

Dźwięk przestrzenny

Projekt Techniki multimedialne 2023Z

Filip Horst 311257

Spis treści

1	Wstęp.....	1
1.1	Informacje dodatkowe do sprawozdania	2
2	Kąt dźwięku.....	2
2.1	Panning (Panoramowanie)	2
2.1.1	Panoramowanie liniowe (Linear Panning)	2
2.1.2	Panoramowanie ze stałą mocą (Constant Power Panning)	3
2.1.3	Porównanie.....	4
2.1.4	Audio 8D	5
2.1.5	Porównanie technik panning-u.....	5
2.1.6	Podsumowanie	6
2.2	Opóźnienie w czasie między kanałami	7
2.3	HRIR	8
2.4	Porównanie metod	9
3	Odległość od źródła	11
3.1	Modyfikacja głośności	11
3.2	Modyfikacja głośności wybranych częstotliwości	12
3.3	Pogłos	12
4	Aplikacja	13
4.1	Instrukcja obsługi.....	13
4.2	Tryb zgadywania	14
5	Źródła.....	15

1 Wstęp

Położenie źródła dźwięku potrafi mieć kluczowy wpływ na jego odbiór przez słuchaczy. Takie zabiegi mogą mieć zarówno powody artystyczne, jak i praktyczne. Odpowiednie zmiany pozwalają na stworzenie unikalnego przedstawienia, które widzowie zapamiętają na dłużej. W przypadku chociażby koncertów na żywo osiągnięcie efektu przemieszczającego się dźwięku nie jest problemem, ponieważ muzycy mogą zwyczajnie przejść w inne miejsce. Przemieszczanie staje się problemem, gdy zadaniem jest przemieszczenie dźwięku bez ruchu odtwarzacza, którym mogą być słuchawki, czy głośniki. W tym przypadku ruch dźwięku ma nie tylko znaczenie artystyczne, ale również i praktyczne. Jednym z przykładów mogą być gry komputerowe i symulacje, gdzie odpowiednie wrażenia dźwiękowe potrafią znacznie zwiększyć immersję użytkownika. Kolejne rozdziały pracy opisują różne techniki uzyskiwania takich efektów oraz przedstawiają program, w którym każdy może przetestować ich działanie.

1.1 Informacje dodatkowe do sprawozdania

W pracy umieszczone są przykładowe odsłuchy. Opisywane pliki dźwiękowe powinny się znajdować w folderze z materiałami załączonymi do projektu.

Wersja 1 z dnia 02.11.2023

2 Kąt dźwięku

2.1 Panning (Panoramowanie)

Panning (inaczej panoramowanie) to technika modyfikacji sygnału dźwiękowego polegająca na modyfikowaniu amplitud obu kanałów według odpowiednich zależności w celu sprawienia wrażenia, że dźwięk dochodzi z dowolnego kąta przed słuchaczem.

Podstawowym problemem tej techniki jest to, że nie jest całkowicie realistyczna. Nie da się jednak ukryć, że jak na swoją banalną prostotę daje na prawdę dobre wyniki.

Istnieje kilka wersji tej operacji, które różnią się zależnościami amplitud obu kanałów od wybranego kąta.

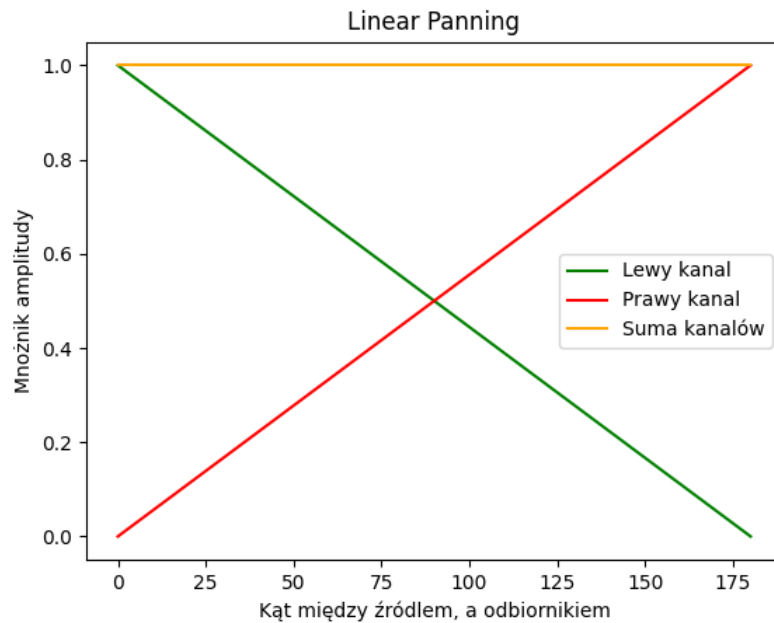
2.1.1 Panoramowanie liniowe (Linear Panning)

W tym podejściu zależność jest, jak nazwa wskazuje, liniowa. Wzory na współczynniki głośności obu kanałów są następujące:

$$amp_L = 1 - \frac{\alpha}{180^\circ}$$
$$amp_R = \frac{\alpha}{180^\circ}$$

Oznacza to, że dla kąta 0° odpowiadającego godzinie 9 na zegarze, amplituda lewego kanału jest mnożona przez współczynnik równy 1, natomiast prawy kanał przez współczynnik równy 0.

Zwiększając stopniowo kąt, czyli idąc zgodnie z ruchem wskazówek, oba współczynniki będą zmieniać się liniowo poprzez wartości 0.5 (lewy) i 0.5 (prawy) dla 90° , aż do wartości 0.0 (lewy) i 1.0 (prawy) dla kąta 180° , czyli godziny 3 na tarczy zegara. Lepiej zobrazowane jest to na poniższym wykresie.



Rysunek 1: Zależności w Linear Panning

Takie podejście charakteryzuje się tym, że suma amplitud obu sygnałów dla każdego punktu jest stała, natomiast suma mocy zmienia się.

Odśłuch 🎵 mowaLP.wav

Odśłuch 1: Krótkie nagranie z syntezy mowy ustawione w pozycji godziny 3 przy pomocy Linear Panning

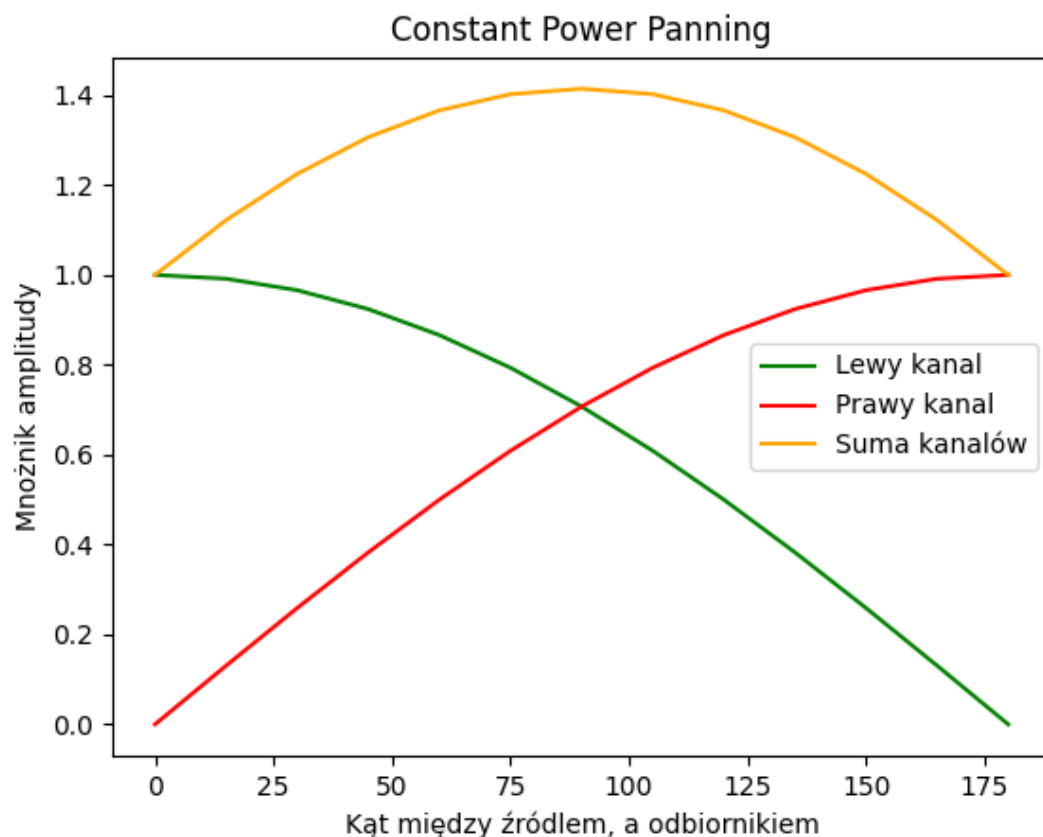
2.1.2 Panoramowanie ze stałą mocą (Constant Power Panning)

Podejście to ma na celu poprawę wyników poprzez modyfikację zależności w taki sposób, że suma mocy obu kanałów dla każdego kąta jest stała. Wzory zapewniające taką zależność prezentują się następująco:

$$amp_L = \cos\left(\frac{\alpha}{2}\right)$$

$$amp_R = \sin\left(\frac{\alpha}{2}\right)$$

Zmiana współczynników nie jest już liniowa, co najlepiej przedstawia wykres:



Rysunek 2: Zależności w Constant Power Panning

Jedyną wadą tego podejścia jest to, że suma amplitud obu kanałów przekracza sumę amplitud tych kanałów przed wykonaniem panning-u. Może to prowadzić do nieumyślnego przekroczenia maksymalnej głośności, dla którego sprzęt jest w stanie odtwarzać czysty dźwięk i zacząć powstawać nieprzyjemne trzaski lub po prostu dźwięk będzie nieprzyjemnie głośny.

Wykorzystanie funkcji trygonometrycznych pozwala osiągnąć zamierzony efekt, co skutkuje lepszymi odczuciami przy odsłuchu przekształconych w ten sposób dźwięków.

Constant Power Panning to aktualnie standard i jest wykorzystywany w zdecydowanej większości programów do obróbki audio.

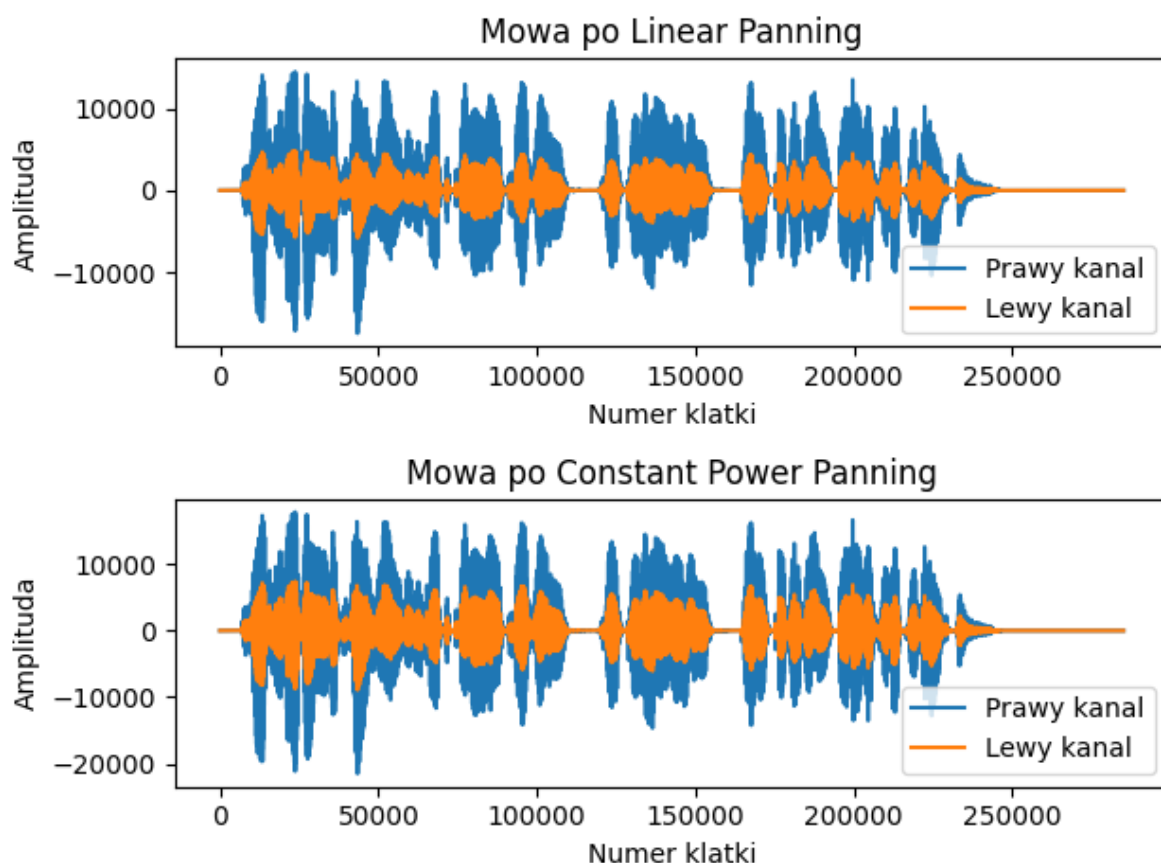
Odsłuch 🎵 mowaCPP.wav

Odsłuch 2: Krótkie nagranie z syntezatora mowy ustawione w pozycji godziny 3 przy pomocy Constant Power Panning

2.1.3 Porównanie

Prawdopodobnie oba odsłuchy brzmiały podobnie, więc która metoda jest lepsza? Okazuje się, że dla pojedynczych kątów, czyli stałego przekształcenia dla całego pliku wejściowego rozpoznanie dokładnego kąta jest równie trudne. Na wykresach amplitud widoczna jest mała różnica w amplitudach (dla CPP amplitudy są około 10-20% wyższe). Prawdziwa różnica będzie widoczna w

porównaniu dla tzw. „dźwięku 8d” w dalszej części sprawozdania.



Rysunek 3: Porównanie amplitud po LP i CPP. Plik wejściowy był mono.

2.1.4 Audio 8D

Określenie Audio 8D odnosi się do specjalnego efektu nakładanego zazwyczaj na muzykę który tworzy wrażenie, że źródło chodzi wokół głowy słuchacza. Mówiąc bardziej technicznie, kąt z którego przybywa dźwięk płynnie zmienia się o stały interwał w stałym kierunku. Oczywiście nie ma to nic wspólnego z 8 wymiarami, ale takie nazewnictwo jest bardzo powszechne dlatego należy o nim wspomnieć. Prawdopodobnym powodem powstania takiego określenia jest wykorzystanie 8 kierunków, między którymi źródło dźwięku wirtualnie przechodzi w trakcie odtwarzania. Muzyka 8D dostępna w popularnych serwisach streamingowych używa większej ilości efektów, ale podstawowe wrażenie obrotu można uzyskać przy użyciu samego panning-u.

Odsłuch 🎵 mowa8dCPP.wav


Odsłuch 3: Krótkie nagranie z syntezy mowy zmieniające swoje położenie w czasie.
Do panning-u użyto techniki CPP.

2.1.5 Porównanie technik panning-u

Audio 8D umożliwia lepsze porównanie różnic między Linear Panning, a Constant Power Panning.

Odsłuch 🎵 mowa8dCPP.wav


Odsłuch 3: Krótkie nagranie z syntezy mowy zmieniające swoje położenie w czasie.
Do panning-u użyto techniki CPP.

Odsłuch  mowa8dLP.wav

Odsłuch 4: Krótkie nagranie z syntezy mowy zmieniające swoje położenie w czasie. Do panning-u użyto techniki LP.


Różnice są subtelne, ale zauważalne. Najważniejsza jest różnica w głośności obu wersji. Plik wejściowy w obu przypadkach był ten sam, ale przy użyciu CPP plik wyjściowy był ogólnie głośniejszy. Jest to spowodowane zależnością opisaną w 2.1.2. Na szczęście w tym przypadku nie spowodowało to żadnych niedoskonałości, chociaż zależy to również od sprzętu, więc u niektórych słuchaczy może być inaczej.

Inną różnicą jest płynność przejścia. Według wielu osób CPP daje bardziej płynne zmiany w czasie, co lepiej odwzorowuje sytuację z głośnikiem chodzącym wokół głowy po okręgu, którą dźwięk 8d stara się zasymulować.

Odsłuch  muzykaMJ8dCPP.wav

Odsłuch 4: Nagranie piosenki Michael Jackson – Beat It zmieniające swoje położenie w czasie. Do panning-u użyto techniki CPP.

W przypadku muzyki lepiej zauważalna jest kolejna wada podstawowego panning-u. Jest nią pewien dyskomfort, kiedy źródło znajduje się w pełni z jednej strony. Nieprzyjemne wrażenie jest spowodowane tym, że w takiej sytuacji amplitudy w jednej słuchawce są wyzerowane, co w rzeczywistości nie zdarza się praktycznie nigdy, ponieważ do drugiego ucha zawsze docierają jakieś odbicia. Aby pozbyć się tego niekomfortowego odczucia można zastosować próg minimalnej głośności. Polega na tym, że gdy mnożnik amplitudy obliczony podstawowym panningiem zejdzie poniżej określonej wartości to jest sztucznie zwiększany.

Odsłuch  muzykaMJ8dCPPprogiMin.wav

Odsłuch 5: Nagranie piosenki Michael Jackson – Beat It zmieniające swoje położenie w czasie. Do panning-u użyto techniki CPP z minimalnym mnożnikiem amplitud (0.15).

Po takiej modyfikacji dyskomfort nie jest już tak odczuwalny, ale w zależności od sprzętu i osobistych preferencji wciąż może występować. Powodem jest wciąż bardzo duży kontrast między kanałami. Aby pozbyć się tego wrażenia można zmniejszyć głośność odtwarzania odsłuchu 5 lub zastosować dodatkowe progi dla wartości maksymalnej. Dokładne wartości progów zależą od osobistych preferencji.

Odsłuch  muzykaMJ8dCPPprogi.wav

Odsłuch 6: Nagranie piosenki Michael Jackson – Beat It zmieniające swoje położenie w czasie. Do panning-u użyto techniki CPP z minimalnym (0.15) oraz maksymalnym (0.92) mnożnikiem amplitud.

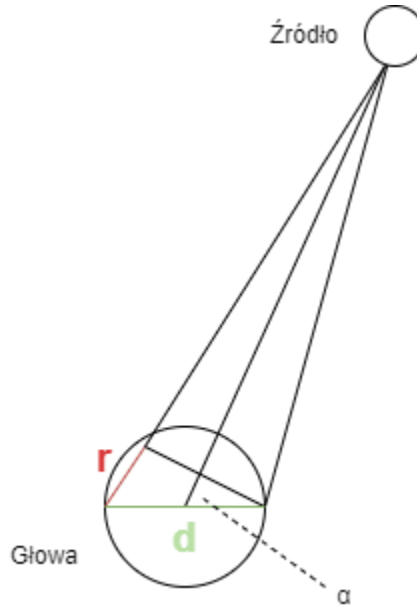
2.1.6 Podsumowanie

Po wykonaniu badań z wykorzystaniem techniki audio 8d udało się ustalić następujące zależności:

- CPP(Constant Power Panning) oferuje lepszą płynność przy wirtualnym zmienianiu lokalizacji źródła
- CPP sprawia problemy z głośnością ze względu na charakter wykorzystywanych zależności
- Części wad CPP można pozbyć się stosując progi mnożników amplitud oraz manipulację ogólną głośnością

2.2 Opóźnienie w czasie między kanałami

Dźwięk ma określoną prędkość, która w pogodny dzień wynosi około 345 [m/s]. Taka prędkość jest istotnie niesamowita, kiedy dotyczy szybkości samochodu, czy motocykla, ale okazuje się że jest na tyle niska, że kiedy dźwięk dochodzi do słuchacza z kierunku innego niż idealnie na wprost to powstaje odczuwalne opóźnienie między jego uszami. Jest to spowodowane różnicą w przebytej odległości, która wynosi maksymalnie tyle, ile średnica głowy na wysokości uszu.



Opóźnienie zależy, oprócz samej pozycji źródła i słuchacza, również od szerokości głowy. Dla człowieka średnia wartość wynosi około 15 cm. Opóźnienie osiąga swoją maksymalną wartość, gdy kąt alfa między słuchaczem, a źródłem wynosi 180°, natomiast minimalną wartość dla kąta 90°. Taką zależność można dobrze przybliżyć poprzez dodanie do kąta z przedziału 0 - 180° kąta 90° i użyciu funkcji sinus.

Ostatecznie funkcję na wartość przesunięcia w klatkach można zapisać w następujący sposób:

$$ndelay = \frac{\sin(\alpha + 90^\circ) * d}{v_s} * f_s$$

, gdzie:

$ndelay$ – liczba klatek opóźnienia

α – kąt między źródłem, a słuchaczem

d – szerokość głowy

v_s – prędkość dźwięku

f_s – częstotliwość (framerate) sygnału

Takie opóźnienie wynosi mniej niż tysięczną część sekundy, ale mimo wszystko mózg człowieka podświadomie wykrywa i wykorzystuje tę informację do zlokalizowania źródła tego dźwięku. Jeśli ktoś posiada dobrej jakości słuchawki lub odpowiedni układ głośników to do wykonania możliwy jest eksperyment polegający na losowym opóźnieniu jednego z kanałów dźwięku. Oczywiście każdy człowiek ma trochę inny zmysł słuchu, ale wiele osób będzie potrafiło bez problemu wskazać, czy dźwięk dochodzi z prawego, czy lewego kierunku. Określenie dokładnego kąta w takim scenariuszu jest bardzo trudne, a prawdopodobnie wręcz niemożliwe, ale już samo wskazanie strony jest wystarczającym dowodem na istnienie takiego mechanizmu.

Prostszą formę takiego eksperymentu można wykonać odsłuchując dwa poniższe pliki dźwiękowe. Odpowiedź, który dobiega z lewej, a który z prawej znajduje się w odwróconym tekście poniżej. W jednym z odsłuchów źródło jest ustawione na godzinie 3, a drugie na godzinie 9.

Odsłuch  mowaDelay1.wav

Odsłuch 7: Nagranie z syntezatora mowy z opóźnionym sygnałem w jednym z kanałów.

Odsłuch  mowaDelay2.wav

Odsłuch 8: Nagranie z syntezatora mowy z opóźnionym sygnałem w jednym z kanałów (innym niż w odsłuchu 7).

Wskazówka: Ruch oczami w płaszczyźnie prawo-lewo może ułatwić wycucie kierunku do źródła.

Rozwiązanie:

W pierwszym pliku dźwięku źródło jest ustawione na godzinie 3, a w drugim na godzinie 9.

2.3 HRIR

HRIR (Head-related transfer function, pl. Funkcja przenoszenia zależna od głowy) – to funkcja charakteryzująca jak ucho odbiera dźwięk. Dokładny kształt jest unikalny dla każdego człowieka, stąd częste porównania słuchu do odciska palca. Do wykonania obliczeń potrzebne jest wykonanie badania polegającego na umieszczeniu najwyższej jakości specjalnych mikrofonów w uszach badanego lub użyciu mikrofonów umieszczonych w obudowie symulującym małżowiny uszne. Następnie funkcja obliczana jest na podstawie porównania dźwięku odtwarzanego z głośników z dźwiękiem zarejestrowanym w mikrofonach. Ze względu na niedoskonałości sprzętowe, obliczenie idealnie dokładnej funkcji HRIR jest niemożliwe, ale przy użyciu nowoczesnych urządzeń wyznaczane wartości są wystarczające, aby umożliwić każdemu słuchaczowi realistyczne doznania.

Funkcja HRIR jest zazwyczaj przechowywana w postaci odpowiedzi impulsowej. Odpowiedź impulsowa to funkcja, która informuje jak badany układ odpowiada na pobudzenie impulsem. W podanym przykładzie układem jest nie tylko ucho ludzkie, ale również pomieszczenie, w którym odbywają się badania, natomiast impulsem jest dźwięk. Odpowiedzi impulsowe służą również do przekształcania dźwięku w celu symulacji różnego typu pomieszczeń lub sytuacji np. pusta hala, wyposażony pokój, dźwięk dochodzący zza ścian.



Rysunek 4: Odpowiedź impulsowa dla przykładowego pliku HRIR.

Użycie odpowiedzi impulsowej jest banalnie proste, ponieważ wystarczy ujednolicić częstotliwości próbkowania między odpowiedzią, a plikiem dźwiękowym (resampling) i wykonać splot.

Odsłuch 🎵 mowaHRIR.wav

Odsłuch 9: Nagranie z syntezy mowy z efektem przestrzennym osiągniętym przez splot z odpowiedzią impulsową HRIR. Źródło ustawione na godzinie 3.

2.4 Porównanie metod

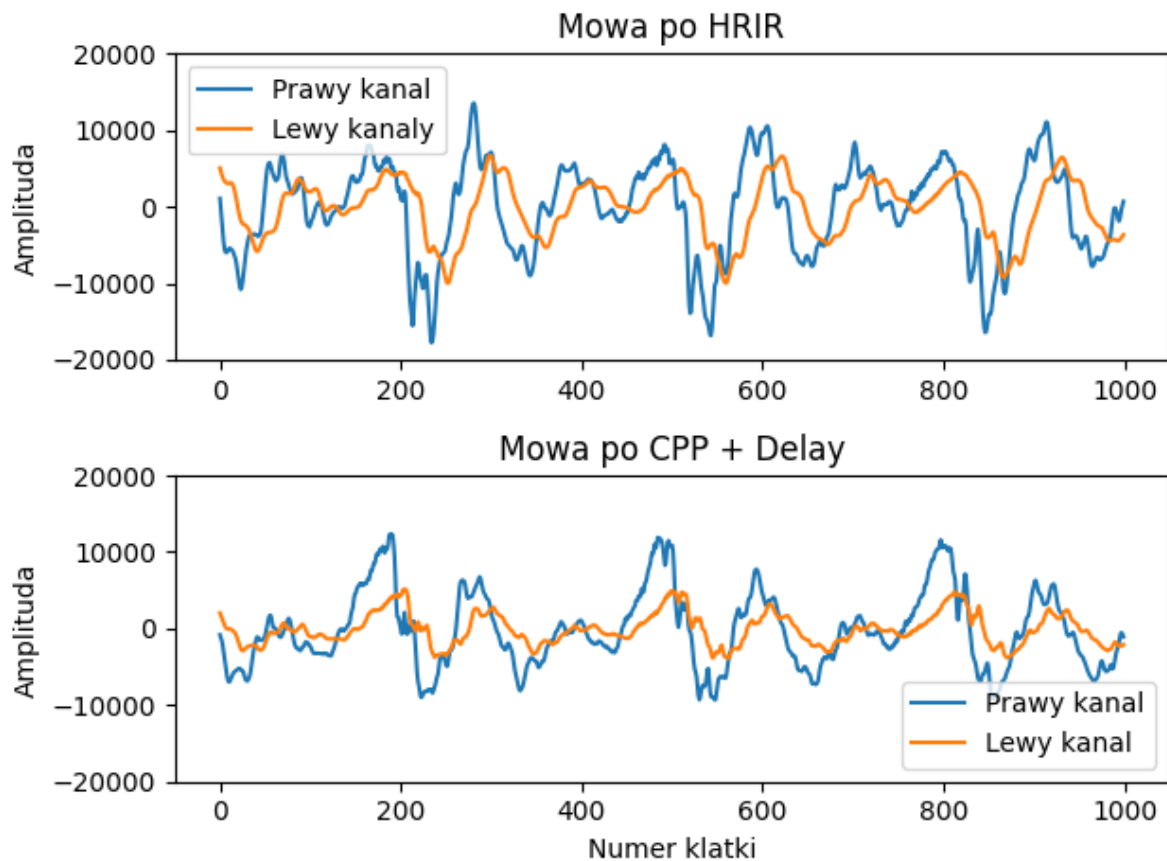
Obiektem porównania będą efekty przesunięcia sygnału mowy z generatora na pozycję odpowiadającą godzinie 3 (45 stopni w prawo od centrum). Pierwszy plik będzie efektem działania splotu z HRIR (wersja bardziej realistyczna), natomiast drugi efektem wykonania Constant Power Panning oraz opóźnienia (wersja będąca prostym w implementacji przybliżeniem).

Odsłuch 🎵 mowaHRIR.wav oraz 🎵 mowaCPPDelay.wav

Odsłuch 10: Nagranie z syntezy mowy z efektem przestrzennym osiągniętym przez 1) splot z odpowiedzią impulsową HRIR 2) Constant Power Panning i opóźnienie między kanałami. Źródło ustawione na godzinie 3.

W tym przypadku dźwięk po HRIR brzmi bardziej realistycznie. Na podstawie niego łatwiej jest określić kierunek, z którego dochodzi dźwięk. Aby sprawdzić, dlaczego tak się dzieje można skorzystać

z wykresu amplitudowego.



Rysunek 5: Porównanie wykresu amplitud sygnałów po modyfikacji HRIR oraz CPP+Delay

Pierwsze co rzuca się w oczy to duża różnica w amplitudach. Jest to spowodowane tym, że w przypadku HRIR odpowiedź impulsowa trochę się różni w zależności od kanału, podczas gdy przy CPP jest ona identyczna w obu przypadkach. Skutkuje to zupełnie innymi odczuciami i lepszym wrażeniem głębi w przypadku HRIR. Opóźnienie między sygnałami jest bardzo podobne. Jak chodzi o kierunkowość dźwięku to największe znaczenie ma stosunek obu kanałów do siebie oraz do oryginału. Prosta analiza daje następujące wyniki:

HRIR

Max Lewa 20774 Stosunek do oryginału 1.0765403948800332

Max Prawa 10708 Stosunek do oryginału 0.5607750720083792

Stosunek lewa/prawa 1.940044826298095

CPP + Delay

Max Lewa 17828 Stosunek do oryginału 0.9238741773332643

Max Prawa 7384 Stosunek do oryginału 0.3866980885048442

Stosunek lewa/prawa 2.414409534127844

Stosunek należy odczytywać w następujący sposób: Jeśli wartość jest bliska 1 to dźwięk sprawia wrażenie, że dobiega z kierunku centralnego, natomiast jeśli wartość jest znacznie wyższą niż 1 to dźwięk dobiega od boku.

W tym przypadku te wartości są liczbowymi dowodami na to, że w przypadku HRIR dźwięk nie jest skierowany tak silnie w prawo.

Eksperyment można powtórzyć również dla innych kątów. Do odsłuchu przygotowany jest przypadek dla 30 stopni, zamiast 45.

Odsłuch 🎵 mowaHRIR30.wav oraz 🎵 mowaCPPDelay30.wav

Odsłuch 11: Nagranie z syntezy mowy z efektem przestrzennym osiągniętym przez 1) splot z odpowiedzią impulsową HRIR 2) Constant Power Panning i opóźnienie między kanałami. Źródło ustawione na kierunku 30 stopni w prawo od centrum.

Podsumowując, metoda CPP w połączeniu z opóźnieniem daje całkiem dobre efekty, ale wciąż odczuwalnie gorsze od realistycznego przypadku. Ma ona jednak wiele zalet. Po pierwsze prostota. Sama implementacja nie jest trudna w żadnym z obu przypadków, ale biorąc pod uwagę metodę pozyskiwania odpowiedzi impulsowych HRIR w tej kategorii zdecydowanie zwycięża prosty panning. Po drugie, ciągłość. Jest to częściowo wina wybranej biblioteki, ale odpowiedzi HRIR są nagrane z interwałem 5 stopni, podczas gdy CPP daje możliwość szybkiego efektu dla każdego kąta, który nawet nie musi przyjmować wartości całkowitej. Ostatnią przewagą jest szybkość. Obliczenie prostych funkcji trygonometrycznych potrzebnych do CPP okazuje się być szybsze od pojedynczego splotu w przypadku HRIR. Cały ten zestaw zalet sprawia, że panning wciąż jest bardzo atrakcyjną ofertą w niektórych przypadkach użycia. Przykładowo w grach komputerowych 3D, które umożliwiają obracanie kamery szybko obliczany panning jest preferowaną opcją, ponieważ gracz obracając wirtualną głowę wciąż jest w stanie idealnie zlokalizować źródło dźwięku. Muzyka, gdzie w większości przypadków sam fakt zmiany kierunku ma większy wpływ niż jego dokładna wartość jest innym dobrym przykładem. Jednak tam gdzie potrzebne są realistyczne efekty nie zależnie od trudności ich osiągnięcia, HRIR zdecydowanie wygrywa.

3 Odległość od źródła

3.1 Modyfikacja głośności

Modyfikacja głośności dźwięku to najprostsza metoda służąca do wirtualnego umieszczenia źródła w danej odległości od słuchacza. Moc dźwięku zmienia się odwrotnie proporcjonalnie do kwadratu odległości źródła od odbiornika. Znacząco to, że jeśli dystans od głośnika zwiększy się dwukrotnie, to moc dźwięku spadnie czterokrotnie. Na podstawie tej zależności utworzona została zasada -6dB, która mówi że po każdym podwojeniu dystansu głośność dźwięku spada o 6 decybeli. Oczywiście w rzeczywistości sytuacja jest o wiele bardziej skomplikowana. Dokładna wartość zależy chociażby od pomieszczenia, czy nawet warunków atmosferycznych, ale zasada jest dobrym przybliżeniem.

Odsłuch 🎵 muzykaMJ.wav

Odsłuch 12: Wycinek piosenki Michael Jackson – Beat It.

Odsłuch 🎵 muzykaMJGain.wav

Odsłuch 13: Wycinek piosenki Michael Jackson – Beat It. Głośność zmieniona na podstawie zasady -6dB. Źródło oddalone o 10 metrów.

Oba odsłuchy warto zostawić otwarte, aby porównać odczucia z kolejną metodą.

Oczywiście sama modyfikacja głośności nie jest wystarczająca, ponieważ głośnik może odtwarzać muzykę z różną mocą nie zmieniając położenia. Z tego powodu, taka modyfikacja jest skuteczna tylko, jeśli dotyczy dobrze znanych mózgowi dźwięków. Przykładem może być dźwięk budzika. Każdy człowiek słyszy go na tyle często, że jest w stanie bardzo dokładnie odgadnąć jak daleko od niego leży telefon.

3.2 Modyfikacja głośności wybranych częstotliwości

Współczynnik utraty mocy w powietrzu zmienia się w zależności od częstotliwości. Wysokie częstotliwości szybciej wygasają, co sprawia że słychać je dużo słabiej z dużych odległości. W przypadku niskich częstotliwości jest dokładnie odwrotnie. Dobrym przykładem takiego zjawiska pochodzącym z prawdziwego życia jest zbliżanie się do sceny, na której odbywa się przemowa. Będąc daleko, słychać praktycznie tylko dźwięki o niskiej częstotliwości, a nie piski, czyli fale o wysokiej częstotliwości.

Zjawisko w środowisku wirtualnym najłatwiej jest osiągnąć przy użyciu filtru typu HighShelf. Zasada działania jest bardzo podobna do filtrów typu dolnoprzepustowych LowPass, ale zmiany amplitudy częstotliwości branych pod uwagę nie są tak duże i dynamiczne. Można więc odpowiednimi ustawieniami zmniejszyć głośność dla wszystkich częstotliwości powyżej wybranego progu o dowolną wartość.

Najlepszym dowodem na to, że taki mechanizm poprawia odczuwanie odległości jest porównanie dwóch dźwięków.

Odsłuch  muzykaMJGainFiltr.wav

Odsłuch 14: Wycinek piosenki Michael Jackson – Beat It. Głośność zmieniona na podstawie zasady -6dB oraz zmianę balansu amplitud niskich i wysokich częstotliwości. Źródło oddalone o 10 metrów.

Efekt uzyskany w ten sposób wydaje się być lekko bardziej realistyczny. Sprawia jednak lekkie wrażenie, że źródło po prostu odtwarza muzykę zmodyfikowaną w taki sposób, aby uwydatnić bass. Inną wadą jest to, że w takim podejściu ciężko jest ocenić o ile dokładnie powinny różnić się amplitudy danych częstotliwości. W implementacji zostało przyjęte przybliżenie polegające na dodatkowym przyciszeniu wyższych częstotliwości o połowę wartości obliczonej przez zasadę -6dB. Przyjęte przybliżenie może nie być całkowicie realistyczne, ale jest wystarczająco dobre do pokazania, że metoda poprawia odczuwanie odległości.

Największym brakiem w takim dźwięku jest brak pogłosu, czy echa. Istnieje jednak pewna wada ich użycia. Mianowicie wprowadzają one charakter pomieszczenia w którym wirtualnie znajduje się słuchacz. Z tego powodu takie efekty powinny być dodawane oddzielnie w kolejnym kroku tak, aby nie było problemu z ich zmianą.

3.3 Pogłos

Ponieważ pogłos jest bardzo ważny w przekazywaniu dystansu to warto wspomnieć, które parametry mają największe znaczenie. Problemem jest to, że algorytmy realizujące pogłos często posiadają małe różnice w swoich parametrach, dlatego te opisane w tym punkcie nie muszą być zgodne z innymi implementacjami. Wykorzystany jest pakiet Spotify Pedalboard.

W kontekście przekazywania odległości najważniejszy jest stosunek Dry/Wet. Czasami nazywa się to Mix, ale w przypadku Pedalboard jest to rozbite na dwa parametry: dry_level oraz wet_level. Dźwięk Wet odpowiada efektowi, czyli odbiciom, natomiast Dry to dźwięk źródła. Mówiąc inaczej ustawienie wet_level = 0 sprawia, że efekt pogłosu w ogóle nie jest aplikowany. Stopniowe zwiększanie wartości

będzie sprawiało, że echa i odbicia będą bardziej zauważalne co sprawi wrażenie, że źródło znajduje się w większej odległości. Nie ma jednej ustalonej zasady co do tego jaki stosunek Wet/Dry odpowiada danej odległości, ponieważ zależy od symulowanego pomieszczenia oraz nie jest tak łatwe do wyznaczenia. W przypadku, gdy problem nie wymaga przeliczania w czasie rzeczywistym to parametry ustawia się metodą prób i błędów oraz z wykorzystaniem nagrań referencyjnych.

Odsłuch 🎵 muzykaMJGainFiltrReverb.wav

Odsłuch 14: Wycinek piosenki Michael Jackson – Beat It. Głośność zmieniona na podstawie zasady -6dB oraz zmianę balansu amplitud niskich i wysokich częstotliwości. Źródło oddalone o 10 metrów. Posiada również pogłos z parametrami `wet_level = 0.5`, `dry_level = 1.0`. Warto zwrócić uwagę na powstały efekt pustej hali.

4 Aplikacja

4.1 Instrukcja obsługi

Parametrami aplikacji steruje się poprzez argumenty wejściowe przy wywołaniu w konsoli (styl CLI). Program posiada automatycznie wygenerowaną instrukcję (help) dostępną poprzez wywołanie programu z opcją -h. Treść instrukcji:

```
usage: main.py [-h] [-auto] [-a [-90 - +90, krok 5]] [-p {cpp,lp}] [-d] [-hrir] [-c8d] [-rd [ROTDUR]] [-ds [DISTANCE]] [-gs]
```

```
    [input_path] [output_path]
```

positional arguments:

input_path	Ściezka do pliku wejściowego
output_path	Ściezka, gdzie ma zostać zapisany wynik

options:

-h, --help	show this help message and exit
-auto, --auto	Automatycznie ustaw parametry dźwięku przestrzennego
-a [-90 - +90, krok 5], --angle [-90 - +90, krok 5]	

Kąt kierunku od słuchacza do źródła. Liczony jako zakres -90 do 90 od godziny 9 do 3 zgodnie z ruchem wskazówek.

-p {cpp,lp}, --pan {cpp,lp}	
-----------------------------	--

Wykonaj panning cpp lub lp

-d, --delay	Dodaj opóźnienie kanału
-hrir, --hrir	Wykonaj panning przez splot HRIR
-c8d, --create8d	Wykonaj panning 8D
-rd [ROTDUR], --rotdur [ROTDUR]	

Czas trwania obrotu w dźwięku 8d

-ds [DISTANCE], --distance [DISTANCE]

Odległość źródła od słuchacza [m]

-gs, --guess Tryb zgadywania kierunku do źródła

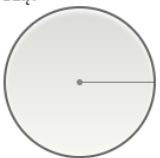
Budowanie komend może być nieintuicyjne, dlatego do programu dołączony jest specjalny dokument html o nazwie kreator.html, który pozwala na łatwe wygenerowanie polecenia. Lepiej prezentuje również pewne ograniczenia i opcje wykluczające się

Kreator poleceń do programu

Plik wejściowy

Plik wyjściowy (wynikowy)

Kąt



Kąt kierunku od słuchacza do źródła

Tryb pracy

- ☒ Modyfikacja pliku
- ☐ Predykcja ustawień pliku

Tryb parametrów

- ☒ Ręczny wybór
- ☐ Automatyczne

Tryb panningu

Opóźnienie między kanałami

- ☒ Tak
- ☐ Nie

Odległość od źródła

Odległość między źródłem, a słuchaczem [m]

Czas trwania obrotu [s]

Czas trwania pełnego obrotu (360) w trybie 8D

Generuj polecenie

Rysunek 6: Zrzut ekranu z generatora poleceń do aplikacji. Graficzna gałka do sterowania kątem nie działa bez połączenia z internetem, ale wartość wciąż może zostać wpisana.

Wynik wystarczy skopiować do konsoli dbając o to by sesja konsoli była w odpowiednim katalogu (zawierającym plik wykonalny aplikacji). Opcje same w sobie są bardzo proste i nie wymagają dodatkowego tłumaczenia. W razie wątpliwości należy zajrzeć do rozdziału teoretycznego sprawozdania.

4.2 Tryb zgadywania

Tryb zgadywania zwraca odpowiedź w postaci czterech przypuszczeń obliczanych różnymi metodami.

Przykładowy wynik dla pliku zmodyfikowanego przy pomocy CPP i przesunięcia dla kąta 30 stopni:

Kąty podane są idąc od godziny 12 na tarczy zegara zgodnie z ruchem wskazówek

Przypuszczenie na podstawie różnic w amplitudach przy założeniu CPP: 30.000516087278967

Przypuszczenie na podstawie różnic w amplitudach przy założeniu LP: 24.115861754998278

Przypuszczenie na podstawie przesunięcia sygnałów: [28.63098984]

Przypuszczenie na podstawie analizy przesunięcia sygnałów alternatywna metoda: [28.63098984]

W tym przypadku program poradził sobie bardzo dobrze. Niestety, w niektórych przypadkach dla sygnałów po splocie z HRIR, program nie radzi sobie z odgadywaniem dokładnych kątów i jest w stanie podać tylko przybliżony kierunek.

5 Źródła

Biblioteka HRIR:

<http://recherche.ircam.fr/equipes/salles/listen/context.html>

Informacje o Panningu:

<https://dsp.stackexchange.com/questions/21691/algorithm-to-pan-audio>

<https://medium.com/klinke-audio/a-detailed-overview-of-panning-functions-dc58f6d94b94>

HRIR:

https://en.wikipedia.org/wiki/Head-related_transfer_function