

POPRAWA ZROZUMIAŁOŚCI MOWY W OBECNOŚCI ZAKŁÓCEŃ Z WYKORZYSTANIEM ALGORYTMU OPARTEGO NA FILTRACJI ADAPTACYJNEJ

Damian ELLWART¹, Andrzej CZYŻEWSKI²

Politechnika Gdańska; Wydział Elektroniki, Telekomunikacji i Informatyki; Katedra Systemów Multimedialnych
ul. G. Narutowicza 11/12, 80-952 Gdańsk,

tel: - fax: - e-mail: ellwart@sound.eti.pg.gda.pl¹, ac@pg.gda.pl²

W pracy opisano nowy sposób wykorzystania filtracji adaptacyjnej do poprawy jakości dźwięków użytecznych nagrywanych w obecności zakłóceń. Przedstawiono stworzony algorytm adaptacji, omówiono możliwości przetwarzania dźwięku dodatkowymi algorytmami, opisano przeprowadzone eksperymenty. Zamieszczono i omówiono wyniki eksperymentów. Zaproponowano sposób integracji opracowanej metody z systemami akustycznego monitorowania aglomeracji miejskiej.

Słowa kluczowe: filtracja, zakłócenia, mowa.

1. WSTĘP

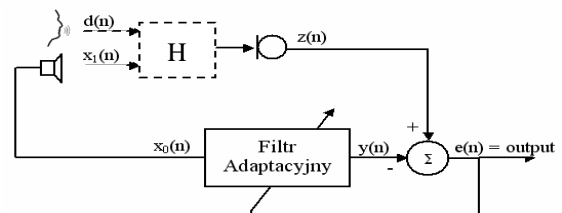
Nagrania sygnałów dźwiękowych mogą być wykonywane za pośrednictwem szerokiego wachlarza dostępnych na rynku urządzeń. Wybór konkretnego narzędzia, dokonywany jest głównie pod kątem przeznaczenia wykonywanych rejestracji oraz uwzględniając warunki, w których nastąpi rejestracja. Zakłócenia wpływające na sygnał użyteczny mogą mieć zróżnicowany charakter. W otwartej przestrzeni niepożądanymi dźwiękami mogą być hałasy pochodzenia drogowego i przemysłowego, w pomieszczeniach natomiast, mogą one pochodzić od działającej klimatyzacji, sprzętów gospodarstwa domowego, maszyn biurowych i wielu innych urządzeń ogólnego użytku. Zdarza się, iż dźwięki, które na co dzień umilają życie, w pewnych warunkach również można traktować jako degradujące sygnał zakłócenie, np. głośna muzyka. W szczególności problem ten można zauważyć w obecnie szeroko rozwijanych systemach identyfikacji głosowej i monitoringu, gdzie wszelkie dodatkowe hałasy wpływają negatywnie na poprawność detekcji i weryfikacji.

2. FILTRACJA ADAPTACYJNA

Filtry adaptacyjne mają szerokie zastosowanie w wielu dziedzinach, a w szczególności w systemach telefonii. W zależności od konfiguracji mogą być wykorzystane do modelowania procesów, takich jak identyfikacja systemu, odwzorowanie odwrotne, likwidacja zakłóceń oraz predykcja [1]. Istnieją liczne rozwiązania algorytmiczne w dziedzinie czasu i częstotliwości, które w większości opierają się na kryterium najmniejszych kwadratów i jego modyfikacjach [2]. Wybór konkretnej metody zależy od zastosowania samego filtru i nałożonych wymagań.

Przykładowo, filtry oparte na algorytmie LMS cechuje niska złożoność obliczeniowa, z kolei rekursywna procedura znajdowania współczynników filtru (RLS) wykazuje się lepszą zbieżnością [3][4].

W przedstawionej pracy zastosowano układ służący minimalizacji zakłóceń zilustrowany na rysunku 1. Filtr adaptacyjny wykorzystany w niniejszym opracowaniu oparty jest na algorytmie estymującym znormalizowany najmniejszy błąd średniokwadratowy (*ang.* NLMS - *normalized least mean squares*). Zadaniem owego filtru jest takie przekształcenie sygnału wejściowego $x_0(n)$, aby wynik filtracji jak najmniej różnił się od sygnału odniesienia $z(n)$ [5]. Kryterium minimalnej wartości błędu (2.1) spełniane jest z wykorzystaniem funkcji gradientu, przedstawionego wzorem 2.2.



Rys. 1. Schemat wykorzystania filtracji adaptacyjnej w systemie redukcji zakłóceń; $d(n)$ – pożądany sygnał mowy, $x_0(n)$ – sygnał zakłócenia przed emisją, $x_1(n)$ – sygnał zakłócenia po emisji, $z(n)$ – nagrany sygnał, $e(n)$ – wyjściowy sygnał błędu

$$J = E[e^2(n)] = E[(d(n) - y(n))^2] \quad (2.1)$$

gdzie: $e(n)$ – sygnał błędu

$$\nabla_n e^2(n) = \frac{\partial J}{\partial w(n)} = \left[\frac{\partial e^2(n)}{\partial w_0(n)}, \frac{\partial e^2(n)}{\partial w_1(n)}, \frac{\partial e^2(n)}{\partial w_0(n)}, \dots, \frac{\partial e^2(n)}{\partial w_M(n)} \right]^T \quad (2.2)$$

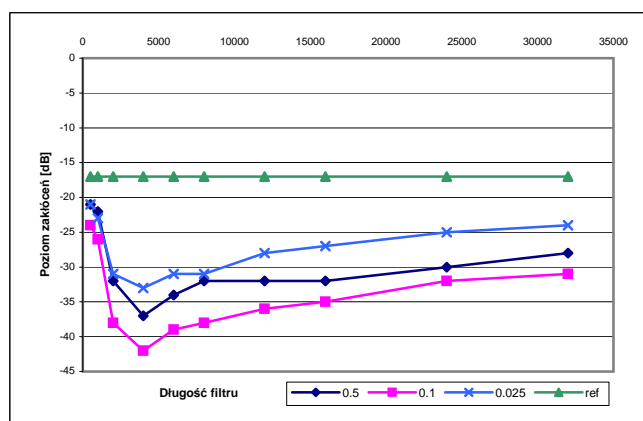
gdzie: $e(n)$ – sygnał błędu, $w(n)$ – współczynniki filtru

Po przekształceniach wzoru 2.2, które można odnaleźć w [5], otrzymana formuła na adaptację współczynników filtru, wyrażana jest w następujący sposób (2.3):

$$w_k(n+1) = w_k(n) + \frac{2\beta}{\sigma + \sum_{k=0}^M x_0^2(n-k)} \cdot e(n)x_0(n-k) \quad (2.3)$$

gdzie: $x_0(n)$ – sygnał zakłócenia przed emisją, $e(n)$ – sygnał błędu, $w(n)$ – współczynniki filtru, β – krok adaptacji, σ – offset

Stała σ , nazywana offsetem ma na celu zapobieganie dzieleniu przez zero lub bardzo małą wartość. Natomiast wprowadzony parametr β , określany mianem kroku adaptacji, wpływa na dokładność tego procesu. Zależność tłumienia zakłóceń od odpowiedniego doboru rozmiaru kroku ilustruje rysunek 2.



Rys. 2. Wpływ rozmiaru kroku adaptacji na stopień tłumienia zakłóceń. Kolorem zielonym oznaczony poziom zakłócenia w nagraniu.

3. ODEJMOWANIE WIDMOWE

Jedną z popularnych metod redukcji szumu jest odejmowanie widmowe. Zasada działania tej metody opiera się na znajomości widma sygnału zakłócającego. Zakładając, że sygnał szumu $n(k)$ jest addytywny względem sygnału mowy $s(k)$, wynik można przedstawić w dziedzinie czasu w postaci wyrażenia (3.1) lub po stronie częstotliwości jak we wzorze (3.2) [6][7][8].

$$x(k) = s(k) + n(k) \quad (3.1)$$

gdzie: $x(k)$ – sygnał zarejestrowany, $s(k)$ – sygnał użyteczny, $n(k)$ – sygnał zakłócenia

$$S(e^{j\omega}) = X(e^{j\omega}) - \alpha N(e^{j\omega}) \quad (3.2)$$

gdzie: $X(e^{j\omega})$ – widmo zarejestrowanego sygnału, $S(e^{j\omega})$ – widmo sygnału użytecznego, $N(e^{j\omega})$ – widmo sygnału zakłócenia, α – współczynnik odpowiedzialny za głębokość odejmowania

Estymowany sygnał poddany redukcji szumu ma postać:

$$\hat{S}(e^{j\omega}) = \begin{cases} [X(e^{j\omega}) - |N(e^{j\omega})|] e^{j\theta_x}, & [X(e^{j\omega}) - |N(e^{j\omega})|] e^{j\theta_x} > 0 \\ 0, & [X(e^{j\omega}) - |N(e^{j\omega})|] e^{j\theta_x} \leq 0 \end{cases} \quad (3.3)$$

W efekcie takiego działania znaczna część szumu zostaje zredukowana, jednak kosztem tego w sygnale mogą pojawić się słyszalne zniekształcenia określane mianem „szumu muzycznego” [2]. Metoda ta z powodzeniem zastosowana została w badanym algorytmie. Rozmiar

analizowanych bloków w procesie odejmowania widmowego, przy ustalonej częstotliwości próbkowania, przyjęty został na 4096 próbek. Do tego celu wykorzystano okno Hanna, z zakładkowaniem 50%.

4. EKSPERYMENTY

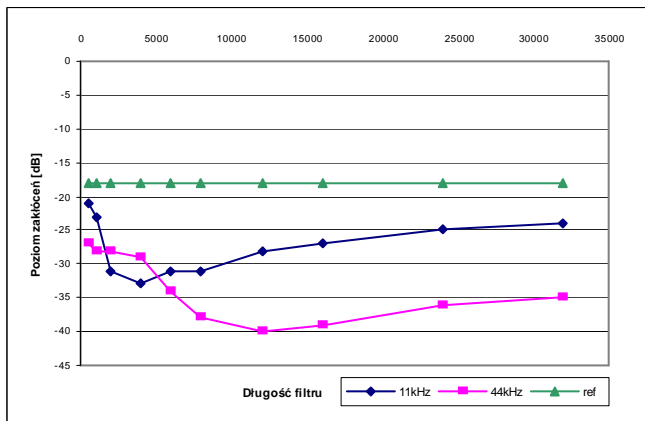
Przeprowadzone testy miały na celu badanie tłumienia sygnałów zakłócających w różnych warunkach wykonywanych nagrań (odmienne pomieszczenia, różny charakter zakłóceń). Zestaw eksperymentalny (rys. 3) składał się z mikrorejestratora firmy Marantz – nagrywającego sygnał z mikrofonu oraz sygnał zakłócenia, bezpośrednio przed emisją. Zakłócenie, w postaci typowej audycji radiowej, emitowane było za pośrednictwem zwyczajnego odtwarzacza w trybie mono. Zapis dźwięku wykonywany był dwukanałowo w formie nieskompresowanej przy próbkowaniu wynoszącym 44kHz oraz rozdzielczości 16 bitów. Przebieg testu polegał na rejestracji mowy w obecności zakłócającej ją muzyki w zmiennych warunkach otoczenia.



Rys. 3. Zestaw testowy złożony z radia, mikrorejestratora oraz wymiennych mikrofonów.

W pierwszym etapie wykonano nagrania z użyciem mikrofonów o różnych charakterystykach. Rezultaty otrzymane z nagrań zarejestrowanych mikrofonem dookólnym wykazywały mały stopień tłumienia zakłóceń. Wynika to z faktu, iż na mikrofon oddziałuje większa energia pochodząca z odbić sygnału zakłócającego w danym pomieszczeniu. W przypadku zastosowania mikrofonu o charakterystyce kardiodoidalnej i nagrania wykonanego w takich samych warunkach wykorzystanie opisanego algorytmu w znaczący sposób poprawiało słyszalność i zrozumiałość mowy. Podobnej poprawy należy spodziewać się w przypadku użycia mikrofonów o jeszcze bardziej kierunkowych charakterystykach.

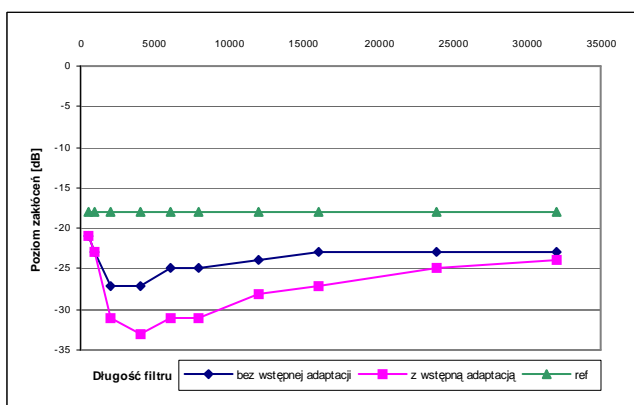
W kolejnym etapie, wykorzystując mikrofon kardiodoidalny, wykonano porównanie nagrań rejestrowanych w pomieszczeniach o różnym czasie pogłosu, a także dla zmiennych parametrów filtru adaptacyjnego. W efekcie z przeprowadzonych badań wynika, iż zwiększenie liczby współczynników filtru adaptacyjnego poprawia jakość otrzymywanego wyniku jednak tylko w ograniczonym zakresie. Istnieje granica, powyżej której dalsze zwiększanie rzędu filtru nie daje spodziewanych efektów (rys. 4), a tylko znacznie wydłuża czas działania algorytmu..



Rys. 4. Wpływ zmiennej długości stosowanego filtra adaptacyjnego na stopień tłumienia zakłóceń dla dwóch częstotliwości próbkowania sygnału. Kolorem zielonym oznaczony poziom zakłócenia w nagraniu.

Na rys. 4 można zauważyć, iż filtracja nagrania o niższej częstotliwości próbkowania skutkuje niższym tłumieniem zakłóceń niż w przypadku nagrania oryginalnego (44kHz), jednak znaczący już dla względnie niedużej ilości współczynników filtru.

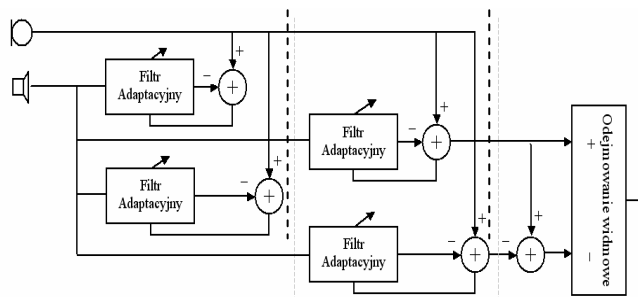
W celu poprawienia jakości otrzymanych wyników, oraz polepszenia zbieżności algorytmu adaptacyjnego, zastosowano podejście dwuetapowe. W pierwszej fazie współczynnik kroku filtru ustalany jest na wartość maksymalną, po czym następuje filtracja wybranego fragmentu nagrania (najlepiej zawierającego wyłącznie zakłócenie), w trakcie której wstępnie i zgrubnie dobierane są współczynniki filtru adaptacyjnego. W drugim etapie z tak oszacowanymi współczynnikami odbywa się filtracja całego nagrania przy zmienionej wartości kroku, pozwalając tym samym na dokładniejszą estymację sygnału. Korzyści wynikające z takiego postępowania prezentowane są na rys. 5.



Rys. 5. Wpływ zastosowania wstępnej adaptacji na stopień tłumienia zakłóceń. Kolorem zielonym oznaczony poziom zakłócenia w nagraniu.

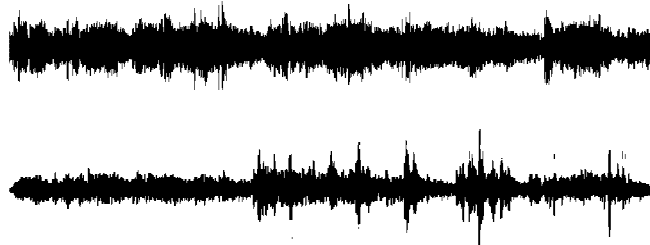
5. ALGORYTM

Opracowany algorytm bazuje na filtracji adaptacyjnej oraz odejmowaniu widmowym opisanym w niniejszej pracy. Schemat blokowy prezentowany na rys. 6 ilustruje zachodzące w nim trój etapowe przetwarzanie sygnału.



Rys. 6. Schemat blokowy opracowanego algorytmu. Można wyróżnić trzy główne części: filtracja wstępna, filtracja właściwa, odejmowanie widmowe

Zasadność wykorzystania dwuetapowej adaptacji został już opisany w poprzedzającym punkcie. Dodatkowe włączenie do systemu bloku odpowiedzialnego za odejmowanie widmowe skutkuje dalszym zwiększeniem tłumienia sygnału zakłócającego. Wynik działania, w postaci przebiegów czasowych, zaprezentowano na rys. 7.



Rys.7. Przebieg czasowy sygnału wejściowego (górną) oraz znormalizowanego sygnału wyjściowego (dolną) – po przetworzeniu algorytmem przedstawionym na rys. 6.

Zaproponowany algorytm wykazuje obiecujące rezultaty, nawet w przypadku, gdy mowa zawarta w zarejestrowanym sygnale jest dobrze maskowana przez zakłócenie. W wyniku zastosowania przedstawionego postępowania, sygnał mowy „wyłania się” na pierwszy plan, obserwowana jest poprawa zrozumiałości, a poziom maskera zostaje zredukowany w znaczącym stopniu. Artefakty wprowadzone do nagrania w wyniku operacji odejmowania widmowego, są tym bardziej słyszalne, im większa wartość współczynnika głębokości. Dalszą poprawę jakości można starać się uzyskać wykorzystując popularnie już do tego celu stosowane algorytmy, np. wybielanie.

6. PODSUMOWANIE

Opisany w niniejszym opracowaniu algorytm oraz dodatkowe możliwości przetwarzania sygnału zawarte w stworzonej do tych celów aplikacji, mogą być z powodzeniem stosowane w celach redukcji sygnału zakłócającego. Skuteczność w tłumieniu zakłócenia jest tym mniejsza im dłuższy czas pogłosu pomieszczenia. Wprowadzenie dodatkowych szumów tła nie wpływa destrukcyjnie na działanie algorytmu – w takim przypadku wynik filtracji zawiera nagraną mowę w towarzystwie hałasów otoczenia. Pomimo wykorzystania względnie prostego obliczeniowo algorytmu adaptacji (NLMS), przy konieczności użycia filtrów wysokich rzędów, zastosowanie prezentowanego rozwiązania możliwe było jedynie w aplikacji w trybie offline. Jednak po wcześniejszym zwiększeniu efektywności czasowej algorytmu, rozwiązanie

to mogłoby stać się blokiem wstępnym większego systemu służącego do akustycznego monitoringu otoczenia bądź automatycznego rozpoznawania mowy. Poza opisaną metodą stworzona aplikacja posiada zaimplementowane algorytmy ślepego rozplotu, wybielania oraz filtracji pasmowej pozwalające na redukcję zakłóceń również o z góry nie znanym charakterze. Dzięki temu może służyć jako doskonałe narzędzie do rekonstruowania nagrań sygnału mowy w obecności zakłóceń.

Podziękowania

Badania są finansowane przez Ministerstwo Nauki i Szkolnictwa Wyższego w ramach grantu nr R00-O0005/3.

7. BIBLIOGRAFIA

1. Haykin S., "Adaptive filter theory", Prentice Hall, New Jersey 2002, ISBN: 0-13-048434-2
2. Martin R, Heute U., Antweiler Ch., "Advances in digital speech transmission", Wiley Interscience 2008, ISBN: 978-0-470-51739-0
3. Iliev G., Kasabov N., "Adaptive filtering with averaging in noise cancellation for voice and speech recognition", ICONIP/ANZIIS/ANNES Workshop, November 1999 Dunedin, s. 71-75
4. Diniz P. S. R., "Adaptive filtering: algorithms and practical implementation", Kluwer Academic Publishers, 2nd ed., July 2002
5. Zieliński P. Tomasz, „Cyfrowe przetwarzanie sygnałów”, Wydawnictwa Komunikacji i Łączności, Warszawa 2005
6. Boll Steven F., "Suppression of acoustic noise in speech using spectral subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-27(2), s. 113-120, April 1979
7. Ayanah G., "Using spectral subtraction to enhance speech and increase performance in automatic speech recognition", Technical report, MERIT program 2005
8. Yektaeian M., Amirfattahi R., "Comparison of Spectral Subtraction Methods used in Noise Suppression Algorithms", International Conference on Information, Communications and Signal Processings, December 2007, s. 1-4, ISBN: 978-1-4244-0983-9

INTERFERED SPEECH INTELLIGIBILITY IMPROVEMENT USING ALGORITHM BASED ON ADAPTIVE FILTERING

Key-words: adaptive filtering, interference, speech

This paper describes a technique of improving the quality of speech signals recorded under interference (adaptive filter based algorithm). Proposed algorithm is described and additional possibilities of speech intelligibility improvement are discussed. Results of the tests are presented. A way of integrating the elaborated method with an agglomeration acoustic monitoring system is proposed. The research is subsidized by the Polish Ministry of Science and Higher Education within Grant No. R00-O0005/3.