząd filtru
  - Stabilność

FILTRY ADAPTACYJNE

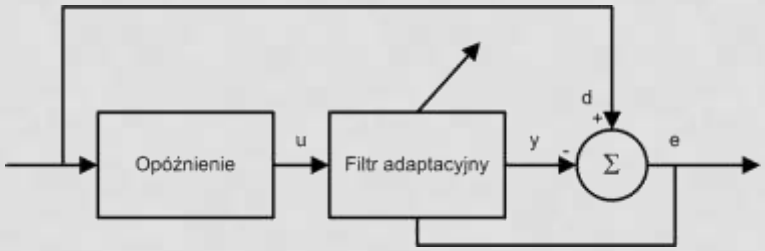
- Konfiguracje wykorzystania filtrów adaptacyjnych
  - Redukcja zakłóceń



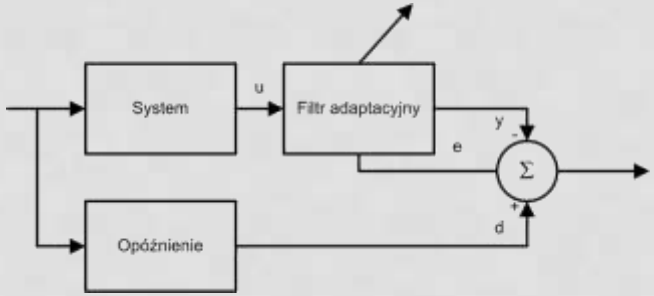
  - Identyfikacja systemów



  - Predykcja

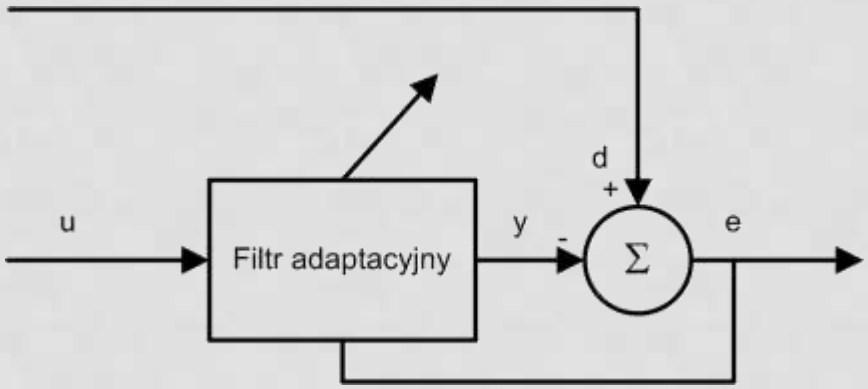


  - Odwzorowanie odwrotne



FILTRY ADAPTACYJNE

- Konfiguracja do redukcji zakłóceń
- Założenie:
  - Addytywny charakter szumu
  - Znany sygnał zakłócenia



zarejestrowany sygnał:

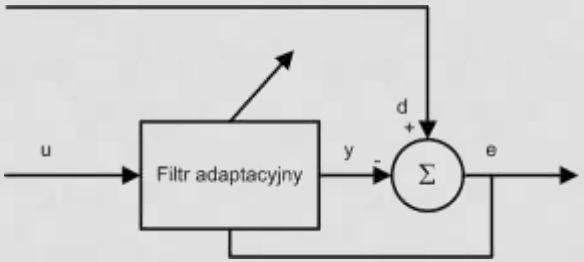
zakłócenie:

sygnał wyj.:

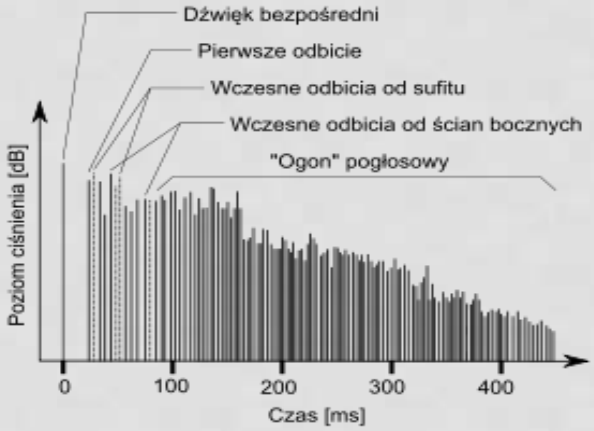
gdzie *d* jest sygnałem wejściowym, *u* sygnałem zakłócenia, natomiast *e* to sygnał wyjściowy

FILTRY ADAPTACYJNE

- Echo – w skrajnych przypadkach może powodować spadek zrozumiałości mowy
- W celu redukcji echa można wykorzystać filtr adaptacyjny

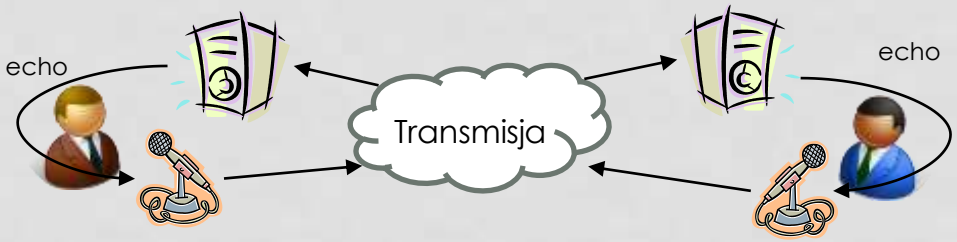


- Jako sygnał referencyjny wykorzystuje się odpowiednio opóźniony sygnał wejściowy

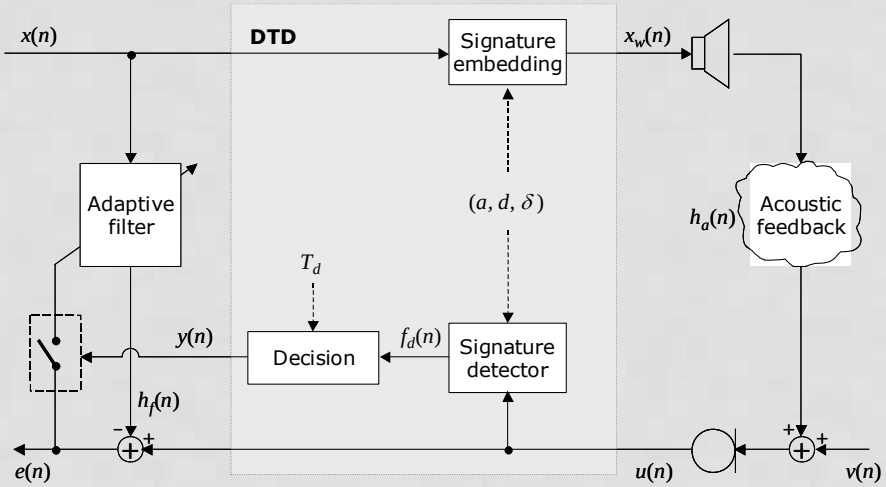


FILTRY ADAPTACYJNE

- Echo w systemach komunikacji głosowej
- Powoduje dyskomfort w trakcie rozmowy
- Generuje niepotrzebny ruch sieciowy



- Redukcja echa z wykorzystaniem filtru adaptacyjnego
- Wsparcie znakowaniem wodnym – kluczowanie adaptacji charakterystyki filtru



ODEJMOWANIE WIDMOWE

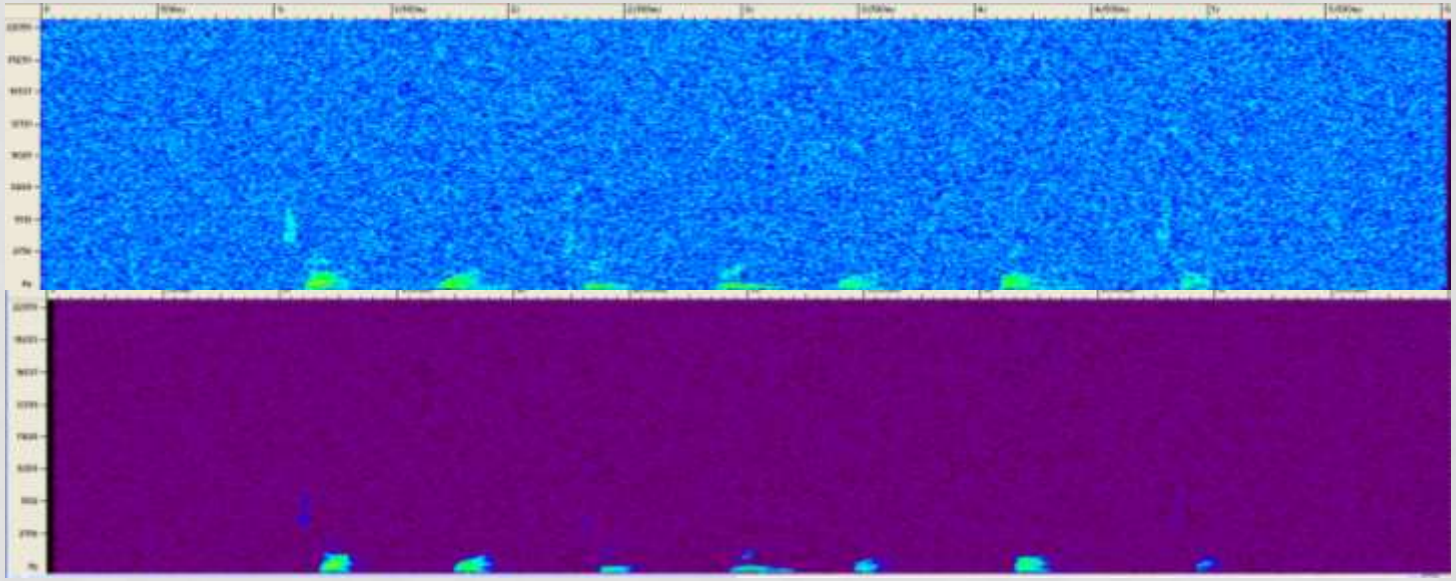
- Wymagane pozyskanie informacji na temat sygnału zakłócenia – np. przez ręczną segmentację sygnału
- Usunięcie z widma amplitudowego sygnału, uśrednionego widma zakłócenia, zgodnie z zależnością:

$$\left|S\left(e^{j\omega}\right)\right| = \left|X\left(e^{j\omega}\right)\right| - \alpha\left|N\left(e^{j\omega}\right)\right| \qquad S_{out}\left(e^{j\omega}\right) = \begin{cases} \left|S\left(e^{j\omega}\right)\right| e^{j\theta_N}, & \left|S\left(e^{j\omega}\right)\right| > 0 \\ 0, & \left|S\left(e^{j\omega}\right)\right| \leq 0 \end{cases}$$

gdzie X to sygnał wejściowy, N reprezentuje zakłócenie, a α oznacza głębokość odejmowania widmowego $\in (0,1)$

ODEJMOWANIE WIDMOWE

- Przykład przetworzenia sygnału mowy



- Wady/zalety: w przypadku dużych wartości parametru α, może pojawić się niepożądany efekt w postaci tzw. szumu muzycznego

EKSPANSJA WIDMA

- Założenie – wyraźny odstęp między poziomem sygnału mowy i szumu
- Wzrost zrozumiałości uzyskuje się przez zwiększenie SNR
- Zastosowanie progu w dziedzinie częstotliwości, np. liniowego:

$$F(n) = \frac{(N-n)a + nb}{N} Y(n)$$

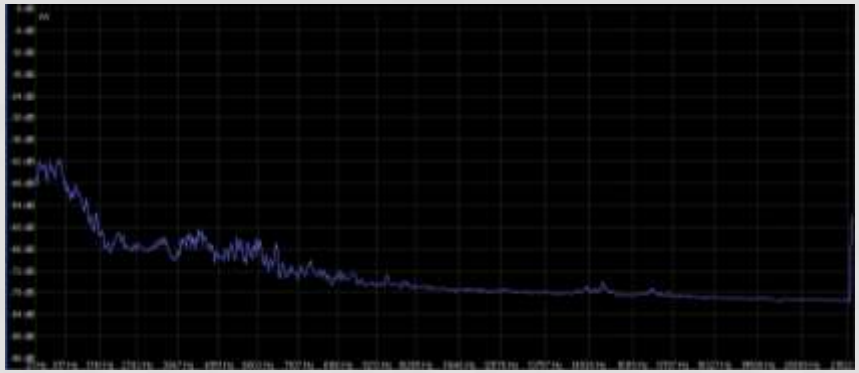
gdzie Y(n) to uśrednione widmo szumu, natomiast a i b to współczynniki opisujące funkcję progu; n oznacza numer próbki widma

EKSPANSJA WIDMA

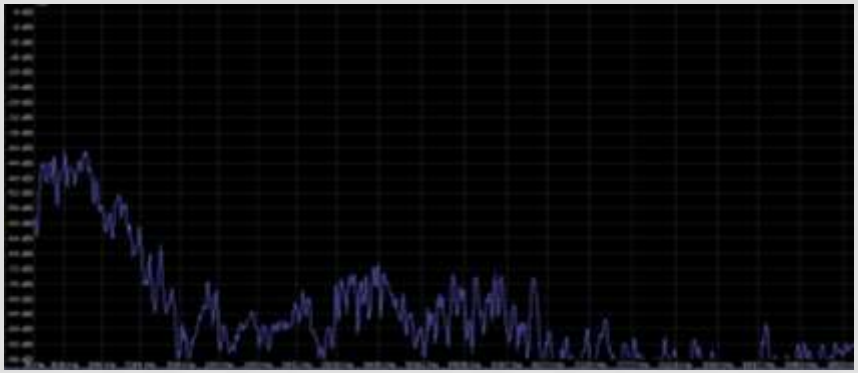
- Po zastosowaniu progu, wynikowy sygnał uzyskuje następującą postać:

$$V = \{\mathrm{Re}[V(n)], \mathrm{Im}[V(n)]\} = \begin{cases} \mathrm{Re}[V(n)] = \dfrac{|X(n)|}{|F(n)|}\mathrm{Re}[X(n)] \wedge \mathrm{Im}[V(n)] = \dfrac{|X(n)|}{|F(n)|}\mathrm{Im}[X(n)], |X(n)| < |F(n)| \\ \mathrm{Re}[V(n)] = \mathrm{Re}[X(n)] \wedge \mathrm{Im}[V(n)] = \mathrm{Im}[X(n)], |X(n)| \geq |F(n)| \end{cases}$$

Mowa zakłócona szumem różowym:

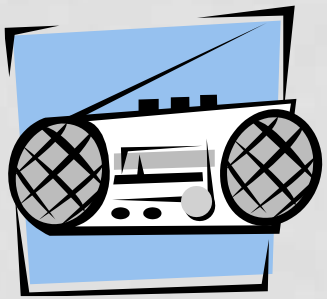


Sygnał po zastosowaniu ekspansji widma:



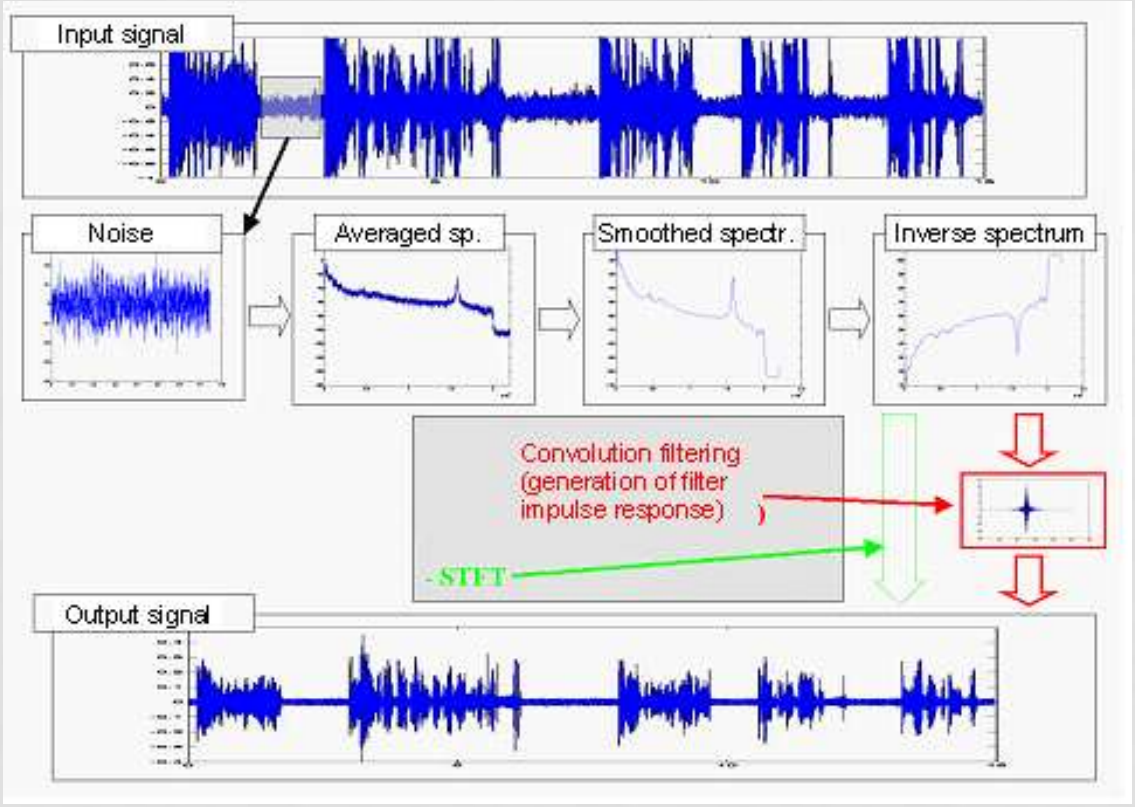
WYBIELANIE

- Założenie – wyrównanie charakterystyki widmowej zakłócenia zmniejszy jego uciążliwość
- Etapy działania algorytmu:
  - Estymacja szumu – automatyczna, bądź manualnie przez wybór segmentów zawierających zakłócenie
  - Obliczenie średniego widma amplitudowego szumu
  - Wygładzenie i odwrócenie otrzymanego widma
  - Wygenerowanie filtru odwrotnego
  - Filtracja sygnału



WYBIELANIE

- Przykład przetworzenia sygnału mowy zakłóconej sygnałem piłokształtnym:



Mowa zakłócona:

Sygnał po operacji wybielenia:

REDUKCJA TRZASKÓW

- Redukcja zakłóceń impulsowych
- Trzaski – krótka lokalna nieciągłość sygnału ~1ms
- Występują często w nagraniach archiwalnych

- Dwuetapowe przetwarzanie sygnału
- Detekcja wystąpień trzasków
  - Progowa analiza sygnału poddanego filtracji górnoprzepustowej
  - Wykorzystanie modelu autoregresywnego i analiza pobudzenia w celu detekcji wartości przekrazczających zadany próg

$$x_n = \sum_{i=1}^{P} a_i x_{n-1} + e_n$$

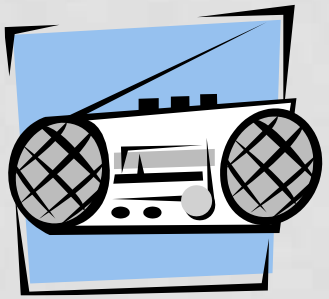
gdzie a_i to współczynniki filtru, P to rząd modelu, e_n oznacza pobudzenie

REDUKCJA TRZASKÓW

- Rekonstrukcja sygnału
  - Zazwyczaj możliwa jest interpolacja do około 100 próbek sygnału (dla f_s=44.1kHz) – najczęściej realizowane w dziedzinie czasu
  - Dla dłuższych fragmentów, częściej wykorzystuje się interpolację w dziedzinie częstotliwości
- Przykładowe algorytmy interpolacji:
  - Filtracja medianowa – prosta ale słabe efekty
  - LSAR – least squares AR
  - MAP – maximum a posterioro AR
  - ARMA – autoregressive-moving-average
  - Audio inpainting

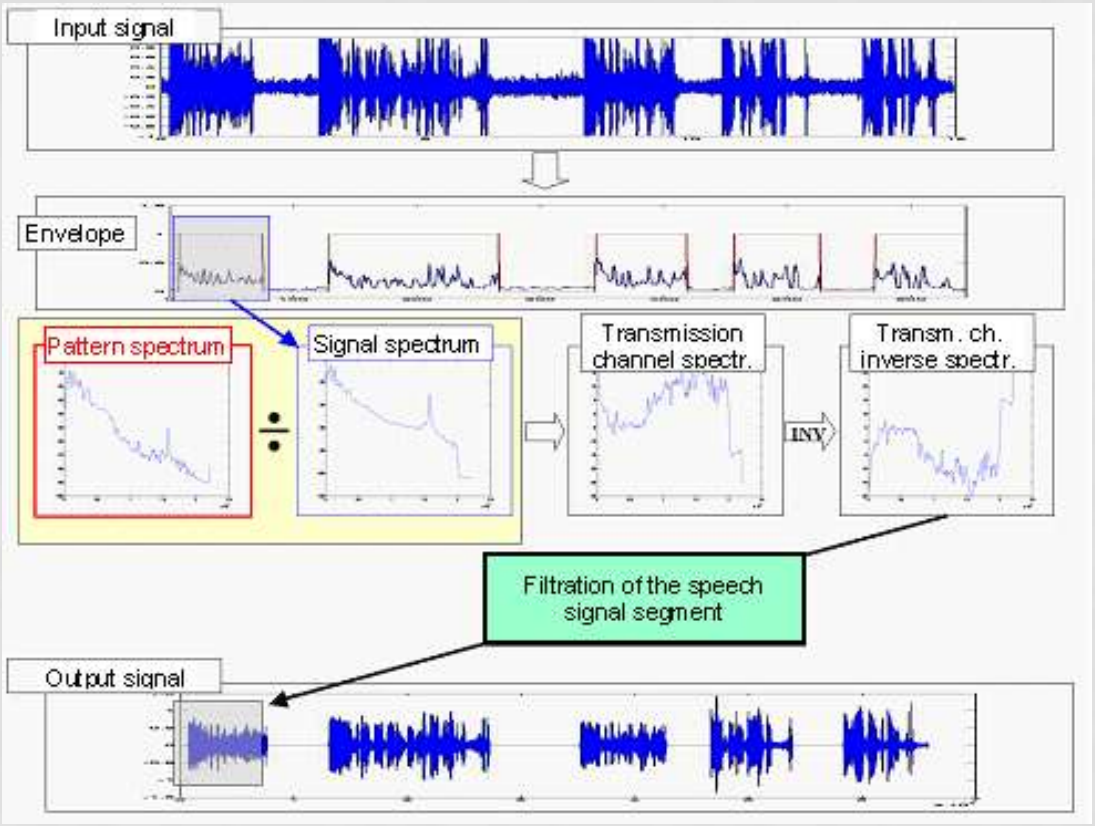
ŚLEPY ROZPLOT

- Założenie – znana charakterystyka sygnału użytecznego
- Redukcja liniowych zniekształceń, np. wynikających z charakterystyki kanału transmisyjnego
- Etapy działania algorytmu:
  - Obliczenie średniego widma sygnału użytecznego
  - Obliczenie średniego widma sygnału zniekształconego w segmentach zawierających mowę
  - Porównanie obu widm i oszacowanie charakterystyki zniekształcenia
  - Wygenerowanie filtru odwrotnego i filtracja



ŚLEPY ROZPLOT

- Przykład przetworzenia sygnału mowy zakłóconej sygnałem piłokształtnym:

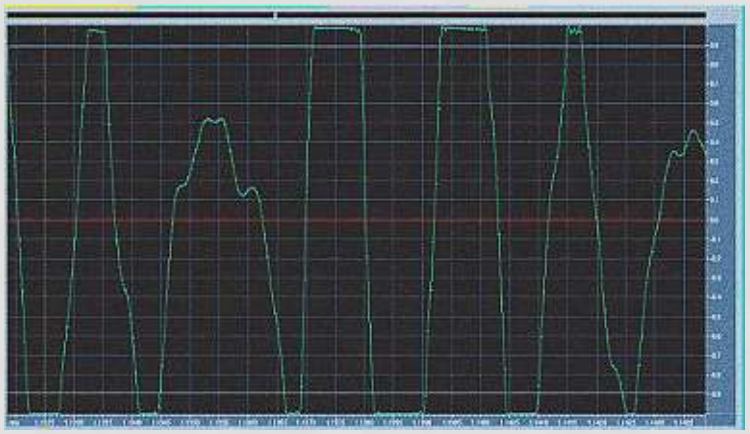


Mowa zniekształcona:

Sygnał po operacji rozplotu:

REDUKCJA PRZESTEROWAŃ

- Rekonstrukcja przesterowanego sygnału
- Przesterowanie związane jest z utratą informacji odnośnie sygnału
- Dwuetapowe postępowanie:
  - Detekcja przesterowań w sygnale – np. na podstawie podobieństwa kolejnych próbek w sygnale
  - Rekonstrukcja sygnału



REDUKCJA PRZESTEROWAŃ

- Rekonstrukcja odbywa się poprzez ekstrapolację nieznieksztalconych próbek sygnału
- Wykorzystanie dwukierunkowej ekstrapolacji (w przód i tył) oraz przetworzenie próbek sygnałów zgodnie z zależnością:

$$z_n = 0.5\left\langle\left[1+\cos\left(\frac{n-j}{k-j}\pi\right)\right]x_n + \left[1-\cos\left(\frac{n-j}{k-j}\pi\right)\right]y_n\right\rangle$$

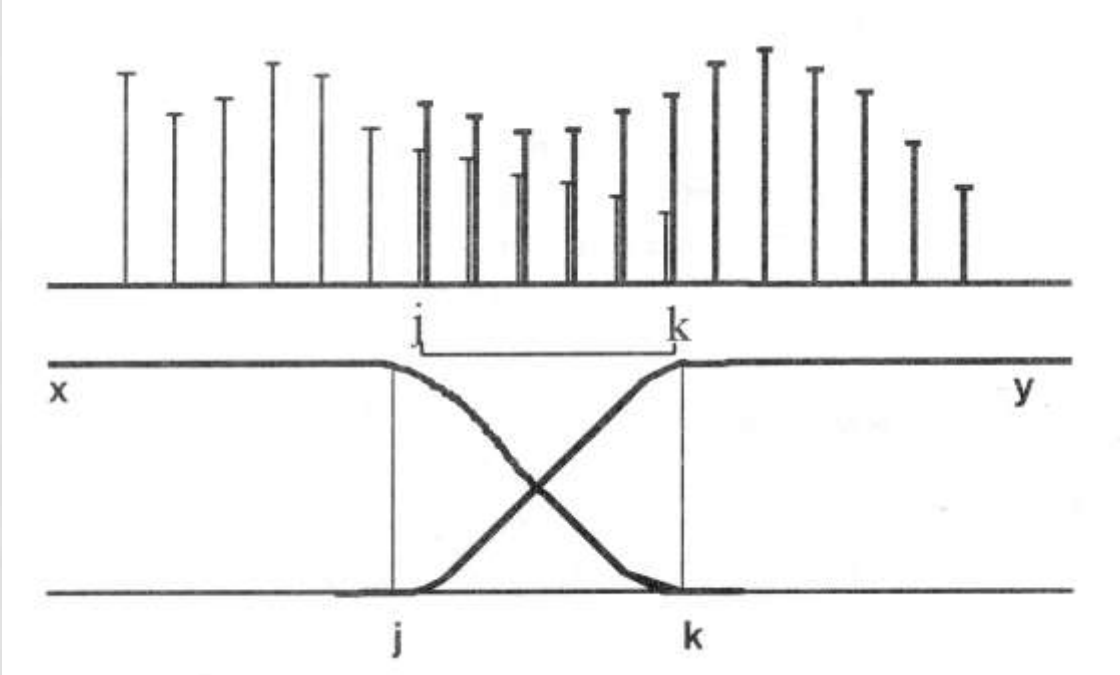
gdzie (j, k) oznacza przedział rekonstrukcji, x_n – reprezentuje ciąg próbek estymowanych do przodu, y_n – to ciąg próbek estymowanych do tyłu, a z_n jest wynikiem operacji

- Końcowe wygładzenie wyniku w oparciu o liniową predykcję

$$\hat{x}(n) = \sum_{k=1}^{p} a_k x(n-k)$$

gdzie x(n) to ciąg próbek, a_k współczynniki predykcji, natomiast p oznacza rząd predykcji

REDUKCJA PRZESTEROWAŃ



REDUKCJA KOŁYSANIA DŹWIĘKU

- W przypadku nagrań archiwalnych może pojawić się efekt drżenia i kołysania dźwięku
  - Niejednostajna prędkość nośników analogowych (taśma, płyta winylowa, cylindry woskowe)
  - Skurcz taśmy

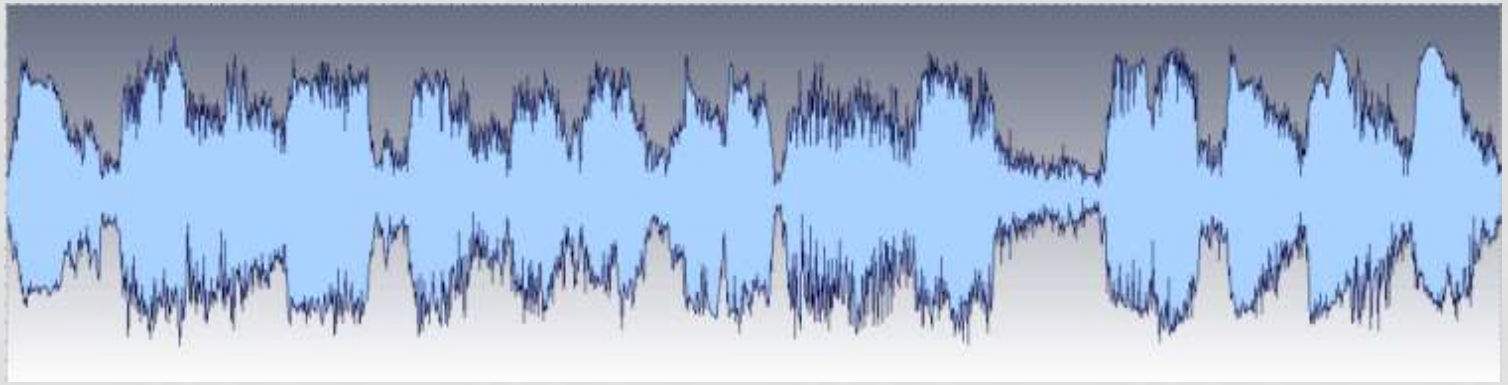
- Algorytm rekonstrukcji
  - Algorytm wyznaczający charakterystykę drżenia opisaną krzywą PVC (Pitch Variation Curve)

$$PVC(t_{org}) = \frac{d[f_w(t_{org})]}{dt_{org}}$$

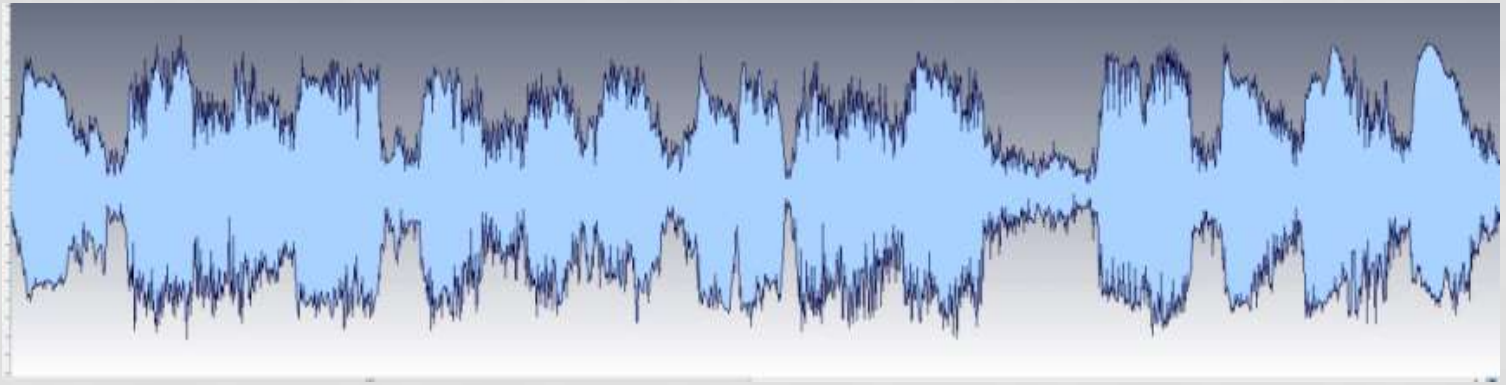
  - Nierównomierne przepróbkowanie sygnału dźwiękowego

REDUKCJA KOŁYSANIA DŹWIĘKU

- Oryginał



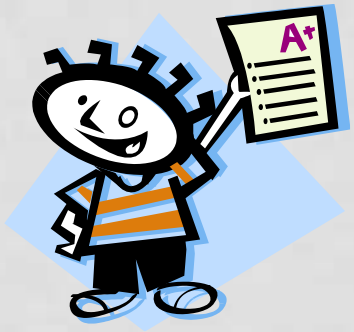
- Po rekonstrukcji



OCENA ZROZUMIAŁOŚCI MOWY

OCENA

- Oceny nagrań oraz ich przetworzonych form dokonuje się w zależności od kontekstu
  - Zrozumiałości mowy
  - Jakość sygnału



OCENA ZROZUMIAŁOŚCI

- Zrozumiałość mowy dla danej rejestracji można szacować wykorzystując obiektywne miary, takie jak:
  - STI – Speech Transmission Index
  - STIPA – STI for Public Addressed Systems – uproszczona wersja STI do specyficznych zastosowań (np. dworce, lotniska)
  - SII – Speech Intelligibility Index
- Miary te charakteryzuje duża korelacja z oceną zrozumiałości mowy
- Zrozumiałość mowy można również określać w sposób subiektywny np. na podstawie oceny wyrazistości logatomowej

OCENA ZROZUMIAŁOŚCI

- Obliczanie przedstawionych parametrów opiera się na pomiarze charakterystyk kanału transmisyjnego (np. pomieszczenia), z uwzględnieniem:
  - Charakterystyki częstotliwościowej kanału
  - Poziomu sygnału mowy
  - Poziomu szumów tła
  - Czasu pogłosu
  - Efektów psychoakustycznych (maskowanie)
  - Inne

ANALIZA MOWY W PROCESACH SĄDOWYCH

- Szczególny rodzaj przetwarzania mowy
- Często zachodzi konieczność przetwarzania sygnałów silnie zdegradowanych
- Zwykle celem analizy jest uzyskanie odpowiedzi na pytania: kto, co, gdzie, jak i kiedy
- Czynności typowo ukierunkowane na zrozumienie treści wypowiedzi
- Dokumentacja wykonywanych czynności jest równie ważna, jak sama analiza, jeśli nie ważniejsza

LITERATURA

1. S. Haykin, "Adaptive filter theory", Prentice Hall, New Jersey 2002, ISBN: 0-13-048434-2
2. R. Martin, U. Heute, Ch. Antweiler, "Advances in digital speech transmission", Wiley Interscience 2008, ISBN: 978-0-470-51739-0
3. G. Iliev , N. Kasabov, "Adaptive filtering with averaging in noise cancellation for voice and speech recognition", ICONIP/ANZIIS/ANNES Workshop, 1999, pp. 71-75
4. P. S. R. Diniz, "Adaptive filtering: algorithms and practical implementation", Kluwer Academic Publishers, 2nd ed., 2002
5. P. T. Zieliński, „Cyfrowe przetwarzanie sygnałów", Wydawnictwa Komunikacji i Łączności, 2005
6. F. Boll Steven, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-27(2), pp. 113-120, 1979
7. G. Ayanah, "Using spectral subtraction to enhance speech and increase performance in automatic speech recognition", Technical report, MERIT program, 2005
8. M. Yektaeian, R. Amirfattahi, "Comparison of Spectral Subtraction Methods used in Noise Suppression Algorithms", International Conference on Information, Communications and Signal Processings, pp. 1-4, 2007, ISBN: 978-1-4244-0983-9
9. A. Czyżewski, M. Dziubiński, J. Kotus, A. Pawlik, A. Rypulak, G. Szwoch, "Multitask noise enhancement system", 26th International Conference: Audio Forensics in the Digital Age, no. 4-1, 2005
10. S. J. Godsill, P. J. W. Reyner, „Digital Audio Restoration – a statistical model based approach", Springer-Verlag, 1998, ISBN: 3-540-76222-1
11. www.sound.eti.pg.gda.pl/denoise
12. https://www.soundonsound.com/techniques/why-forensic-audio-isnt-audio-engineering
13. https://theconversation.com/dont-believe-your-ears-enhancing-forensic-audio-can-mislead-juries-in-criminal-trials-113844
14. https://web.stanford.edu/~jurafsky/

Dziękuję