# Abc

# Kąt dźwięku

## Panning (Panoramowanie)

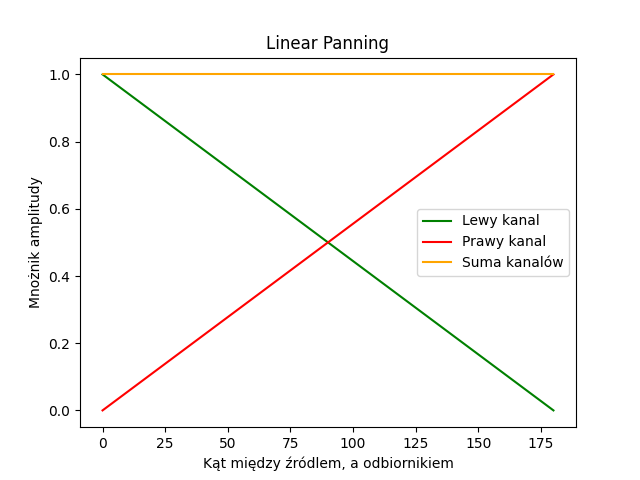
Panning (inaczej panoramowanie) to technika modyfikacji sygnału dźwiękowego polegająca na modyfikowaniu amplitud obu kanałów według odpowiednich zależności w celu sprawienia wrażenia, że dźwięk dochodzi z dowolnego kąta przed słuchaczem.   
Podstawowym problemem tej techniki jest to, że nie jest całkowicie 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 zależnością amplitud obu kanałów od wybranego kąta.

### Panoramowanie liniowe (Linear Panning)

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

Oznacza to, że dla kąta 0° odpowiadającego godzinę 9 na zegarze, amplituda lewego kanalu jest mnożona przez współczynnik równy 1, natomiast prawy kanal 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: Zależnosci w Linear Panning

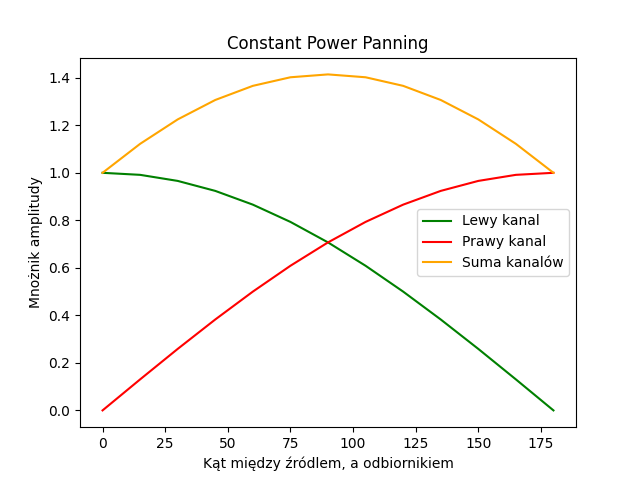
Takie podejście charakteryzuje się tym, że suma amplitud obu sygnałów dla każdego punktu jest stała, natomiast suma mocy zmienia się.

Odsłuch 🎵 mowaLP.wav  
Odsłuch 1: Krótkie nagranie z syntezatora mowy ustawione w pozycji godziny 3 przy pomocy Linear Panning

### Panoramowanie ze stalą mocą (Constant Power Panning)

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

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



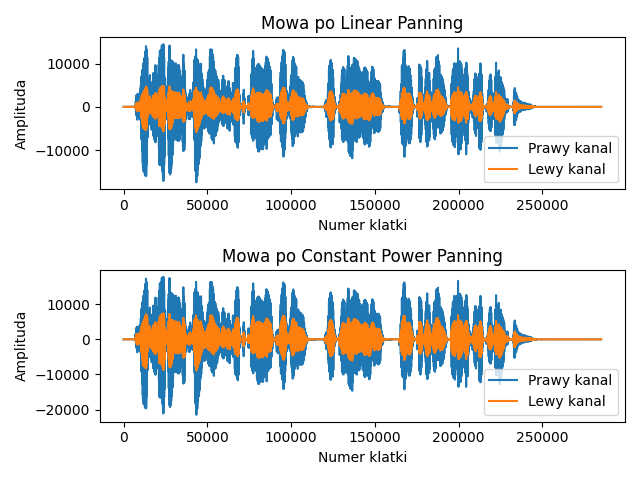
Rysunek 2: Zależnosci w Constant Power Panning

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 zaczną 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

### 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 kata jest rownie 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.

## Audio 8D

Określenie Audio 8D odnosi się do specjalnego efektu nakładanego zazwyczaj na muzykę który tworzy wrażenie, ze źródło chodzi wokół głowy słuchacza. Mówiąc bardziej technicznie, kąt z którego przybywa dźwięk płynnie zmienia się o stały interwal 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, miedzy 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 syntezatora mowy zmieniające swoje położenie w czasie. Do panning-u użyto techniki CPP.

### Porównanie technik panning-u

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

Odsłuch 🎵 mowa8dCPP.wav  
Odsłuch 3: Krótkie nagranie z syntezatora mowy zmieniające swoje położenie w czasie. Do panning-u użyto techniki CPP.

Odsłuch 🎵 mowa8LP.wav  
Odsłuch 4: Krótkie nagranie z syntezatora 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 opisana w 2.1.2 . Na szczęście w tym przypadku nie spowodowało to żadnych niedoskonałości, chociaż zależy to również od sprzętu, wiec u niektórych słuchaczy może być inaczej.  
Inną różnicą jest płynność przejścia. Według wielu osób CPP daje bardziej płynne zmiany w czasie, co lepiej odwzorowuje sytuacje z głośnikiem chodzącym wokół głowy po okręgu, która 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, ze 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 miedzy kanałami. Aby pozbyć się tego wrażenia można zmniejszyć głośność odtwarzania odsłuchu 5 lub zastosować dodatkowe progi dla wartości maksymalnej. Dokładne wartości progów zalezą 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.

### Podsumowanie

Po wykonaniu badań z wykorzystaniem techniki audio 8d udało się ustalić nastepujace zaleznosci:

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

## 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 ze jest na tyle niska, że kiedy dźwięk dochodzi do słuchacza z kierunku innego niż idealnie na wprost to powstaje opóźnienie między jego uszami. Jest to spowodowane różnicą w przebytej odległości, która wynosi maksymalnie tyle, ile średnica głowy na wysokości uszu.

Rysunek 😊

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 opóźnianiu 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 kata 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).

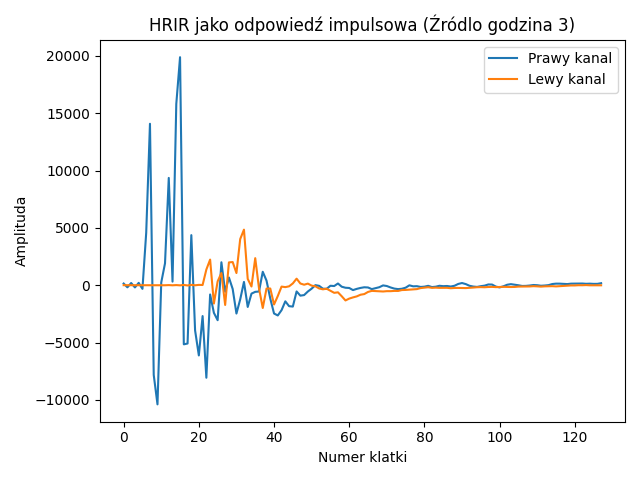
Wskazówka: Ruch oczami w płaszczyźnie prawo-lewo może ułatwić wyczucie kierunku do źródła.  
Rozwiązanie:

ɾǝʍɐɹd z nʞᴉld ɯᴉƃnɹp ʍ ɐ 'ʎuoɹʇs ɾǝʍǝl z ćᴉzpoɥɔop uǝᴉuᴉʍod ʞǝᴉʍzp nʞᴉld ɯʎzsʍɹǝᴉd M

## 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, stad 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 miedzy odpowiedzią, a plikiem dźwiękowym (resampling) i wykonać splot.

Odsłuch 🎵 mowaHRIR.wav  
Odsłuch 9: Nagranie z syntezatora mowy z efektem przestrzennym osiągniętym przez splot z odpowiedzią impulsową HRIR. Źródlo ustawione na godzinie 3.

Odsluch: Przykladowy HRIR 😊

## Porównanie metod

# Odległość od źródła

## 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. Znaczy to, ze jeśli dystans od głośnika zwiększy się dwukrotnie, to moc dźwięku spadnie czterokrotnie. Na podstawie tej zależności utworzona została zasada -6dB, która mówi ze 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.

Odsluch: Cicha muzyka 😊

Oczywiście sama modyfikacja głośności nie jest wystarczająca, ponieważ głośnik może odtwarzać muzykę z rożną 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, ze jest w stanie bardzo dokładnie odgadnąć jak daleko od niego leży telefon.

## Modyfikacja głośności wybranych częstotliwości

Współczynnik utraty mocy w powietrzu zmienia się w zależności od częstotliwości. Wysokie częstotliwości szybciej wygasają, co sprawia ze słychać je dużo słabiej z dużych odległości. W przypadku niskich częstotliwości jest dokładnie odwrotnie. Dobrym przykładem takiego zjawiska pochodzącym z prawdziwego zżycia 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 LowPass, ale zmiany amplitudy częstotliwości branych pod uwagę nie są tak duże i dynamiczne. Można wiec odpowiednimi ustawieniami zmniejszyć głośność dla wszystkich częstotliwości powyżej wybranego progu o dowolna wartość.

Najlepszym dowodem na to, że taki mechanizm poprawia odczuwanie odleglosci jest porównanie dwóch dzwiekow. Oba z nich maja taka amplitudę, ale tylko na jednym z nich uzyto filtru. To wlasnie ten drugi powinien wydawac się być dalej od słuchacza.

Odsluch: muzyka na niższej glosnosci 😊

Odluch: muzyka na niższej glosnosci + filtr 😊