

Juliusz MAĆKOWIAK

Politechnika Śląska, Instytut Elektroniki

## FILTRACJA SYGNAŁU MOWY Z WYKORZYSTANIEM ALGORYTMU SVD

**Streszczenie.** Technika filtracji bazująca na estymacji widma szumu, z wykorzystaniem rozkładu na wartości szczególne, jest jedną ze skutecznych metod filtracji szumu. Rozkład na wartości szczególne, SVD, jest techniką, która umożliwi szybsze obliczenia, gdzie istnieje konieczność odwracania jedynie macierzy diagonalnych.

## SPEECH FILTRATION BASED ON SVD

**Summary.** Filtration's techniques using estimation of the spectrum of the clear speech from noisy speech is based on Singular Value Decomposition is one of the efficient method of the denoising. SVD requires less computing burden that is limited to inversion of the only diagonal matrixes.

### 1. Wstęp

W wielu technikach telekomunikacyjnych takich, jak audio konferencje, telefonia komórkowa, przetwarzany sygnał zawiera znaczną ilość szumu. Ta sytuacja często jest spowodowana tym, iż źródło dźwięku jest oddalone od mikrofonu, co sprzyja nagrywaniu szumu wraz z sygnałem użytecznym. Obecnie telefonu używa się coraz częściej na ulicy oraz w miejscach gdzie jest duże natężenie hałasu. Szum tła powoduje zniekształcenia sygnału, które mogą prowadzić do kompletnej niemożliwości zrozumienia informacji zawartej w sygnale mowy. Skuteczne metody filtracji mają istotny wpływ na rozwój tej gałęzi telekomunikacji.

W tym artykule chciałbym przedstawić informacje o filtracji z wykorzystaniem dekompozycji na wartości szczególne (SVD).

## 2. Informacje o sygnale mowy

System słuchowy człowieka wraz z mózgiem jest najdoskonalszym systemem percepcji, przetwarzania, analizy i szeroko pojętego rozpoznawania wrażeń akustycznych. Mimo to układ ten nie zawsze jest w stanie poradzić sobie z dużą ilością obcych dźwięków.

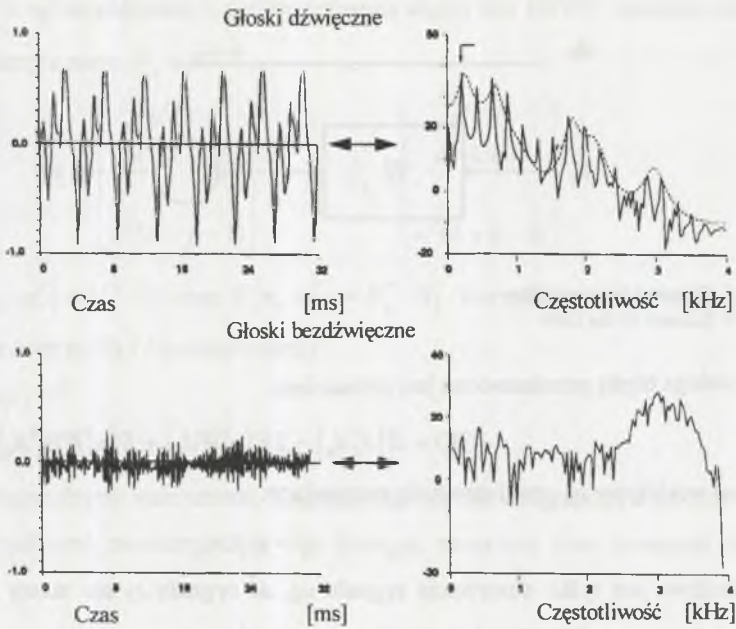
Sygnał mowy jest niestacjonarny w okresie długoterminowym, lecz w krótkich okresach 5-20 ms może być uważany za pseudostacjonarny. Statystyczne i spektralne własności mowy są zdefiniowane w krótkich segmentach czasu. Mowa może być generalnie podzielona na głoski dźwięczne<sup>1</sup>, bezdźwięczne<sup>2</sup> lub mieszane. Wykresy w dziedzinie czasu i częstotliwości są przedstawione na rysunku 2.1. Głoski dźwięczne są pseudookresowe w dziedzinie czasu i o harmonicznej strukturze w dziedzinie częstotliwości, natomiast głoski bezdźwięczne mają charakter losowy. Dodatkowo energia głosek dźwięcznych jest przeważnie większa od energii głosek bezdźwięcznych.

Widmo krótkookresowe dla głosek dźwięcznych jest określone przez kształt struktury formantowej zależnej od źródła sygnału strun głosowych, i traktu głosowego, który składa się z gardzieli i jamy ustnej. Struktura formantowa jest charakteryzowana przez zbiór wierzchołków, które noszą nazwę formantów. Dla częstotliwości do 5 kHz jest od trzech do pięciu formantów. Amplitudy i umiejscowienie pierwszych trzech formantów zwykle występują poniżej 3 kHz. Właściwości mowy są związane z fizycznym systemem generującym mowę. Głoski bezdźwięczne są wytwarzane przez pobudzenie traktu głosowego kwaziokresowymi impulsami powietrza wzmocnianymi w głośni. Częstotliwość impulsów jest odniesiona do częstotliwości fundamentalnej<sup>3</sup>, która charakteryzuje cechy osobnicze mówcy i zwykle wynosi 100-140 Hz. Głoski bezdźwięczne są wytwarzane przez wypychanie powietrza poprzez przewężenia w trakcie głosowym. Dźwięki nosowe są uzyskiwane przez łączenie traktu nosowego z głosowym, natomiast głoski zwarte przez gwałtowne uwalnianie sprężonego powietrza [1, 2].

<sup>1</sup> Na przykład „a”, „i”

<sup>2</sup> Na przykład „sz”

<sup>3</sup> Na rys. 1 częstotliwość oznaczona strzałką.



Rys. 2. 1. Charakterystyki głosek dźwięcznych i bezdźwięcznych oraz ich widma [2]

Fig. 2. 1. Voiced and unvoiced segments and their short-time spectra [2]

### 3. Metoda SVD

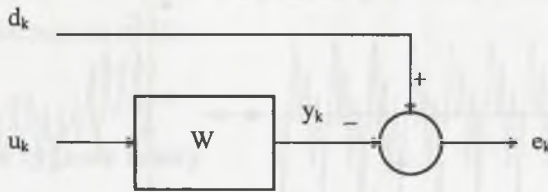
Metoda SVD<sup>4</sup> [3] - rozkład na wartości szczególne, jest algorytmem, który znacznie upraszcza podejście do problemu wyznaczania współczynników filtru bazującego na macierzy autokorelacji. Schemat filtru został przedstawiony na rys 3.1.

Sygnal  $u_k$  jest to sygnał wejściowy:

$$u_k = [u(k) \quad u(k-1) \quad \dots \quad u(k-N+1)]^T,$$

a sygnał wyjściowy przedstawia się równaniem  $y_k = u_k^T W$ , gdzie  $W \in \mathfrak{R}^{N \times N}$ , natomiast  $d_k$  i  $e_k$  są to sygnały odniesienia i błędu.

<sup>4</sup> Ang. Singular Value Decomposition.



Rys. 3.1. Schemat blokowy filtra  
Fig. 3.1. Scheme of the filter

Funkcja błędu przedstawiona jest równaniem:

$$J(W) = E\{d_k^T d_k\} - 2E\{u_k^T W d_k\} + E\{u_k^T W W^T u_k\} \quad (1)$$

Sygnał wejściowy  $u_k$  przedstawia się następująco:

$$u_k = s_k + n_k \quad (2)$$

Możliwa jest tylko obserwacja sygnału  $u_k$ , do sygnału  $s_k$  nie mamy dostępu, natomiast sygnał szumu jest dostępny podczas przerw w mowie.

Celem naszych działań jest odzyskanie sygnału  $s_k$  z  $u_k$  za pośrednictwem filtra liniowego przedstawionego na rys. 3.1. W ujęciu optymalnego filtra sygnał  $d_k = s_k$ , ale sygnału  $d_k$  nie można bezpośrednio zaobserwować.

Musimy poczynić dwa założenia:

- krótkoterminowa stacjonarność szumu  $E\{n_k \cdot n_n^T\} = E\{n_k \cdot n_n^T\}$ ,
- statystyczna niezależność mowy i szumu  $E\{s_k \cdot n_n^T\} = 0$ .

Pierwsze założenie umożliwia nam estymację macierzy autokorelacji szumu  $E\{n_k \cdot n_n^T\}$  podczas przerw w sygnale mowy. Korzystając z drugiego założenia łatwo potwierdzić, że  $E\{u_k \cdot u_n^T\} = E\{s_k \cdot s_n^T\} + E\{n_k \cdot n_n^T\}$  oraz że  $E\{u_k \cdot s_n^T\} = E\{s_k \cdot s_n^T\}$ , a więc równanie optymalnego filtra przedstawia się następująco:

$$E\{u_k \cdot u_n^T\} w^o = E\{u_k \cdot u_n^T\} - E\{n_k \cdot n_n^T\} \quad (3)$$

Można wykorzystać własność macierzy  $E\{u_k \cdot u_n^T\}$  oraz  $E\{n_k \cdot n_n^T\}$  - symetrię, aby je uprościć.

$$E\{u_k \cdot u_n^T\} = X \text{diag}\{\sigma_i^2\} X^T$$

$$E\{n_k \cdot n_n^T\} = X \text{diag}\{\eta_i^2\} X^T$$



W praktyce  $X$ ,  $\sigma_i^2$  i  $\eta_i^2$  są obliczane z wykorzystaniem algorytmu GSVD<sup>5</sup> macierzy danych  $U_k \in \mathfrak{R}^{p \times N}$  i macierzy szumu  $N_k \in \mathfrak{R}^{q \times N}$ ,

$$U_k = \begin{bmatrix} u^T(k) \\ u^T(k+1) \\ \vdots \\ u^T(k+p-1) \end{bmatrix} \quad N_k = \begin{bmatrix} n^T(k) \\ n^T(k+1) \\ \vdots \\ n^T(k+q-1) \end{bmatrix}$$

w ten sposób  $E\{u_k \cdot u_n^T\} \approx U_k^T \cdot U_k$  oraz  $E\{n_k \cdot n_n^T\} \approx N_k^T \cdot N_k$ . Korzystając z algorytmu GSVD w odniesieniu do macierzy  $N_k$  i  $U_k$  otrzymujemy:

$$U_k = U \text{diag}\{\sigma_i^2\} X^T$$

$$N_k = N \text{diag}\{\eta_i^2\} X^T,$$

gdzie  $U$  i  $V$  są ortogonalnymi macierzami,  $X$  odwracalna ale nie ortogonalna a  $\sigma_i^2$  oraz  $\eta_i^2$  są wartościami szczególnymi macierzy  $U_k$  i  $N_k$ . Stosując powyższe dwa równania do (2) otrzymujemy:

$$W^0 = X^{-T} \cdot \text{diag}\left\{\frac{\sigma_i^2 - \eta_i^2}{\sigma_i^2}\right\} \cdot X^T \quad (3)$$

Macierz  $X$  może być interpretowana jak bank filtrów, który dokonuje analiz – transformuje z przestrzeni czasu w przestrzeń transformaty, natomiast  $X^T$  jest to bank syntezujący, z powrotem w dziedzinę czasu.

#### 4. Algorytm rozkładu na wartości szczególne

Rozkład macierzy  $A \in \mathfrak{R}^{m \times n}$  na wartości szczególne możemy przedstawić za pomocą następującego równania [4]:

$$A = U \Sigma V^T, \text{ gdzie } U \in \mathfrak{R}^{m \times m}, V \in \mathfrak{R}^{n \times n}$$

a  $\Sigma$  jest to diagonalna macierz, na której przekątnej znajdują się wartości szczególne. Macierz  $U$  i  $V$  są to macierze ortogonalne, a ich kolumny są wektorami własnymi odpowiednio  $AA^T$  oraz  $A^T A$ . Macierz diagonalna  $\Sigma$  wartości szczególnych składa się z kwadratów wartości własnych macierzy  $AA^T$  lub  $A^T A$ .

<sup>5</sup> Ang. Generalized singular value decomposition- uogólniona dekompozycja na wartości szczególne.

Uogólniony rozkład na wartości szczególne GSVD macierzy  $A$  i  $B$  wyraża się następującymi równaniami:

$$A = U \cdot C \cdot X^T, \quad \text{gdzie } A \in \mathfrak{R}^{m \times n} \text{ a } U \in \mathfrak{R}^{m \times m},$$

$$B = V \cdot S \cdot X^T, \quad \text{gdzie } B \in \mathfrak{R}^{p \times n}, V \in \mathfrak{R}^{p \times p} \text{ a } X \in \mathfrak{R}^{n \times q}$$

$C^T \cdot C + S^T \cdot S = I$ , gdzie  $C \in \mathfrak{R}^{m \times n}$  i  $S \in \mathfrak{R}^{p \times n}$  są to macierze diagonalne z uogólnionymi wartościami szczególnymi na przekątnej. Macierze  $U$  i  $V$  są ortogonalne natomiast macierz  $X$  jest macierzą nieosobliwą.

Uogólnione wartości szczególne mogą być wyznaczone z wykorzystaniem rozkładu na wartości szczególne SVD, a mianowicie:

$$A^T A = X C^T C X^T \quad \text{oraz} \quad B^T B = X S^T S X^T. \quad \text{Po zsumowaniu otrzymujemy:}$$

$$A^T A + B^T B = X C^T C X^T + X S^T S X^T, \quad \text{a następnie}$$

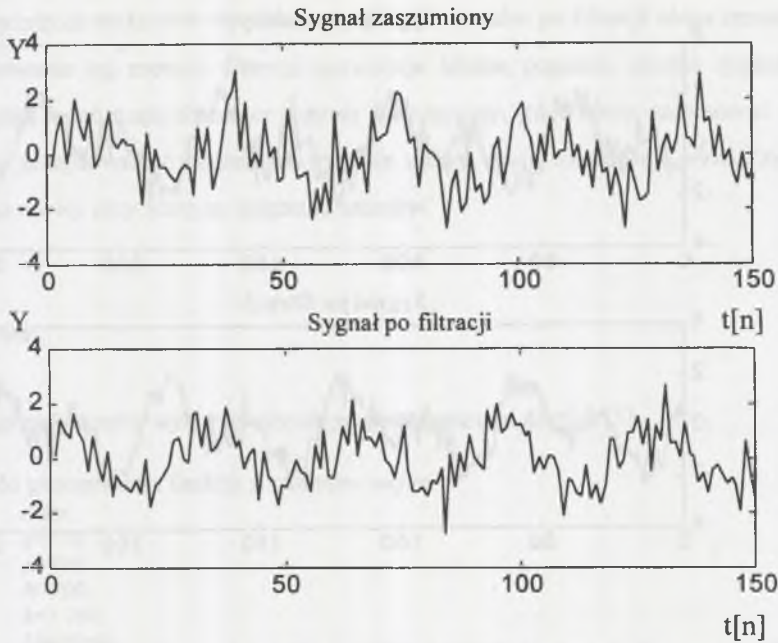
$$A^T A + B^T B = X \{C^T C + S^T S\} X^T, \quad \text{korzystając z } C^T C + S^T S = I \text{ otrzymujemy}$$

$$A^T A + B^T B = X X^T.$$

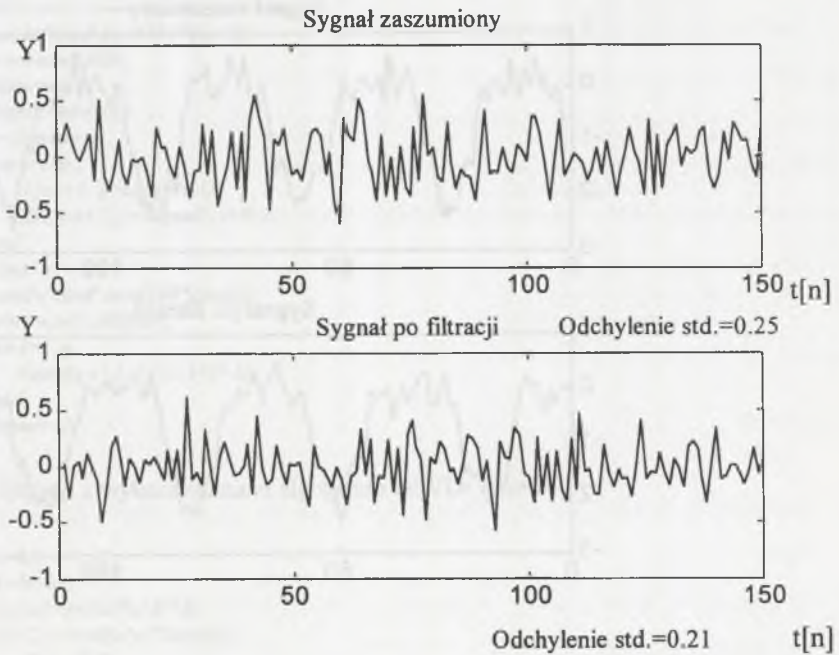
Poddając macierz  $XX^T$  SVD, otrzymujemy:  $XX^T = U_1 \Sigma_1 \Sigma_2 U_1^T$ . Chcemy odtworzyć  $X$  z iloczynu tej macierzy z macierzą transponowaną i korzystamy z rozkładu macierzy  $X$  na wartości szczególne otrzymując:  $X = V_1 \Sigma U_1^T$ . Po podstawieniu i przekształceniu otrzymujemy rozwiązanie:  $\Sigma^{-1} U_1 A = U \cdot C \cdot V_1$  oraz  $\Sigma^{-1} U_1 B = V \cdot S \cdot V_1$ , a po ponownej dekompozycji SVD lewej strony równań otrzymujemy potrzebne macierze  $S$ ,  $C$  oraz  $X$ .

## 5. Wyniki badań

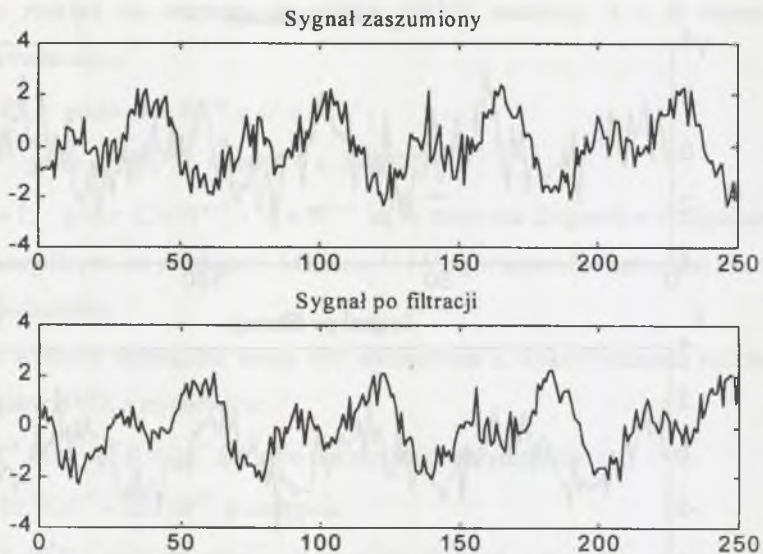
Algorytm przedstawiony w punkcie trzecim i czwartym został zaimplementowany w języku wewnętrznym interpretera poleceń *Matlaba* [5], a jego wydruk znajduje się w dodatku. Wyniki obliczeń przykładowych funkcji z szumem o rozkładzie normalnym zostały przedstawione na rysunkach od 4.1 do 4.4. Badania przeprowadzone na kilku wybranych funkcjach. Wartość odchylenia standardowego szumu po filtracji była mniejsza o około 20-30 procent od wartości przed filtracją, zależnie od wariancji szumu, jak i od kształtu funkcji.



Rys. 5.1. Wyniki filtracji sygnału sinusoidalnego z szumem o wariancji równej amplitudzie sinusoidy  
 Fig. 5.1. Issue of the filtration of the sinusoidal signal with equal noise

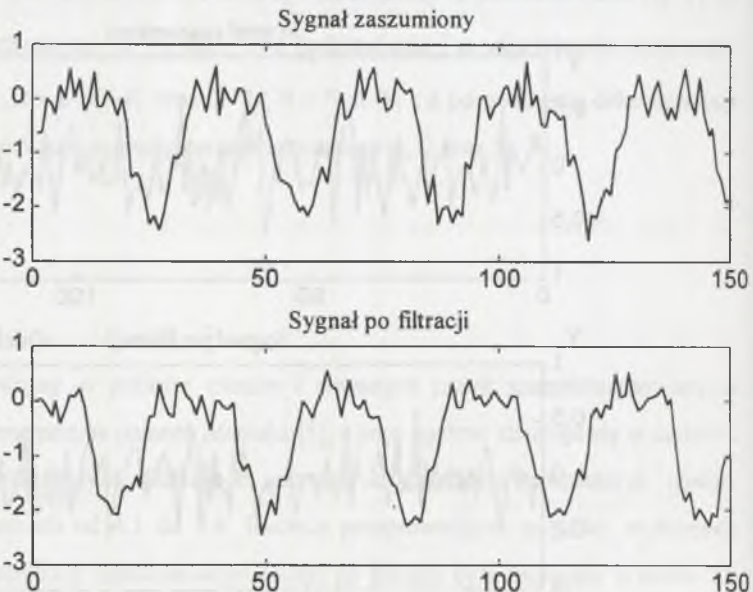


Rys. 5.2.. Wyniki filtracji funkcji liniowej o wartości zero z szumem  
 Fig. 5.2. Results of the filtration of the noise alone



Rys. 5.3. Filtracja sumy dwóch sinusoid z szumem o wariancji równej 0,25

Fig. 5.3. Filtration of the double sine with noise having variance 0,25



Rys. 5.4. Suma sygnału sinusoidalnego i wartości bezwzględnej z sinusoidy z szumem o wariancji 0,13

Fig. 5.4. Sum sine, absolute of the sine and noise having variance 0,13



Z powyższych wykresów wynika, że amplituda szumów po filtracji ulega zmniejszeniu, a więc stosowanie tej metody filtracji spowoduje istotną poprawę jakości sygnału mowy. Wyniki tych badań mają charakter jedynie orientacyjny, gdyż chcąc zastosować tę metodę należałoby przeprowadzić badania na sygnale mowy wielu mówców i wyznaczyć poziom rozumienia mowy przy różnym natężeniu szumów.

## 6. Dodatek

Listingi programów wykonywalnych pod interpreterem *Matlab* [5].

Program do generowania funkcji z szumem- *wej.m*.

```
clear;
q=100;
p=600;
N=100;
k=1:700;
U(p,N)=0;
Noise(q,N)=0;
SN=4; % stosunek sygnał szum

%-----
randn('seed',sum(100*clock));
n=randn(k)/SN;
Szum=n;
Signal=sin(k/5);
u=Signal+n;
for i=1:p
    U(p-i+1,:)=u(i:i+N-1);
    SBS(p-i+1,:)=Signal(i:i+N-1);
end
clear n;
randn('seed',sum(100*clock));
n=randn(1,200)/SN;
for i=1:q
    Noise(q-i+1,:)=n(i:i+N-1);
end
Input=U;
```

Program do filtracji z wykorzystaniem algorytmu GSVD- *filter.m*.

```
a=U;
b=Noise;
[u,l,u]=svd(a'*a+b'*b);
[U,C,v]=svd(a*u*1/sqrt(l));
x=u*sqrt(l)*v;
[V,S,v]=svd(b*u*1/sqrt(l));
k=rank(S);
for i=0:fix(k/2)
    pom=S(k-i,k-i);
```

```

    S(k-i,k-i)=S(i+1,i+1);
    S(i+1,i+1)=pom;
end;
l=(C'*C-S'*S)*inv(C'*C);
W=inv(x')*l*x';
e=Noise*Noise'/N*W;
Emin=abs(e(1,1));
Index=1;
for j=1:N
    e(j,j);
    if Emin>abs(e(j,j))
        Emin=abs(e(j,j));
        Index=j;
    end
end
for i=1:N
    W(:,i)=W(:,Index);
end

y=Input*W;

```

*%-- Odchylenie Standardowe*

```

'Std input'
std1=min(std(Input))
'Std y'
std2=min(std(y))
sum(std(Input)-std(y))/rank(y)

```

*%-----rysowanie-----*

```

subplot(2,1,1)
h1=plot(Input(1:150,1));
title('Sygnał zaszumiony');
%text(20,1,'Odchylenie standardow Std='Std1);

```

```

subplot(2,1,2)
h2=plot(y(1:150,1));
title('Sygnał po filtracji');

```

## Literatura

1. Basztura C.: Źródła, sygnały i obrazy, Wydawnictwo Komunikacji i Łączności, Warszawa 1988.
2. Spanias A.S.: Speech coding: A tutorial review, pp. 1541-1563, Proceedings of the IEEE, vol. 82, no 10,1994.

3. Doclo S., Moonen M.: SVD-based optimal filtering with application to noise reduction in speech signals, Proc. 1999 IEEE Workshop on Application of Signal Processing to Audio and Acoustics, New Paltz, New York, Oct. 17-20, pp. 143-146, 1999.
4. Soederstroem T., Stoica P.: Identyfikacja procesów, Wydawnictwo PWN, Warszawa 1997.
5. Mrozek B., Mrozek Z.: Matlab uniwersalne środowisko do obliczeń naukowo technicznych, Wydawnictwo PLJ, Warszawa 1996.

Recenzent: Dr hab.inż. Krystyna Macek-Kamińska

#### Abstract

In this paper a class of SVD-based signal improvement procedures is described, which amount to a specific optimal filtering technique for the case where the so-called 'desired response' signal cannot be observed. It is shown that this optimal filter can be written as a function of the generalized singular values and vectors. A number of simple symmetry properties of the optimal filter are derived, which are valid for the white noise case as well as for the colored noise case. Also the averaging step of the standard one-microphone SVD-based noise reduction techniques is investigated, leading to serious doubts about the necessity of this averaging step.