

TRANSLACJA CYFROWEGO OBRAZU ŚCIEŻKI DŹWIĘKOWEJ NA DŹWIĘK CYFROWY

ADAM KUPRYJANOW

Katedra Systemów Multimedialnych, Politechnika Gdańska
ul. G. Narutowicza 11/12, 80-952 Gdańsk
minister_forest@o2.pl

ABSTRACT

The aim of this work is to present algorithm for the analog optical sound track translation to the digital sound. It can be use to make restoration of the old film soundtracks. Instead of repair sound, digital pictures of the sound track can be corrected, and then translated it to restored digital sound. This method can also be useful in cases when the film tape is damaged and it can't be play in traditional way because it will split. In this document was featured method of the optical sound track digitalization, algorithm for the picture to sound translation and and results of it work.

Optyczna ścieżka dźwiękowa, gęstościowy zapis dźwięku, powierzchniowy zapis dźwięku, cyfrowy odczyt analogowego zapisu, rekonstrukcja archiwalnych optycznych ścieżek dźwiękowych

1. WPROWADZENIE

Praca ta przedstawia metodę translacji cyfrowego obrazu optycznych ścieżek dźwiękowych na dźwięk cyfrowy. Proces ten może zostać wykorzystany do rekonstrukcji filmowej ścieżki dźwiękowej znajdującej się na archiwalnych taśmach filmowych. Rekonstrukcja dźwięku może odbywać się poprzez rekonstrukcję cyfrowego obrazu ścieżki optycznej. Po translacji zrekonstruowanego obrazu ścieżki otrzymuje się zrekonstruowany dźwięk. Możliwe jest także wykorzystanie algorytmu do odczytu dźwięku zapisanego na taśmach, które poprzez długoletnie składowanie uległy osłabieniu i zrywają się podczas odtworzenia na projektorze. Jeżeli taśma jest jedyną istniejącą kopią danego materiału filmowego jej uszkodzenia podczas odtwarzania mogą spowodować brak możliwości digitalizacji niezniekształconego dźwięku.

2. ANALOGOWE OPTYCZNE ŚCIEŻKI DŹWIĘKOWE

Ścieżki z optycznym zapisem dźwięku rozpowszechniły się w technice filmowej już na początku lat 20-tych XX wieku. Dźwięk umieszczany był na tej samej taśmie co obraz filmowy, jednak wyprzedzał on obraz o kilkanaście klatek. Ścieżka dźwiękowa, w zależności od szerokości taśmy (35mm, 16mm czy 8mm), występowała pomiędzy perforacją a klatką obrazu lub bezpośrednio na krawędzi taśmy nie oddzielonej przez perforację. Mogła ona znajdować się zarówno po lewej jak i po prawej stronie taśmy filmowej.

Pierwszym typem zapisu optycznego był zapis gęstościowy. Amplituda sygnału dźwiękowego była w nim reprezentowana poprzez stopień zaczernienia ścieżki dźwiękowej. Pełne zaczernienie ścieżki odpowiadało brakowi sygnału, a ścieżka całkowicie przezroczysta reprezentowała maksymalną do zarejestrowania amplitudę sygnału.

Drugą metodą zapisu optycznego był zapis powierzchniowy. W tej metodzie poziom sygnału odwzorowywany był przez stosunek powierzchni pola czarnego do powierzchni pola przezroczystego. Jeżeli ścieżka była całkowicie czarna, oznaczało to, że na filmie rejestrowana jest cisza. W momencie, gdy pole przezroczyste zajmowało całą szerokość ścieżkę dźwiękową sygnał zapisywany miał maksymalny możliwy poziom. Zapis powierzchniowy posiadał kilka odmian. Bez względu na odmianę zapisu powierzchniowego zasada dotycząca reprezentacji poprzez stosunek zaczernienia pól czarnych i białych na taśmie była zachowana. Na rysunku 1. przedstawiono klatki filmu z optycznymi ścieżkami dźwiękowymi.

Odczyt optycznej ścieżki dźwiękowej przy użyciu projektora filmowego polegał na prześwietleniu taśmy filmowej światłem o stałym natężeniu w miejscu występowania zapisu. Światło po przejściu przez taśmę padało na fotodetektor umieszczony po drugiej stronie taśmy. Napięcie wytworzone przez ten fotoelement reprezentowało amplitudę sygnału dźwiękowego. Nie zależnie od rodzaju zapisu optycznego odczyt sygnału dźwiękowego był taki sam[3][5].



Rysunek 1. a) przykład zapisu gęstościowego na taśmie 35mm b) przykład zapisu powierzchniowego symetrycznego pojedynczego na taśmie 35mm c) przykład zapisu powierzchniowego symetrycznego podwójnego na taśmie 16mm

W celu stworzenia algorytmu pozwalającego na translacje obrazu ścieżek dźwiękowych na dźwięk cyfrowy konieczne było ustalenie formatu obrazu, który zostawał poddawany przetwarzaniu. Materiał filmowy przetwarzany był do postaci cyfrowej przy użyciu skanera Plustek OpticFilm 7200i. Jest to skaner przeznaczony do skanowania klisz i negatywów. Przyjęto, że optymalną rozdzielczością skanera będzie rozdzielczość 3600 dpi, gdyż jest to najmniejsza rzeczywista rozdzielczość skanera, która nie jest mniejsza od rozdzielczości taśmy filmowej (ziarnistość taśmy wynosi około 60-80 linii/mm, tak więc rozdzielczość taśmy wynosi 1524-2032 dpi). Od rozdzielczości obrazu ścieżki dźwiękowej zależy częstotliwość próbkowania dźwięku cyfrowego uzyskanego w procesie translacji.

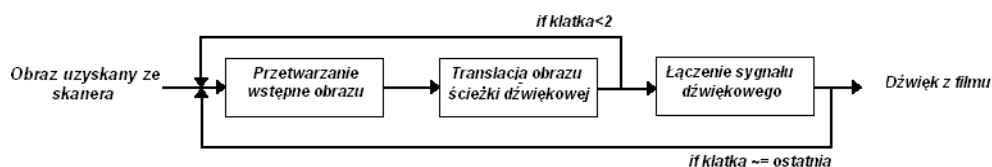
Dźwięk uzyskany z obrazu ścieżki dźwiękowej o rozdzielczości 3600 dpi pochodzącej z filmu 35mm ma częstotliwość próbkowania równą 68 kHz.

Zakres dynamiki cyfrowego sygnału dźwiękowego jest zależny od liczby możliwych poziomów przyjmowanych przez sygnał. Oznacza to, że im więcej różnych odcieni szarości występuje w obrazie ścieżki optycznej zapisu gęstościowego tym większa jest dynamika sygnału odczytanego z obrazu. W przypadku zapisu powierzchniowego wpływ liczby bitów reprezentujących kolor na dynamikę uzyskanego sygnału jest mniejszy (wynika to ze specyfiki zapisu). Przyjęto, że należy skanować obraz z maksymalną dostępną dla skanera głębią bitową. Wynosi ona 16 bitów na kolor. Teoretycznie powinno to pozwolić na uzyskanie dynamiki sygnału 96 dB (zapis gęstościowy).[1] W rzeczywistości dynamika ta nie jest osiągana z tego powodu, że sygnał zapisany na taśmie ma dynamikę około 40 dB. W przypadku zapisu powierzchniowego wpływ na dynamikę ma rozdzielczość obrazu ścieżki. Przy rozdzielczości 3600 dpi dynamika dźwięku powinna wynosić około 48 dB.

Istotne jest nie wprowadzanie w procesie skanowania korekcji żadnego z parametrów obrazu takich jak wyrównania histogramu, zmiany gammy lub jasności pikseli. W celu umożliwienia lepszego dopasowania sygnału z kolejnych klatek filmu obraz skanowany jest nadmiarowo. Oznacza to umieszczanie w obrazie jednej klatki, fragmentów klatki poprzedniej i następnej (rysunek 1). Obraz zapisywany jest w bezstratnym formacie tiff.

3. REALIZACJA ALGORYTMU

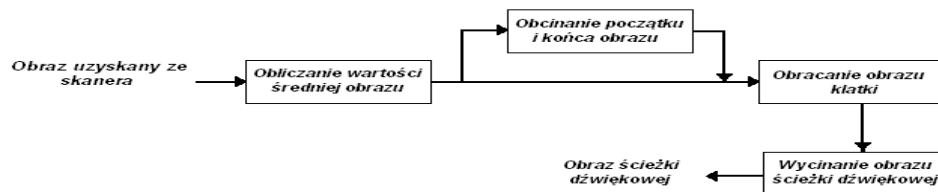
Proponowany algorytm służący do translacji obrazu ścieżki dźwiękowej na dźwięk cyfrowy składa się z trzech głównych bloków. Pierwszym z nich jest przetwarzanie wstępne, którego zadaniem jest usunięcie z obrazu klatki filmowej wszelkich zbędnych elementów i pozostawienie jedynie obrazu ścieżki dźwiękowej. W kolejnym kroku dokonywana jest translacja ścieżki dźwiękowej z pojedynczych klatek na sygnał cyfrowy. Ostatnim krokiem jest połączenie sygnałów reprezentujących poszczególne klatki w jeden ciągły sygnał odpowiadający ścieżce dźwiękowej dla danego fragmentu filmowego. Należy wspomnieć, iż algorytm został zoptymalizowany do przetwarzania obrazów zeskanowanych zgodnie z wcześniej przedstawionymi założeniami. Ogólny schemat blokowy algorytmu przedstawiono na rysunku 2.



Rysunek 2. Ogólny schemat blokowy algorytmu do translacji obrazu optycznej ścieżki dźwiękowej na dźwięk cyfrowy

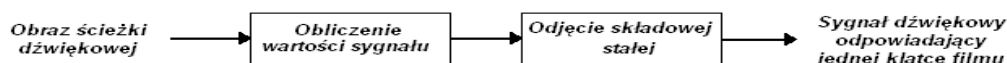
Przetwarzanie wstępne obrazu składa się z czterech etapów. Pierwszym jest obliczanie wartości średniej obrazu klatki filmowej. Na podstawie przebiegu wartości średniej obrazu sprawdzana jest poprawności obrazu wejściowego. Jeżeli na początku lub/i na końcu obrazu klatki występuje pole czarne oznacza to, że obraz wejściowy jest

niepoprawny i musi nastąpić obcięcie niepotrzebnego fragment. Kolejnym krokiem jest obrócenie klatki tak aby ścieżka dźwiękowa znalazła się po jej lewej stronie. Obrót wykonywany jest po to, aby możliwe było przetwarzanie ścieżek dźwiękowych pochodzących z różnych typów filmu. Ostatni krok polega na wycięciu obrazu ścieżki dźwiękowej. Schemat blokowy przetwarzania wstępnego umieszczono na rysunku 3.



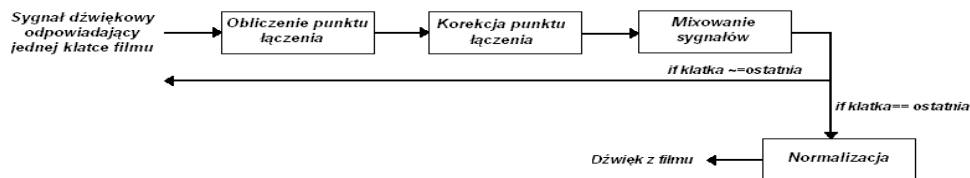
Rysunek 3. Schemat blokowy przetwarzanie wstępnego obrazu

Na podstawie obrazu ścieżki dźwiękowej obliczany jest sygnał dźwiękowy odpowiadający jednej klatce obrazu. Wykonywane jest to w bloku nazwanym blokiem translacji obrazu. Schemat translacji obrazu na dźwięk umieszczono na rysunku 4. Wartość sygnału liczona jest tu jako wartość średnia z pikseli obrazu w danym rzędzie dla obrazów czarno-białych lub jako wartość średnia luminancji obrazu ścieżki w rzędach dla obrazów kolorowych. Wartość ta jest równoważna wartości natężenia światła jakie padałoby na fotodetektor po przejściu przez taśmę filmową. Od uzyskanego sygnału dźwiękowego odejmowana jest składowa stała, która znacząco różni się w każdej klatce.



Rysunek 4. Schemat blokowy translacji obrazu jednej klatki

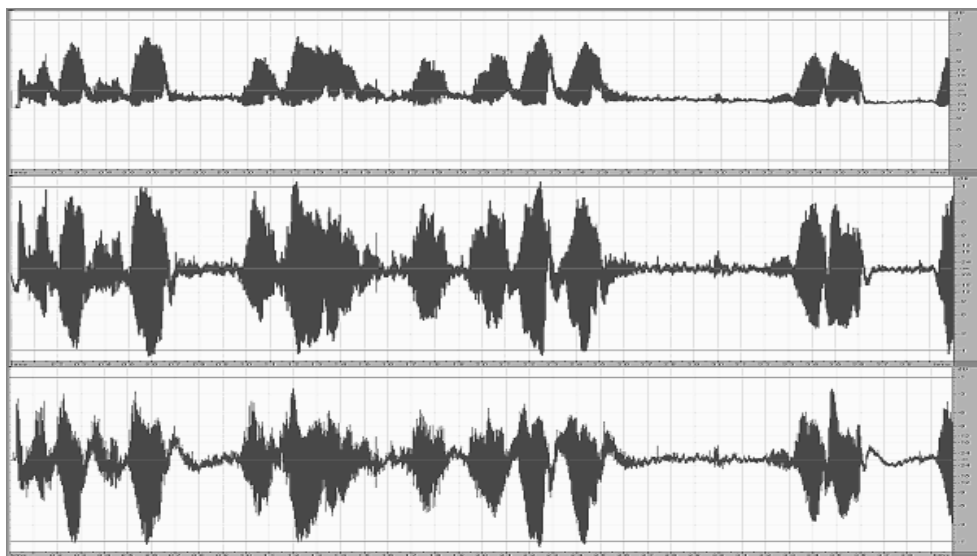
Sygnał dźwiękowy reprezentujący ścieżkę filmową uzyskuje się poprzez połączenie sygnałów z poszczególnych klatek filmu. Proces łączenia fragmentów sygnału przedstawiono na rysunku 5. Dzięki stosowaniu nadmiarowego skanowania obrazu w sygnałach dźwiękowych uzyskanych z dwóch kolejnych klatek koniec wcześniejszej klatki i początek późniejszej są prawie identyczne. Na podstawie analizy obrazu klatki obliczany jest punkt w którym klatki zaczynają być różne. Obliczenie tego punktu polega na policzeniu korelacji skrośnej wartości średniej obrazu wzdłuż ścieżki dźwiękowej. Liczba próbek o jaką przesunięte jest maksimum funkcji korelacji skrośnej względem jej środka określa miejsce, w którym sygnały zaczynają być podobne. Następnie poprzez obliczenie korelacji skrośnej dwóch teoretycznie podobnych sygnałów dźwiękowych, koryguje się punkt cięcia [4]. W celu zredukowania różnicy poziomów sygnału pochodzącego z kolejnych klatek filmu wykonuje się miksowanie podobnych fragmentów z końca jednej klatki i początku drugiej [2]. W momencie, gdy sygnały ze wszystkich klatek zostaną połączone wykonywana jest normalizacja poziomów sygnału. Polega ona na przeskalowaniu wartości sygnału z przedziału od 0 do 2^{16} do przedziału od -1 do 1 .



Rysunek 5. Schemat blokowy łączenia sygnałów odpowiadających jednej klatce filmu

4. WYNIKI DZIAŁANIA ALGORYTMU

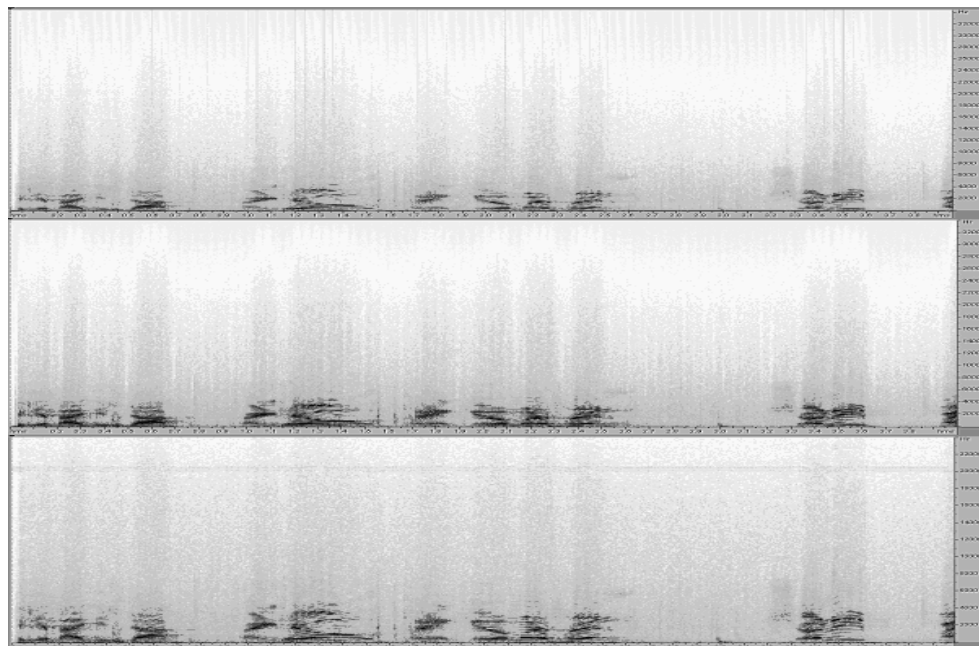
W pierwszej wersji algorytmu nie dokonywano odjęcia składowej stałej dla każdej klatki co skutkowało brakiem symetrii przebiegu czasowego dźwięku. Dzięki wprowadzeniu usuwania składowej stałej dla każdej klatki przebieg czasowy stał się symetryczny. Na rysunku 6 przedstawiono przebiegi czasowe sygnałów otrzymanych ze ścieżek dźwiękowych z optycznym zapisem gęstościowym. Pierwszy wykres to przebieg czasowy sygnału, w którym podczas translacji nie wykonywano usuwania składowej stałej dla każdej klatki. Widoczny jest wyraźny brak symetrii sygnału. Drugi przebieg przedstawia sygnał z usuwaną składową stałą. Widoczna jest znacząca różnica w stosunku do przebiegu pierwszego. Sygnał stał się symetryczny. Na ostatnim przebiegu pokazany jest dźwięk uzyskany bezpośrednio z projektora filmowego. Ma on kształt podobny do przebiegu drugiego. Pojawiające się różnice w sygnałach mogą być spowodowane błędami skanera.



Rysunek 6. Przebiegi czasowe sygnałów po translacji i sygnału pochodzącego z projektora

Na rysunku 7 przedstawiono spektrogramy dźwięku poddanego translacji oraz odczytanego bezpośrednio z projektora. Pierwszy spektrogram pochodzi od dźwięku

powstałego przy wykorzystaniu algorytmu, w którym nie było stosowane cięcie ścieżki dźwiękowej z dodatkową korekcją cięcia oraz sygnały końca jednej klatki i początku drugiej nie były miksowane. Na spektrogramie widoczne są trzaski powstałe w miejscach niedokładnego połączenia klatek. Drugi spektrogram przedstawia sygnał powstały poprzez translację obrazu przy użyciu algorytmu wykorzystującego korekcję punktu cięcia oraz miksowanie. Ostatni spektrogram reprezentuje sygnał uzyskany bezpośrednio z projektora. Można zaobserwować, że spektrogramy drugi i trzeci są bardzo podobne. W spektrogramie 2 nie występuje jednak pewna składowa częstotliwościowa w okolicach 20 kHz występująca w sygnale oryginalnym. Jest ona najprawdopodobniej związana z charakterystyką urządzenia.



Rysunek 7. Spektrogramy sygnałów po translacji i sygnału pochodzącego z projektora

5. WNIOSKI I PODSUMOWANIE

Przedstawiony algorytm translacji pozwala na uzyskanie dźwięku o jakości porównywalnej z dźwiękiem powstającym podczas odczytu dźwięku z projektora filmowego. Dynamika sygnału oryginalnego (odczytanego z projektora) jest wyższa o około 1-2dB od dynamiki dźwięku odczytanego z obrazu. Zaletą translacji sygnału z obrazu jest to, że dźwięk w górnym paśmie częstotliwości ma niższy poziom szumu od dźwięku oryginalnego. Algorytm może być pomocny w budowie urządzenia służącego do bezdotykowej digitalizacji ścieżek dźwiękowych. Taki sposób archiwizacji nagrań pozwalałby na rekonstruowanie dźwięku bez konieczności ingerencji w oryginalną taśmę.

Istotnym problemem jest długi czas translacji obrazu na dźwięk. Związane to jest z dużą liczbą danych obrazu, które muszą zostać przetworzone. Jedna minuta filmu czarno-

białego zajmuje około 50GB, aby uzyskać dźwięk pochodzący z tego materiału algorytm zaimplementowany w środowisku Matlab przetwarza ten dane ponad godzinę. Dlatego konieczne jest przeprowadzenie testów związanych z wpływem kompresji obrazu na jakość dźwięku uzyskanego podczas translacji obrazu.

PODZIĘKOWANIA

Opisane eksperymenty badawcze zostały wykonane w ramach prac w Europejskim projekcie: "PRESTOSPACE – Preservation towards storage and access. Standardized Practices for Audiovisual Contents Archiving in Europe", Integrated Project No. FP6-507336.

BIBLIOGRAFIA

- [1] CZYŻEWSKI A.: Dźwięk Cyfrowy, wyd. EXIT, Warszawa 1998.
- [2] KAUPPINEN I., ROTH K.: Audio signal extrapolation – theory and applications, proc. of the 5th int. conference on digital audio effects (DAFx –02), Hamburg, 2002
- [3] MAŁECKI I.: Technika Nagrywania i Odtwarzania Dźwięków, PWT, Warszawa 1953.
- [4] MAZIEWSKI P., SZCZUKO P. : System Synchronizacji nagrań wielokanałowych, Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej Nr.19, Seria: Zastosowanie Komputerów w Nauce i Technice 2003, str. 79 - 84, Gdańsk 2003.
- [5] URBAŃSKI B.: Technika Zapisywania i Odczytywania Dźwięków, wyd. kom. i łącz., Warszawa 1978