Rafał Korycki

Wybrane metody poprawy jakości sygnału mowy

Coraz częściej można zetknąć się w codziennym życiu z rozwiązaniami technicznymi zarezerwowanymi dotychczas jedynie dla wąskiej grupy specjalistów. Dotyczy to także tych dziedzin, które są związane z analizą i przetwarzaniem sygnałów fonicznych. Szczególnie w komercyjnych zastosowaniach powiązanych z szeroko pojętą telekomunikacją coraz bardziej powszechne jest wykorzystywanie zaawansowanych technik redukcji szumu i zakłóceń. Wiele z tych rozwiązań funkcjonuje zarówno w układach czasu rzeczywistego, jak i w postaci aplikacji umożliwiających korekcję zarejestrowanych nagrań. Często oparte są na rozwiązaniach dedykowanych do konkretnych zastosowań i przystosowanych do usuwania zakłóceń o określonych parametrach. Poniżej przedstawione zostaną wybrane metody poprawy jakości nagrań monofonicznych oraz omówione możliwości wynikające z wielokanałowej rejestracji.

Jednokanałowe systemy poprawy jakości nagrań stosowane są wszędzie tam, gdzie jedynym dostępnym sygnałem jest sygnał mowy rejestrowany przy pomocy jednego mikrofonu w obecności sygnałów zakłócających. W takim przypadku nie można mówić o „usuwaniu" szumu, lecz jedynie o jego „redukcji" przy pomocy estymacji[1)] stosunku sygnału do szumu (ang. *Signal To Noise Ratio* – SNR) widma analizowanego sygnału. Systemy jednokanałowe zależą od statystycznych modeli sygnału i szumu, które mogą być estymowane w czasie braku aktywności mówcy lub dekodowane na podstawie wytrenowanych modeli mowy i szumu. Przykładem takiego rozwiązania jest aplikacja jednokanałowego systemu poprawy jakości sygnału mowy w telefonii mobilnej.

W przypadku systemów wielokanałowych sygnały zawierające zarówno informacje użyteczne, jak i zakłócenia rejestrowane są przy pomocy wielu mikrofonów. Przykładami takich rozwiązań są systemy adaptacyjnego „usuwania" szumu, szyki mikrofonowe z adaptacyjnym formowaniem wiązki, czy systemy usuwania echa akustycznego MIMO.[2)] W tych aplikacjach niezwykle istotne jest odpowiednie umieszczenie mikrofonów w taki sposób, aby maksymalnie wykorzystać możliwości danego rozwiązania.

Systemy jednokanałowe

W systemach jednokanałowych jedynym dostępnym sygnałem wejściowym jest zakłócony sygnał użyteczny. Przyjmując założenie, że sygnał mowy $s(n)$ oraz sygnał szumu $w(n)$ są addytywne,[3)] „zaszumione" nagranie można opisać w następujący sposób:

$$(1) \qquad y(n) = s(n) = w(n)$$

[1)] Estymacja to szacowanie rozkładów cech w populacji na podstawie cech jednostek wchodzących w skład losowo dobranej próby (źródło: Encyklopedia PWN)

[2)] Technologia MIMO (ang. *Multiple-Input, Multiple-Output*) to rozwiązanie zwiększające przepustowość sieci bezprzewodowych dzięki wykorzystaniu wielu anten zarówno po stronie nadawczej, jak i odbiorczej.

[3)] Sygnały addytywne to takie, które można ze sobą łączyć (sumować).

gdzie *n* oznacza indeks czasu dyskretnego, czyli numer próbki sygnału.[4] Zakłada się, że sygnał mowy nie jest skorelowany z szumem, co ma uzasadnienie w większości przypadków, w których sygnały generowane są przez różne źródła. Systemy redukcji szumu znacznie różnią się od siebie pod względem złożoności algorytmów, jak i skuteczności przekładającej się bezpośrednio na stopień poprawy zrozumiałości przetwarzanej mowy. W większości z nich można wyróżnić wspólne rozwiązania.[5]

Pierwszym etapem przetwarzania jest segmentacja, czyli podział sygnału na fragmenty o długości ok. 20-30 milisekund, a następnie wymnożenie ich przez tzw. funkcję okna (np. Blackmana, Hanna itp.). Tak przygotowane segmenty poddaje się Dyskretnej Transformacji Fouriera w celu otrzymania próbek widma. Kolejnym etapem jest tworzenie modeli, detektorów i dekoderów sygnałów mowy i szumu w celu zapewnienia skutecznego odtworzenia sygnału użytecznego. Ważnym elementem jest tu detektor mowy wykorzystywany do estymacji i adaptowania modeli szumu na podstawie sygnału wejściowego, poprzez wykrywanie momentów, w których nie występują głosy mówców. Ponadto, dzięki takiemu rozwiązaniu istnieje możliwość redukcji zakłóceń w przerwach między wypowiedziami. Następnie wyznaczana jest estymata widma mowy niezakłóconej. Wymaga to modyfikacji widma sygnału wejściowego stosownie do estymowanego stosunku sygnału do szumu dla każdej dyskretnej wartości częstotliwości. Ostatnim etapem przykładowej realizacji układu redukcji zakłóceń jest wyznaczenie odwrotnej transformacji Fouriera i odtworzenie fazy sygnału.

Odejmowanie składników widma

Jedna z najprostszych metod pozwalająca na poprawę jakości nagrań mowy została szeroko opisana w literaturze już trzy dekady temu[6,7,8,9] i polega na odejmowaniu składników szumowych widma (*spectral noise subtraction method*). To rozwiązanie sprowadza się do odjęcia estymaty widma szumu z widma sygnału mowy, przyrównania ujemnych różnic do zera, odtworzenia nowego widma z oryginalną wartością przesunięcia fazowego oraz rekonstrukcji przebiegu czasowego sygnału. Proces można opisać w następujący sposób:

$$D(\omega) = P_S(\omega) - P_N(\omega) \tag{2}$$

$$P_S^{'}(\omega) = \begin{cases} D(\omega), & jeśli \quad D(\omega) > 0 \\ 0 & w\ innym\ przypadku \end{cases}$$

[4] Popularne określenie „sygnały cyfrowe" odnosi się do sygnałów spróbkowanych w dziedzinie czasu i skwantowanych w dziedzinie wartości. Oznacza to, że sygnały posiadają swoją reprezentację jedynie w określonych chwilach czasu, których liczba zależy od częstotliwości próbkowania. Przykładowo dla częstotliwości próbkowania 44100 Hz na każdą sekundę sygnału przypada $n = 44100$ próbek.

[5] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons Ltd., England, Chichester, West Sussex, England, 2006.

[6] H. Suzuki, J. Igarashi, and Y. Ishii, *Extraction of Speech in Noise by Digital Filtering*, J. Acoust. Soc. of Japan, Vol. 33, No. 8, Aug. 1977, pp. 105 - 411.

[7] R.A. Curtis, R.J. Niederjohn, *An Investigation of Several Frequency - Domain Methods for Enhancing the Intelligibility of Speech in Wideband Random Noise*, ICASSP, April 1978, pp. 602 - 605.

[8] S. Boll, *Suppression of acoustic noise in speech using spectral subtraction*, IEEE Transactions on Acoustics Speech and Signal Processing, ASSP-27(2) pp 113-120, 1979.

[9] M. Berouti, R. Schwartz, J. Makhoul, *Enhancement of speech corrupted by acoustic noise*, IEEE ICASSP'79, Washington, 1979, pp. 208-211.

W powyższej zależności $P_S(\omega)$ jest widmem zakłóconego sygnału wejściowego, $P_N(\omega)$ stanowi wygładzoną estymatę widma szumu, natomiast $P_S'(\omega)$ jest zmodyfikowanym widmem sygnału. $P_N(\omega)$ otrzymuje się w dwóch etapach. Najpierw uśredniane jest widmo szumu z kilku kolejnych segmentów sygnału, w których nie występuje sygnał mowy, a następnie tak otrzymane widmo jest wygładzane.

W przypadku większości metod poprawy jakości sygnału mowy przyjmuje się założenie, że widmo sygnału zakłóconego przez nieskorelowany szum jest równe sumie widm: sygnału oryginalnego oraz szumu. Założenie to jest prawdziwe jedynie w sensie statystycznym i może być spełnione przy zastosowaniu widma wyznaczonego przy wykorzystaniu krótkiego okna.[10] Ze względu na fakt, że sygnał mowy oraz sygnał zakłócający nie zawsze spełniają warunek braku wzajemnej korelacji, niektóre składowe $P_S'(\omega)$ mogą być ujemne. W związku z powyższym przyrównywane są do zera.[11,12,13]

Poważnym mankamentem opisanej metody jest powstawanie dodatkowych sygnałów odbieranych przez słuchaczy jako „dzwonienie" lub „świergotanie". Wynika to z występowania szczytów i dolin w krótkookresowym widmie szumu białego.[14] Ich amplituda i położenie w dziedzinie częstotliwości mają charakter losowy i zmieniają się z ramki na ramkę. Po odjęciu wygładzonej estymaty widma szumu z bieżącego widma szumu, wszystkie maksima widmowe są redukowane, podczas gdy minima są zerowane (2). Na skutek tej operacji pozostają maksima obwiedni widma szumu. Szersze odbierane są przez słuchacza jako wąskopasmowy sygnał szumu. Węższe natomiast brzmieniem przypominają sygnały tonalne o częstotliwościach zmiennych w czasie, które można określić mianem „szumu muzycznego". Jest to jeden z najczęściej pojawiających się efektów podczas korzystania z popularnych aplikacji służących do redukcji szumu i zakłóceń. Jedną z metod jest rozwiązanie zaproponowane m.in. przez Berouti'ego:[15]

$$D(\omega) = G[P_S^{\gamma}(\omega) - \alpha P_N^{\gamma}(\omega)] \tag{3}$$

$$P_S'(\omega) = \begin{cases} D^{\frac{1}{\gamma}}(\omega), & jeśli \quad D^{\frac{1}{\gamma}}(\omega) > \beta P_N(\omega) \\ \beta P_N(\omega), & w\ innym\ przypadku \end{cases}$$

W powyższych zależnościach (3) zaleca się zachowanie następujących warunków: $\alpha \geq 1$, $0 < \beta << 1$; gdzie α jest współczynnikiem redukcji, natomiast β jest związany z poziomem progowym widma. Dzięki tym parametrom możliwa jest zarówno lepsza redukcja szerszych maksimów, jak i zmniejszenie głębokości minimów obwiedni.

---

[10] Przyjmuje się, że sygnał mowy jest w przybliżeniu stacjonarny gdy analizowane segmenty mają długość ok. 20-30 milisekund.

[11] H. Suzuki, J. Igarashi, and Y. Ishii, *Extraction of Speech in Noise by Digital Filtering*, J. Acoust. Soc. of Japan, Vol. 33, No. 8, Aug. 1977, pp. 105 - 411.

[12] R.A. Curtis, R.J. Niederjohn, *An Investigation of Several Frequency - Domain Methods for Enhancing the Intelligibility of Speech in Wideband Random Noise*, ICASSP, April 1978, pp. 602 - 605.

[13] M. Berouti, R. Schwartz, J. Makhoul, *Enhancement of speech corrupted by acoustic noise*, IEEE ICASSP'79, Washington, 1979, pp. 208-211.

[14] Długookresowe widmo mocy szumu białego jest płaskie.

[15] M. Berouti, R. Schwartz, J. Makhoul, *Enhancement of speech corrupted by acoustic noise*, IEEE ICASSP'79, Washington, 1979, pp. 208-211.

Ponadto γ jest współczynnikiem redukcji widma mocy, natomiast G współczynnikiem normalizacji. Dla ustalonej wartości α proces odejmowania widm dla wartości parametru $\gamma < 1$ skutkuje większą zmiennością składowych widmowych niż dla $\gamma = 1$. Proponowane w literaturze realizacje algorytmu zakładały następujące kombinacje parametrów: $\gamma = 0{,}5$ (dla $\alpha = 1$, $\beta = 0$) [16] oraz $\gamma = 1$.[17,18]

Wśród algorytmów pozwalających na znaczne zwiększenie wydajności rozwiązań polegających na odejmowaniu składników widma znajdują się Bayesowskie metody estymacji widma amplitudowego. Wykorzystuje się w nich funkcje gęstości prawdopodobieństwa sygnału i szumu oraz minimalizuje koszt funkcji błędu. Przykładem takiego rozwiązania jest estymacja minimalnego średniego błędu kwadratowego (ang. *Minimum Mean Squared Error – MMSE*) krótkookresowego widma amplitudowego (ang. *Short Time Spectral Amplitude – STSA*).[19,20,21]

Systemy wielokanałowe

Sygnały zawierające zarówno mowę jak i zakłócenia mogą być rejestrowane przy pomocy wielu sensorów, a następnie filtrowane w celu redukcji szumu, echa, pogłosu, czy też głosu innych osób. Wykorzystywany jest tu fakt, że nawet w przypadku niewielkich odległości między poszczególnymi mikrofonami, docierające do nich sygnały różnią się od siebie pod względem natężenia oraz charakterystyki częstotliwościowej i fazowej. Dzieje się tak m.in. ze względu na różnice w czasie dotarcia sygnałów do poszczególnych odbiorników. Redukcja i usuwanie zakłóceń w systemach wielokanałowych realizowane są np. poprzez identyfikację odpowiedzi impulsowej pomieszczenia, w którym rejestrowana jest mowa, na podstawie zarówno samego sygnału użytecznego, jak i zakłócającego. Istotna jest również szybkość adaptacji („przystosowania") współczynników stosowanych filtrów do zmieniających się warunków akustycznych.

Filtracja adaptacyjna

Projektowanie filtrów cyfrowych polega na wyborze struktury,[22] rzędu[23] oraz doborze wartości współczynników. Tak stworzony układ cechuje się odpowiednią charakterystyką częstotliwościową, dzięki czemu umożliwia wzmacnianie lub tłumienie

---

[16] S. Boll, *Suppression of acoustic noise in speech using spectral subtraction*, IEEE Transactions on Acoustics Speech and Signal Processing, ASSP-27(2) pp 113-120, 1979.

[17] H. Suzuki, J. Igarashi, and Y. Ishii, *Extraction of Speech in Noise by Digital Filtering*, J. Acoust. Soc. of Japan, Vol. 33, No. 8, Aug. 1977, pp. 105 - 411.

[18] R.A. Curtis, R.J. Niederjohn, *An Investigation of Several Frequency - Domain Methods for Enhancing the Intelligibility of Speech in Wideband Random Noise*," ICASSP, April 1978, pp. 602 - 605.

[19] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

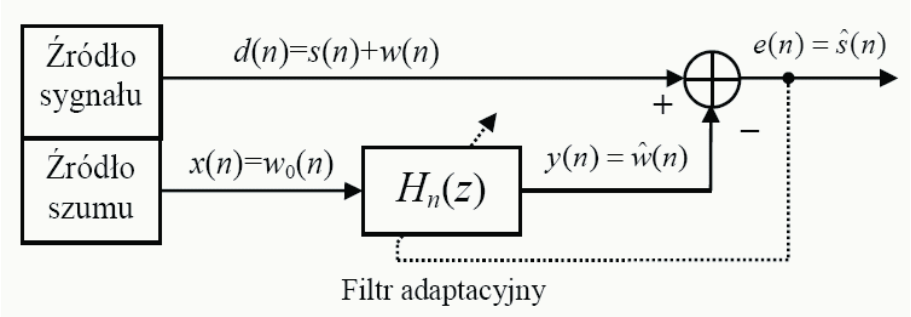
[20] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean square error log-spectral amplitude estimator*, IEEE Trans. on Acoust., Speech, Signal Processing, vol. ASSP-33, pp. 443-445, Apr. 1985.

[21] R. Martin, *Speech Enhancement Using MMSE Short Time Spectral Estimation with Gamma Distributed Speech Priors*, IEEE ICASSP'02, Orlando, Florida, May 2002.

[22] Rozróżnia się filtry FIR (ang. *Finite Impulse Response*) o skończonej odpowiedzi impulsowej oraz filtry typu IIR (ang. *Infinite Impulse Response*) o nieskończonej odpowiedzi impulsowej.

[23] Rząd filtra określa złożoność układu; im większy rząd, tym więcej współczynników i elementów opóźniających tworzy filtr.

określonych zakresów częstotliwości. W przypadku dokonywania rejestracji w zmiennych warunkach akustycznych, istnieje konieczność dopasowania charakterystyki filtra do sygnałów zakłócających w taki sposób, by móc je sprawnie eliminować. Filtry adaptacyjne pozwalają na modyfikację ich charakterystyk poprzez automatyczną aktualizację wartości współczynników. Proces adaptacji polega na minimalizacji błędu między sygnałem wyjściowym a sygnałem docelowym (lub pożądanym). Rys. 1 ilustruje ogólny schemat układu filtracji adaptacyjnej wykorzystywany do usuwania szumu przy wykorzystaniu źródła sygnału referencyjnego.



Rys. 1. Schemat układu filtracji adaptacyjnej służącego do usuwania szumu przy wykorzystaniu źródła sygnału referencyjnego.

Sygnał wejściowy $d(n)$ jest sumą sygnałów użytecznego oraz zakłócającego. Na wyjściu układu otrzymuje się estymatę sygnału zakłócającego. Po odjęciu od siebie tych dwóch sygnałów uzyskuje się sygnał błędu. W zależności od własności sygnału zakłócającego, przybliżeniem $s(n)$ może być sygnał błędu lub sygnał wyjściowy. Proces adaptacji (zilustrowany na rys. 1 przerywaną linią) uzależniony jest od zastosowanego algorytmu aktualizacji współczynników filtra. Algorytmy filtracji adaptacyjnej stanowią dość liczną grupę. Dwa najprostsze i jedne z najczęściej stosowanych w praktyce to

$$(1) \qquad w(n+1) = w(n) + \mu[y(n)e(n)]$$

$$(5) \qquad w(n) = w(n-1) + k(n)e(n)$$

$$\text{gdzie:} \quad k(n) = \frac{\lambda^{-1} R_{yy}^{-1}(n-1)y(n)}{1 + \lambda^{-1} y^{\tau}(n) R_{yy}^{-1}(n-1)y(n)}$$

LMS (ang. *Least Mean Squares*) i RLS (ang. *Recursive Least Squares*).

Zależności (4) i (5) przedstawiają odpowiednio algorytm aktualizacji współczynników filtra metodą „najmniejszej średniej kwadratowej" (LMS) oraz rekursywny algorytm aktualizacji współczynników filtra metodą najmniejszych kwadratów (RLS). Wektor w to wektor wartości wag współczynników filtra, μ jest współczynnikiem szybkości adaptacji, R_{yy} oznacza macierz autokorelacji, natomiast λ pełni funkcję współczynnika zapominania.

Zaletą algorytmu LMS jest jego prostota, jednakże dla sygnałów o szerokim spektralnym zakresie dynamiki, odznacza się nierównomiernym i niskim współczynnikiem zbieżności. Ponadto jeśli analizowany sygnał jest dodatkowo niestacjonarny, a taki jest m.in. sygnał mowy, algorytm LMS nie jest dobrym narzędziem do tworzenia układów redukcji szumu. Można jednak wykorzystać jego własności do usuwania wolnozmiennych

i wąskopasmowych sygnałów zakłócających. Z kolei algorytm RLS cechuje się szybszą zbieżnością i większą odpornością na zmiany parametrów sygnału wejściowego.

Ślepa separacja źródeł

W rejestrowanych nagraniach często spotyka się mieszaniny wielu sygnałów, np. kilku mówców, muzyki, szumu tła itp. Występują one jednocześnie, podobnie jak w tzw. efekcie cocktail party. Niektóre z tych sygnałów są dźwiękami pożądanymi, natomiast inne stanowią jedynie zakłócenia. Ślepa Separacja Źródeł (ang. *Blind Source Separation – BSS*) jest rozwiązaniem pozwalającym na segregację sygnałów źródłowych z kilku mieszanin tych sygnałów. Innymi słowy pozwala na odtworzenie oryginalnych sygnałów źródłowych z sygnałów rejestrowanych przez mikrofony.

Metoda analizy składowych niezależnych (ang. *Independent Component Analysis – ICA*) jest statystyczną techniką dekompozycji złożonych grup danych na niezależne podgrupy. W przypadku gdy dwa zarejestrowane sygnały są od siebie niezależne, tzn. obserwacja jednego z nich nie pozwala na znalezienie informacji na temat drugiego, dzięki Ślepej Separacji Źródeł możliwe jest rozdzielenie sygnałów tworzących superpozycję. Przedstawiając powyższy problem w sposób ogólny, dąży się do rozwiązania zależności (6):

$$x = As \tag{6}$$

gdzie s jest dwuwymiarowym wektorem zawierającym niezależne sygnały źródłowe, A jest macierzą mieszającą o wymiarach 2x2, natomiast x jest wektorem zawierającym obserwowane (zmiksowane) sygnały.[24] Istotne jest, aby nie więcej niż jeden komponent poddawany separacji miał charakter szumu gaussowskiego (o normalnym rozkładzie prawdopodobieństwa). Ponadto liczba obserwowanych składników mieszaniny sygnałów (np. liczba mówców) musi być mniejsza lub równa liczbie zastosowanych sensorów (mikrofonów).[25] Pierwszym krokiem większości algorytmów wykorzystujących metodę analizy składowych niezależnych jest wybielenie danych, tzn. dążenie do tego, by analizowane sygnały były nieskorelowane i ich wariancje były równe 1. Poszukiwana jest zatem taka liniowa transformacja w^T (7,8), aby macierz kowariancji $E\{ss^T\}$ była równa macierzy identycznościowej I (o składowych równych 1 na głównej przekątnej i zerami w pozostałych miejscach).

$$s = w^T x, \tag{7}$$

$$E\{ss^T\} = I \tag{8}$$

Połączone komponenty mogą zostać odseparowane poprzez maksymalizację „nie-gaussowości”[26] wektora w^T.[27]

---

[24] A. Hyvarinen, *Survey of Independent Component Analysis*, http://www.cs.helsinki.fi.

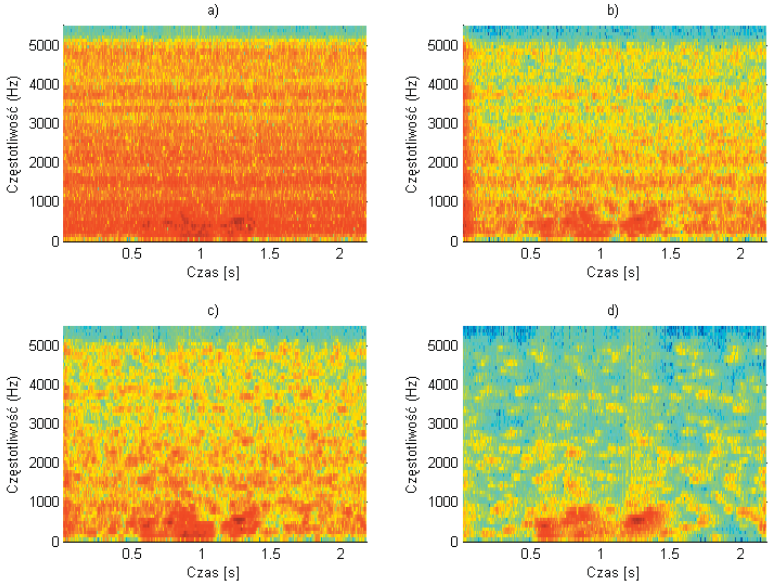
[25] N. Murata, S. Ikeda, A. Ziehe, *An Approach to Blind Source Separation based on Temporal Structure of Speech Signal*.

[26] Miarą podobieństwa rozkładu do rozkładu normalnego jest *kurtoza*.

[27] A. Hyvarinen, *Survey of Independent Component Analysis*, http://www.cs.helsinki.fi.

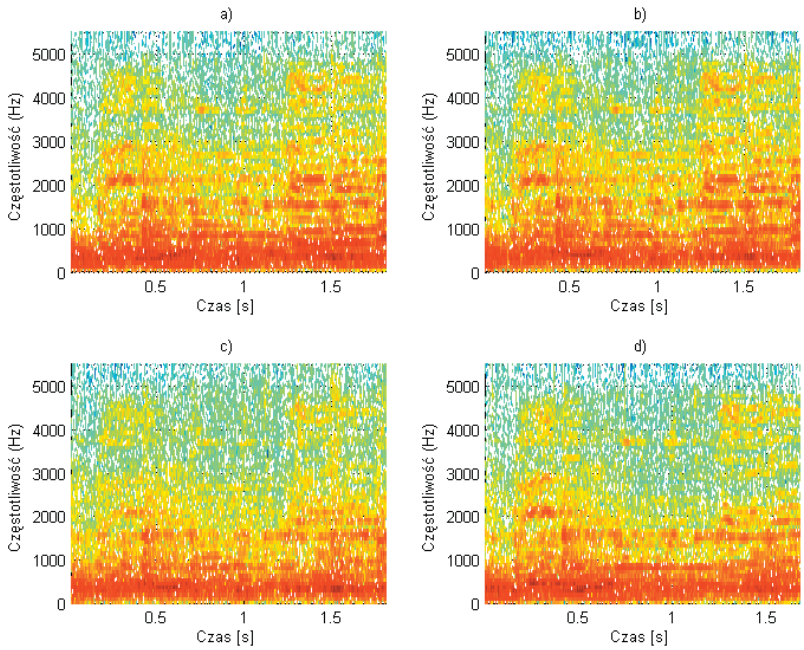
Przykłady zastosowań i podsumowanie

Wykorzystując opisane rozwiązania wykonano poglądowe symulacje, pozwalające na wstępną ocenę stosowanych metod i algorytmów. Nagrania zarejestrowano w pomieszczeniu biurowym bez dodatkowej adaptacji akustycznej. Rejestracja odbywała się przy pomocy urządzeń cyfrowych przy częstotliwości próbkowania 11025 Hz. Sygnałem zakłócającym był szum biały generowany przy pomocy przenośnego odtwarzacza CD (rys. 2a). Dokonano redukcji zakłóceń wykorzystując implementacje w środowisku MATLAB estymacji minimalnego średniego błędu kwadratowego krótkoczasowe-



Rys. 2. Porównanie algorytmów poprawy jakości sygnału mowy. Spektrogramy przedstawiają:
a) sygnał mowy zakłóconej,
b) sygnał po korekcji przy wykorzystaniu algorytmu opisanego w: Y. Ephraim and D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Transactions on Acoustics, Speech, and SignalProcessing, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984 oraz Y. Ephraim and D. Malah, Speech enhancement using a minimum mean square error log-spectral amplitude estimator, IEEE Trans. on Acoust., Speech, Signal Processing, vol. ASSP-33, pp. 443-445, Apr. 1985,
c) sygnał po zastosowaniu metody opisanej w: P. Scalart and J. Vieira-Filho, Speech enhancement based on a priori signal to noise estimation, 21st IEEE Int. Conf. Acoust. Speech Signal Processing, Atlanta, GA, May 1996, pp. 629–632 , d) sygnał po korekcji metodą przedstawioną w: I. Cohen, Speech Enhancement Using a Noncausal A Priori SNR Estimator, IEEE Signal Processing Letters, Vol. 11, No. 9, Sep. 2004, pp. 725-728.

go widma amplitudowego, opisanej przez Ephraima i Malaha[28,29] (rys. 2b) oraz dwóch algorytmów wykorzystujących estymację *a priori* stosunku sygnału do szumu: przedstawionego przez Scalarta i Vieira-Filho[30] (rys. 2c) oraz opisanego przez Cohena (rys. 2d).[31] W przypadku pierwszego rozwiązania można było zauważyć niewielką poprawę jakości nagrania, przy minimalnych zniekształceniach wprowadzanych przez program. Najlepsze efekty uzyskano stosując dwie ostatnie metody, jednakże szczególnie w przypadku zilustrowanym na rys. 2c zaznaczył się wpływ „szumu muzycznego".



Rys. 3. Spektrogramy ilustrujące zastosowanie poszczególnych filtrów adaptacyjnych: a) sygnał wejściowy zawierający sygnał mowy oraz sygnał zakłócający w postaci muzyki, b) sygnał na wyjściu filtra z algorytmem LMS, c) sygnał po przetworzeniu przez filtr wykorzystujący algorytm RLS, d) sygnał po dokonaniu adaptacyjnej filtracji w dziedzinie częstotliwości FDAF.

[28] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

[29] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean square error log-spectral amplitude estimator*, IEEE Trans. on Acoust., Speech, Signal Processing, vol. ASSP-33, pp. 443-445, Apr. 1985.

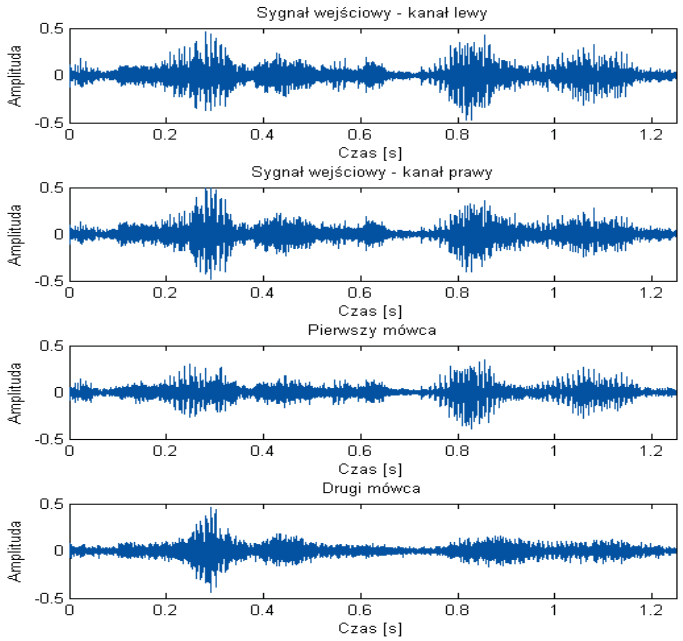
[30] P. Scalart and J. Vieira-Filho, *Speech enhancement based on a priori signal to noise estimation*, 21st IEEE Int. Conf. Acoust. Speech Signal Processing, Atlanta, GA, May 1996, pp. 629–632.

[31] I. Cohen, *Speech Enhancement Using a Noncausal A Priori SNR Estimator*, IEEE Signal Processing Letters, Vol. 11, No. 9, Sep. 2004, pp. 725-728.

W podobnych warunkach akustycznych przetestowano również opisane filtry adaptacyjne oraz porównano je z dostępnym w środowisku MATLAB adaptacyjnym algorytmem filtracji w dziedzinie częstotliwości FDAF. Tym razem jako sygnał zakłócający wykorzystano utwór muzyczny, który rejestrowano równolegle przy pomocy mikrofonu oddalonego o ok. 20 cm od odtwarzacza. Na rys. 3a przedstawiono spektrogram sygnału wejściowego, będącego sumą sygnału mowy oraz sygnału zakłócającego w postaci muzyki. Kolejne spektrogramy (rys. 3b-d) ilustrują sygnały wyjściowe filtrów adaptacyjnych, w których wykorzystano algorytmy: LMS, RLS oraz pracujący w dziedzinie częstotliwości FDAF.

Warto zauważyć, że największą poprawę zrozumiałości mowy uzyskano stosując dwa ostatnie z demonstrowanych rozwiązań. Algorytm LMS ze względu na wolną zbieżność nie przyniósł zadowalających efektów, co było do przewidzenia.[32,33,34)]

Skuteczność metody analizy składowych niezależnych także sprawdzono w tym samym pomieszczeniu. Nagrania dokonano przy pomocy stereofonicznego rejestratora, a pomiary wykonano dla sygnału o częstotliwości próbkowania 16000 Hz, wykorzystując przy tym rozwiązanie opisane w: R. Prasad, H. Saruwatari, A. Lee, K. Shikano, *A fixed-point ICA algorithm for convoluted speech Signac separation*, 4th Symp. on ICA and BSS (ICA2003), Nara, Japan, Apr. 2003. Na rys. 4 i 5 przedstawiono odpowiednio przebiegi czasowe oraz spektrogramy sygnałów pochodzących z rejestratora oraz sygnały wyjściowe, stanowiące odseparowane głosy dwóch osób mówiących jednocześnie.

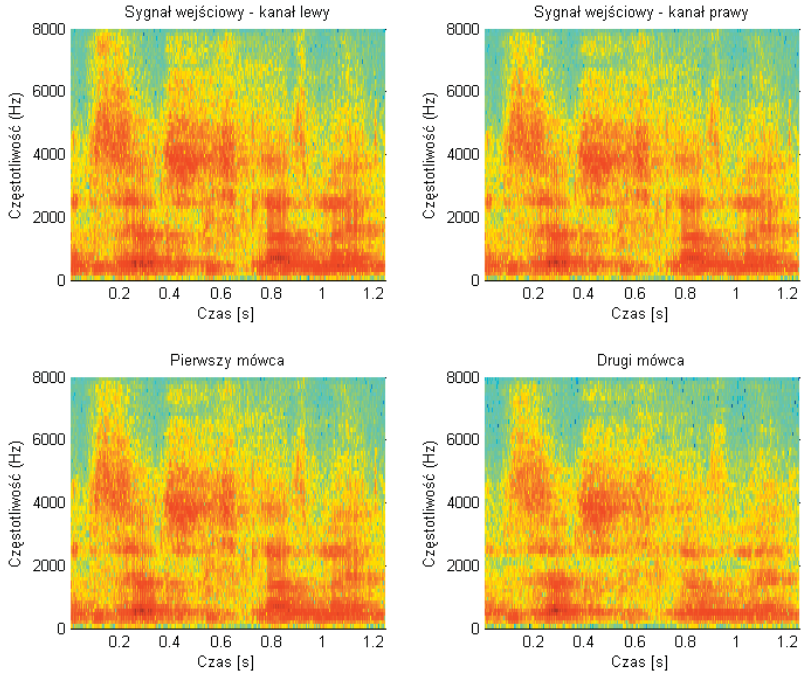


Rys. 4. Przebiegi czasowe ilustrujące zastosowanie metody analizy składowych niezależnych do separacji mówców.

32) S. Haykin, *Adaptive Filter Theory*, Prentice-Hall International, Inc. USA, 1991.
33) L. Rutkowski, *Filtry adaptacyjne i adaptacyjne przetwarzanie sygnałów*, WNT Warszawa, 1994.
34) P. Zieliński, *Cyfrowe Przetwarzanie Sygnałów, od teorii do zastosowań*, WKŁ Warszawa, 205, 2007.

Zgodnie z opisem zamieszczonym w literaturze – A. Hyvarinen, *Survey of Independent Component Analysis*, http://www.cs.helsinki.fi oraz N. Murata, S. Ikeda, A. Ziehe, *An Approach to Blind Source Separation based on Temporal Structure of Speech Signal* – w nagraniach wynikowych pojawił się sygnał mowy drugiej osoby. Wynika to z faktu występowania pogłosu w pomieszczeniu, w którym prowadzono rejestrację. Innymi słowy program „dokonał" redukcji sygnału mowy drugiej osoby, jednakże „nie zdołał" usunąć sygnałów odbitych, docierających do rejestratora z opóźnieniem.

Zaprezentowane przykłady algorytmów stanowią jedynie niewielki fragment dostępnych rozwiązań, pozwalających na redukcję zakłóceń i poprawę zrozumiałości mowy. Znajdują zastosowanie m.in. w telefonii: mobilnej i internetowej, systemach telekonferencyjnych, zestawach głośnomówiących, aparatach słuchowych, czy w studiach nagrań. Szczególną uwagę warto zwrócić na filtry adaptacyjne, które umożliwiają redukcję nawet silnych zakłóceń, o ile oczywiście jest do dyspozycji sam sygnał zakłócający. Dzięki temu wykorzystywane są m.in. w kabinach pilotów w samolotach. Ciekawymi rozwiązaniami są także techniki ślepej separacji źródeł. Umożliwiają nie tylko rozdzielenie mówców prowadzących rozmowę w tym samym czasie, w szczególnych warunkach także dźwięków instrumentów w nagraniach muzycznych.



Rys. 5. Spektrogramy ilustrujące zastosowanie metody analizy składowych niezależnych do separacji mówców.