

XIV Seminarium
ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE' 2004
Oddział Gdański PTETiS

**WSTĘPNE PRZETWARZANIE DŹWIĘKU PRZEZNA-
CZONEGO DO ODSŁUCHU WIELOKANAŁOWEGO**

Przemysław MAZIEWSKI

Katedra Systemów Multimedialnych, Wydział Elektroniki Telekomunikacji i Informatyki,
Politechnika Gdańska, ul. G. Narutowicza 11/12, 80-952 Gdańsk
tel: (058) 347-23-01 e-mail: p_maziewski@sound.eti.pg.gda.pl,

W pracy opisano opracowane metody wstępnego przetwarzania plików dźwiękowych przeznaczonych do odsłuchu w systemie wielokanałowym. Umożliwiają one synchronizację i normalizację głośności. Normalizacja uwzględnia wpływ położenia głośników w panoramie odsłuchowej. Opis metod poprzedza krótka analiza głównych mechanizmów odpowiedzialnych za lokalizację dźwięku przez człowieka.

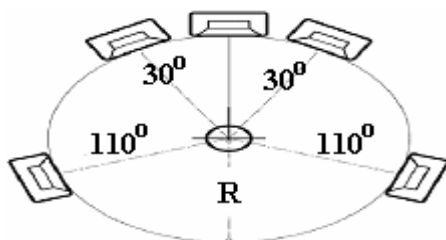
1. WSTĘP

Opisane poniżej metody synchronizacji i normalizacji głośności potrzebne są ze względu na założenia asynchroniczności i różnicy poziomów w rejestracji dźwięku. Taka sytuacja ma miejsce w przypadku rejestracji komunikacji radiowej za pomocą radiostacji ulokowanych w różnych miejscach. Każda z nich rejestruje nadawany dźwięk z innym poziomem wysterowania. Może to być spowodowane różnym tłumieniem fal radiowych odbieranych w poszczególnych radiostacjach. Dodatkowo, ze względu na różne prędkości propagacji fali dźwiękowej w kierunku poszczególnych odbiorników i ich różną odległość od nadajnika momenty rozpoczęcia rejestracji są różne. Otrzymane w ten sposób wersje nagrania są więc asynchroniczne i posiadają różny poziom głośności. Przy ich prezentacji za pomocą głośników wielokanałowego systemu odsłuchowego wystąpi problem przesunięcia się miejsca najlepszego odsłuchu. By umożliwić prawidłowy odsłuch nagrań niezbędne jest przetworzenie sygnałów w celu wyrównania ich głośności oraz momentów rozpoczęcia i zakończenia. Dodatkowo prawidłowy odsłuch pozwala wyeliminować wpływ nieskorelowanych zakłóceń z poszczególnych kanałów rejestracji.

2. WYBRANE ASPEKTY ODSŁUCHU WIELOKANAŁOWEGO

Jako odsłuch wielokanałowy rozumiany jest odsłuch przy pomocy pięciu głośników ustawionych w przestrzeni odsłuchowej zgodnie ze schematem z rysunku 1.

Zgodnie z zaleceniami ITU odległość pomiędzy poszczególnymi głośnikami powinna być odpowiednio dobrana. Jeśli odległość pomiędzy prawym i lewym głośnikiem przednim

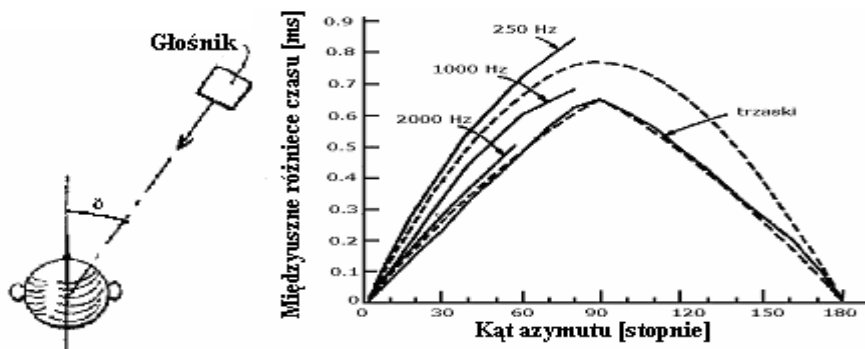


Rys. 1 Położenie głośników w pięciokanałowym systemie odsłuchowym [1]

oznaczymy jako B to promień okręgu R powinien mieć długość równą $0,9B$. Wartość B mieści się w przedziale od 2m do 4m. W przypadku takiej konfiguracji miejscem prawidłowego odsłuchu jest współśrodkowy okrąg o długości promienia do 0,8m. Znajduje się on wewnątrz dużego okręgu o promieniu R [1].

Wśród głównych mechanizmów odpowiedzialnych za lokalizację dźwięku w miejscu prawidłowego odsłuchu, należy wymienić: międzyuszne różnice czasu (ang. *ITD - Interaural Time Differences*), międzyuszne różnice w poziomach (ang. *IID - Interaural Intensity Differences*) oraz efekty filtrujące głowy, ramion, torsu i ucha [2].

Pierwszy mechanizm jest wynikiem różnic w czasie percepcji danego zdarzenia dźwiękowego pomiędzy dwoma uszami. Po lewej stronie rysunku 2 pokazano rzut z góry na sytuację odsłuchu pojedynczego źródła dźwięku. Wraz z przesuwanym się źródła dźwięku w skrajnie prawe pozycje, rośnie różnica w czasie docierania fali dźwiękowej pomiędzy uchem lewym i prawym. Obrazuje to wykres przedstawiony po prawej stronie rysunku.

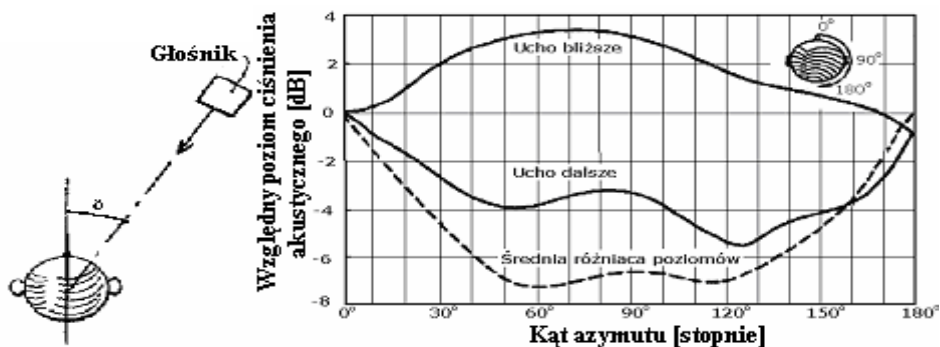


Rys. 2 Wykres przedstawiający międzyuszne różnice czasowe (ITD) w funkcji zmian kąta azymutu dla wybranych częstotliwości i sygnałów akustycznych [3]

ITD są znaczącym czynnikiem lokalizacyjnym dla częstotliwości dźwięku nie większych niż 1500Hz [2]. Zważywszy na opisany mechanizm, asynchroniczne odtwarzanie plików dźwiękowych przez poszczególne głośniki systemu wielokanałowego spowoduje niekontrolowane przesunięcie się miejsca prawidłowego odsłuchu. W celu uniknięcia takiej sytuacji zaproponowano metodę synchronizacji. Została ona opisana w następnym punkcie pracy.

Kolejnym mechanizmem lokalizacyjnym są różnice w natężeniu danego dźwięku, rejestrowane w obu uszach. Rysunek 3 zawiera wykres obrazujący względny poziom ciśnienia

nia akustycznego w obu uszach dla różnych kątów padania fali dźwiękowej. Wykres obrazuje ponadto przebieg średniej różnicy poziomów w funkcji kąta padania fali. Zakres uwzględnionych w badaniach częstotliwości akustycznych wynosił od 500 do 2000 Hz.



Rys. 3 Względny poziom ciśnienia dźwięku w obu uszach i średnia różnica poziomów w zależności od kąta padania fali w zakresie częstotliwości 500-2000 Hz [4]

Wraz z przesuwaniem się źródła dźwięku w skrajnie prawe pozycje rośnie różnica ciśnień fali dźwiękowej odbieranych przez ucho prawe i lewe. Widoczna na poprzednim wykresie symetria jest na tym wykresie złamana (na rysunku 2 krzywe czasowe od 0° do 90° i od 90° do 180° są symetryczne). Wynika to zapewne z kolejnego czynnika lokalizacyjnego jakim są efekty filtrujące głowy, ramion, torsu i samego ucha (np. małżowiny usznej). Fala dźwiękowa odbija się, załamuje i jest tłumiona przez te elementy. Proces zniekształcania (filtracji) jest dobrze widoczny na wykresie z rysunku 4. Prezentowane są na nim odpowiedzi impulsowe zarejestrowane w lewym (ciemna linia) i prawym (jasna linia) uchu przy pobudzeniu odtwarzanym z prawej strony słuchacza [2].



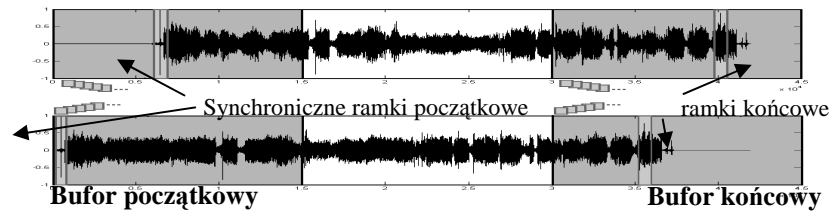
Rys. 4 Odpowiedzi impulsowe zarejestrowane w uchu lewym (ciemna linia) i prawym (jasna linia) przy pobudzeniu ulokowanym z prawej strony słuchacza

Zważywszy na opisane mechanizmy, jednoczesne odtwarzanie plików dźwiękowych z różnym poziomem w wielokanałowym systemie odsłuchowym spowoduje niekontrolowane zaburzenie miejsca prawidłowego odsłuchu. Proponowana metoda normalizacji głośności stosowana w celu uniknięcia takiej sytuacji została opisana w punkcie numer cztery.

3. SYNCHRONIZACJA PLIKÓW DŹWIĘKOWYCH

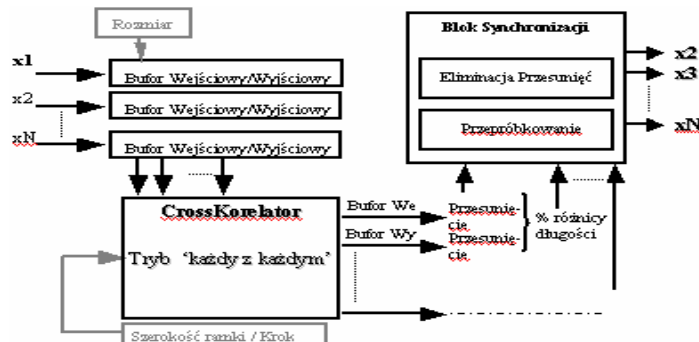
Synchronizacja jest niezbędna w przypadku wielokanałowego odtwarzania kilku wersji nagrania pochodzących np. z różnych stacji rejestrujących. Proponowana metoda opiera się na analizie funkcji korelacji. W wydzielonych buforach początkowych i końcowych wyszukiwane są podobieństwa pomiędzy poszczególnymi plikami dźwiękowymi. Lokalizacja najbardziej podobnych fragmentów nagrania pozwala określić wartości opóźnień i różnice w długościach plików. Następnie za pomocą odpowiedniego algorytmu można dokonać próby wyrównania plików na osi czasu.

Początkowy i końcowy bufor referencyjnego pliku dźwiękowego są porównywane z odpowiednikami z synchronizowanych plików dźwiękowych. Porównanie to obejmuje badanie maksimów funkcji korelacji skrośnej wyznaczonych dla ramek wybieranych z buforów. Maksima funkcji korelacji wskazują po dwie synchroniczne ramki w buforze początkowym i końcowym. Następnie wyrównywana jest długość plików (przepróbkowanie) i wykonywana eliminacja opóźnień. Maksymalna wartość korelacji interpretowana jest jako stopień poprawności wykonanych operacji. Rysunek 4 przedstawia proces wyszukiwania synchronicznych ramek w dwóch przykładowych plikach dźwiękowych [5, 6].



Rys.4. Proces wyszukiwania synchronicznych ramek w dwóch przykładowych plikach dźwiękowych

Praktyczna realizacja opisanej metody może przebiegać zgodnie z algorytmem przedstawionym schematycznie na rysunku 5. Pierwszy plik traktowany jest jako referencyjny ($x1$). Pozostałe będą do niego wyrównywane ($x2 .. xN$). W każdym pliku wyznaczone są bufor początkowy i końcowy. Następnie zgodnie z procedurą opisaną powyżej badane są przesunięcia pomiędzy plikiem referencyjnym a kolejnymi (blok CrossKorelator). Ostatnim krokiem jest eliminacja wykrytych przesunięć oraz różnic w długościach plików.

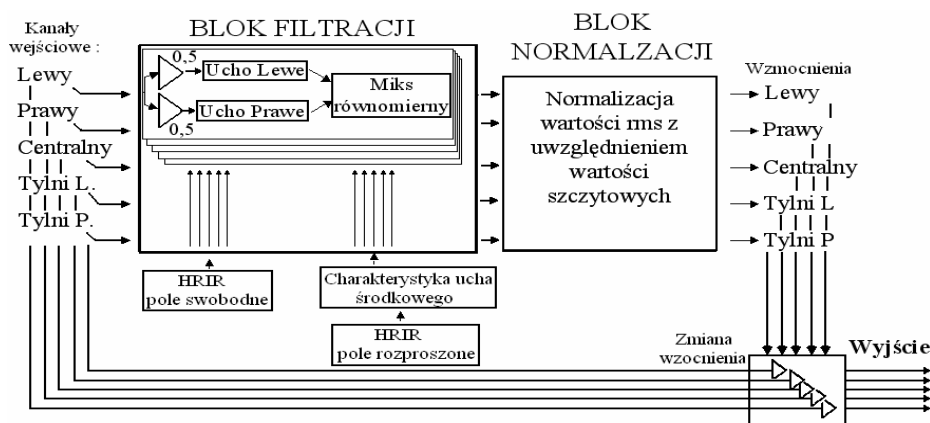


Rys. 5. Przykładowy schemat algorytmu synchronizacji

4. NORMALIZACJA GŁOŚNOŚCI PLIKÓW DŹWIĘKOWYCH

W przypadku posiadania kilku nagrań można je odtworzyć za pomocą oddzielnych głośników systemu wielokanałowego. Nagrania te powinny charakteryzować się zbliżonym poziomem głośności, tak by panorama dookólna tworzona przez odgrywane dźwięki była prawidłowa. W proponowanej metodzie ujednoczenie głośności plików dźwiękowych realizowane jest w trzech etapach. Odpowiadają one głównym blokom schematu przedstawionego na kolejnym rysunku (rys. 6).

Jako obiektywny parametr opisujący głośność skończonego sygnału wybrano jego wartość skuteczną. Ujednoczenie poziomów głośności następuje dzięki wyrównaniu ich wartości skutecznych. Poprzedzone jest ono wstępną filtracją symulującą efekty filtrujące opisane w punkcie 2.



Rys. 6 Schemat blokowy algorytmu normalizacji głośności [7].

Pierwszy etap przetwarzania polega więc na wstępnej filtracji kopii sygnałów. Ma ona na celu uwzględnienie wpływu czynników takich jak: położenie źródła dźwięku w stosunku do słuchacza, efekty filtrujące ramion, torsu oraz małżowiny i kanału usznego (patrz punkt 2). W filtracji wykorzystywane są odpowiedzi impulsowe, rejestrowane w komorze bezekhowej za pomocą mikrofonów umieszczonych w kanałach usznych manekina (ang. *HRIR* – *Head Related Impulse Response*). Wybór zestawu *HRIR* podyktowany jest położeniem głośników w pomieszczeniu odsłuchowym (powinien być zgodny z zaleceniami ITU-R [1]). W „bloku filtracji” monofoniczne ścieżki dźwiękowe są równomiernie dzielone. Kolejnym krokiem jest ich filtracja w oparciu o charakterystyki częstotliwościowe otrzymane na podstawie wybranych odpowiedzi *HRIR*. Wcześniejszy podział ścieżki na ucho lewe i prawe, pozwala na symulację słyszenia dwuosznego. Całość przetwarzania w „bloku filtracji” zakończona jest równomiernym miksem sygnałów.

Drugi etap polega na wyrównaniu wartości skutecznych przetworzonych wcześniej kopii. W tym celu w bloku normalizacji liczone są wzmocnienia potrzebne do uzyskania jednolitego poziomu *rms* wszystkich przefiltrowanych sygnałów. By nie dopuścić do ewentualnych przesterowań korygowany jest maksymalny poziom szczytowy.

W trzecim końcowym etapie, poziom oryginalnych (nieprzefiltrowanych) sygnałów modyfikowane są zgodnie z otrzymanymi współczynnikami wzmocnień [7].

5. REALIZACJA

Obie metody (w postaci algorytmów) zostały zaimplementowane jako części składowe systemu rekonstrukcji nagrań korespondencji pilotów lotniczych [8]. System został wykonany w Katedrze Systemów Multimedialnych Politechniki Gdańskiej. Obecnie testowany jest w Wyższej Oficerskiej Szkole Sił Powietrznych w Dęblinie.

6. BIBLIOGRAFIA

1. ITU Recommendation BS. 775-1: Multichannel stereophonic sound system with and without accompanying picture, Geneva 1997.
2. Moore, B.C.J.: Wprowadzenie do psychologii słyszenia, Warszawa – Poznań PWN 1999, ISBN 830112766X.
3. Benson, K.B.: Audio Engineering Handbook, New York Mc Graw-Hill 1988, ISBN 0070047774.
4. Steinberg, .C., Snow, W.B.: Physical factors in auditory perspective, Journal of SMPTE, Vol. 61, 1953, p. 420-430.
5. 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, Gdańsk 2003, s. 79 – 84.
6. Szczuko P., Maziewski P.: System synchronizacji nagrań wielokanałowych, Prace Naukowe Instytutu Telekomunikacji i Akustyki Politechniki Wrocławskiej, Seria: Konferencje - X Sympozjum Inżynierii i Reżyserii Dźwięku ISSET 2003, Wrocław 2003, s. 23-28.
7. Maziewski P.: Algorytm normalizacji poziomów głośności dźwięku zarejestrowanego w plikach, Zeszyty Naukowe Wydziału ETI Politechniki Gdańskiej Nr. 3, Seria: Technologie Informacyjne, Gdańsk 2004, s. 161-168.
8. Czyżewski A., Kaczmarek A., Kotus J., Pawlik A., Rypulak A., Żwan P.: A System for Multitask Noisy Speech Enhancement, 116 Audio Engineering Society Convention., Berlin 2004.

SOUND PREPROCESSING FOR MULTICHANNEL AUDIO MONITORING

Methods of initial sound processing for multichannel audio monitoring are presented. They allow sound synchronization and loudness normalization with regards to loudspeaker placement in the listening area. Additional short analysis of human sound localization mechanism is also included.