Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej Nr 19

XIII Seminarium
ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE 2003
Oddział Gdański PTETiS

TECHNIKA WIRTUALIZACJI DŹWIĘKU WYKORZYSTUJĄCA ODPOWIEDZI IMPULSOWE ZAREJESTROWANE ZA POMOCĄ SZTUCZNEJ GŁOWY W KOMORZE BEZECHOWEJ

Przemysław MAZIEWSKI

Politechnika Gdańska, ul. G. Narutowicza 11/12, 80-952 Gdańsk
tel: (058) 3472301 fax: (058) 3471114 e-mail: przemas@sound.eti.pg.gda.pl

Przedstawiono opracowany w Katedrze Systemów Multimedialnych WETI PG, komputerowy system przetwarzania sygnałów. Zadaniem systemu jest prawidłowe umieszczenie wirtualnego źródła dźwięku w panoramie dookólnej. Odsłuch dźwięku następuje za pomocą słuchawek stereofonicznych. Rozwiązanie wykorzystuje techniki splotu cyfrowego. Splotowi podlegają dany materiał dźwiękowy i odpowiedzi impulsowe zarejestrowane za pomocą sztucznej głowy w komorze bezechowej. W ostatniej części pracy pokazano uzyskane wyniki testów dokładności lokalizacji wirtualnych źródeł dźwięku. Dodatkowo, w pierwszym punkcie, znajduje się krótki opis metody wirtualizacji dźwięku za pomocą głośników.

1. IDEA WIRTUALIZACJI DŹWIĘKU PRZESTRZENNEGO

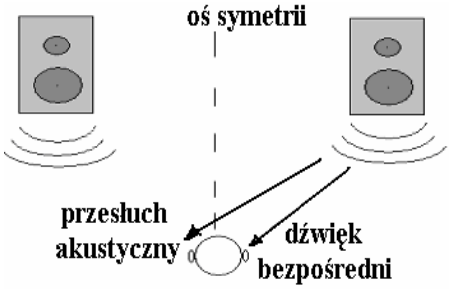
1.1. Wstęp

W niniejszej pracy za reprezentację dźwięku przestrzennego uważany jest materiał foniczny zapisany w jednym z licznych formatów wielokanałowych. Wirtualizacja rozumiana jest jako próba poprawnej reprodukcji dźwięku przestrzennego w panoramie dookólnej za pomocą odsłuchu dwukanałowego. Problem wirtualizacji dźwięku ma duże znaczenie w dziedzinach takich jak: telekonferencje, systemy wirtualnej rzeczywistości, symulatory odsłuchu wielokanałowego. Wirtualizery dźwięku można podzielić na dwie główne grupy: głośnikowe i słuchawkowe.

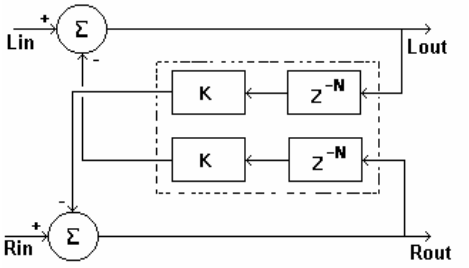
1.2 Reprodukcja dźwięku przestrzennego za pomocą głośników stereofonicznych

Dobrze znaną techniką wirtualizacji dźwięku, jest metoda opisana w 1936 r. przez M. Schroedera i B.S. Atala [2]. Podstawą tego rozwiązania jest założenie, iż proces wirtualizacji da pożądane wyniki, jeśli sygnał akustycznego przesłuchu pomiędzy danym głośnikiem a przeciwnym do niego uchem (np.: prawy głośnik i lewe ucho) zostanie odpowiednio osłabiony. Praktyczna realizacja następuje poprzez, wprowadzony elektronicznie do toru przeciwnego głośnika, sygnał redukujący przesłuch (jest to odwrócony w fazie, odpowiednio stłumiony i opóźniony sygnał z aktualnego głośnika).

Przesłuch akustyczny (rys. 1) jest sygnałem rejestrowanym w uchu, będącym odpowiedzią na pobudzenie pochodzące z przeciwnego do danego ucha głośnika.



Rys. 1 Schemat obrazujący powstawanie przesłuchów akustycznych



Rys. 2 Prosty układ redukcji przesłuchów akustycznych [2]

Sprzyjającym jest fakt, iż proces powstawania przesłuchów akustycznych jest „w przybliżeniu" liniowy i niezmienny w czasie [2]. W rezultacie sam proces może być symulowany jako liniowy, czterokońcówkowy filtr (dwa wejścia i dwa wyjścia jak na rys.2), którego wejściami są akustyczne wyjścia głośników, a wyjścia filtru prowadzą do uszu słuchacza. Funkcje przejścia procesu powstawania przesłuchów akustycznych mogą być scharakteryzowane przez pomiar efektywnej odpowiedzi impulsowej z każdego głośnika w każdym uchu, a następnie wyrażone jako kwadratowa macierz transmitancji [2].

Systemy redukcji przesłuchów mogą być wykonane jako algorytmy (lub układy elektroniczne) składające się z przyczynowych filtrów FIR. Zasada działania takich algorytmów oparta jest na następującej analizie. Podstawową różnicą pomiędzy akustycznymi sygnałami docierającymi do ucha bliższego i dalszego w stosunku do danego głośnika, jest fakt, iż sygnał docierający do ucha dalszego jest opóźniony i osłabiony. Generacja sygnału redukującego przesłuch akustyczny polega więc na odejmowaniu od kanału przeciwnego (do tego, który jest źródłem przesłuchu) sygnału odpowiednio opóźnionego i osłabionego. Oczywiście sygnał redukujący przesłuch, stanie się w chwili wyemitowania go przez głośnik, źródłem nowych przesłuchów. Dlatego również on musi być redukowany przez odpowiedni sygnał w drugim głośniku a pochodzący z toru emitującego go pierwszego głośnika (kanału). W konsekwencji - proces redukcji przesłuchów powoduje powstawanie nowych przesłuchów będących przyczyną długich odpowiedzi impulsowych, co z kolei utrudnia proces implementacji komputerowej oparty na filtrach FIR.

Schemat przedstawiony na rys. 2 pokazuje prosty przykład sposobu redukcji przesłuchów akustycznych. Opóźnienie z^{-N}, typowo jest wartością około 140 μs dla głośników rozstawionych +/- 15° w stosynku do osi symetrii. Wartość opóźnienia można rówież wyrazić w próbkach - w przybliżeniu 6 próbek (przy szybkości próbkowania 44100 próbek na sekundę [Sa/s]). Wartość współczynnika tłumienia K wynosi zazwyczaj 0,9. Wejście każdej z dwóch ścieżek algorytmu redukcji przesłuchów (jedna dla każdego kanału) jest wyprowadzone z wyjścia odpowiedniego sumatora miedzykanałowego sprzężenia zwrotnego [2].

Kolejnym czynnikiem, jaki należy uwzględnić w prezentowanym schemacie, jest wpływ częstotliwości na proces powstawania przesłuchów w dalekim polu akustycznym. Opis tego zjawiska można znaleźć w publikacjach panów Davisa i Fellersa [2].

MOST WIEDZY Pobrano z mostwiedzy.pl

Zaprezentowana w pierwszej części niniejszego podpunktu technika stosowana jest przede wszystkim w celu rozszerzenia bazy w reprodukcji stereofonicznego materiału źródłowego. Otrzymywane rezultaty są zazwyczaj różne dla poszczególnych materiałów źródłowych. W systemie *Dolby Surround* procesy kodowania i dekodowania są wzajemnie jednoznaczne (poszczególne ścieżki dźwiękowe dedykowane są konkretnym głośnikom w systemie odsłuchowym). Oznacza to, że odtworzenie materiału audio zapisanego w tym formacie nie może polegać jedynie na poszerzaniu panoramy dźwięku. Musi nastąpić rzeczywiste odwzorowanie materiału źródłowego w panoramie dookólnej (poszczególne ścieżki dźwiękowe muszą być lokalizowane w miejscu odpowiadającym położeniu głośnika).

W świetle powyższego akapitu przedstawione rozważania należałoby uzupełnić o zagadnienia poprawnego przetwarzania i końcowego miksu pięciu albo sześciu (w zależności od systemu) kanałów charakterystycznych dla systemów dookólnych.

1.3 Reprodukcja dźwięku wielokanałowego za pomocą słuchawek stereofonicznych

Jedną z technik reprodukcji dźwięku wielokanałowego za pomocą słuchawek stereofonicznych jest metoda oparta na funkcjach *Head Related Impulse Response* – odpowiedzi impulsowych zarejestrowanych w kanałach usznych słuchacza, dalej określanych skrótem HRIR [1, 3, 7, 9].

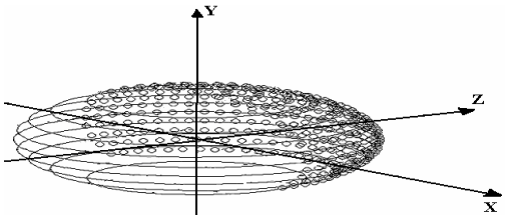
Idea HRIR opiera się na właściwościach splotu odpowiedzi impulsowej danego pomieszczenia z wybranym sygnałem akustycznym. Dysponując odpowiedzią impulsową przykładowego pomieszczenia, można uzyskać efekt wirtualnego pomieszczenia odsłuchowego o zbliżonych parametrach akustycznych. W tym celu dany sygnał akustyczny splata się z odpowiedzią impulsową danego pomieszczenia [5, 8]. Odpowiedni zestaw odpowiedzi impulsowych HRIR nagrywa się w komorze bezechowej [5]. Charakterystyki częstotliwościowe HRTF (ang. *Head Related Transfer Function*) są to transformaty Fouriera funkcji HRIR. W dalszej części skróty HRTF i HRIR będą używane zamiennie, jako jednoznacznie określające daną odpowiedź impulsową. Pojedyncza funkcja HRIR odpowiada danemu położeniu (kąty: α - elewacja, czyli odchylenie kątowe w pionie, β - azymut, czyli odchylenie kątowe w poziomie, w stosunku do odpowiednich osi symetrii) głośnika (pobudzenia) w stosunku do mikrofonu (rejestratora odpowiedzi impulsowej) usytuowanego wewnątrz kanału usznego [5]. Splatając uzyskaną w komorze bezechowej odpowiedź HRIR z sygnałem akustycznym i odtwarzając wynik za pomocą słuchawek stereofonicznych, uzyskuje się efekt wirtualnego źródła dźwięku usytuowanego w panoramie dookólnej w miejscu wyznaczonym przez kąty α i β. Co najważniejsze, słuchacz powinien lokalizować takie wirtualne źródło „na zewnątrz" głowy, a nie w „jej środku" – co ma miejsce w „normalnym" stereofonicznym odsłuchu słuchawkowym. Jako „normalny" traktowany jest odsłuch bez wcześniejszej wirtualizacji dźwięku. Dużą zaletą techniki opartej na HRTF jest to, że zachowuje ona całościowy wzorzec różnic „międzyuszych": ITD – ang. *Interaural Time Differences*, czyli „międzyuszne" różnice w czasie percepcji danego pobudzenia, i IID ang. *Interaural Intensity Differences* – „międzyuszne" różnice w poziomie percepcji danego pobudzenia. Dodatkowo zachowane są informacje o wpływach efektów filtrujących: małżowiny usznej (ang. *pinna efect*), głowy, ramion i torsu [7].

MOST WIEDZY Pobrano z mostwiedzy.pl

2. SYSTEM WIRTUALIZACJI DZWIĘKU – PRAKTYCZNA REALIZACJA

2.1 Opis systemu

Przedstawiona w pp. 1.3 metoda splatania odpowiedzi impulsowych HRIR leży u podstaw przedstawianego systemu. Został on zaimplementowany w środowisku Matlab. Bazuje on na zestawie funkcji HRIR zarejestrowanych w komorze bezechowej przez panów Keitha i Gardnera [3]. Rys. 3 przedstawia zestaw funkcji HRIR prezentowany w zobrazowaniu trójwymiarowym. Poszczególne punkty (kółka) na wykresie oznaczają „położenie" HRIR (wyznaczone przez kąty α i β) w przestrzeni. Definicje płaszczyzn w przestrzeni pokazuje rys. 4.



Rys. 3 Pełen zestaw funkcji HRIR w przestrzeni 3D

Rys. 4 Definicja płaszczyzn w przestrzeni 3D

Wybrane odpowiedzi impulsowe splatane są z przykładowymi monofonicznymi próbkami dźwiękowymi. W systemie wykorzystywane jest kilka metod realizacji splotu cyfrowego. Są to m.in. blokowe metody *overlap-save* i *overlap-add* [6].

W zaprojektowanym systemie użytkownik może, korzystając z interfejsu graficznego, wybrać pożądaną lokalizację źródła w wirtualnej przestrzeni odsłuchowej. Następnie system dobiera odpowiednią odpowiedź impulsową (z pośród dostępnych w zestawie) znajdującą się najbliżej w przestrzeni 3D, w stosunku do wymagań użytkownika. W kolejnym kroku następuje splot wejściowej próbki dźwiękowej, dostarczonej przez użytkownika, z wybraną funkcją HRIR. W konsekwencji, użytkownik otrzymuje przetworzony stereofoniczny plik dźwiękowy gotowy do odsłuchania za pomocą słuchawek.

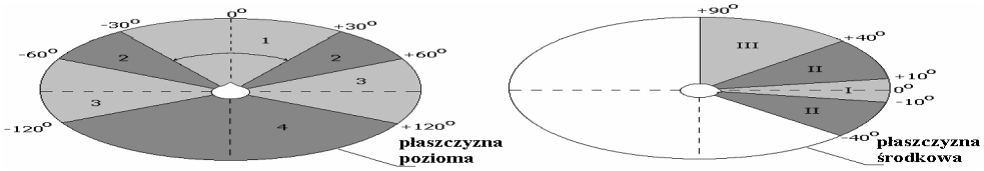
2.2 Przeprowadzone badania

W celu zbadania dokładności lokalizacji pozornych źródeł dźwięku opracowano odpowiedni test dźwiękowy. Łączy on w sobie cechy metody sekwencji wymuszonych i kluczowania swobodnego [4]. Test wykonywany jest automatycznie w postaci aplikacji komputerowej rejestrującej odpowiedzi badanych osób. W teście jako sygnały prezentowane są instrumenty muzyczne takie jak: gitara akustyczna, skrzypce, gitara basowa, flażolet, jak i syntetyczne próbki dźwiękowe - tony o częstotliwościach 60Hz i 1000Hz oraz szum różowy. Rodzaje błędów zdefiniowanych na potrzeby testu obrazuje rys.5.

Rys. 5 Rodzaje błędów

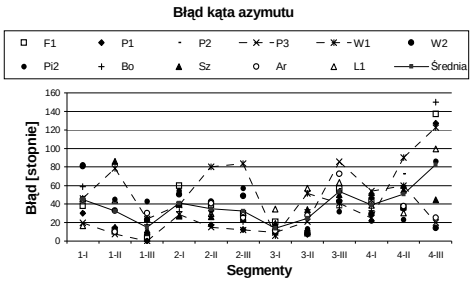
MOST WIEDZY Pobrano z mostwiedzy.pl

Duża liczba funkcji HRTF oraz potrzeba uproszczenia analizy wyników wymusiła konieczność ograniczenia liczby użytych w teście odpowiedzi HRIR. W tym celu płaszczyznę poziomą i środkową podzielono na kilka części, zgodnie z rys. 6.

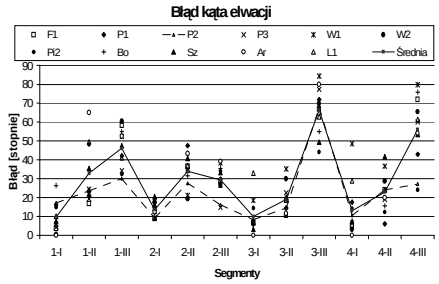


Rys. 6 Segmentacja płaszczyzny poziomej i środkowej

Prezentowane poniżej wykresy przedstawiają uzyskane średnie wartości błędu kąta azymutu (rys. 7) i kąta elewacji (rys. 8) w poszczególnych sektorach.



Rys. 7 Wykres średnich wartości błędu azymutu w poszczególnych sektorach dla badanych osób



Rys. 8 Wykres średnich wartości błędu elewacji w poszczególnych sektorach dla badanych osób

Otrzymane wyniki błędu kąta azymutu i elewacji pokazują wyraźne zależności. Precyzja określenia elewacji maleje wraz z jej odchyleniem od płaszczyzny poziomej. Świadczy o tym rosnąca średnia wartość błędu (czarna ciągła linia na rys. 8). Trzeba jednak zaznaczyć, iż wraz ze wzrostem (co do wartości bezwzględnej) kąta elewacji maleje średni błąd kąta azymutu (czarna ciągła linia na rys. 7). Zależność ta widoczna jest w sektorach znajdujących się przed słuchaczem. W sektorach bocznych widoczna jest odwrotna zależność. Precyzja określenia kąta azymutu maleje ze wzrostem bezwzględnej wartości kąta elewacji. Sektory kątowe znajdujące się z tyłu użytkownika charakteryzują się dużymi wartościami błędów kąta azymutu i elewacji.

Porównanie otrzymanych wyników z rezultatem badań innych autorów wykazuje liczne podobieństwa [1, 7, 9] . Wartości średniego kąta elewacji w sektorach 1-I, 2-I i 3-I wahają się od dziesięciu do kilkunastu stopni. Podobne wartości odnotowują Senova, McNally i Martin [7]. Również Begault i Wenzel w swych badaniach możliwości lokalizacyjnych ludzkiej mowy określają średnią wartość błędu elewacji równą 15° w warunkach zbliżonych do opracowanego testu [1]. Wyniki prezentowane przez Begaulta i Wenzela określają średnią wartość błędu azymutu równą 24°. Odnotowują oni jednak duży rozrzut wyników. W zależności od badanej osoby średnia wartość błędu waha się od 18° do 50°. Podobny rozrzut średnich wartości błędu widoczny jest w prezentowanych rezultatach.

MOST WIEDZY Pobrano z mostwiedzy.pl

3. WNIOSKI KOŃCOWE - PODSUMOWANIE

W referacie przedstawiono opis techniki wykorzystywanej do wirtualizacji dźwięku przestrzennego. Praktyczna realizacja opisanej techniki została wykonana w środowisku Matlab. Użyte metody splotu pozwalają jednak na łatwą implementację poszczególnych części składowych systemu w innych środowiskach, np. w C++.

Opisany system i eksperymenty zostały wykonane w Katedrze Systemów Multimedialnych (dawnej Katedrze Inżynierii Dźwięku i Obrazu). Bardziej szczegółowy opis przeprowadzonych testów i otrzymanych wyników znajduje się w innej pracy autora [5].

4. BIBLIOGRAFIA

[1] Begault, D. R., Wenzel, E. M., „Direct comparision of the impact of head tracking, reverberation, and individualized head-related transfer function on the spatial perception of a virtual speech source", J. Audio Eng. Soc., Vol. 49, No. 10, pp. 904-916, October 2001.
[2] Davis, M. F., Fellers, M. C., „Virtual presentation of dolby ac-3 and pro logic signals", 103rd AES Convention, Preprint No. 542, New York, September, 1997.
[3] Keith M., Gardner B., „Hrtf measurements of a kemar dummy-head microphone", Report 280, MIT Media Lab, May 1994.
[4] Łętowski, T., „Słuchowa ocena sygnałów i urządzeń", Akademia Muzyczna w Warszawie, Warszawa, 1984.
[5] Maziewski, P., „System przetwarzania sygnałów do celu odbioru dźwięku wielokanałowego za pomocą słuchawek stereofonicznych", X Sympozjum Inżynierii i Reżyserii Dźwięku, Wrocław 2003.
[6] Proakis, G. J., Manolakis, D. G., „Digital Signal Processing Principes, Algorithms, and Applications", Prentice Hall, New Jersey, 1996.
[7] Senova M. A., McNally K. I., Martin R. L., „Localization of virtual sound as a function of head-related impulse response duration", J. Audio Eng. Soc., Vol. 50, No. 1/2, pp. 57-65, January/February 2002.
[8] Smith, S. W., „The Scientist and Engineer's Guide to Digital Signal Processing", California Technical Publishing, 1997.
[9] Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman F. L. „Localization using nonindividualized head-related transfer functions", J. Acoust. Soc. Am., Vol. 94, No. 1, pp. 111-123, July, 1993.
[10] Wightman, F. L., Kistler, D. J., „Headphone simulation of free-field listening I: stimulus synthesis", J. Acoust. Soc. Am., Vol. 85, No. 2, pp. 858-867, February 1989.

SOUND VIRTUALISATION TECHNIQE BASED ON HRIR CONVOLUTION

System for presenting a virtual sound source in a surround panorama is described. It was developed in Multimedia Systems Department on ETI faculty of Gdansk University of Technology. The system is dedicated for usage with headphones. System uses convolution techniques. Input sound is being processed with the correct impulse response. The impulse responses were recorded with a dummy head microphones in an anechoic chamber. Results illustrating the degree of localization accuracy are included. In addition existing virtualisation technique over stereo loudspeakers is also mentioned.

MOST WIEDZY Pobrano z mostwiedzy.pl