

DOBROWOLSKI Andrzej P.<sup>1</sup>  
KAŻMIERCZAK Jakub<sup>2</sup>  
SUCHOCKI Michał<sup>3</sup>

## **METODY POPRAWY JAKOŚCI DŹWIĘKU W SYSTEMACH IDENTYFIKACJI MÓWCÓW**

*Przy ocenie skuteczności systemów identyfikacji mówców, ważną rolę odgrywa nie tylko prawidłowa i właściwie dokonana ekstrakcja cech sygnału mowy oraz dobór odpowiednich algorytmów identyfikujących mówcę, ale także prawidłowe pozyskanie i zarejestrowanie dźwiękowych sygnałów głosowych w jak najmniejszym stopniu zniekształconych w celu porównania go później ze wzorcem. W referacie przedstawiono opis toru dźwiękowego pomiędzy ustami mówcy a systemem decyzyjnym, przedstawiono klasyfikacje źródeł zakłóceń i zniekształceń towarzyszących procesowi rejestracji dźwięku oraz klasyfikację metod i algorytmów poprawy jakości sygnału mowy.*

## **SPEECH ENHANCEMENT METHODS IN SPEAKERS RECOGNITION SYSTEMS**

*In evaluation of efficiency of speakers recognition systems, a proper extraction of the features of the speech signal, choice of appropriate algorithms of speaker recognition, as well as an appropriate, introducing possibly small distortions, acquisition and logging of the voice signals are of great significance. The paper presents a description of the voice path, from the speaker's mouth to the decision system, a classification of various sources of interferences and distortions connected with the sound logging, as well as classification of methods and algorithms of improvement of the voice quality.*

### **1. WSTĘP**

Współczesne systemy identyfikacji i weryfikacji mówcy bazują na coraz bardziej złożonych procesach decyzyjnych. Jednak niezależnie od stosowanych metod czy algorytmów identyfikujących rozmówcę, rzeczą niezmiernie ważną jest dostarczenia do

---

<sup>1</sup> Wojskowa Akademia Techniczna, Wydział Elektroniki; 00-908 Warszawa; ul. Gen. Sylwestra Kaliskiego 2  
Tel: +48 22 683 7534, Fax: +48 22 683 9444, E-mail: andrzej.dobrowolski@wat.edu.pl

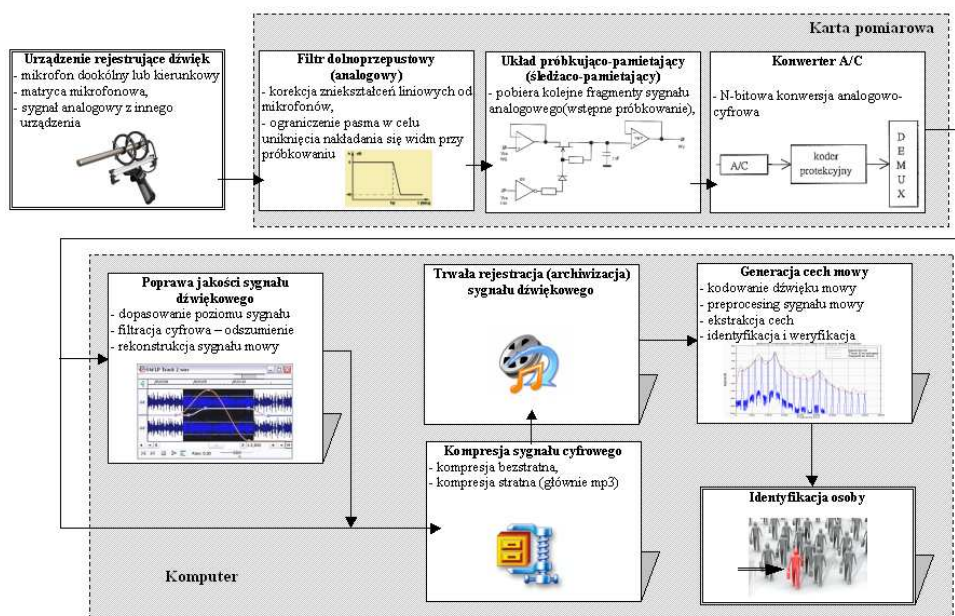
<sup>2</sup> Wojskowa Akademia Techniczna, Wydział Elektroniki; 00-908 Warszawa; ul. Gen. Sylwestra Kaliskiego 2  
Tel: +48 22 683 9233, Fax: +48 22 683 9444, E-mail: jakub.kazmierczak@wat.edu.pl

<sup>3</sup> Wojskowa Akademia Techniczna, Wydział Elektroniki; 00-908 Warszawa; ul. Gen. Sylwestra Kaliskiego 2  
Tel: +48 22 683 7534, Fax: +48 22 683 9444, E-mail: michal.suchocki@gmail.com

takiego systemu sygnału (mowy, fragmentu wypowiedzi) w jak najmniejszym stopniu zniekształconego, który najwierniej będzie odzwierciedlał cechy osobowe mówcy.

Na wysoką jakość i możliwość wiernego porównania rejestrowanego dźwięku ze wzorcem wpływ ma wiele elementów. Dźwięk wytwarzany przez mówcę po wyemitowaniu przez usta, zanim zostanie wprowadzony do systemu rozpoznawania mówcy, jest transmitowany złożonym torem. Tor ten stanowi w większości przypadków łańcuch zawierający część akustyczną od ust mówcy do przetwornika elektroakustycznego, następnie analogową część elektryczną do przetwornika analogowo-cyfrowego, a dalej część cyfrową [1]. Części toru może być jednak więcej, mogą występować urządzenia pośrednie, np. rejestrujące lub transmitujące drogą radiową, światłowodową czy też przewodami elektrycznymi. Każda część toru może wносить specyficzne zniekształcenia liniowe, nieliniowe oraz szum.

Na rys. 1 przedstawiono poglądowy schemat blokowy systemu identyfikacji mówcy.



Rys.1. Schemat blokowy systemu identyfikacji mówcy

## 2. ŹRÓDŁA ZAKŁÓCEŃ I ZNIEKSZTAŁCENŃ DŹWIĘKU W PROCESIE JEGO REJESTRACJI

Podczas rejestracji mowy należy uwzględnić warunki w jakich dokonywana jest rejestracja. W środowisku naturalnym sygnał mowy rzadko zdarza się w czystej formie, często towarzyszy mu szereg dodatkowych dźwięków. W otwartej przestrzeni niepożądanymi dźwiękami mogą być naturalne hałasy otoczenia (szum uliczny, przemysłowy, dźwięki przyrodnicze), a w pomieszczeniach mogą one pochodzić od pracujących urządzeń mechanicznych (klimatyzacja, maszyny biurowe, sprzęt komputerowy lub gospodarczy) oraz od urządzeń nagłaśniających (megafony, głośniki).

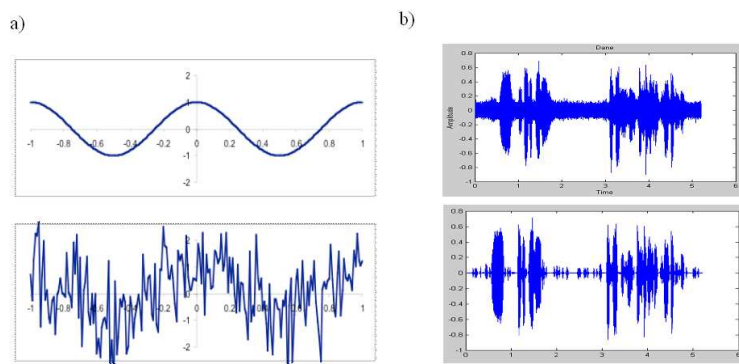
Kolejnym elementem zakłócającym rejestrację mowy jest echo akustyczne i pogłos. Elementem negatywnie wpływającym na jakość przechwytywanej mowy mogą być także sprzężenia akustyczne, zwłaszcza wzbudzenie się układów wzmacniających w przypadku występowania źródeł w pobliżu mikrofonów.

Ostatnim elementem utrudniającym prawidłowe pozyskiwanie dźwięków mowy jest obecność wielu osób w pobliżu miejsca rejestracji, odgłosy drobnych czynności przez nie wykonywanych - rozmowy, szepty, chodzenie, szuranie, postukiwanie, szelest, korzystanie z telefonu czy odtwarzaczy muzycznych, przemieszczanie niewielkich przedmiotów jak walizki, wózki itp.

Wszystkie te szумы tła (otoczenia) można podzielić na:

- ✓ **Szum kierunkowy**, który dociera do mikrofonu z jednego kierunku w przestrzeni, dlatego może być łatwo zlokalizowany w oparciu tylko o informację słuchową.
- ✓ **Szum rozproszony**, który dociera do mikrofonu z wielu kierunków w przestrzeni dlatego jest bardzo trudny do zlokalizowania. Szum ten pojawia się w wysoce pogłosowych otoczeniach lub też pochodzi od urządzeń o dużych rozmiarach.
- ✓ **Szum cocktail party lub "babble"** (z ang. *paplanina*), to mieszanka wielu ludzkich głosów podobnie jak na przyjęciu, w poczekalniach, restauracji czy kolejce.
- ✓ **Szum stacjonarny**, to szum, którego charakterystyki (średni poziom, skład częstotliwościowy itp.) są stałe i nie zmieniają się w czasie (np. szum wentylatora).
- ✓ **Szum niestacjonarny lub fluktuujący**, który ma zmienne w czasie charakterystyki i objawia się zmienną głośnością, nieregularnym występowaniem i intensywnością (przypadkowe zakłócenia typu impulsowego).

Tego typu zakłócenia, często potocznie zwanymi szumami, skutecznie mogą utrudnić, a nawet uniemożliwić prawidłowe rozpoznanie mowy czy identyfikację rozmówcy, poprzez nałożenie się na właściwy sygnał [2] (rys. 2).



Rys. 2. Sygnał oczyszczony i sygnał zakłócony a) harmoniczny, b) sygnał mowy

Dlatego zjawiska zakłóceń akustycznych stały się w ostatnich latach ważnym elementem uwzględnianym przy projektowaniu systemów rejestrujących i analizujących mowę ludzką (systemy rozpoznawania mowy, systemy rozpoznawania mówcy oraz

systemy identyfikacji osób). Obecność szumu i zakłóceń nie tylko obniża zrozumiałość mowy ale potrafi znacząco utrudnić jej prawidłową rejestrację i późniejszą analizę.

Mimo, że problem ten nie jest zupełnie nowy, a techniki redukcji zakłóceń i szumu rozwijane są od wielu lat, jest on wciąż bardzo aktualny. Rozwój nowych technologii oraz możliwości sprzętowych skłania badaczy do podejmowania nowych, trudniejszych wyzwań projektowych, których celem jest uzyskanie wysokiej jakości zarejestrowanego sygnału mowy w celu dalszej analizy. Opracowywane są coraz skuteczniejsze metody i algorytmy uzdatniania sygnału mowy, zdolne do pracy w warunkach wyjątkowo silnych zakłóceń i zniekształceń akustycznych.

Głównym celem systemów poprawy jakości sygnału mowy jest usunięcie niepożądanych sygnałów zakłócających (szumu, echa, pogłosu) przy jednoczesnym zachowaniu jak najlepszej jakości sygnału użytecznego. W przypadku, gdy jedyną dostępną informacją jest sygnał zakłócony problem przestaje być trywialny.

Tłumieniu tych zakłóceń prawie zawsze towarzyszy wprowadzanie zniekształceń do sygnału użytecznego. Aby uzyskać w pełni satysfakcjonujące rezultaty należy wykorzystać odpowiednie metody poprawy jakości sygnału, najczęściej będące kompromisem pomiędzy jak największą eliminacją sygnałów zakłócających, a jak najmniejszym zniekształceniem cech wymaganych przez właściwe systemy analizy mowy. Metody odszumiające sygnał pod kątem uzyskania najlepszej zrozumiałości mowy najczęściej nie dają już takich rezultatów dla innych systemów głosowych, np. identyfikacji rozmówcy (gdyż w procesie tym eliminacji ulega część elementów niezbędnych do prawidłowej identyfikacji). Dla systemów identyfikacji rozmówcy mniejszą wagę przywiązuje się do treści językowej, natomiast skupia się nad informacją o tożsamości, zawartości prozodycznej oraz zmienności cech mowy. Dlatego w tego typu systemach tak dużą rolę przywiązuje się do wstępnej obróbki rejestrowanego sygnału.

### 3. KLASYFIKACJA METOD I ALGORYTMÓW POPRAWY JAKOŚCI SYGNAŁU MOWY

Dotychczas opracowano wiele procedur i metod poprawy jakości sygnałów akustycznych. Część z tych procedur znalazła szerokie zastosowanie w wielu urządzeniach powszechnego użytku, inne, z pewnych względów są mniej popularne gdyż posiadają pewne uwarunkowania co do zastosowań. Biorąc pod uwagę rodzaj analizowanego sygnału i jego pochodzenie oraz znajomość informacji o elementach zakłócających sygnał, metody poprawy jakości sygnałów dźwiękowych można klasyfikować wg kilku różnych kryteriów:

► ze względu na urządzenie realizujące zmniejszanie wpływu sygnałów zakłócających na sygnał użyteczny, poprawa jakości sygnału akustycznego może być realizowana w oparciu o:

- filtrację analogową (filtry analogowe w dziedzinie czasu),
- filtrację w oparciu o procesory sygnałowe (DSP i FPGA),
- zaawansowaną filtrację cyfrową.

Filtracja analogowa w systemach rejestracji dźwięku (mowy) zapewnia wstępną obróbkę sygnału i zazwyczaj przygotowuje sygnał do dalszej, zaawansowanej analizy cyfrowej. Głównie wykorzystywana jest do redukcji i korekcji zniekształceń liniowych wnoszonych przez elementy toru akustycznego (fonicznego) i ograniczenia pasma w celu uniknięcia nakładania się widm przy próbkowaniu sygnału.

Poprawa jakości sygnału (redukcja zniekształceń dźwięku) wykorzystująca procesory sygnałowe DSP (i układy FPGA) znalazła szerokie zastosowanie w systemach nie wymagających dużej wierności sygnału odtworzonego. Rozwój technologii elektronicznej sprawił, że wykorzystywanie procesorów DSP w filtracji staje się coraz powszechniejsze. Wpływa na to ciągła miniaturyzacja, uzyskiwanie coraz lepszych efektów rekonstrukcji sygnału i przede wszystkim możliwość realizacji w czasie rzeczywistym. Tego typu filtracja bardzo powszechnie wykorzystywana jest w systemach telekomunikacji mobilnej (telefonii komórkowej czy zestawach głośnomówiących), gdzie największy nacisk położony jest na poprawę zrozumiałości mowy i minimalizację wpływu zewnętrznych źródeł hałasu na prowadzone rozmowy.

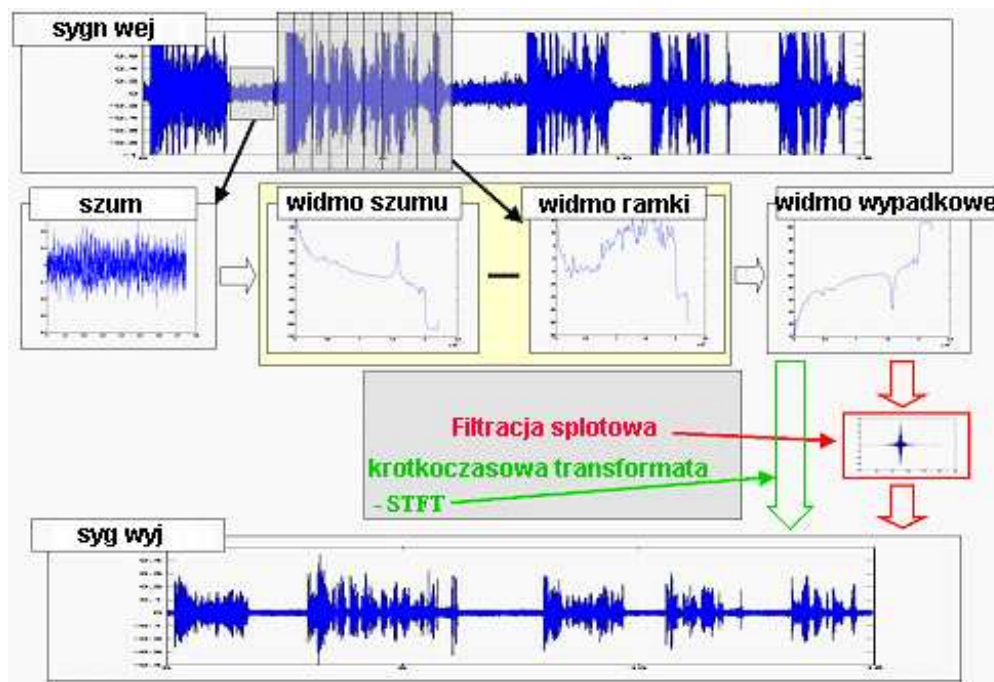
Zaawansowana filtracja cyfrowa w pełni wykorzystuje właściwości całego procesu cyfrowej analizy sygnałów. Dzięki wykorzystaniu złożonych algorytmów i mechanizmów analizy osiąga najwyższą skuteczność rekonstrukcji sygnałów. Niestety wysoka poprawa jakości sygnału wymaga czasu i dużej mocy obliczeniowej. Podstawową przewagą filtracji cyfrowej nad analogową jest możliwość otrzymywania bardzo dobrych parametrów charakterystyk częstotliwościowych. Poziomy tłumienia osiągane w filtrze cyfrowym są dużo większe niż możliwe do osiągnięcia w filtrze analogowym. Pasma przejściowe w filtrach cyfrowych jest bardzo wąskie. Ponadto w filtrach cyfrowych można uzyskać pasmo przejściowe pozbawione jakichkolwiek zafalowań. Kolejną zaletą filtrów cyfrowych jest możliwość uzyskania filtru o liniowej fazie w paśmie przepustowym, co powoduje, że taki filtr nie wprowadza zniekształcenia fazy w sygnale użytecznym. Filtry te cechuje duża elastyczność w projektowaniu, a czas projektowania jest znacznie krótszy niż w przypadku filtrów analogowych.

► ze względu na dziedzinę (domenę), w której realizowana jest rekonstrukcja sygnału mowy [3]:

- metody modyfikacji widmowej – dziedzina częstotliwości,
- analiza podprzestrzeni sygnału – dziedzina mieszana
- algorytmy filtrowania z wykorzystaniem dziedziny czasu,
- algorytmy bazujące na metodach statystycznych,
- inteligentne systemy decyzyjne – sieci neuronowe i wnioskowanie rozmyte.

Przetwarzanie w dziedzinie częstotliwości w głównej mierze opiera się na wykonaniu odpowiednich operacji na widmie dźwięku (ang. *spectral modification method*), najczęściej poprzez odejmowanie określonych składowych widma od zniekształconego sygnału lub sygnału szerokopasmowego o wartości energii szumu (tzw. *odejmowanie widmowe*). Charakterystyki widma szumu otrzymuje się poprzez analizę tych fragmentów, w których sygnał użyteczny jest nieobecny (np.: w sygnale mowy są to momenty przerwy między poszczególnymi wyrazami, zdaniem, fragmentami rozmowy). Ponieważ z założenia sygnał użyteczny i szum są nieskorelowane, w związku z tym widmo mocy sygnału użytecznego oblicza się, jako różnicę widma sygnału i szumu pomnożonego przez pewien współczynnik  $\alpha$ , dobierany w zależności od stosunku sygnału do szumu (ang. *SNR - Signal to Noise Ratio*). Charakterystyka szumu i współczynnik  $\alpha$  są określane dla kolejnych prążków częstotliwościowych lub filtrów pasmowych. Sygnał odszumiony (wyjściowy) uzyskuje się po zastosowaniu odwrotnej transformaty Fouriera (ang. *IFT - Inverse Fourier Transform*). W procesie tym zmiany amplitudy widma sygnału przynoszą zadawalające efekty (zwłaszcza w odniesieniu do zrozumiałości mowy), niestety faza sygnału nie jest

zachowana, co prowadzi do pojawienia się zniekształcenia zwanego szumem muzycznym. Graficzną prezentację zasady działania tej metody przedstawiono na rys. 3.



Rys. 3. Graficzna interpretacja metody odejmowania widmowego

Alternatywną metodą do metod w dziedzinie widmowej jest metoda analiz podprzestrzeni sygnału (ang. *sub-space analysis signal*). To co odróżnia je od klasycznych metod w dziedzinie częstotliwości, to teoretycznie optymalna dekorelacja sygnału uzyskiwana przy pomocy transformaty Karhunen-Loeve'a (modelem sygnału w tym przypadku jest suma losowych sinusoid na tle szumu o znanej funkcji kowariancji). Istotą tej metody jest dekompozycja przestrzeni sygnału zakłóconego na podprzestrzeń sygnału (główną) i podprzestrzeń szumu. Ponieważ rzut sygnału zakłóconego na podprzestrzeń szumu nie zawiera informacji o sygnale użytecznym, przetwarzana jest wyłącznie ta część sygnału zakłóconego, która „znajduje się” w podprzestrzeni głównej. Niestety operacje te wymagają dużej mocy obliczeniowej. Drugim przykładem algorytmu analizy czasowo-częstotliwościowej (dziedziny mieszanej) jest krótkoczasowa transformata Fouriera (ang. *STFT - Short-Time Fourier Transform*). Umożliwia ona wydobyć z sygnału informacji o tym, jak zmienia się jego widmo w czasie, czyli jednoczesną obserwację jego właściwości zarówno w dziedzinie czasu jak i częstotliwości. Wycinek sygnału przeznaczony do analizy jest sukcesywnie dzielony na segmenty, z których każdy podlega analizie widmowej niezależnie. Przesuwając okno w czasie, wzdłuż sygnału, wyznacza się jego zawartość widmową wewnątrz przedziału czasowego, którego długość jest określona szerokością okna.

Dosyć skuteczne okazują się także cyfrowe metody przetwarzania sygnałów realizowane w dziedzinie czasu, zwłaszcza dla krótkich sygnałów szerokopasmowych obciążonych zniekształceniami po zastosowaniu operacji nieliniowych. Do takich metod filtrowania dźwięku w dziedzinie czasu można zaliczyć filtrację Wienera, która usuwa z sygnału czasowego składową szumu oraz filtrację Kalmana estymującą składowe szumu i sygnału pozbawionego szumu. Algorytmy takich filtrów wykrywają sekwencje szumopodobne w obecności zakłóceń. Nieliniowe operacje polegają na odrzuceniu ujemnych wartości funkcji splotu i nieliniowym wzmocnieniu sygnału.

Metody filtracji oparte na modelach statystycznych wykorzystują statystyczne porównanie zarejestrowanych danych z pożądanym, pozbawionym szumu sygnałem. Prekursorami statystycznego ujęcia filtracji byli Wiener i Kołmogorow, którzy niezależnie zdefiniowali pojęcie procesu losowego, charakteryzującego się pewnymi właściwościami statystycznymi. Wydobycie sygnału użytecznego z szumu traktuje się jako estymację jednego sygnału na podstawie drugiego. Dostępny, zaszumiony sygnał jest przetwarzany w celu uzyskania możliwie najlepszej estymaty sygnału, o którym zakłada się, że stanowi niezaszumioną część sygnału wejściowego. Dla każdej z chwil czasu, dla której wyznaczona jest estymata, optymalne wagi filtru (parametr charakterystyczny dla filtracji Wienera) muszą spełniać warunek minimalizacji błędu średniokwadratowego. Do algorytmów wykorzystujących statystykę można zaliczyć także metody przetwarzania sygnałów w oparciu o statystyki wyższego rzędu: metodę separacji zmieszanych niezależnych źródeł-sygnałów (ang. *blind separation*) lub metodę usuwania wpływu nieokreślonego filtru (ang. *blind deconvolution*).

W ostatnim czasie można zaobserwować dynamiczny rozwój nowych metod analizy i przetwarzania danych, bazujących na inteligentnych procesach decyzyjnych wykorzystujących sztuczną inteligencję. Zastosowanie inteligentnych systemów decyzyjnych do usuwania szumów i zakłóceń z sygnałów stało się możliwe dzięki połączeniu wnioskowania rozmytego (ang. *fuzzy logic*) z systemem decyzyjnym opartym na sieci neuronowej (ang. *neural network*) [4]. Takie połączenie pozwala skutecznie wyizolować właściwą, niezniekształconą informację z rejestrowanego sygnału. Metoda ta jest stosunkowo nowa, ciągle się rozwija i wiąże się z nią duże nadzieje.

► ze względu na rodzaj eliminowanych zniekształceń sygnału dźwiękowego przed jego obróbką, poprawę jakości sygnału realizuje się przez:

- minimalizowanie i usuwanie zakłóceń impulsowych,
- redukowanie szumu i zakłóceń szerokopasmowych,
- filtracja echa i pogłosu (redukowanie zniekształceń liniowych),
- redukowanie sprzężeń,
- redukowanie szumu i zakłóceń wąskopasmowych,
- normalizacja poziomu, mocy i pasma sygnału akustycznego.

Kolejność wymienionych czynności rekonstruowania sygnału dźwiękowego nie jest przypadkowa. Niezwykle istotne znaczenie dla jakości uzyskiwanych wyników końcowych ma właśnie prawidłowy porządek kolejnych etapów minimalizacji zakłóceń i zniekształceń.

► ze względu na ilość kanałów pozyskiwania informacji sygnału użytecznego wraz z informacjami o warunkach w jakich ta rejestracja następuje (np. z tzw. sygnałem odniesienia), wyróżnia się:

- jednokanałową redukcję szumu
  - wybielanie i odejmowanie widmowe,

- progowanie współczynników falkowych,
- algorytm minimalizacji błędu średniokwadratowego MMSE – STSA,
- filtracja iteracyjna Wienera lub Kalmana,
- dwukanałową redukcję szumu
  - dwukanałowa filtracja adaptacyjna,
  - metoda formowania wiązek dla *beamformera* dwumikrofonowego,
  - redukcja pogłosu oparta na funkcji koherencji,
- wielokanałową redukcję szumu
  - adaptacyjna filtracja przestrzenna (ang. *spatial filtering*),
  - metody formowania wiązek (ang. *acoustic beamforming*),
  - metoda ślepej separacji sygnałów.

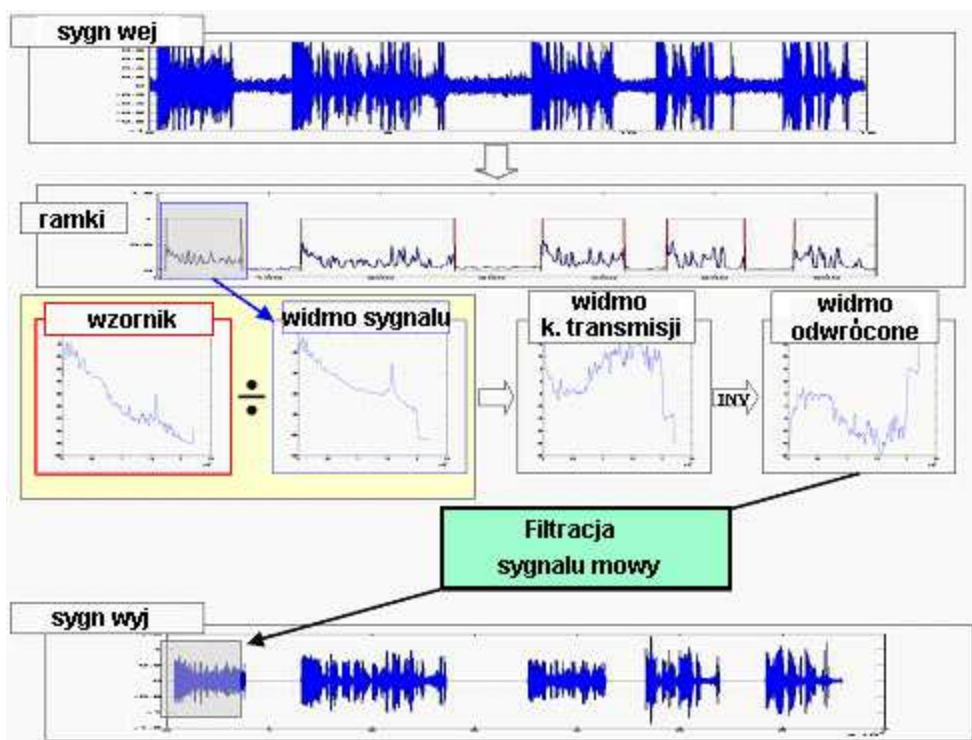
Algorytmy jednokanałowej redukcji szumu (ang. *single-channel noise reduction*) wykorzystują jedno źródło rejestracji dźwięku (mikrofon). Ten rodzaj redukcji zakłóceń bazuje jedynie na czasowych, widmowych i statystycznych cechach rejestrowanego sygnału i przynosi zadawalające efekty jeśli można przyjąć, że szum jest stacjonarny i nie jest skorelowany ze źródłem sygnału. Do eliminacji szumu wykorzystuje się odszumianie w oparciu o metody widmowe i estymacje szumu wykorzystujące kryteria minimalizacji błędu średniokwadratowego (ang. *Minimum Mean Square Estimators* – MMSE) zwanego też metodą Ephraima i Malaha (prekursorów tego kryterium). W ten sposób uzyskuje się efekt wyraźnej poprawy średniego stosunku sygnału do szumu analizowanego materiału. Niestety metody te powodują oprócz pewnego zredukowania poziomu szumu także zmniejszenie poziomu sygnału użytecznego w poszczególnych fragmentach, gdyż chwilowe wartości zarówno sygnału, jak i szumu mogą odbiegać od wartości średnich. Dlatego też metody te wymagają pewnego kompromisu pomiędzy pozostawieniem części szumu w sygnale a zachowaniem jeszcze poprawnych właściwości tego sygnału. Ponadto wymagany jest dodatkowy element systemu – detektor aktywności głosowej (ang. *voice activity detector* – VAD), który wskazuje momenty kiedy rozmowa nie następuje (np. pomiędzy wyrazami, chwilami pauzy, itp.).

Dwukanałowa redukcja szumu bazuje głównie na adaptacyjnym redukowaniu szumu. Metody te zazwyczaj obok tradycyjnego czujnika (mikrofonu) rejestrującego sygnał użyteczny wykorzystują drugi mikrofon do pozyskania informacji o warunkach w jakich dokonywana jest rejestracja dźwięku. Ze względu na istotę metody korzystne jest, aby drugi mikrofon był ulokowany blisko źródła zakłóceń (hałasu) i dość daleko od źródła dźwięku.

Wielokanałowa redukcja szumu wykorzystuje przestrzenną informację z miejsca rejestracji dźwięku do zredukowania szumu tła i eliminacji towarzyszących rejestracji zakłóceń. Jako układy filtrujące wykorzystuje się filtry adaptacyjne LMS (ang. *Least Mean Square*), filtry adaptacyjne RLS (ang. *Recursive Least Squares*), metody formowania wiązki (ang. *beamforming*) polegające na przestrzenno-czasowym przetwarzaniu sygnału pochodzącego z wielu źródeł (matrycy czujników) oraz metody ekstrakcji szumu metodami ślepej separacji sygnałów lub źródeł (ang. *Blind Signal Separation*). Ostatnia z metod odnosi się do problemu „odzyskiwania” pojedynczego sygnału z kilku zarejestrowanych mieszanin sygnałów [5]. Jest to metoda, która segreguje sygnały źródłowe z kilku mieszanin tych sygnałów metodą wyszukiwania źródeł „po omacku”, tj. może odtworzyć oryginalne sygnały źródłowe z wielu różnych sygnałów zarejestrowanych przez mikrofony.



Algorytm tej metody opiera się na porównaniu widm fragmentów mowy z widmem wzorca mowy, który powinien zawierać sygnał bez zakłóceń. Najwięcej problemu nastęrcza znalezienie prawidłowego wzorca, który właśnie poszukiwany jest trochę „na ślepo” spośród wielu dostępnych kanałów rejestracji sygnałów dźwiękowych. Graficzną prezentację zasady działania tej metody przedstawiono na rys. 4.



Rys. 4. Graficzna interpretacja metody ślepej separacji sygnałów

#### 4. WNIOSKI

Skuteczność oraz szybkość działania współczesnych systemów identyfikacji mowy zależą w bardzo dużym stopniu od zastosowanych algorytmów i metod rekonstrukcji zniekształconego lub zakłóconego sygnału mowy. Wymagania w tych systemach odnośnie jakości pozyskanego do analizy sygnału mowy są znacznie większe niż w tradycyjnych systemach rozpoznających mowę. Tutaj zwraca się szczególną uwagę na wychwycenie (niezatrącenie) cech mowy, aby na ich podstawie system prawidłowo mógł zidentyfikować mówcę na podstawie danych zgromadzonych w bazie. Różnorodność źródeł zakłóceń sygnału użytecznego, rodzaj i jakość sprzętu użytego do rejestracji czy złożone warunki w jakich pozyskiwany jest sygnał mowy stawia bardzo wysokie wymagania procesom filtracyjnym.

Prawidłowo przeprowadzona rekonstrukcja sygnału mowy w znaczący sposób zwiększa wiarygodność i skuteczność systemów identyfikacji mowy. Dlatego tak dużą wagę

przywiązuje się w tych systemach do znalezienia właściwych metod i algorytmów redukcji zniekształceń i zakłóceń, implementując w nich nawet bardzo złożone algorytmy i metody w celu uzyskania jak najlepszej skuteczności.

## 5. BIBLIOGRAFIA

- [1] Czyżewski A.: *Dźwięk cyfrowy – wybrane zagadnienia teoretyczne, technologia, zastosowania*, Akademicka Oficyna Wydawnicza EXIT, 2001
- [2] Maithani S., Tyagi R.: *Noise Characterization and Classification for Background Estimation*, International Conference on Signal Processing, Anna University Chennai India 2008.
- [3] Loizou Y., Loizou C.: *A Comparative Intelligibility Study of Speech Enhancement Algorithms*, IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu 2007.
- [4] Wierzbicki J., Augustyn G., *Usuwanie szumu i zakłóceń z wykorzystaniem inteligentnych systemów wnioskowania rozmytego*, Process Control Club 2001.
- [5] Czyżewski A., Kotus J., Szwoch G., Dziubiński M.: *Multitask Noisy Speech Enhancement System*, AES 26th International Conference, Denver, USA 2005.