

Restauracja nagrań dźwiękowych – usuwanie zakłóceń impulsowych

Leszek GRAD

Zakład Automatyki, Instytut Teleinformatyki i Automatyki WAT,
ul. Kaliskiego 2, 00-908 Warszawa

STRESZCZENIE: W artykule zostało przedstawione zagadnienie restauracji nagrań dźwiękowych w aspekcie usuwania zakłóceń impulsowych. Ponadto zamieszczono krótki przegląd oprogramowania przydatnego w restauracji nagrań dźwiękowych.

SŁOWA KLUCZOWE: restauracja nagrań dźwiękowych, zakłócenia impulsowe, sieci neuronowe

Wstęp

Zagadnienie restauracji nagrań, podobnie jak restauracji obrazów, cieszy się od lat dużym zainteresowaniem. Stan archiwizowanych nagrań na nośnikach analogowych bywa bardzo często niezadowalający. Powodów tego jest kilka. Płyty analogowe w trakcie eksploatacji podlegały zarówno zwykłemu zużyciu, jak i wypadkom „losowym” w postaci zarysowań. Taśmy magnetyczne straciły swoje właściwości najpierw z powodu ścierania (przesuwanie po rolkach), a następnie traciły stopniowo właściwości magnetyczne. Z powyższych powodów nie wystarczy jedynie dokonać akwizycji nagrań do postaci cyfrowej, lecz w miarę możliwości przywrócić ich pierwotne brzmienie. Te właśnie działania obejmuje termin restauracja nagrań¹.

Przenoszenia cennych zasobów z nośników analogowych na nośniki cyfrowe można dokonywać już od dłuższego czasu. Zostały opracowane metody sprzętowo-programowe (urządzenia, programy) umożliwiające przeniesienie danych z nośników analogowych na cyfrowe. Na rynku dostępnych jest wiele urządzeń i programów. Zadaniem urządzeń jest dokonanie wstępnej obróbki

¹ Restauracja z łac. *restoratio* – odbudowa.

analogowej², konwersji analogowo-cyfrowej oraz wykonanie przetwarzania w celu uzyskania pierwotnego brzmienia nagrania.

Podstawowym celem niniejszego artykułu jest przybliżenie problemów rozwiązywanych w procesie restauracji dźwięku ze szczególnym uwzględnieniem eliminacji zakłóceń impulsowych. Został także przedstawiony krótki przegląd urządzeń i oprogramowania wykorzystywanych do restauracji nagrań dźwiękowych.

1. Pojęcie restauracji nagrań dźwiękowych

Pojęcie restauracji nagrań obejmuje metody przetwarzania (filtracji) sygnałów umożliwiające uzyskanie pierwotnego brzmienia nagrań (odnowienie). Metody te są dość ściśle związane z charakterem zniekształceń, zakłóceń, jakie przedostały się do zapisu cyfrowego. Podstawowymi operacjami³ realizowanymi w ramach restauracji są: usuwanie zakłóceń impulsowych (trzasków, klików, popów itp.), redukcja szumu (poprawa stosunku sygnału do szumu), w tym redukcja wpływu tła (urządzeń odtwarzających). Nierzadko trafiają się także nagrania przesterowane wymagające korekcji, nagrania z ubytkami, wymagające rekonstrukcji całych ścieżek (w przypadku płyty winylowej) lub odtworzenia nagrania z wyciętego fragmentu taśmy (możliwe na przykład w przypadku refrenu piosenki). W dużej części zadania te mogą zostać zrealizowane w drodze filtracji, jednak nie wszystkie, niektóre wymagają zastosowania metod z dziedziny sztucznej inteligencji. Celem niniejszego artykułu jest przedstawienie problemu zakłóceń impulsowych oraz metod ich usuwania.

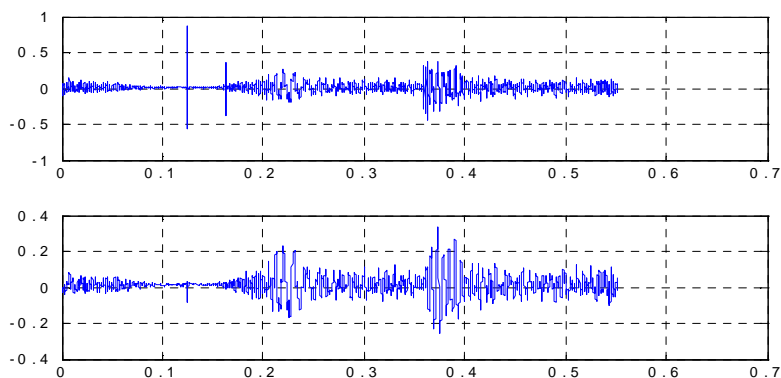
2. Usuwanie zakłóceń impulsowych

Zakłóceniami impulsowymi nazywamy wszelkiego rodzaju trzaski, kliki, które w sygnale znalazły się w wyniku np. mechanicznego uszkodzenia płyty w postaci poprzecznej rysy. Zakłócenia impulsowe cechują się krótkim czasem trwania i wyraźną amplitudą (są słyszalne). Natomiast ich widmo jest rozłożone w długim przedziale częstotliwości, a co za tym idzie amplitudy poszczególnych harmonicznych są niewielkie. Automatyczne usuwanie (redukcję) zakłóceń impulsowych można prowadzić dwoma podstawowymi sposobami. Pierwszy polega na zastosowaniu filtracji dolnoprzepustowej (liniowej lub nieliniowej,

² Np. ograniczenie pasma, korekcja widma sygnału.

³ Na podstawie funkcjonalności oferowanych przez wyspecjalizowane programy do restauracji dźwięku.

np. medianowej). Metody te dość skutecznie redukują zakłócenia impulsowe. Mają jednak niekorzystny wpływ na cały rekonstruowany sygnał poprzez usunięcie z niego wszystkich składowych wysokoczęstotliwościowych. Przykład zastosowania filtra medianowego przedstawiony został na rys. 1. Widoczne na przebiegu w górnym oknie zakłócenia impulsowe pomiędzy 0.1 sekundy i 0.2 sekundy zostały mocno zredukowane, lecz sam sygnał pozbawiony został istotnych „detali”.

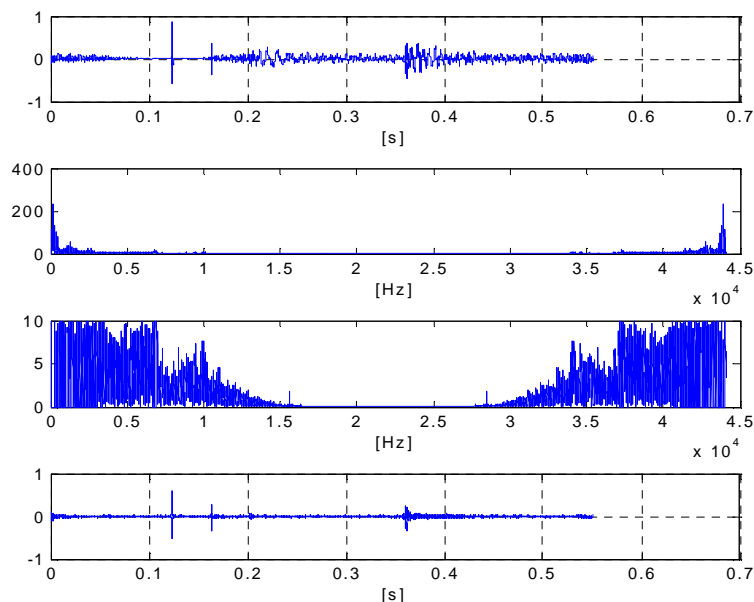


Rys. 1. Nagranie oryginalne (przebieg górny) i wynik filtracji medianowej (przebieg dolny). Widoczna redukcja trzasków występujących w przedziale 0.1- 0.2 s (opracowanie własne)

Drugi ze stosowanych sposobów polega na postępowaniu dwuetapowym. W pierwszym dokonywana jest lokalizacja zakłóceń impulsowych, a dopiero w drugim ich usunięcie. W przypadku ręcznej procedury wygląda to następująco: w edytorze graficznym wyszukiwany jest uszkodzony fragment, następnie zostaje usunięty, a w jego miejsce wstawiony fragment z bezpośredniego sąsiedztwa (z zachowaniem zasady łączenia fragmentów w punktach montażowych). W przypadku procedury automatycznej detekcji trzasków o krótkim czasie trwania zazwyczaj analiza jest prowadzona w oknach czasowych o długości od 256 do 2048 próbek (kilkanaście do kilkudziesięciu milisekund nagrania), przesuwanych o liczbę próbek mniejszą niż wynosi szerokość okna (sąsiednie okna nakładają się na siebie). Po zrekonstruowaniu sąsiadujących ramek w rejonie ich nakładania dokonywana jest operacja miksowania (np. przez uśrednienie próbek).

Jedną z podstawowych metod automatycznej detekcji trzasków jest metoda oparta na analizie widma amplitudowego transformaty Fouriera [8]. Z sygnału usuwane są harmoniczne o amplitudzie większej od przyjętego progu. Z uwagi

na to, że trzaski posiadają widmo rozproszone o stosunkowo małej amplitudzie, pozostaną w sygnale i będą dobrze „widoczne” (rys. 2). Pozwoli to na zlokalizowanie trzasków w nagraniu. Do detekcji zakłóceń impulsowych dobrze nadają się także sztuczne sieci neuronowe [8].



Rys. 2. Detekcja trzasków z wykorzystaniem analizy widma amplitudowego. Na wykresach przedstawiono (od góry): sygnał z trzaskami znajdującymi się w przedziale 0.1-0.2 s, widmo amplitudowe, widmo amplitudowe po wyzerowaniu składowych przewyższających próg, sygnał po wykonaniu odwrotnej transformaty Fouriera z wyeksponowanymi trzaskami (opracowanie własne)

Po detekcji przedziałów zawierających zakłócenia impulsowe dokonywana jest ich rekonstrukcja. Na podstawie przeglądu literatury można powiedzieć, że właśnie to zagadnienie jest kluczowe w procesie restauracji i cieszy się bardzo dużym zainteresowaniem badaczy. Starają się oni różnymi metodami dokonywać rekonstrukcji możliwie najdłuższych zdegradowanych odcinków. Działania te intensyfikuje fakt, że rekonstrukcja, oprócz restauracji, ma zastosowanie na przykład w przypadku utraty pakietów przy przesyłaniu dźwięku przez sieć komputerową. Do oceny jakości rekonstrukcji wykorzystywane są: stosunek sygnału do szumu SNR [17], błąd

średniokwadratowy MSE [7] oraz miara subiektywna PAQM⁴ [5], [3]. SNR i MSE są oczywiście możliwe do wyznaczenia przy rekonstrukcji fragmentów niezdegradowanych. Algorytmy rekonstrukcji oceniane są także pod względem złożoności obliczeniowej. Ma to duże znaczenie, nie tyle w restauracji, ile w rekonstruowaniu utraconych pakietów, gdzie operacja powinna być realizowana w czasie rzeczywistym.

Najczęściej stosowanym podejściem do zagadnienia rekonstrukcji sygnału jest ekstrapolacja z wykorzystaniem modelu AR⁵ [2], [3], [8], [17] lub interpolacja wielomianami [8]. Model AR pozwala na rekonstrukcję przedziałów o szerokości do 20 ms⁶ (ok. 1000 próbek przy częstotliwości próbkowania 44,1 kHz) [17]. Model ten jest wykorzystywany do ekstrapolacji: „w przód” oraz „w tył”. Ostateczny wynik rekonstrukcji jest wypadkową obu ekstrapolacji. W celu wyznaczenia parametrów modelu AR należy dysponować liczbą próbek co najmniej trzykrotnie większą od długości rekonstruowanego przedziału.

W literaturze można znaleźć algorytmy, które bazując na modelu AR, pozwalają na rekonstrukcję dłuższych przedziałów. W pracy [3] autorzy zastosowali model AR w podpasmach, wykorzystując bank filtrów. Wykazali, że analiza w tym przypadku wymaga zastosowania modeli AR niższych rzędów. Z kolei w artykule [17] zastosowano wielokanałowy model AR, uwzględniający sygnały z innych kanałów przesunięte w czasie. Otrzymane wyniki były zadowolające z uwagi na silną korelację pomiędzy kanałami. Uzyskano dobrej jakości rekonstrukcję odcinka nagrania o długości 240 ms.

Jak wspomniano wyżej, można do rekonstrukcji zastosować interpolację wielomianami, pamiętając jednak, że interpolacja wielomianami niskich rzędów może być stosowana do rekonstrukcji jedynie krótkich przedziałów⁷.

Oprócz wyżej wymienionych technik do rekonstrukcji utraconych próbek stosowane są sztuczne sieci neuronowe [21], [27]. Za ich pomocą można

⁴ Perceptual Audio Quality Measure.

⁵ ARMA (*autoregressive moving average*), model uwzględniający występowanie w sygnale szumu o znanej wariancji, jest definiowany jako:

$$u(k) = \sum_{i=1}^p a(i)u(k-i) + \sum_{j=0}^r b(j)e(k-j),$$

gdzie: $e(k)$ – szum, $a(i)$, $b(j)$ – współczynniki modelu.

AR (*autoregressive*), model wykorzystujący korelację istniejącą w sygnale, zakłada się w nim stacjonarność sygnału szumu, znany jest także pod nazwą LPC (*linear predictive coding*), model AR jest równoważny modelowi ARMA przy $r = 0$, $b(0) = 1$.

⁶ Co odpowiada odcinkowi czasu, w którym można przyjąć, że sygnał jest stacjonarny.

⁷ Przykładowo w systemie Cedar do interpolacji krótkich przedziałów wykorzystuje się interpolację wielomianem pierwszego rzędu (odcinkiem) [8].

dokonywać ekstrapolacji na drodze predykcji nieliniowej [7], [8], [24]. W pracach [7] i [21] pokazano zastosowanie neuronowej sieci nieliniowej do realizacji predykcji sygnałów w podpasmach. W pierwszej z prac wykorzystano sieć z sigmoidalną funkcją aktywacji (z wyjściową warstwą liniową). Otrzymano zadowalające wyniki interpolacji w przedziałach o długości do 113 ms (5000 próbek). W drugiej pracy zastosowano sieć z neuronami radialnymi. Przedstawiono zadowalające wyniki dla rekonstruowanych przedziałów o długości 1000 próbek. Ciekawy algorytm rekonstrukcji długich przedziałów zaprezentowano w [19]. Do określania samopodobieństwa w sygnale audio wykorzystano podejście znane z analizy tekstur w obrazach. Idea metody sprowadza się do zastępowania brakującego fragmentu najlepiej dopasowanym niezdegradowanym fragmentem pochodzącym z tego samego nagrania. Uzyskano dobre wyniki dla rekonstrukcji przedziałów o długości rzędu 1 sekundy.

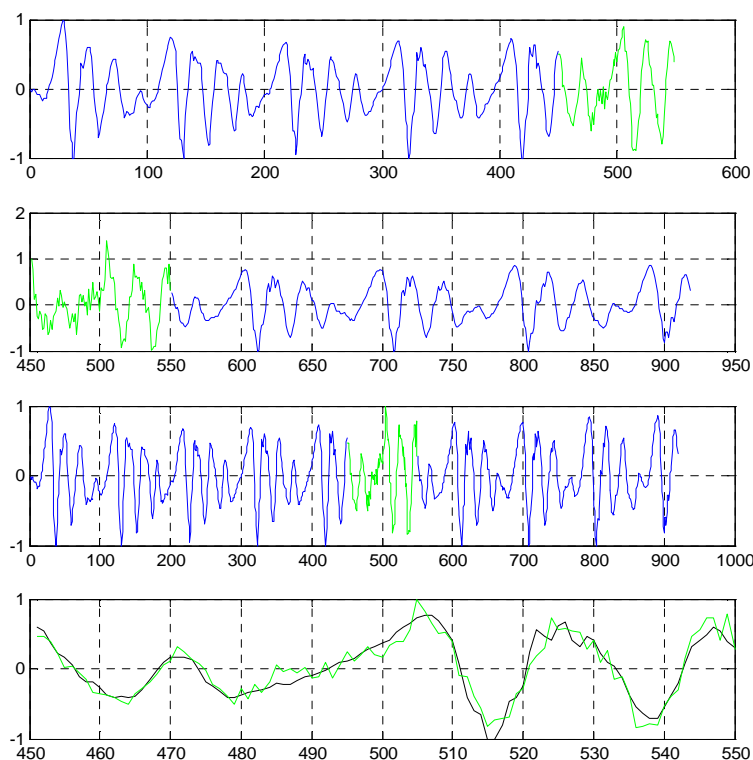
Na rys. 3 przedstawiony został przykład interpolacji sygnału metodą sieci neuronowych. Nie zastosowano analizy pasmowej. Wynik został osiągnięty na drodze rozwiązania zadania ekstrapolacji „w przód” (wykres górny, prognozę zaznaczono jaśniejszym odcieniem) oraz ekstrapolacji „w tył” (wykres drugi od góry, prognozę zaznaczono jaśniejszym odcieniem), złożenie obu prognoz i ostateczny wynik rekonstrukcji sygnału przedstawiono na wykresie trzecim od góry (ekstrapolowany przedział oznaczono jaśniejszym odcieniem). Na ostatnim wykresie przedstawiony został wynik interpolacji (jaśniejszy odcień) na tle sygnału oryginalnego. Przedział, w którym dokonywana była interpolacja, liczył 100 próbek. Zadanie zrealizowano przy pomocy dwuwarstwowej jednokierunkowej sieci neuronowej o 40 sigmoidalnych neuronach w warstwie pierwszej i 100 neuronach liniowych w warstwie drugiej. Każdy neuron warstwy wyjściowej dokonywał predykcji na jeden okres w przód. Rząd predykcji⁸ został dobrany empirycznie i przyjął wartość 120. Złożenia wyników ekstrapolacji dokonano, stosując kombinację wypukłą:

$$y(i) = a(i)y_p(i) + (1 - a(i))y_t(i), \quad i = 1, 2, \dots, L,$$

gdzie:

- L – szerokość interpolowanego przedziału,
- y_p – wynik ekstrapolacji w przód,
- y_t – wynik ekstrapolacji w tył,
- a – współczynniki kombinacji, $0 \leq a(i) \leq 1$.

⁸ Rozmiar sygnału podawanego na sieć, liczba próbek sygnału, na podstawie której wyznaczana jest prognoza.



Rys. 3. Interpolacja metodą nieliniowej sieci neuronowej – opis w tekście (opracowanie własne)

3. Przegląd funkcji i oprogramowania do restauracji dźwięku⁹

Do restauracji dźwięku można wykorzystać programy dedykowane do tego celu lub takie, których ogólnie użyteczne funkcje filtracji można wykorzystać w procesie restauracji dźwięku (redukcja szumu, redukcja przesterowania).

Do pierwszej grupy można zaliczyć programy (tabela 1): Cedar, reNOVator, iZotope, Dart Pro¹⁰. Programy z tej grupy oferują, wśród innych

⁹ Wyłącznie do celów informacyjnych.

¹⁰ Strony internetowe producentów oprogramowania do restauracji dźwięku:
<http://www.adobe.com/products/audition/index.html>, <http://www.algorithmix.com>,
<http://www.cedaraudio.com>, <http://www.dartpro.com>, <http://www.izotope.com>,
<http://www.sonycreativesoftware.com/soundforge>.

funkcji przetwarzania sygnałów, wyspecjalizowane moduły do restauracji dźwięku. Realizują one zadania detekcji i usuwania zakłóceń impulsowych (trzaski, kliki, popy), rekonstrukcji sygnału w długich oraz krótkich przedziałach, redukcji szumów, naprawy efektu przesterowania.

Tabela 1. Wybrane programy do obróbki sygnału audio z rozbudowanymi elementami do restauracji dźwięku (opracowanie własne na podstawie stron internetowych producentów)

Nazwa	Opis	Cena
Cedar	Profesjonalny pakiet sprzętowo-programowy. Jest uznany za jeden z mocniejszych systemów restauracji dźwięku pod względem wydajności i skuteczności. Najnowsze algorytmy restauracji dźwięku są odnoszone do wyników otrzymywanych w tym programie. Pakiet posiada budowę modułową. Do modułów z zakresu restauracji należą: Mauual Declick (usuwanie klików i trzasków), Dethump oraz system redukcji szumów RetouchT.	kilkanaście tysięcy USD
reNOVator	Program firmy Agorithmix. Program wyróżnia się dobrą wydajnością, inteligentnym algorytmem usuwania zakłóceń impulsowych.	ok. 2500 USD
iZotope RX	Program firmy iZotope. Likwiduje zakłócenia impulsowe (kliki, popy). Redukuje szum w szerokim paśmie częstotliwości. Redukuje efekty przesterowania (clipping), likwiduje przydźwięki sieciowe.	350-1200 USD
DartPro MT	Program firmy Dart Pro. Likwiduje zakłócenia impulsowe. Realizuje rekonstrukcję sygnału w długich i krótkich przedziałach. Redukuje szum w wąskim oraz szerokim paśmie częstotliwości. Wykorzystuje do tego model psychoakustyczny człowieka.	ok. 300 USD

Do popularnych programów z grupy drugiej można zaliczyć takie programy jak: Adobe Audition (dawniej Cool Edit), Sound Forge czy Matlab. Adobe Audition posiada opcję Klik/Pop/Crackle Eliminator do odnowy nagrań oraz funkcję redukcji szumów. Matlab jest bardzo rozbudowanym programem (środowiskiem z własnym językiem programowania) przeznaczonym do obliczeń naukowo-technicznych. Możliwość jego wykorzystania do restauracji dźwięku wynika z tego, że posiada bardzo bogate biblioteki funkcji standardowych, między innymi z dziedziny cyfrowego przetwarzania sygnałów oraz sieci neuronowych. Umożliwia to realizację szerokiej gamy eksperymentów.

Podsumowanie

W artykule omówiono problem usuwania zakłóceń impulsowych występujących w nagraniach archiwalnych. Z badań literaturowych wynika, że należy on do zasadniczych problemów restauracji dźwięku. Przedstawiono podstawowe podejścia do eliminacji zakłóceń impulsowych i wskazano konkretne rozwiązania opisane w literaturze tematu. Jedno z nich, oparte na sztucznej sieci neuronowej, zostało wykorzystane do realizacji przykładowego rozwiązania rekonstrukcji zdegradowanego fragmentu nagrania. Na zakończenie dokonany został krótki przegląd oprogramowania stosowanego do restauracji dźwięku.

Literatura

- [1] BARNWELL T.P., *Recursive Windowing for Generating Autocorrelation Coefficients for LPC Analysis*, IEEE Trans. ASSP, Vol. 29, no. 5, 1981, pp. 1062-1066.
- [2] BISCAINHO L.W.P., DINIZ P.S.R., ESQUEL P.A.A., *A Model for an ARMA Process Split in Sub-Bands*, in Proc. 2000 IEEE Intern. Symposium on Circuits and Systems, vol. III, IEEE May 2000, pp. 97-100.
- [3] BISCAINHO L.W.P., DINIZ P.S.R., ESQUEL P.A.A., *ARMA Process in Sub-Bands with Application to Audio Restoration*, Circuit and Systems, 2001 ISCAS, 2001 IEEE International Symposium on Volume 2, 6-9 May 2001, pp. 157-160.
- [4] BILIŃSKI J., *Odnawianie starych nagrań dźwiękowych*, Estrada i Studio, 2002-07.
- [5] BEERENDS J.G., STEMERDING J.A., *A perceptual audio quality measure based on a psychoacoustic sound representation*, J.Audio Eng. Soc., vol. 40, Dec. 1992, pp. 963-978.
- [6] CANANZA S., DE POLI G., MIAN G.A., *Restoration of audio documents by means of Extended Kalman Filter*, 2009 IEEE Intern. Draft.
- [7] COCCHI G., UNCINI A., *Subbands Audio Signal Recovering using Neural Nonlinear Prediction*, Acoustics, Speech, and Signal Processing, 2001 IEEE International Conference, vol. 2, 7-11 May 2001, pp. 1289-1292.
- [8] CZYZEWSKI A., *Dźwięk cyfrowy. Wybrane zagadnienia teoretyczne, technologia, zastosowanie*, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2001.
- [9] CZYZEWSKI A., *Learning Algorithms for Audio Signal Enhancement, Part 1: Neural Network Implementation for the Removal of Impulse Distortions*, Journ. of Audio Engineering Society, vol. 45, No. 10, Oct. 1996, str. 815-831.
- [10] GODSILL S.J., RAYNER P.J.W., *Frequency- Based Interpolation of Sampled Signals with Applications in Audio Restoration*, Acoustics, Speech, and Signal

- Processing, 1993 IEEE International Conference, vol. 1, April 1993, 1993 IEEE Intern., pp. 209-212.
- [11] GODSILL S.J., RAYNER P.J.W., *A Bayesian Approach to the Restoration of Degraded Audio Signals*, 1995 IEEE Intern. IEEE Trans. On Speech and Audio Processing, vol. 3, No. 4, July 1995, pp. 267-268.
- [12] GODSILL S.J., RAYNER P.J.W., *Robust Noise Modelling with Application to Audio Restoration*, Applications of Signal Processing to Audio and Acoustics, 1995, IEEE ASSP Workshop, Oct. 1995, pp. 143-146.
- [13] GRAD L., *Badania porównawcze zastosowania liniowego i nieliniowego modelu predykcji w analizie sygnału mowy*, Biuletyn IAIr, nr 10, 1999.
- [14] HEINES R., *Cyfrowe przetwarzanie dźwięku*, Mikom, Warszawa, 2002.
- [15] KAUPPINEN I., ROTH K., *Improved Noise Reduction In Audio Signals Using Spectral Resolution Enhancement With Time- Domain Signal Extrapolation*, 2005 IEEE intern., IEEE Trans. On Speech and Audio Processing, Vol. 13, No 6, November 2005.
- [16] KWIATKOWSKI W., *Wstęp do cyfrowego przetwarzania sygnałów*, Warszawa, 2001.
- [17] LIN H., GODSILL S., *The Multi-channel AR Model for Real-time Audio Restoration*, 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2005, pp. 335-338.
- [18] LLOYD R.G., *Wprowadzenie do cyfrowego przetwarzania sygnałów*, WKŁ, Warszawa, 2000.
- [19] LU L., MAO Y., WENYIN L., ZHANG H.J., *Audio Restoration by Constrained audio Texture Synthesis*, Multimedia and Expo, 2003 ICME. Proceedings. International Conference, July 2003, vol. 3, pp. 405-408.
- [20] O RUANAIDH J.J.K., FITZGERALD W.J., *Interpolation of missing Samples for audio restoration*, IEE Electronic Letters, April 1994, Vol. 30, No. 8, pp. 622-623.
- [21] OSOWSKI S., *Sieci neuronowe do przetwarzania informacji*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2000.
- [22] PAPULIS A., *Probability, Random Variables and Stochastic Processes*, section 10.4, New York, USA: McGraw-Hill, 3 ed., 1991 {model arma – źródło}, pp. 332-336.
- [23] *Przedwzmacniacz gramofonowy RIAA*, Elektronika dla Wszystkich, 2003-10.
- [24] SENG K.P., HUI L.E., MING T.K., *Multimedia Signal Processing Using AI*, Communications, APPC 2003 Asia-Pacific Conference, Vol. 2, pp. 825-829.
- [25] SCOTT H.R.R., WILSON R., *A Multiresolution Audio Restoration Algorithm*, Applications of Signal Processing to Audio and Acoustics, 1995. IEEE ASSP Workshop, Oct. 1995, IEEE Intern., pp. 151-154.

- [26] SCOTT H.R.R., WILSON R., *Multiresolution Techniques in Audio Restoration*, PhD thesis, Department of Computer Science, The University of Warwick, UK, March 1995.
- [27] ŻURADA J., BARSKI M., JĘDRUCH W., *Sztuczne sieci neuronowe*, Wydawnictwo Naukowe PWN, Warszawa, 1996.
- [28] VASEGHI S.V., RAYNER P.J., *The Effects of Non-Stationary Signal Characteristics on The Performance of Adaptive Audio Restoration Systems*, Acoustics, Speech, and Signal Processing, 1989, ICASP-93, 1993 IEEE International Conference, vol. 1, April 1993, pp. 209-212.
- [29] VASEGHI S.V., *Algorithms for Restoration of Archived Gramophone Recordings*, PhD thesis, University of Cambridge, 1988.

Restoration of audio recordings – impulsive noise removing

ABSTRACT: This paper presents problems of audio recordings restoration, especially impulsive noise suppression. Additionally, short review of computer programs for audio restoration was included.

KEYWORDS: audio restoration, impulsive noise suppression, neural networks

Praca wpłynęła do redakcji: 22.12.2009.