AKADEMIA GÓRNICZO – HUTNICZA

IM. STANISŁAWA STASZICA W KRAKOWIE

WYDZIAŁ INŻYNIERII MECHANICZNEJ I ROBOTYKI

PRACA DYPLOMOWA

magisterska

Techniki rejestracji i reprodukcji w kontekście map dźwiękowych

Surround sound recording and reproduction techniques in the context of sound maps

Autor: **Aleksandra Żebrowska**

Kierunek studiów: Inżynieria akustyczna

Opiekun pracy: **dr inż. Jacek Wierzbicki**

Kraków, rok 2019r.

STRESZCZENIE

Tematyka map dźwiękowych nie miała dotychczas opracowania wykorzystującego aktualnie dostępne, najświeższe technologie, do ich aspektu kulturalno-społecznego. Przy tym, już sama nomenklatura może być myląca. Szeroko omawiane są zwykle mapy akustyczne, zajmujące się pomiarem hałasu. Mapy dźwiękowe natomiast to opis przestrzeni dźwiękowej konkretnego miejsca na mapie. Dokładniejsze rozwinięcie tego pojęcia, wraz z przykładami takich map, stanowi rozdział pierwszy. W dalszej części pracy przedstawione są aspekty techniczne, począwszy od opisu mechanizmu słuchu, w szczególności pod kątem słyszenia przestrzennego, w rozdziale drugim. Kolejne dwa rozdziały opisują techniki rejestracji i reprodukcji dźwięku, z naciskiem na dźwięk przestrzenny - od zarysu historycznego, do najnowszych dostępnych rozwiązań. Rozdział piąty poświęcony jest formatom zapisu dźwięku. Opisane są najpopularniejsze z nich, rozwinięty jest wątek kompresji plików dźwiękowych oraz dokładnie omówiony jest format wielokanałowego dźwięku przestrzennego, wprowadzony na rynek komercyjny stosunkowo niedawno. Następny rozdział opisuje sposób wykonania projektu, a więc niejako jest podsumowaniem dokonanych wyborów spośród uprzednio przedstawionych możliwości, na rzecz najlepszego możliwego efektu. W rozdziale zamykającym pracę, opisane są wnioski i możliwe ścieżki rozwoju projektu.

SUMMARY

Subject of sound maps had to date no publications applying the latest available technologies to their cultural and social aspects. Even the nomenclature can be misleading. Extensively discussed are acoustic maps, concerned with noise measurement. In contrast, sound maps describe sound space of a specific place. More detailed description, along with several examples, constitute chapter one. The following sections include technical aspects, starting with hearing mechanism, especially in terms of spatial hearing, in chapter two. Two following chapters depict techniques of recording and playback, with an emphasis on spatial sound – from the historical outline to the latest available solutions. Chapter five is dedicated to audio formats. The most popular ones are described, topic of sound compression is developed, included is in-depth description of multi-channel spatial sound format launched recently on the commercial market. Next chapter depicts thesis realization, it is a substantiated summary of the choices made from among previously presented possibilities, with an objective of achieving best possible outcome. Closing chapter includes conclusions and possible further development paths.

Spis treści

Wstęp

Mapy akustyczne, za sprawą dyrektywy 2002/49/WE, są obowiązkowo wykonywane dla miast powyżej 100 tysięcy mieszkańców od 2002 roku. Na podstawie pomiarów hałasu miały zostać opracowane programy ochrony środowiska. Ale już w latach 60. XX wieku padło pytanie- czym w ogóle jest hałas? Czy otaczające nas dźwięki nie są już częścią naszej kultury? Tak powstała nowa, interdyscyplinarna dziedzina- **mapy dźwiękowe.** A więc mapy akustyczne mówią o tym jak głośne jest dane otoczenie, ale jakie to dźwięki i jaką tworzą one przestrzeń to domena map dźwiękowych, o których mowa w tej pracy. Jest to zatem archiwizacja akustycznego krajobrazu, rejestracja charakterystycznych na danego obszaru dźwięków, tworzących tożsamość regionu. Takie mapy były tworzone w różnych miastach na świecie, również w Polsce, jednak wciąż nie weszły w aktualną technologię, powszechnie już dostępną na najpopularniejszych platformach jak Facebook, YouTube czy Google. Niewykorzystywane w tej dziedzinie przestrzenne formaty dźwięku (w formie szerszej niż stereofonia), w połączeniu z przestrzennym obrazem dawałyby znacznie większe możliwości i podniosły atrakcyjność odbioru. Dźwięki ambientu (brzmienie otoczenia) w dwuwymiarowej przestrzeni nie dają tak wyraźnej wizualizacji. W formie dźwięku 3D połączonego z obrazem w 360^{o} dostajemy znacznie więcej informacji i bardziej realne przeżycia.

Tworzony przeze mnie projekt ma na celu zbudowanie bazy nagrań miast (i nie tylko) z jak największego obszaru- Polski, Europy, a docelowo z całego świata. W pracy przedstawiony zostanie tor realizacji takiej mapy. Kamera umieszczona w kilku najbardziej charakterystycznych miejscach rejestruje otoczenie z jego naturalnym dźwiękiem, zapisując fragment naszej teraźniejszości. Dzięki temu, z jednej strony możemy wybrać się w wirtualną podróż po mapie i usłyszeć charakterystyczny klimat danych miejsc. Z drugiej strony, nagranie jest bezosobowe, więc może również być archiwalnym zapisem tego co teraz nas otacza, a tak dynamicznie w dzisiejszych czasach się zmienia. Zachęcająca, ciekawa forma takiej mapy dźwiękowej, może być dobrą wizytówką miasta, reklamą i wierniejszą wersją archiwizacji.

Cel pracy i zakres pracy

Mapy dźwiękowe nie były rozwijane przy użyciu nowszych technologii od prawie 10 lat. Celem tej pracy jest wykorzystanie aktualnych możliwości i splecenie ich w jedno, ciekawe w odbiorze rozwiązanie. Systemy dźwięku przestrzennego w zestawieniu z obrazem dadzą wyraźniejszy odbiór i lepszą lokalizację źródeł charakterystycznych dla otoczenia dźwięków. Nagrania wykonywane są w formie kilkominutowego filmu obejmującego wizję 360º i dźwięk dookólny. Docelowo zostaną umieszczone na stronie internetowej dedykowanej projektowi i udostępnione.

W niniejszej pracy opiszę rozwój technologii, która dała podstawy do realizacji tego projektu w najwyższej możliwej jakości i najbardziej optymalnej wersji gotowej do udostępnienia w sieci.

W rozdziale pierwszym przybliżona jest tematyka map dźwiękowych z przykładami istniejących map na świecie i w Polsce. W rozdziale drugim nakreślona jest zasada działania narządu słuchu w ogólności oraz pod kątem słyszenia przestrzennego, a w szczególności mechanizmy wykorzystywane w technicznych rozwiązaniach systemu audio. Rozdział trzeci poświęcony jest technikom rejestracji dźwięku, od rodzajów mikrofonów do konfiguracji technik mikrofonowych, kończąc na rejestracji dźwięku przestrzennego i mikrofonie ambisonicznym. W rozdziale czwartym omówione są techniki reprodukcji- rozwój systemów audio od stereofonii do najnowszych, rozbudowanych systemów wielokanałowych. Tematem rozdziału piątego są formaty dźwięku i systemy kompresji, a ostatecznie- wielokanałowy dźwięk przestrzenny. Rozdział szósty podsumowuje niejako wszystkie poprzednie, opisując dokonany na rzecz projektu wybór, a zatem: mikrofon, kamerę, system rejestracji, format dźwięku i sposób publikacji projektu (a więc i reprodukcji), kompleksowo, na podstawie uprzednio przybliżonej tematyki. W rozdziale tym opisana jest część praktyczna pracy, sposób realizacji, zwarta instrukcja pozwalająca odtworzyć wykonany tok pracy. Wyciągnięte z pracy wnioski zawarte są w rozdziale siódmym.

Rozdział 1: Mapy dźwiękowe

„Yes, but is it music?"

Takim pytaniem swoją książkę „The New Soundscape" rozpoczyna Raymund Murray Schafer- kanadyjski kompozytor i pedagog, prekursor map dźwiękowych. On jako pierwszy podstawił pytanie „jak brzmi miasto?", czym jest hałas, czym różni się od muzyki i czy w ogóle go zauważamy. W latach 60. XX wieku zwrócił uwagę na pogarszającą się kondycję dźwiękową naszej codziennej przestrzeni. Drastyczny rozwój technologii sprawił, że szumy, brzęczenie i stukot maszyn momentalnie stał się wszechobecny. A my bardzo szybko się do niego przyzwyczailiśmy. Staliśmy się obojętni na dźwięki tła i na to czym jest dla nas cisza. Schafer w swojej książce punktuje aspekty tej ignorancji. Mimo iż minęło już 50 lat, problemy tam opisane są wciąż aktualne, a może nawet „ jeszcze aktualniejsze". Do jego tekstów i badań do dziś odwołują się badacze zajmujący się przestrzenią dźwiękową. Juhani Pallasmaa w swoich tekstach stwierdza nawet, że nasz zmysł słuchu został „oślepiony". Współczesna architektura mocno ogranicza przestrzeń dźwiękową. Ulice nie dają echa, a mieszkania projektuje się tak aby były wytłumione. Brak naturalnej przestrzenności kumuluje hałas i cenzuruje dźwięk. Schafer podkreśla jednak, że hałas nie jest jednoznacznym pojęciem. Dla jednych przejeżdżający Harley-Davidson z dźwiękiem swojego 62-konnego silnika jest niemalże muzyką, dla innych tylko nieznośnym hukiem. „Czymże jest muzyka" pyta zatem swoich studentów, usiłując ułożyć z nimi odpowiednią do dzisiejszych czasów definicję. [13] Z jednej strony kompozytorskie eksperymenty, wynikające z możliwości tworzenia nowych brzmień w świecie muzycznym- dzięki elektronice i komputerom. Z drugiej, rozwijająca się, nieustanna, losowa „muzyka miasta". Granice pomiędzy tymi dwoma twórcami coraz bardziej się zacierają. Wniosek może być jeden- całość tworzy dźwiękowy pejzaż naszych czasów.

Rejestracja tegoż pejzażu nie jest zbyt rozpowszechniona, ale ustaliła się w tej branży pewna terminologia: [19]

Audiosfera- to sfera dźwięków nas otaczających, przestrzeń dźwiękowa;

Field recording- nagrywanie audiosfery, nagrania terenowe;

Pejzaż dźwiękowy (Soundscape)- wszystkie dźwięki docierające do nas, charakterystyczne dla danego otoczenia;

Soundmark- dźwięk przypisany do danego miejsca, unikatowy, rozpoznawalny w społeczności i kojarzony z tym miejscem.

Na uniwersytecie w Burnaby w Kanadzie, Schafer opracował koncepcję pejzażu dźwiękowego oraz stworzył fundamenty i ideologię do nowej dziedziny- **ekologii muzyki**. We wspomnianej książce „The New Soundscape" opisuje pejzaż dźwiękowy jako akustyczne środowisko człowieka. Odnosi się do tego jak człowiek odbiera dźwięk, nie tylko pod kątem fizjologii , ale również społecznie i historycznie. Twierdzi, że obraz ten został zaburzony, „zaśmiecony" i wskazuje kierunek badań oraz działań jakie należy podjąć by poprawić ten stan. Owocem tych badań był projekt rozpoczęty w 1969 roku- *The Soundscape Project*. W obszarze zainteresowań znalazł się człowiek w środowisku dźwiękowym, jego relacja z tym środowiskiem i jego znaczeniem. Zespół badawczy wykonywał field recording, analizował i porównywał zebrane dane. Celem było stworzenie kompletnego archiwum audiowizualnego. Tym sposobem powstały pierwsze mapy dźwiękowe, a w 1977 roku, Schafer opublikował wyniki badań w książce „The Tuning Of The Word". [13, 19]

Praca Schafera była kontynuowana przez jego zespół, jednak taka forma „podróżowania" nie stała się zbyt popularna. Warto przyrównać popularność map dźwiękowych do Google Street View, które stało się modną zabawą i ciekawą formą poznawania różnych zakątków naszej planety. A przecież to tylko statyczne, rozmazane zdjęcia! Oczywiście, na dźwięk zwracamy mniejszą uwagę niż na obraz, ale bezsprzecznie, samo zdjęcie nie oddaje całości atmosfery danego miejsca. Wykonywane dotychczas mapy dźwiękowe były ukierunkowane z kolei głównie na

dźwięk. W większości są to punkty na mapie z podpiętym plikiem dźwiękowym (w stereo/binaural) i w najlepszym razie załączonym pojedynczym zdjęciem z miejsca nagrania lub jego panoramą. Map tego typu jest bardzo wiele, od globalnych projektów obejmujących nagrania z całego świata do map tworzonych w obrębie miasta. Poszukiwania najciekawszych zaczęłam od haseł „sound maps", „sound travelling" i „ambisonics sound map". Przy tym ostatnim żadna strona (z pierwszych kilku stron wyszukiwarki) nie nawiązywała do tematyki map dźwiękowych. Pierwsze dwa hasła otwierają szeroki przegląd wielu map dźwiękowych tworzonych w podobnej konwencji, dlatego przytoczę kilka najciekawszych czy też najbardziej rozbudowanych. Oto przegląd wybranych map dźwiękowych [na dzień 20.11.2019r.]:

1.1 Przegląd map dźwiękowych na świecie

- https://aporee.org/maps/ - strona istnieje od 2006 roku i kolekcjonuje amatorskie nagrania z całego świata. Autorzy strony zachęcają do wysyłania nagrań w formacie WAV lub MP3 wysokiej przepływności. Zachęcają do dzielenia się nieskompresowanymi plikami, opisują jak wykonać rejestrację aby była dobrej jakości. Nagrania powinny być utrzymane w standardzie 44,1 lub 48 kHz i 16 lub 24 -bitowej rozdzielczości, o długości minimum 1 minuty i maksymalnej wielkości pliku do 100 MB. Wykorzystana jest mapa satelitarna Google Maps z umieszczonymi na niej punktami z nagraniami. Strona daje możliwość stworzenia własnej playlisty, a nawet miksowania dźwięków. Jest na bieżąco aktualizowana i rozwijana o nowe nagrania.
- http://www.naturesoundmap.com/ -kolekcjonuje wyjątkowe nagrania dźwięków natury z całego świata wykonane przez profesjonalistów. Jest rozwijana przez *Wild Ambience Nature Sound* z Australii. Dedykowana mapa zawiera pinezki z odnośnikami do nagrań. Każde opatrzone jest w zdjęcie i notkę opisującą pochodzenie dźwięku. Nagrania są udostępniane przez platformę SoundCloud, istnieje możliwość stworzenia własnej playlisty. Strona jest bardzo czytelna i spójna. Bardzo podobną formę, również nastawioną na archiwizację otaczających nas dźwięków, ale pod kątem urbanizacji prezentuje strona

https://citiesandmemory.com/ . Obie strony do nagrań stereofonicznych dołączają fotografie.

- https://www.world-sounds.org/map/ - to nagrania z Europy, Azji i jednego punktu w Afryce opublikowane na SoundCloud i przypięte do mapy własnego projektu z załączonym szerokim opisem miejsca i okoliczności nagrania (przypominające blog). Do dźwięku dołączona jest fotografia. Strona dzieli dźwięki na kategorie takie jak miasto, natura czy kościoły. Jest bardzo przejrzysta, a nagrania są wysokiej jakości. Wszystkie w formacie stereo. Twórcy wymieniają dokładnie wszystkie elementy użytego wyposażenia- podają modele mikrofonów, rodzaje baterii, a nawet futerały.
- http://www.glasgow3dsoundmap.co.uk/soundmap.html - wchodzi na następny poziom. Opiera się na dźwięku 3D w formacie binauralnym. Niektóre nagrania zawierają dodatkowo zdjęcie panoramiczne, niektóre widok z Google Street View. Twórcy skupiają się głównie na hałasie, projekt ma być pomocny przy jego ocenie. Zawiera nagrania z kilku miejsc na każdym kontynencie.
- Inną perspektywę proponuje „Ambisonic- City Life" na stronie https://www.asoundeffect.com/sound-library/ambisonic-city-life/ . Jest to biblioteka nagrań ambisonicznych dostosowanych do wykorzystania na potrzeby VR/AR, a zatem gier, filmów czy właśnie wirtualnych podróży. Całość obejmuje 67 minut nagrań w 24 plikach w formacie AmbiX i stereo. Cena- około 60$.

1.2 Przegląd map dźwiękowych w Polsce

W Polsce także pojawiły się mapy dźwiękowe kilku miast. Wynikało to raczej z nawiązania do sztuki czy ekologii. Również skupiano się wyłączenie na dźwięku. Na tą chwilę, dostępną do ogólnego użytku poprzez strony internetowe, mapą dźwiękową miasta mogą się pochwalić:

- Wrocław- jako część projektu „Pejzaż Dźwiękowy" jest projektem badawczym mającym na celu archiwizację audiosfery miasta. Nagrania tworzone są od 2011 roku i archiwizowane w Bibliotece Kulturoznawstwa i Muzykologii Uniwersytetu Wrocławskiego. Liczy około 800 plików dźwiękowych (stan z oficjalnej strony- ostatnia aktualizacja 2012 rok; ostatnie nagranie dodane

w 2015 roku) o długości 1-2 minuty. Do nagrań dołączone są fotografie, informacje o czasie i warunkach atmosferycznych oraz gdzieniegdzie komentarze opisujące sytuację nagrania. Wykorzystano rejestrator ZOOM H4n i Tascam DR-100 do nagrań stereofonicznych w formacie WAV, a do publikacji skompresowano je do MP3. Wedle strony, w planie było wydanie płyty z nagraniami w 2014 roku. Mapa zamieszczona jest na dedykowanej stronie www . Pliki dźwiękowe są umieszczone w formie pinesek na mapie Google.

- Lublin- w charakterze historyczno-społecznym stworzył mapę dźwiękową w ramach projektu „Opowieści o dzielnicach Lublina". Jego celem jest upowszechnienie krajobrazu dźwiękowego miasta. Stworzone także na mapie Google odnośniki, przekierowują nas do nakładki z dźwiękiem i fotografią. Dostępne są również dwie dodatkowe- jedna z nagraniami audycji radiowych o Lublinie, a druga z czytanymi wierszami o Lublinie (nawiązującym do danego miejsca). Twórcy skupiają się przede wszystkim *soundmarkach* oraz dźwiękach, które mogą zaniknąć w niedalekiej przyszłości. Spróbowano nawet zremiksować rekonstrukcję dźwiękową starej dzielnicy żydowskiej. Strona jest częścią większej całości, w której możemy znaleźć makiety średniowiecznego Lublina, panoramy, a nawet interaktywny przewodnik 3D z animowaną postacią Józefa Czechowicza. Nagrania są publikowane w formacie MP3 w stereo. Twórcy nie podają szczegółów technicznych.

- Katowice- stworzyły swoją mapę dźwięków przy okazji warsztatów z *field recordingu* prowadzonych w ramach projektu „Medialab Katowice" przez Marcina Dymitera. Rejestrowane są głównie dźwięki z terenów zielonych miasta. Twórcy postarali się o własną szatę graficzną interaktywnej mapy. Nagrane są minimalnie dwa pliki audio z każdego z zaznaczonych miejsc oraz zmierzony jest poziom natężenia dźwięku, co daje obraz skali zjawiska dźwiękowego

- Legnica- kieruje do swojej „mapy" dźwiękowej w formie playlisty na portalu soundcloud.com. Twórcy zachęcają do wysyłania nagrań i wspólnego poszerzania zbiorów. Plik jest opisany miejscem nagrania. Nie dołączono żadnej mapy. Do niektórych nagrań dołączono zdjęcie.

- Kraków- pod fundacją Radiokit stworzono stronę Stowarzyszenia Radiofonia z mapą, na której umieszczono kilka nagrań z dźwiękami miasta oraz nagrania opowieści o charakterystycznych punktach Krakowa. Dołączone są symboliczne zdjęcia dotyczące miejsca o których mowa. Twórcy nie podają żadnych szczegółów co do nagrań. Krótka notka na stronie informuje, że celem projektu jest zebranie różnego typu historii na temat Krakowa, a nagrania terenowe są uzupełnieniem opisywanego obrazu.

W Krakowie stworzono także mapę dźwiękową z dokładnymi opisami i interaktywną mapką w ramach pracy dyplomowej na wydziale architektury Politechniki Krakowskiej. Niestety strona nie jest już dostępna, znaleźć można jedynie artykuł autorki Martyny Kozak na temat map dźwiękowych, z którego dowiadujemy się, że inicjatywę stworzenia takiej mapy miasta podjęto także w Poznaniu (obecnie już niedostępna; prawdopodobnie pierwsza mapa dźwiękowa w Polsce), Łodzi (niedostępna; wedle strony internetowej organizowane są terenowe wycieczki pod podobnym tytułem i historyczno-społecznym przesłaniem) oraz Toruniu (niedostępna; Tonopolis- był bardzo rozbudowanym projektem stworzonym przez Ewę i Jacka Doroszenko; obecnie dostępna jest jedynie dokumentacja i opis projektu wykonany przez Rafała Kołackiego z Uniwersytetu Mikołaja Kopernika w Toruniu). Z inicjatywy fundacji UTU z Torunia powstała strona zrzeszająca nagrania z kilku obszarów. Jako twórcy wymienieni są między innymi Marcin Dymiter i Rafał Kołacki zatem zapewne jest to rozwinięcie poprzednich projektów. Działania kierowane były głównie do dzieci i uczestników warsztatów terenowych w wymiarze artystyczno-społecznym. Mapa prezentuje same nagrania (z jednozdaniowym opisem) podpięte do pinesek

w wybranych miejscach Torunia i okolic, a także różnych punktów na Mazurach. Do zakładki z wybranej okolicy dołączony jest krótki opis poszczególnych projektów oraz blog o nazwie „Dźwiękospacery". Ostatnie nagrania dodano w 2012 roku. Co ciekawe był to z jakiegoś powodu szczególny rok dla map dźwiękowych. Większość projektów powstała właśnie wtedy, wspomniane artykuły zostały stworzone właśnie wtedy. Wszystkie formy map dźwiękowych w Polsce składają się (i składały) wyłącznie z nagrań w technice mono lub stereo, krótkiego opisu i ewentualnie pojedynczej fotografii.

Rozdział 2: Dźwięk przestrzenny

2.1 Biologia dźwięku

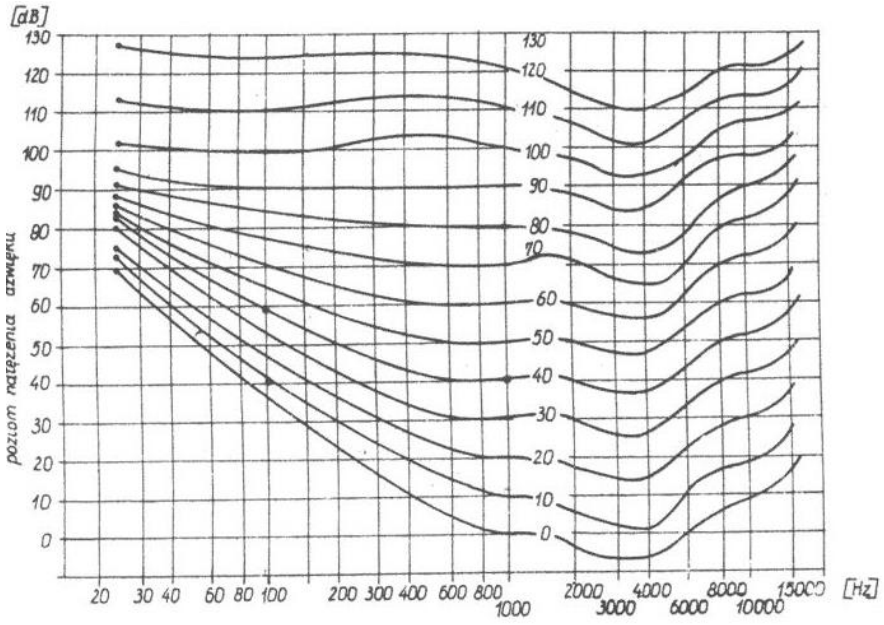
Zmysł słuchu nie jest dla nas zagadką już od dawna. Zrozumienie niesamowitych, skomplikowanych zjawisk jakie zachodzą od małżowiny usznej aż po odpowiednie ośrodki w mózgu, dało nam już nawet możliwość manipulacji odbieranymi bodźcami. W świecie cyfrowych danych, a więc i dźwięku, wiedza z zakresu psychoakustyki okazałą się bardzo przydatna.

Jesteśmy naturalnie dostrojeni do zakresu ludzkiej mowy. Dlatego zakres częstotliwości jakie jesteśmy w stanie usłyszeć mieści się w granicach 20 Hz- 20 kHz. W większości. Zdarza się spotkać stwierdzenie, że dolna granica może zejść do 16 Hz (u noworodków przyjmuje się taką granicę). Młodzi ludzie słyszą lepiej wysokie częstotliwości, więc i górny zakres nie jest stały. Jednak te widełki zostały przyjęte za słuszne dla ogółu i do nich jest dostosowany w większości sprzęt audio. Dodatkowo, ponieważ domyślnie ludzka mowa interesuje nas najbardziej, nasz słuch jest szczególnie wyczulony w zakresie 1- 3 kHz (język polski jest akurat przykładem szerokiego wykorzystania wyższych częstotliwości poprzez takie głoski jak sz, ś, ć czy ź). Na podstawie licznych badań stwierdzono, że ucho ma swoją charakterystykę częstotliwościową tzn. odbiera dźwięki z różną czułością.

Wynikiem tych badań jest między innymi prawo Webera- Fechnera i miara decybelowa, a także wiedza o słyszeniu przestrzennym. Prawo Webera- Fechnera mówi, że zauważalny przyrost intensywności bodźca odnosi się, w skali logarytmicznej, do już działającego.[9] Dlatego podwojenie natężenia dźwięku, da nam wzrost poziomu głośności o 3dB, ale dopiero wzrost poziomu o 6dB da nam wrażenie podwojenia głośności. Stąd też wynika logarytmiczna skala i „decy"-bele w akustyce. Dzięki tak przyjętej skali próg słyszalności (zerowe natężenie dźwięku) mamy wygodnie określony jako 0dB, w związku ze wzorem na poziom natężenia dźwięku (β), gdzie I_0 to poziom odniesienia (próg słyszalności):

$$\beta = \log \frac{I}{I_0}$$

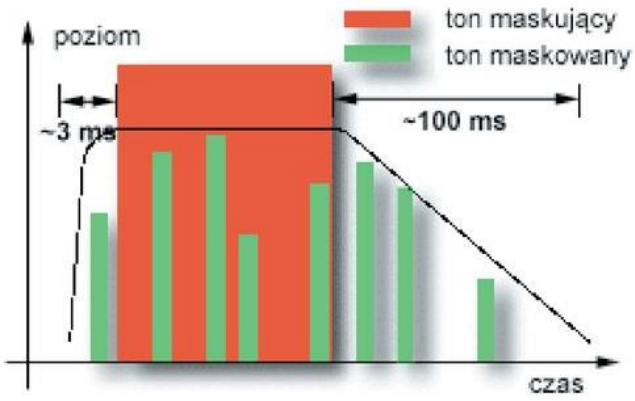
Dzięki badaniom odsłuchowym potwierdziło się działanie prawa odkrytego przez Webera (a które Fechner opisał matematycznie), również w przypadku ludzkiego słuchu. Stworzono izofony czyli krzywe jednakowej głośności. Obrazują one jak nieliniowo odbieramy dźwięk. Krzywe nakreślają odbierane przez nas natężenie dźwięku (jako poziom głośności w fonach) w zakresie słyszalnych częstotliwości. Fon jest jednostką percepcyjnej głośności dźwięku, w odniesieniu do wartości natężenia w dB dźwięku o częstotliwości 1000 Hz. Izofony przedstawia rysunek poniżej. [4, 10]



Rys. 2.1 Izofony- krzywe jednakowej głośności [23]

Bardzo istotnym, powszechnie wykorzystywanym zjawiskiem, wynikającym z budowy ludzkiego ucha jest **efekt maskowania dźwięków**. Polega on na ignorowaniu niejako dźwięków cichszych na koszt głośniejszego o bliskiej częstotliwości. Wynika to z budowy ślimaka i znajdującej się w nim błony podstawnej, która „pasmowo" reaguje pobudzeniem na odbieraną częstotliwość. W związku z tym częstotliwości pobliskie, lecz cichsze, nie wywołają już pobudzenia w tym obszarze, w tym samym (w przybliżeniu) momencie. Dzieje się tak, ponieważ błona podstawna dzieli się na tzw. **pasma krytyczne**, określające rozdzielczość

częstotliwościową naszego słuchu. Takich pasm jest około 24, każde zajmujące około 1,3 mm (ok. 1300 neuronów), ale nie są one równo podzielone. Ponad 2/3 długości błony jest zarezerwowane dla częstotliwości od dolnej granicy do 5 kHz. Pozostała niecała 1/3 zajmuje się wyższymi tonami. Wiąże się to z wieloma konsekwencjami- maskowanie wynika z pewnej bezwładności błony. Tuż po głośnym dźwięku, cichsze nie wywołają pobudzenia, gdy błona nie zdążyła jeszcze wrócić do stanu pierwotnego. Mowa tu o czasach rzędu 2-5 ms. Nawet dowolnej częstotliwości ciche dźwięki występujące nieznacznie wcześniej niż głośny, nie zostaną odebrane.



Rys. 2.2 Maskowanie czasowe dźwięku (czarna linia to próg słyszalności) [24]

Efekt maskowania jest tym wyraźniejszy im bliższe siebie są częstotliwości i kierunki z których dochodzą. Dźwięk maskujący powoduje podniesienie progu słyszalności i im bardziej jest on intensywny tym szersze pasmo maskuje. Chętnie jest to wykorzystywane w metodach kompresji dźwięku- o których w rozdziale *Formaty dźwięku*. Co więcej, wykorzystuje się również przewidywalność naszej wyobraźni. **Tony różnicowe** to pojawiające się brzmienie tonu podstawowego powstające pomiędzy harmonicznymi jednego wielotonu. Wynika to z nieliniowości słuchu związanej z budową ślimaka. Gdy dwa tony odległe od siebie o mniej niż kwintę brzmią równocześnie z równym natężeniem, słyszymy niższy od nich dźwięk należący do szeregu harmonicznego. Zjawisko to jest od dawna wykorzystywane- pomocne przy ograniczaniu pasma częstotliwości przesyłanego dźwięku (np. w telefonie analogowym), zastosowane przy konstrukcji piszczałek organowych

(zamiast stawiania bardzo dużej basowej piszczałki używa się dwóch mniejszych w odpowiednich proporcjach). Już w XIII wieku skrzypek G. Tartini uczył swoich uczniów wykorzystania tonów różnicowych do strojenia skrzypiec.

W wyniku wszystkich tych badań powstały **modele psychoakustyczne**, dzięki którym mamy uśrednioną wiedzę na temat ludzkiego słuchu i możemy wykorzystać ją w praktyce- w systemach audio, systemach kompresji cyfrowej, systemów rozpoznawania mowy lub tworzeniu nowych brzmień muzycznych czy metodach terapeutycznych. Są to modele opisujące system słyszenia z uwzględnieniem mechanizmów percepcji odbiorcy, opracowane na podstawie pomiarów odsłuchowych. Słuchacz ocenia wrażenia wywołane sygnałami testowymi, a sygnał jest równocześnie przetwarzany przez model, tak aby na jego wyjście było predykcją subiektywnych odczuć badanych. W ten sposób dostajemy matematyczną formułę dla sposobu odbioru słuchowego przeciętnego człowieka. Dźwięk może być już nie tylko metodą komunikacji i artystycznego wyrazu, ale i narzędziem technicznym. Możliwości te poszerzają się dodatkowo przez świadomość przestrzenności jaką człowiek przez niego odbiera.

2.2 Kompas dźwiękowy

Od narodzin jesteśmy „zanurzeni" w dźwiękach więc w naturalny sposób odbieramy je przestrzennie. Mimo że, w porównaniu do wzroku, słuch daje nam mniej informacji o lokalizacji to wykształciliśmy sobie szereg sposobów, aby wywnioskować jak najwięcej z dochodzącego do nas dźwięku. Mózg człowieka jest w stanie określić w ten sposób kierunek z dokładnością do 2 stopni u młodych ludzi i dokładność ta spada z wiekiem do 5-15 stopni. Tyczy się to przedniej płaszczyzny. Z tyłu głowy dokładność drastycznie spada. Zwykle, w warunkach laboratoryjnych, możemy określić kąt dochodzenia dźwięku z dokładnością 3-4 stopnie, ale tylko w poziomie. I nawet w poziomie, jeśli dźwięk będzie pojawi się dokładnie na wprost lub z tyłu (0 lub 180 stopni), możemy nie odróżnić czy źródło znajduje się przed czy

za nami. Ale do takich, poziomych dźwięków, jesteśmy naturalnie przyzwyczajeni. Normalnie radzimy sobie z tym automatycznie- obracając głowę, aby rozwiać niejasności. Jednak w osi pionowej poradzimy sobie ze zmianami kąta dopiero powyżej 9 stopni, a nad głową dokładność spada do 20 stopni. Skąd ta różnica i jak w ogóle osiągamy taką dokładność? Wszystko to dzięki zdolności mózgu do obliczania różnic natężeniowych i czasowych przy odbiorze dźwięków docierających do prawego i lewego ucha. Gdy dźwięk dociera do nas pod pewnym kątem, powstają różnice, wynikające z efektu maskowania głowy, czyli jej budowy- umiejscowienia uszu po dwóch stronach.

Ta „niewielka" odległość między nimi determinuje różnice w czasie odbioru dźwięku między uchem lewym a prawym w zakresie od 0 do 690 mikrosekund. W ten sposób, w zależności od kąta z którego dochodzi dźwięk, odbieramy go jako przesunięty w fazie. Najmniejsze opóźnienie, które zinterpretujemy jako zmiana kierunku to 6 mikrosekund. Największe opóźnienie- gdy dźwięk pada pod kątem 90 stopni, to około 650 mikrosekund (w zależności od szerokości głowy). Jednak różnice czasowe najistotniejsze do tej interpretacji, te z którymi mózg radzi sobie najlepiej, mieszczą się w zakresie od 6 do 400 mikrosekund. To znaczy, że najważniejsza jest dla nas lokalizacja tego co znajduje się w zasięgu naszego wzroku. Wspomagamy obraz dźwiękiem. Na temat tego co dzieje się za nami, mamy w gruncie rzeczy tylko informację- że dzieje się za nami.

Najłatwiej zinterpretować różnice fazy będące porównywalne z odległością między „punktami pomiarowymi", czyli naszymi uszami. Dlatego opóźnienie (różnice fazowe) daje nam informacje o lokalizacji najdokładniej w zakresie częstotliwości do 1 kHz. Wyższe częstotliwości, około 2-3 kHz, nie tworzą dla nas już tak wyraźnych różnic dlatego na przykład doskonale słyszymy sygnały alarmowe, ale ciężko nam określić kierunek ich dochodzenia. Z drugiej strony, poniżej 100 Hz też jest to już trudne, szczególnie w pomieszczeniach, ponieważ fala mająca wtedy długość około 3-4 metry zdąży się już odbić od ścian i osąd o kierunku dźwięku bezpośredniego jest zaburzony. Dla jeszcze niższych częstotliwości, od około 20 Hz, fala jest na tyle długa, że różnice czasowe w jej odbiorze są pomijalne. Dla takich częstotliwości

kierunku już nie określimy. Tą właściwość wykorzystano tworząc subwoofer. Ograniczenia dotyczące częstotliwości oraz pasma, na które jesteśmy szczególnie wrażliwi wynikają z preferencji naszego mózgu i układu słuchowego, który jest dostrojony do zakresu naszej mowy. W środku tego pasma, około 1 kHz, nasz słuch jest szczególnie wyczulony. Do przestrzennej lokalizacji dźwięku również wykorzystany zostaje ten wzmocniony zakres. Najlepiej lokalizujemy dźwięki z zakresu od 400 do 1500 herców. [14, 4, 3]

Jednak zakres częstotliwości daje też inne informacje o przestrzenności, a konkretnie o odległości. Barwa dźwięku także jest brana pod uwagę przy tej złożonej interpretacji. Zawartość wyższych częstotliwości w paśmie odbieranego sygnału zależy od odległości- wysokie częstotliwości są szybciej tłumione. Zatem im dalej znajduje się źródło tym bardziej przytłumiona będzie barwa, dźwięk będzie zawierał mniej wysokich składowych. Jest to jednak różnica zauważalna dopiero przy bardzo złożonych, widmowo skomplikowanych sygnałach.

Inną ważną cechą dla określenia kierunku jest różnica natężeń dźwięku docierającego do uszu. Także w tym przypadku wyższa częstotliwość daje wyraźniejszą informację. Przy częstotliwości około 300Hz dla kąta 90 stopni różnica natężeń będzie wynosiła 2-3 dB, natomiast dla 1-2kHz jest to już 20 dB. Na podstawie natężeń najlepiej możemy lokalizować źródło, którego dźwięk ma częstotliwość powyżej 400 Hz. Gdy różnica w natężeniu pomiędzy uszami wynosi powyżej 15 dB to lokalizujemy je jako „z boku". [10,4]

Do tego dochodzi jeszcze niejako próba oceny wielkości źródła dźwięku. Im bardziej strome jest czoło fali docierające do odbiory (atak dźwięku jest ostrzejszy) rozumiemy je jako bliższe. Na podstawie tych przesłanek próbujemy umiejscowić w przestrzeni źródło dźwięku. W przypadku pomieszczeń pojawia się kolejny czynnik- siła odbić od ścian pomieszczenia. Do zestawu danych do interpretacji dochodzi różnica czasów docierających do odbiorcy sygnałów bezpośrednich i odbitych. Im bliżej dźwięku jesteśmy tym silniejszy jest dźwięk bezpośredni w stosunku do odbić. Mimo wszystko ocena odległości jest mało precyzyjnym

procesem. Można jednak wykorzystać tą wiedzę do symulacji przestrzeni dźwiękowej, umiejscowienia źródeł i geometrii pomieszczenia.

Pewną informację o kierunku dochodzenia dźwięku uzyskujemy także dzięki małżowinie usznej. W jej zagłębieniach zachodzi tłumienie i wzmacnianie średnich i wysokich tonów, zależne od kąta padania dźwięku i jego częstotliwości.

Mózg potrafi zanalizować sygnały z obu uszu, przeliczyć na bieżąco różnice między nimi i zinterpretować to jako kierunek dochodzącego dźwięku. Jest to niewątpliwie imponujący proces, bardzo złożony, jak opisano powyżej. Mimo to lokalizacja słuchowa jest o dwa rzędy wielkości słabsza od wzorkowej. Łatwiej nam sztucznie stworzyć realistyczne wrażenia słuchowe. Jest to oczywiście szeroko wykorzystywane przez twórców filmów, nagrań muzycznych i systemów audio oraz wciąż dopracowywane.

2.3 Technika słuchu

Już dawno doskonale zdawano sobie sprawę z przestrzennego charakteru zmysłu słuchu człowieka, jednak praktyczne wykorzystanie tej wiedzy rozpoczęło się dopiero pod koniec XIX wieku. Wtedy to prowadzono badania nad przetwarzaniem stereofonicznego sygnału. Jednakże ówczesna technologia dawała możliwość wykorzystania jedynie źródeł monofonicznych- dźwięk był przesyłany jednym torem, więc odwzorowanie przestrzeni było niemożliwe. Ponadto początkowo, z powodu błędnych założeń badawczych, wykorzystywano jedynie tony czyste (sinusoidalne), które są akurat najtrudniejsze do oceny kierunku. Dopiero użycie złożonych dźwięków (które są też zdecydowanie bardziej dla nas naturalne i powszechne) dało pewne rezultaty. Próby opisania przestrzennego słyszenia człowieka wytarły sobie pewne ścieżki i schematy łączące niejako fizykę z biologią i technologią.

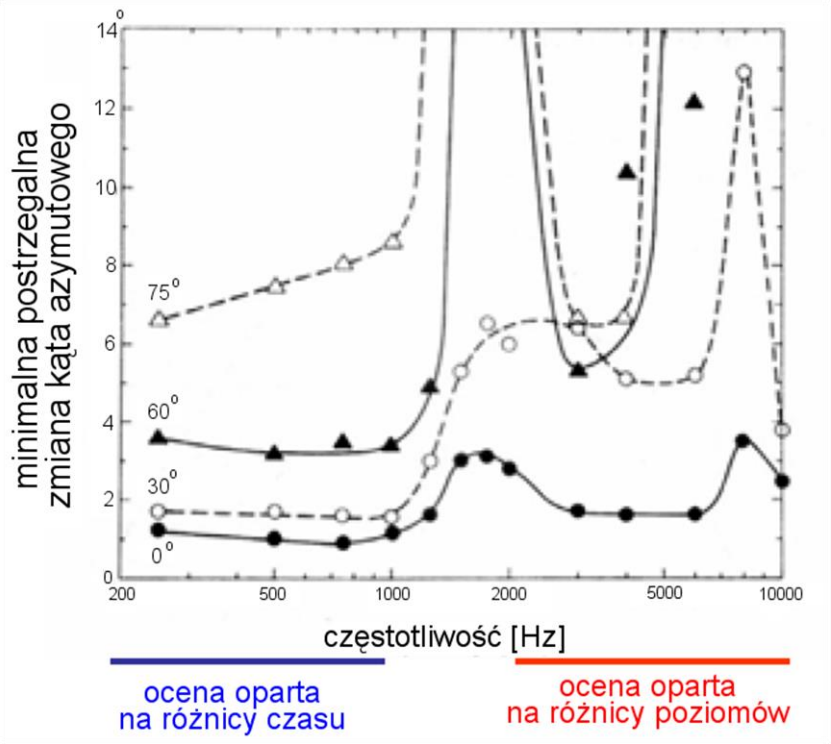
Gdy dźwięk dociera do nas pod pewnym kątem jedno ucho odbierze go bezpośrednio, a drugie zostanie objęte **cieniem akustycznym** głowy (wspomniane „maskowanie głowy”). Powstające w ten sposób różnice intensywności i fazy są

wykorzystane do lokalizacji w konkretnych procedurach zachodzących w mózgu. Mechanizm lokalizacji źródła na podstawie różnicy intensywności dźwięku określa się skrótem **IID (Interaural Intensity Difference)**. Funkcjonuje on najlepiej powyżej częstotliwości 2000 Hz, kiedy długość fali jest istotnie mniejsza od średnicy głowy i efekt jej tłumienia jest zauważalny. Minimalna różnica poziomów możliwa do zarejetrowania to 0,5-1 dB. Mechanizm interpretacji różnicy czasowej to **ITD (Interaural Time Difference)**. Opiera się on na analizie różnicy fazy fali dźwiękowej i funkcjonuje dobrze dla częstotliwości, których długość fali jest co najmniej dwukrotnie większa od średnicy głowy. Minimalna, dostrzegalna różnica czasowa to 10-20 mikrosekund. Przekłada się to bezpośrednio na minimalną, możliwą do wychwycenia, różnicę fazy, a więc i kąta (w osi azymutalnej).

$$\Delta t = \Delta \emptyset \, (2\pi f)$$

Nie są to jednak wystarczające informacje by zlokalizować źródło. Wiemy jedynie, że znajduje się ono na powierzchni stożka wytyczonego przez oboje uszu. Szczególnie trudna jest lokalizacja na podstawie częstotliwości około 1500 Hz, ponieważ okres fali jest bliski naturalnemu ITD. Różnica fazy jest więc mała i w związku z tym błąd lokalizacji duży. Dotyczy to jednak czystych tonów sinusoidalnych, które poza laboratorium raczej nam się nie zdarzają. Działanie mechanizmów IID i ITD przeplata się, wykorzystując całe widmo docierającego dźwięku. Poglądowo, na podstawie samych tonów prostych współpracę tą można by rozgraniczyć jak na rys. 2.3.

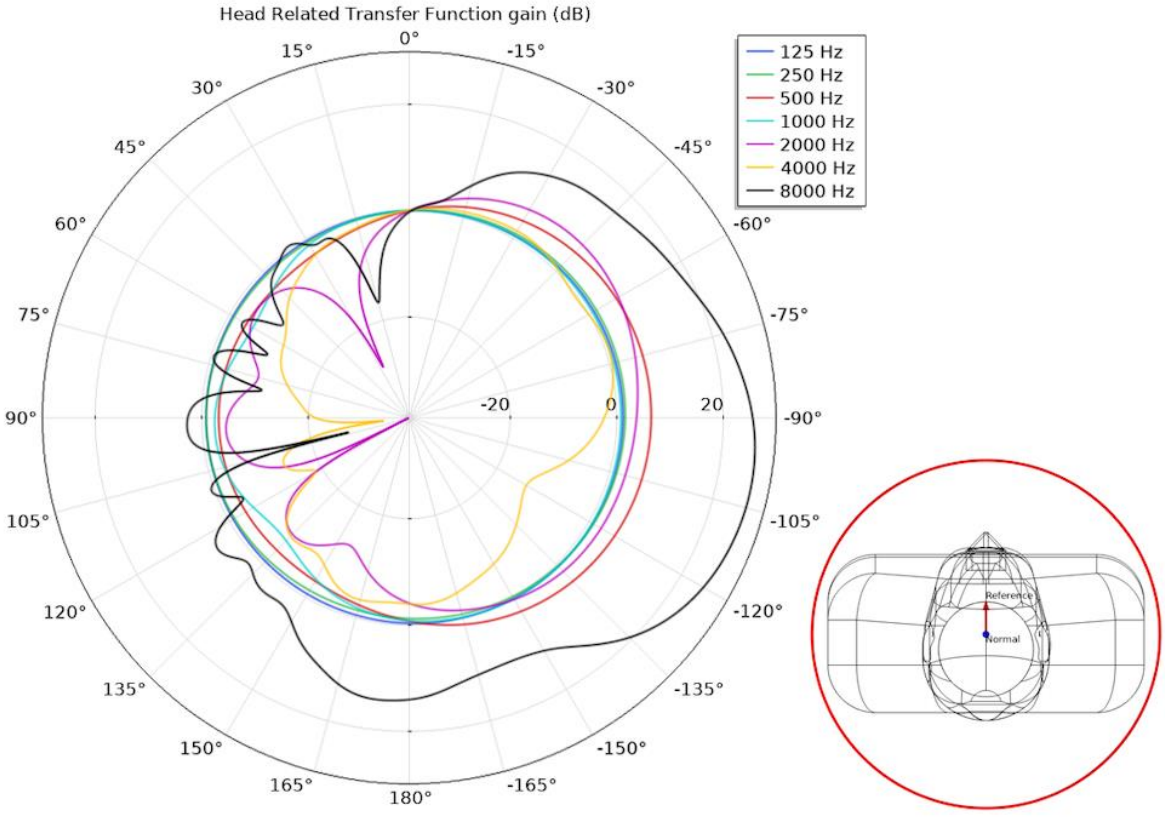
Sumując, wskazówki otrzymane przez IID i ITD otrzymujemy bardzo dobrą lokalizację azymutalną źródła dźwięku, ale nie odgrywają one znaczącej roli przy określaniu wysokości źródła. Pomocnym, wspomnianym wcześniej, elementem wpływającym na odbiór, jest małżowina uszna. Teoria Batteau mówi, że ma ona znaczącą rolę w określaniu lateralizacji i kąta podniesienia źródła. Wykonane przez niego badania, wskazują, że charakterystyka częstotliwościowa małżowiny usznej jest znacznie bardziej wrażliwa na kierunek góra- dół, niż prawo- lewo. Mamy zatem kolejny, swojego rodzaju filtr stworzony przez małżowinę uszną, która tworzy odbicia i tłumienia docierającego do nas dźwięku. [3, 15]



Rys. 2.3 Dokładność oceny kąta azymutalnego w stosunku do częstotliwości i kąta dochodzenia dźwięku. Dla 1500 Hz okres fali jest bliski naturalnemu ITD [25]

Z oceną odległości źródła jest związany inny mechanizm- **DRR (Direct to Reverberant Ratio)**, który mierzy stosunek głośności sygnału bezpośredniego do odbitego. Co więcej, odpowiednia interpretacja tych danych wynika z naszego doświadczenia. Odległość dźwięku wnioskujemy na podstawie porównania go z wzorcem, który już znamy. Podobna „wprawa" jest potrzebna poniekąd w przypadku pozostałych parametrów, ponieważ każdy z nas ma zupełnie inną małżowinę uszną- są one absolutnie unikatowe, jak odcisk palca. Dlatego, u każdego z nas inaczej odbity, wytłumiony i zmodyfikowany zostanie dźwięk przed dotarciem do bębenka wewnątrz ucha. Stąd potrzeba uśredniania tzw. funkcji HRTF, jeśli chcemy z nich skorzystać. Nie bez powodu też poszczególnym mechanizmom nadano nazwy. Sumę tych wszystkich oddziaływań- filtr głowy i małżowiny, wynikające z nich IID i ITD oraz inne modyfikacje sygnału spowodowane fizjologią słuchacza mogą być opisane przez funkcje **HRIR (Read-Related Impulse Response)**. Opisują one relację

źródła dźwięku i słuchacza, odpowiedź impulsową naszego układu słuchowego. Funkcje są wyróżnione oddzielnie dla prawego i lewego ucha. W praktyce mówi się zwykle o HRIR po transformacie Fouriera i opisuje terminem **HRTF (Head-Related Transfer Function)** czyli „funkcjami przejścia głowy". Zawierają informacje o właściwościach przestrzennych i po splocie z sygnałem akustycznym dają taki dźwięk jaki zarejestrowałoby ucho człowieka. Uśrednione funkcje HRTF wykorzystuje się do modeli psychoakustycznych, do do uprzestrzennienia dźwięku w systemach audio i nie tylko. Ponieważ funkcje są uśrednione, efekt również jest dostosowany do przeciętych parametrów (anatomii) słuchacza. Aby dokładnie odtworzyć wrażenia przestrzenne HRTF musiałby być dostosowany (zmierzony) konkretnie do ucha słuchacza. [3,12]



Rys 2.4 Wykres biegunowy funkcji HRTF dla komputerowo wymodelowanego słuchacza [26]

Rozdział 3: Rejestracja dźwięku przestrzennego

Przestrzenność jest nieodzowną cechą dźwięku. Trudność w jej rejestracji wynikają ze złożoności problemu. Ludzkie ucho doskonale sobie z tym radzi, ponieważ ma za sobą wspaniałą maszynę obliczeniową, która składa wszystkie przesłanki na temat dźwięku, w całość. Aby odpowiednio spleść ten proces i przekazać wirtualnie wszystkie wrażenia słuchowe potrzebny jest rozbudowany łańcuch urządzeń rejestrujących i kalibrujących dźwięk. Wszystko zaczyna się oczywiście od mikrofonu.

3.1 Mikrofony

Pierwsze pojawiły się w połowie XIX wieku. Były to mikrofony kwasowe, później węglowe (stykowe). Rozwój tej techniki był ukierunkowany na telefonię. Pierwsze nagranie płytowe z udziałem mikrofonu wykonano dopiero w roku 1925. Obecnie rodzajów mikrofonów jest bardzo wiele. Mamy możliwość dobrania odpowiedniego do warunków w jakich nagrywamy, źródła dźwięku czy po prostu własnych upodobań co do brzmienia nagrania. Te różnice wynikają z ich parametrów oraz często, biorąc pod uwagę na przykład już samą barwę brzmienia, z konstrukcji mikrofonu.

Zagłębiając się z w różnice konstrukcyjne, technikę jaka stoi za takim a nie innym brzmieniem, możemy wyróżnić kilka typów mikrofonów.

Mikrofony węglowe (stykowe) czyli najstarsze, działające skutecznie rozwiązanie. Stoi za nim T.A. Edison i D. E Hughes. Zaproponowane przez Wheatstone'a, w latach 70. XIX wieku, pionerskie rozwiązanie zakładało, że membrana odbierała drgania na skutek docierającego dźwięku i przekazywała je do igły zanurzonej w rozcieńczonym kwasie. Edison zastąpił kwas granulkami węgla, które pod wpływem ciśnienia zmieniały swoją rezystancję. Pomysł został wykorzystany w telefonie Bella. Zauważono dużą czułość granulatu na zmiany ciśnienia. Włączając go w obwód zasilany baterią, doskonale odbieramy zmiany natężenia prądu

wywołane falą akustyczną. Takie mikrofony są tanie, odporne mechanicznie i nie wymagają wzmacniaczy. Minusem jest wąski zakres częstotliwości z jakim sobie radzą- węższy niż zakres ludzkiej mowy, a także duży poziom szumów i niestabilność pracy. Obecnie w zasadzie nie są stosowane, ale swojego czasu, poza wykorzystaniem w słuchawkach telefonicznych były też wizytówką stacji BBC- dawały charakterystyczną, ciepłą barwę głosu.

Warto wspomnieć o innych, rzadszych rozwiązaniach, takich jak na przykład **mikrofony piezoelektryczne**. W tym podejściu wykorzystane są właściwości materiałów ceramicznych i kryształów. Piezoelektryczna płytka, stanowiąca membranę mikrofonu, reaguje na ciśnienie napięciem elektrycznym. Jeszcze do niedawna wykorzystywano je w słuchawkach telefonicznych, dobrze sprawdzają się również jako czujniki ultradźwięków ze względu na wysoką czułość na wysokie tony. Poza tym, w związku z wieloma ograniczeniami (czułość na zmiany temperatury i wilgoć) raczej wykorzystywane są już tylko w sprzętach amatorskich, ze względu na dużą skuteczność i niską cenę.

Poza tym istnieją jeszcze **mikrofony magnetostrykcyjne** działające na podobnej zasadzie (odkształcają się pod wpływem pola magnetycznego) oraz **elektromagnetyczne** (wykorzystujące zmiany oporu magnetycznego w obwodzie). Pozostają one jednak ciekawostką techniczną, raczej niewykorzystywaną w praktyce.

Przechodząc w końcu do rozwiązań aktualnych i najbardziej powszechnych, można powiedzieć, że na co dzień w obszarze naszych zainteresowań znajdują się mikrofony dynamiczne oraz pojemnościowe. Jedne i drugie znajdą kilka odmian jednak główna zasada działania pozostaje ta sama.

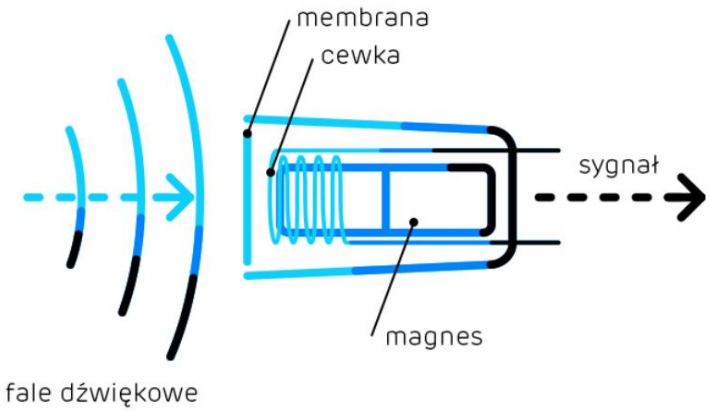
Mikrofony dynamiczne czy też magnetoelektryczne, w ogólności, wykorzystują fakt iż ruch w polu magnetycznym (a więc jego zmiana) indukuje siłę elektromotoryczną i tym samym zmiany napięcia elektrycznego.

Jednym z takich mikrofonów jest **mikrofon wstęgowy**. W tym przypadku drgającym elementem, odbierającym zmiany ciśnienia spowodowane dźwiękiem, jest cienka aluminiowa wstęga o grubości 2-5 mm, szerokości około 0,5 cm i długości 4-7 cm.

Folia ta jest umieszczona w szczelinie pomiędzy nabiegunnikami magnesu. Poruszając się w polu magnetycznym magnesów, folia wytwarza zmiany pola magnetycznego, które następnie przekładają się na sygnał elektryczny. Mikrofony wstęgowe są lubiane ze względu na swoją ciepłą barwę, niski poziom szumów i szerokie pasmo przetwarzanych częstotliwości. Nie potrzebują zasilania, ale mają bardzo delikatną budowę. Są bardzo wrażliwe na ruch powietrza, szczególnie podmuchy wiatru, dlatego wykorzystuje się je tylko w odpowiednich warunkach.

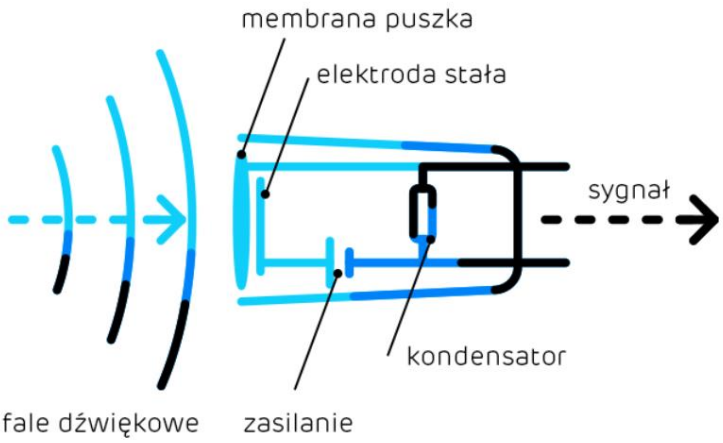
Konstrukcja **mikrofonów dynamicznych cewkowych** została opatentowana w 1931 roku i jest analogiczna do budowy typowego głośnika. Do cienkiej elastycznej membrany zostaje przymocowana cewka w silnym polu magnetycznym. Pod wpływem dźwięku membrana drga, a z nią cewka. Zgodnie z prawem Faraday'a, w cewce poruszającej się w polu magnetycznym wytwarza się prąd elektryczny odpowiadający częstotliwości drgań i natężeniu dźwięku. Mocniejsza fala akustyczna (o większym ciśnieniu) powodowałaby większe wychylenie membrany więc i większą prędkość drgań. Tak też wyższe częstotliwości wywołają „gęstszy" ruch. Aby zachować płaską charakterystykę amplitudową mikrofonu dla całego zakresu częstotliwości, drgania membrany będą malały ze wzrostem częstotliwości.

Taka architektura mikrofonu daje małą wrażliwość na wysokie tony, ale w wielu przypadkach może być to plusem, kiedy akurat chcemy je ograniczyć. Nie uzyskamy też dużej czułości. Jednak to także może być pożądane kiedy chcemy wyeliminować dalsze tło dźwiękowe. W związku z tym świetnie nadają się do nagrań z bliska. Dobrze zbierają sygnał z określonego kierunku. Wiążą się z tym także pewne minusy. Bardzo łatwo przejmują sygnały indukowane z zewnętrznego pola magnetycznego więc trzeba uważać na przykład na duże moce światła przy nagraniach. Są dość duże i ciężkie. Jednakowoż wady tego rozwiązania są zupełnie akceptowalne. A ponieważ takie mikrofony są też wytrzymałe i odporne na sygnały o dużej mocy oraz niedrogie, weszły do kanonu każdego realizatora dźwięku. Na Rys. 3.1 przedstawiono schematycznie konstrukcję mikrofonu dynamicznego.



Rys. 3.1 Schemat działania mikrofonu dynamicznego [27]

Niezbędnym punktem nr. 2 w studiu nagraniowym jest **mikrofon pojemnościowy**. Jego konstrukcja również wykorzystuje ruch w polu magnetycznym, jednak z zupełnie innego punktu widzenia jeśli chodzi o fizykę. Tu chodzi przede wszystkim o elektrostatykę. Taki mikrofon jest w kondensatorem, którego jedna z okładek jest nieruchomą płytką, a druga lekką ruchomą membraną. Generalnie rzecz ujmując membrana poruszana dźwiękiem zmienia pojemność kondensatora, a to znów skutkuje interesującymi nas zmianami napięcia. Membrana (o grubości rzędu 25 mikrometrów) jest umieszczona w odległości od 25 do 50 mikrometrów od nieruchomej okładki (stałej elektrody) z rowkami. Dzięki rowkom zwiększa się podatność przestrzeni powietrznej i zmniejsza opór tarcia powietrza. Masa membrany powinna być jak najmniejsza, aby zmaksymalizować możliwe wychylenia. Ponieważ jednak i tak są one niewielkie, dla właściwego odbioru sygnału potrzebny jest przedwzmacniacz mikrofonowy. Napięcie na zaciskach mikrofonu rzadko przekracza 1V, przedwzmacniacz podnosi ten sygnał. Jego odpowiednia kalibracja wpływa na poziom szumów w nagraniu i przesterowania. Do regulacji czułości przedwzmacniacza stosuje się potencjometr GAIN. Przedwzmacniacz może być także źródłem pewnego napięcia dla mikrofonu. Pojawia się w tym miejscu kilka rozwiązań, w związku z tym owe mikrofony możemy podzielić na elektretowe i z polaryzacją zewnętrzną. Mowa tu o wkładce mikrofonowej i w zasadzie materiale z jakiego została wykonana.



Rys. 3.2 Schemat działania mikrofonu pojemnościowego [27]

Mikrofony pojemnościowe z **polaryzacją zewnętrzną** charakteryzuje połącznie okładek kondensatora, za pośrednictwem dużej rezystancji, z zaciskami baterii. Bateria jest tutaj źródłem napięcia polaryzującego i zazwyczaj, w praktyce, wykorzystany zostaje do tego mikser lub przedwzmacniacz. Stosuje się tutaj wzmacniacz przymikrofonowy, który zmienia małą impedancję wkładki mikrofonowej na dużą impedancję wyjściową mikrofonu, dopasowaną do przedwzmacniacza. Dzięki temu pojemność rezystora nie zostaje zaburzona. Jest to istotne, gdyż jego rola jest kluczowa. Rezystor pełni dwojaką funkcję: gdy mikrofon nie odbiera sygnału, zapobiega rozładowywaniu się kondensatora; kiedy odbierana jest fala dźwiękowa, ruchoma okładka drga, zmieniając pojemność kondensatora, a przez rezystor płynie od źródła prąd ładowania (i rozładowania) usiłujący zniwelować te zmiany. Tym samym między zaciskami rezystora obserwujemy zmiany napięcia adekwatne do padającego dźwięku. Membranę (ruchomą okładkę) możemy potraktować jako źródło siły elektromotorycznej, która zależna jest od napięcia polaryzującego i amplitudy wychyleń w stosunku do nieruchomej okładki. Niestety nie możemy dowolnie zwiększać tych parametrów (chcąc osiągnąć duże wychylenia i tym samym większą skuteczność mikrofonu), gdyż w pewnym momencie skutkowałoby to wyładowaniami iskrowymi. Podsumowując, prędkość drgań membrany powinna być proporcjonalna do

częstotliwości aby skutkowała odpowiednią siłą elektromotoryczną i odpowiednim napięciem.

Odrobię inne rozwiązanie jest zastosowane w **mikrofonach elektretowych**. Elektrety to dielektryki (izolatory) wytwarzające zewnętrzne pole elektryczne. Wynika to z uporządkowania ładunków poprzez polaryzację silnym polem elektrycznym- są to niejako elektrostatyczne odpowiedniki magnesów. Takie właściwości wykazuje wosk karnauba, selen czy siarka, a także bardziej nas interesujące: polimetakrylan metylu, polioctan winylu i inne. Utrzymują niezależność od temperatury i wilgoci, a także dużą stałość w czasie. W mikrofonie elektretowym wykorzystuje się je na ruchomą okładkę kondensatora- membranę. Zastosowanie elektretu eliminują konieczność polaryzacji zewnętrznym zasilaniem. Okładka jest stale spolaryzowana i wytwarza pole elektryczne w szczelinie powietrznej dając napięcie około 180 V. Mikrofony elektretowe są niewrażliwe na wilgoć i odporne mechanicznie. Nie ma ryzyka elektrycznego przebicia szczeliny między okładkami. Poza tym, potrzebują dużo mniejszego napięcia zasilania, ale również wymagają przedwzmacniacza transformującego impedancję. [5,22]

3.2 Parametry mikrofonów

Jak widać, zasada działania mikrofonu, to jak przetwarza falę dźwiękową na sygnał elektryczny, może mieć duże znaczenie przy jego wyborze. W zależności od tego czy nagrywamy w terenie, jak wysoki mamy poziom szumu w tyle, czy wilgoć może być problemem przy nagraniu, wybieramy odpowiedni mikrofon do naszego nagania. Podejmując jednak tą decyzję nadal mamy do wyboru bardzo wiele modeli. Aby je porównać najłatwiej posłużyć się parametrami, charakterystycznymi dla każdego mikrofonu. Podstawowe cechy mikrofonów to: [4]

- charakterystyka częstotliwościowa,
- charakterystyka kierunkowa,
- czułość mikrofonu,
- odstęp sygnału użytecznego od szumu,

- ciśnienie graniczne,

W dalszej części rozdziału zostanie dokładnie przedstawiona cała charakterystyka jednego mikrofonu- ZOOM H3-VR. Najpierw jednak warto zapoznać się z tymi pojęciami.

Charakterystyka częstotliwościowa, pasmo przenoszenia czy też odpowiedź częstotliwościowa (wprost z angielskiego *frequency response*), mówi o tym z jakim zakresem i w jaki sposób mikrofon sobie poradzi. Jest to zwykle pierwsze na co zwraca się uwagę przy wyborze. Poglądowo, te charakterystyki umieszczone zostały też powyżej w opisie rodzajów mikrofonów. Przedstawione w formie wykresu dane, mówią nam nie tylko o tym jaki zakres częstotliwości fal akustycznych mikrofon przetworzy na sygnał elektryczny, ale też nakreśla sposób w jaki to zrobi. Widać jakie zakresy są wzmacniane przez mikrofon, a jakie wytłumiane. Dzięki temu, mniej więcej, możemy sobie wyobrazić jego brzmienie.

Charakterystyka kierunkowa czy też odpowiedź osiowa to parametr wskazujący kierunek, z którego mikrofon najlepiej zbiera dźwięk. Producenci podają charakterystykę mierzoną na wprost mikrofonu (pod kątem 0^o) czyli w jego osi podłużnej, gdzie czułość jest najwyższa. Dodatkowo zwykle załącza się charakterystykę pozaosiową. Pomiar nie zawsze odbywa się w tej samej odległości, co oczywiście wpływa na wyniki, ale chodzi tu przede wszystkim o kształt wykresu. Może on być przedstawiony w formie wykresu X-Y lub biegunowego, przy czym ten drugi wydaje się być bardziej obrazowy. Zewnętrzny okrąg wykresu biegunowego wskazuje zwykle źródło tonu sinusoidalnego o częstotliwości 1 kHz, a każdy kolejny okrąg w głąb, spadek poziomu o 5 dB. Naniesione linie odpowiadają kolejnym częstotliwościom pomiarowym. Wykres powinien być symetryczny, a linie możliwie jak najbardziej wyrównane i nieprzecinające się. Dzięki temu wiemy jak ukierunkowany jest mikrofon, jak dokładnie zbiera częstotliwości z wybranego kierunku i jak wrażliwy jest na źródła poza osią. Często załącza się również trzecią charakterystykę- odpowiedź w polu rozproszonym. Ilustruje ona jak odbierany jest dźwięk gdy nie kierujemy go wprost w stronę mikrofonu, a w większej przestrzeni dodatkowo zdąży się odbić od ścian i dociera do odbiornika wzmocniony, z różnych

kierunków. Pomiar ten ma największe znaczenie dla mikrofonów wszechkierunkowych, które mają szczególną zdolność do wyłapywania niższych częstotliwości i słabiej reagują na wyższe. Dzięki takiej charakterystyce wiemy jakie proporcje zachodzą w tym odbiorze. Zatem charakterystyka kierunkowa, mówi nam o tym jak zorientowany jest mikrofon, tym samym wprowadzając kolejny podział na:

- mikrofony wszechkierunkowe
- mikrofony dwukierunkowe
- mikrofony jednokierunkowe

Mikrofony **wszechkierunkowe** to inaczej (z technicznego punktu widzenia)- ciśnieniowe. W tym przypadku, membrana reaguje na falę akustyczną tylko z jednej strony i najlepiej gdy rozmiar mikrofonu jest znacznie mniejszy niż rozmiar tej fali. Ciśnienie akustyczne wywierane wtedy na membranę, jest takie samo ze wszystkich kierunków i niezaburzone- takie jak przed postawieniem mikrofonu w jej polu. Kierunek na wprost membrany jest zawsze uprzywilejowany. Aby mikrofon wprowadzał możliwie najmiejsze zaburzenie pola akustycznego robi się je możliwie jak najmniejsze.

Mikrofony dwukierunkowe, o charakterystyce dwustronnej, to inaczej **dwubiegunowe (ósemkowe).** W tym przypadku fala akustyczna jest odbierana z obu stron i wypadkowa siła jest różnicą ciśnień. Elementem odbierającym może być jedna membrana lub dwie odsunięte od siebie na pewną odległość (odbierające sygnał od zewnętrznej strony ich powierzchni) lub wstęga odsłonięta z obu stron- czyli jak w przypadku mikrofonu wstęgowego, który zawsze ma charakterystykę ósemkową. Mikrofony te nazywa się też gradientowymi, w związku z tym, że sygnał, który ostatecznie otrzymujemy, jest różnicą ciśnień między dwoma punktami fali (tzn. gradientem fali). Gdy dźwięk jest na tyle wysoki by długość fali była porównywalna z wielkością mikrofonu, to ruch membrany jest porównywalny z prędkością cząstki. W związku z tym mówi się także o takich mikrofonach „prędkościowe". Określenie „ósemkowy" odnosi się do kształtu charakterystyki, co widać na rys. 3.3.

Specyficznym połączeniem tych mikrofonów jest mikrofon **jednokierunkowy (kardioidalny)**. Będzie to zatem mikrofon ciśnieniowo- gradientowy. Szeregowe połączenie mikrofonu wszechkierunkowego z ósemkowym, o takich samych skutecznościach, da nam charakterystykę nerkową- jak na rysunku. Zmieniając stosunek tych skuteczności wpływamy na szerokość i kształt „pola widzenia" mikrofonu- między kołową, a ósemkową charakterystyką. Jak widać, najlepiej odbierany jest dźwięk z kąta 0^o , a na sygnał z kąta 180^o nie jest odbierany. Kształt charakterystyki sugeruje nazwę- kardioida. Zastosowanie dodatkowych ustrojów, takich jak reflektory akustyczne czy rury z bocznymi otworami, które kierują falę akustyczną wprost na przetwornik daje charakterystykę ostro jednokierunkową, w kształcie wydłużonej połówki ósemki. Są to mikrofony **superkierunkowe**, stosowane do nagrań, przy których bardzo nam zależy na wyeliminowaniu echa, pogłosu i niepożądanych dźwięków.

Rys. 3.3 Charakterystyki kierunkowe [27]

Kolejnym parametrem jest **czułość (skuteczność) mikrofonu**. Jest to zdolność przetwarzania sygnału akustycznego na napięcie elektryczne. Można podać ją na dwa sposoby. Wedle normy IE 268-4 czułość mikrofonu powinna być podawana w miliwoltach wygenerowanych na każdy paskal ciśnienia (dla częstotliwości 1 kHz; dla mikrofonów pomiarowych- 250 Hz). Zgodnie z amerykańską tradycją podalibyśmy czułość w decybelach w stosunku do 1 V/Pa- co daje nam wartość ujemną. Parametr czułości mikrofonu mówi nam o tym jaki będzie stosunek napięcia na wyjściu mikrofonu do padającego ciśnienia. Jakie napięcie pojawi się na zaciskach mikrofonu jeśli padająca fala wywrze ciśnienie 1 Pa (tzn. około 94 dB). Zwykle czułość mikrofonów dynamicznych waha się pomiędzy 0,7-3,5 mV/Pa. Mikrofony pojemnościowe osiągają dziesięciokrotnie wyższą czułość, zwykle w zakresie 7-20 mV/Pa. Mikrofon o wysokiej czułości będzie wymagał mniejszego wzmocnienia, ponieważ daje mocny sygnał o dużym napięciu. Zastosujemy go również, do stosunkowo cichych nagrań, co pozwoli zachować większy odstęp od szumów własnych- z czym wiąże się następny parametr.

Odstęp sygnału użytecznego od szumu, czy też **szumy własne** (equivalent noise level) to informacja o dolnej granicy zakresu mikrofonu. Wynika ona z szumów, zakłóceń elektromagnetycznych, powstających przy przepływie prądu przez urządzenie. Dlatego dotyczy to szczególnie mikrofonów ze aktywnych (wymagających doprowadzenia zasilania) czyli pojemnościowych czy wstęgowych. To także jest szczególnie istotne przy cichych nagraniach, kiedy istnieje ryzyko, że poziom szumów przewyższy poziom sygnału użytecznego. Istnieją standardy w jakich podaje się ów parametr: skala *dB(A) ważona* i skala *CCIR 268-1.* Obie wykorzystują pewien sposób ważenia, a w skali dB(A) ustosunkowuje się go także czułości ludzkiego słuchu.

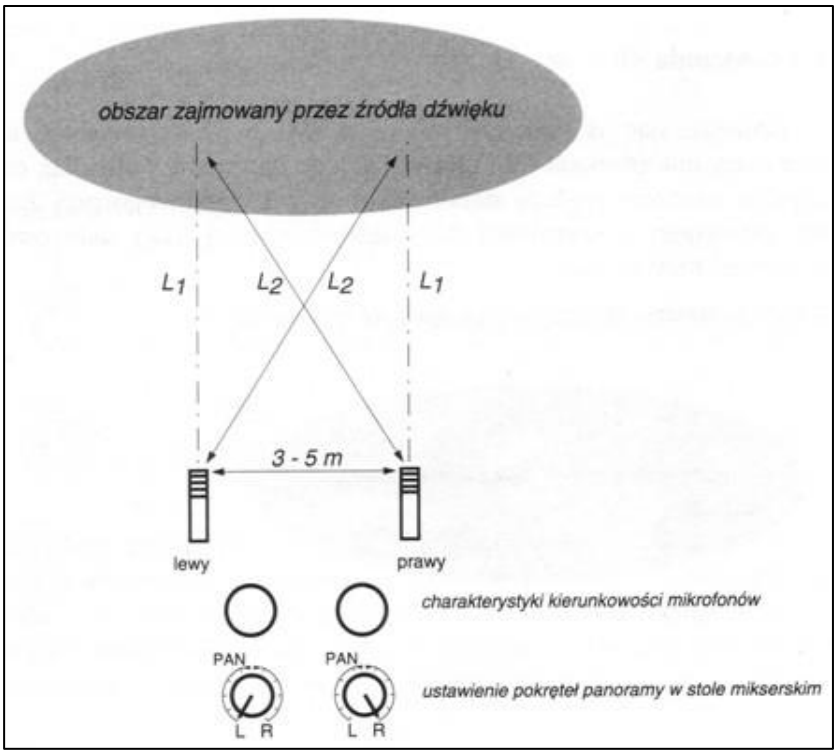
Ciśnienie graniczne to inaczej odporność mikrofonu na ciśnienie dźwięku. Innymi słowy jest to maksymalny poziom ciśnienia dźwięku (SPL- Sound Pressure Level) jaki zniesie mikrofon. Zwykle przy nagraniach muzycznych ogólnie przyjęty średni poziom zostaje bez trudu znacznie przekroczony, dlatego podaje się informację jaka jest górna granica tolerancji mikrofonu. Na pewnym poziomie,

w nagraniu pojawią się zniekształcenia harmoniczne (THD- Total Harmonic Distortion) i przesterowania, co skutkuje obcięciem szczytów fali i rejestracji jej w formie fali kwadratowej. Przyjmuje się, że poziom zniekształceń, które można zmierzyć (ale jeszcze ich nie słyszymy) to 0,5-1%. Podawana wartość THD powinna odnosić się do całości mikrofonu- zarówno kapsuły jak i przedwzmacniacza. [5,6,11,22]

3.3 Techniki nagraniowe

Zmierzając w kierunku dźwięku przestrzennego, owe mikrofony należy odpowiednio ze sobą zestawić. Wykorzystanie kilku mikrofonów nie wystarczy do uprzestrzennienia nagrania. Potrzebne są osobne tory akustyczne (kanały), aby do każdego głośnika doprowadzić inny sygnał. Pierwszym etapem jest dwa tory czyli stereofonia. W tym celu na etapie nagrania wykorzystuje się dwa mikrofony i dwa niezależne kanały. Wyróżnia się dwie podstawowe techniki rejestracji dźwięku stereofonicznego: AB oraz XY.

System AB powstał na podstawie doświadczeń specjalistów z radia francuskiego. Wykorzystane w nim dwa mikrofony ustawione są równolegle, zwykle w odległości około 1 m od siebie. Obraz stereofoniczny uzyskuje się przez różnice fazowe i czasowe między nimi. Jednak ta metoda jest obarczona dużą niezgodnością w fazach kanałów. Skutkuje to znoszeniem się sygnałów w przeciwnych fazach i powstawaniem „dziury" w środku obrazu stereofonicznego. Natomiast próba odtworzenia nagrań na urządzeniach monofonicznych, wymagająca sumowania sygnałów, skutkuje zanikiem części pasma. Powoduje to niekompatybilność nagrań z urządzeniami monofonicznymi takimi jak radioodbiorniki czy starsze telewizory. Dlatego system AB, nazywany też *fazowym*, jest rzadziej wykorzystywany.

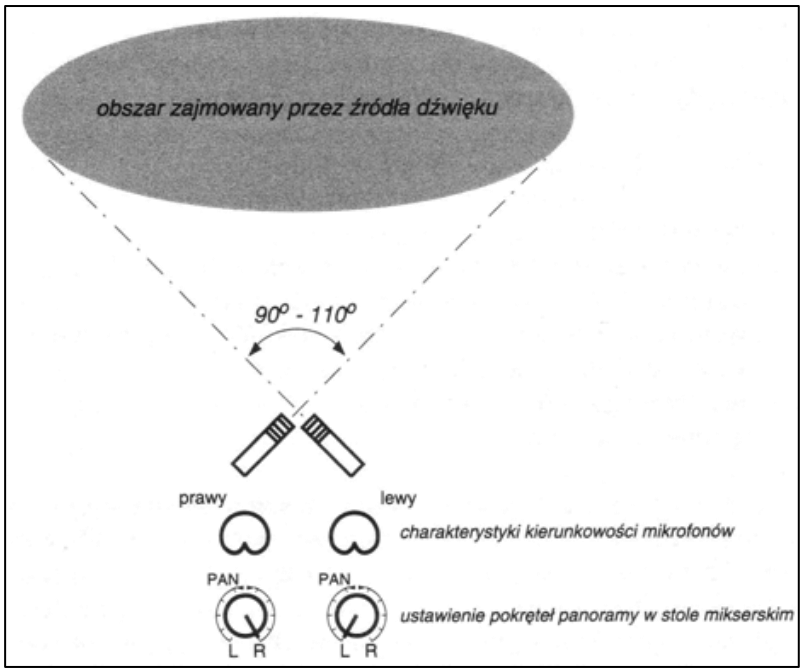


Rys 3.4 Technika nagrywania AB [28]

System XY jest rozwiązaniem pozbawionym tych wad. Również są to dwa mikrofony, jednak ustawione jeden nad drugim w małej odległości, oba o tej samej charakterystyce- nerkowej lub ósemkowej. Ich osie największej skuteczności powinny być rozchylone o 45 stopni w prawo i w lewo w stosunku do źródła dźwięku, a kąt między ich osiami wzajemnie nie mniejszy niż 90 stopni. W takim systemie różnice czasowe i fazowe nie odgrywają specjalnej roli. Istotna jest tutaj różnica w natężeniu sygnału lewego i prawego kanału, toteż nazywa się go też systemem *natężeniowym*.

Systemem wyrastającym z układu XY jest **system koincydencyjny**. Z wykorzystanych w tej metodzie mikrofonów, jeden powinien mieć charakterystykę kardioidalną, a drugi ósemkową. Kardioida jest ustawiona wprost na źródło dźwięku, natomiast mikrofon ósemkowy w poprzek. W ten sposób pierwszy odbiera sygnał bezpośredni, a drugi sumę sygnałów dochodzących z boków. Takie rozwiązanie jest zawarte już w jednym mikrofonie z odpowiednimi wkładkami

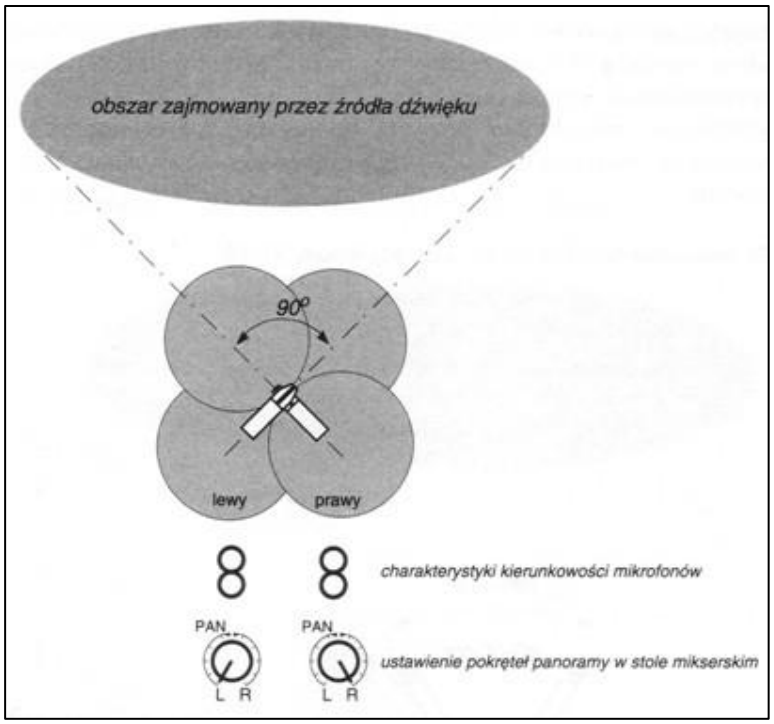
o danych charakterystykach. Sygnał z takiego mikrofonu koincydencyjnego odpowiada sygnałom MS.



Rys. 3.5 Technika nagrywania XY [28]

Obecnie, w praktyce nie stosuje się pojedynczej pary mikrofonów, tworzy się nagrania wielomikrofonowe. Wykorzystuje się w nich mikrofony stereofoniczne, monofoniczne (tzw. *podpórki*), a do nadania kierunku również panoramy. Nazywa się to stereofonią *pan-pot*. Kompatybilność z odbiornikami monofonicznymi nie jest już potrzebna. Stereofonia zawładnęła studiami nagrań i jest realizowana w radiu i telewizji. Jest to już podstawowy format w powszechnym użytku.

Następną techniką wyrastającą z XY (i również koincydencyjną) jest **Technika Blumleina**. Wykorzystuje mikrofony z charakterystyką ósemkową, najczęściej wstęgowe, w ustawieniu pod kątem 90 stopni i zgodnie skierowaną polaryzacją. Dzięki ósemkowej charakterystyce, mikrofony zbierają więcej dźwięków odbitych od ścian pomieszczenia, przez co osiągamy wrażenie lepszej przestrzenności. Na początku rozwoju stereofonii, w latach 30. XX wieku była to bardzo popularny sposób mikrofonowania.



Rys. 3.6 Technika nagrywania Blumleina [28]

Radio France w 1960 roku opracowało również technikę łączącą zalety AB i koincydencyjnej. **Technika ORTF** (Office de Radiodiffusion Television Francaise), określa małą odległość między kapsułami mikrofonów, co nawiązuje do angielskiej nazwy *near-coincident pair*. Wykorzystuje zarówno pomiar różnicy czasowej jak i natężeniowej. Są to dwa mikrofony o kardioidalnej charakterystyce, rozstawione pod kątem 110 stopni, oddalonych od siebie o około 17 cm (mniej więcej jest to rozstaw uszu człowieka). Takie ustawienie daje tą samą wartość ITD (Interaural Time Difference), więc lepiej sprawdza się przy odsłuchu słuchawkowym i eliminuje przesłuchy.

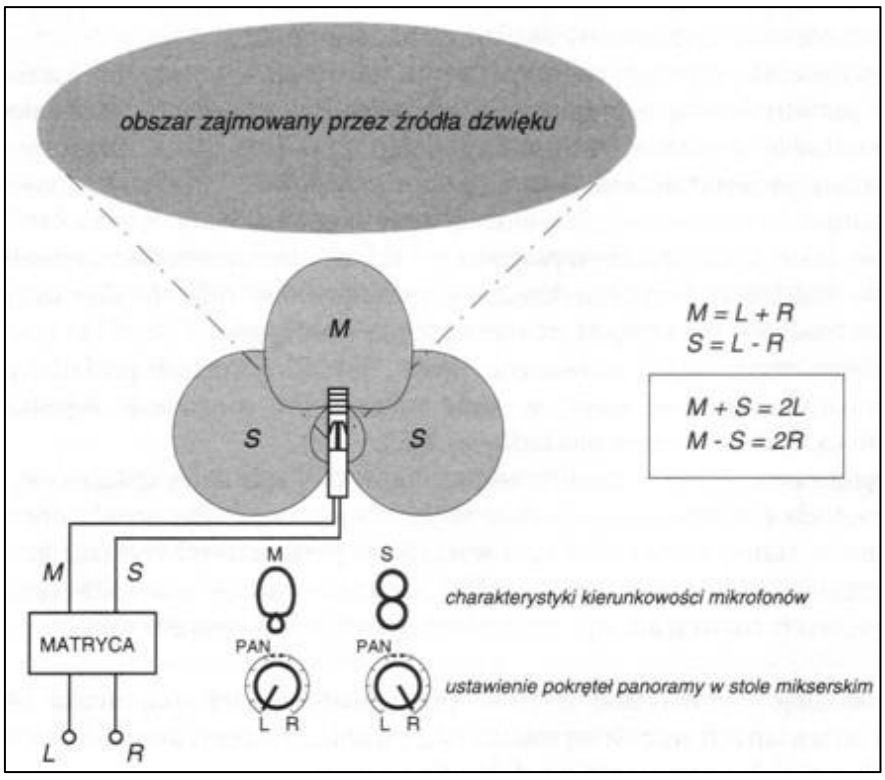
Odmienne rozwiązanie, zaproponowane przez niemieckie radio w 1957 roku to **system MS**. Jest to zestawienie mikrofonu wszechkierunkowego (później także kardioidalnego) z ósemkowym, koincydencyjnie. Mikrofon dookólny zbiera sygnał M (middle) ustalony jako monofoniczny, a mikrofon ósemkowy zbiera sygnały boczne S (side) i tym samym jest informacją o kierunkowości. Największą zaletą jest

wzajemna wymienność systemów MS i XY. W zależności od potrzeb możemy uzyskać odpowiedni sygnał w oparciu o proste wzory:

$$X = \frac{\mathrm{M} + \mathrm{S}}{2}$$

$$Y = \frac{\mathrm{M} - \mathrm{S}}{2}$$

Zamianę sygnału może wykonywać podstawowy dekoder wtyczki VST (Visual Studio Technology) zawarty w standardzie DAW, lub można to zrobić samodzielnie. Głównym powodem wykorzystania MS jest umożliwienie odtwarzania nagrań w systemie monofonicznym np. w radiu. W takim przypadku, w odbiorniku brany jest pod uwagę sygnał M, a sygnał S jest odrzucany. Natomiast w odbiorniku stereofonicznym zachodzi konwersja MS na XY, według powyższych wzorów i sygnał jest odtwarzany na dwóch kanałach.



Rys. 3.7 Technika nagrywania MS [28]

W technice **Double MS** użyte są dwie pary mikrofonów MS (ale ponieważ dla sygnału S odbiorniki się pokrywają), w praktyce- dwa mikrofony kardioidalne i jeden ósemkowy zbierający sygnał S. Kardioidy są skierowane w przeciwnych kierunkach: M_{front} i M_{rear} . Taki sposób rejestracji nie jest już techniką stereofoniczną. Dostajemy trzy kanały, które możemy przekodować na przykład do systemu 5.1 zgodnie ze wzorami:

$$L = M_{front} + S$$

$$R = M_{front} - S$$

$$C = M_{front}$$

$$Ls = M_{rear} + S$$

$$Rs = M_{rear} - S$$

Dzięki temu dostajemy obraz dźwiękowy w osi azymutalnej, nie tylko wyliczony przez system z nagrania stereo, ale zarejestrowany w szerszym planie.

Dalszy rozwój wymusił kolejne rozwiązania, bardziej rozbudowane połączenia. Powstała zupełnie nowa, odmienna technika nagrywania- **ambisonia**. Procedura została opracowana na przełomie lat 60. i 70. XX wieku, między innymi przez Michaela Gerzona. Ideą jest rejestracja i reprodukcja pełnego pola dźwiękowego- *sound field*. Jest to zastosowanie koincydencyjnej metody rejestracji Blumleina w trzech wymiarach. W wersji podstawowej- Ambisonii pierwszego rzędu, dźwięk jest rejestrowany w formie A-formatu, a następnie kodowany do czterech kanałów: W, X, Y, Z- tzw. B-formatu. Kanał W odpowiada sygnałowi z mikrofonu wszechkierunkowego. Kanały X, Y i Z odpowiadają sygnałom z mikrofonów ósemkowych skierowanych odpowiednio: w przód, w lewo i w górę. Dodanie kanału Z umożliwia oddanie pełnej trójwymiarowej przestrzeni dźwiękowej (w odróżnieniu od innych systemów).

Sztandarowym mikrofonem, opatentowanym przez Michaela Gerzona jest Soundfield- mikrofon ambisoniczny, rejestrujący A–format. Tym samym stworzony został standard konstrukcji mikrofonów tego typu. Są to cztery przetworniki

umieszczone na planie czworościanu foremnego. Aby nie pojawił się problem przesunięć fazowych, wszystkie cztery kapsuły powinny znajdować się w tym samym punkcie. Jedna z kapsuł skierowana jest przód-tył (sygnał X), druga prawo-lewo (sygnał Y), trzecia góra dół (sygnał Z), a wszechkierunkowa kapsuła (sygnał W) daje informacje o ogólnej amplitudzie docierającej do mikrofonów kierunkowych. Nagrania ambisoniczne nie mają problemów z fazą. Można je przetransponować na dowolną liczbę kanałów lub nawet wirtualnych mikrofonów. Nagranie musi jednak być wykonane bardzo precyzyjnie. Jeśli mikrofon choćby minimalnie zmieni swoją pozycję to obraz ulegnie przesunięciu, co ma silny negatywny wpływ na nagranie, dużo większy niż w nagraniu stereo.

Rys. 3.8 Mikrofon Soundfield

3.4 Nagrania binauralne

Odwrotnym podejściem jest tworzenie nagrań do systemów binauralnych. Eksperymenty rozpoczęto w berlińskim radiu w latach 70-tych XX w. System miał być przydatny w subiektywnych ocenach akustyki pomieszczeń i działać jako metoda ich porównywania. Jednym z rozwiązań jest **system sztucznej głowy**- metoda pomiarów i nagrań oparta na koncepcie Shunke'a. Podstawą jest umieszczenie dwóch mikrofonów dookólnych w odległości 15-20 cm (na szerokość

ludzkich uszu) i próba oceny jak w naturalny sposób dźwięk zostałby odebrany przez słuchacza. Dla dokładniejszego pomiaru stosuje się sztuczną głowę, która ma za zadanie odbieranie dźwięku tak jak zrobiłoby to ludzkie ucho, uwzględniając odbicia od ciała. Z materiału reagującego na dźwięk podobnie jak ludzkie ciało, tworzy się manekina (zwykle od pasa w górę). Uszy są odlewem konkretnej małżowiny lub mają uśredniony kształt. Mikrofony umieszcza się wewnątrz ucha. W takich nagraniach wyjątkowo dobrze oddane są wrażenia przestrzenne w osi pionowej. Można także zastosować technikę nagrań binauralnych, w których nie wykorzystujących sztucznej głowy, a mikrofony douszne (podobnych do słuchawek typu „pchełki") zakładane przez nagrywającego. Tego typu mikrofony tworzy między innymi firma Soundman. Twórcami sztucznych głów (ang. *dummy head*) są przede wszystkim firmy Neumann i Head Acoustics.

Rys. 3.9 Sztuczna głowa firmy Neumann

Otrzymana z nagrania funkcja HRTF zależy od głowy manekina lub nagrywającego. Nawet nagrania na uśrednionym manekinie mogą nie dawać poprawnego efektu, jeśli budowa ciała słuchacza będzie się znacznie różniła. Może powodować to efekt „dźwięku wewnątrz głowy" lub rozmijania się wirtualnego źródła z wizualnym odbiorem. Zwykle jednak, dzięki uśrednionym HRTF-om dostajemy zbliżony odbiór przez większość słuchaczy.

Rozdział 4: Reprodukcja dźwięku przestrzennego

4.1 Stereofonia

Stereofonia zaistniała historycznie w roku 1881 w Paryżu, gdzie na wystawie elektryczności Clement Ader zaprezentował audycję koncertu z opery paryskiej wykorzystując układ dwóch niezależnych torów fonicznych. Zastosowano oddzielne mikrofony i głośniki, sygnał był przekazywany przez dwie linie telefoniczne. Była to tylko namiastka, daleka od ideału, ale zainspirowała do dalszych badań. Użytkowe systemy stereofoniczne pojawiły się w latach dwudziestych XX wieku- uzyskano pierwszy patent na urządzenie tego typu. Podjęto również próbę transmisji- w 1993 roku laboratorium Bella dokonało transmisji próbnej z Filadelfii do Waszyngtonu. Transmisja była trójkanałowa, a jej jakość (jak na tamte możliwości) bardzo wysoka. W tym czasie, w Wielkiej Brytanii, opatentowano też stereofoniczną płytę gramofonową.

Pierwszą spektakularną prezentację systemów stereo można było przeżyć w kinie. Do tej pory wykorzystywano jeden głośnik umieszony zazwyczaj centralnie za ekranem. Wprowadzenie systemu stereo dawało oszałamiające efekty relacji dźwięku z obrazem. Dwa głośniki umieszczano w równej odległości wokół osi symetrii. Dzięki regulacji natężenia sygnałów podawanych na oba głośniki, zmiany proporcji między nimi, mamy możliwość określenia położenia pozornego źródła dźwięku. Jeśli proporcje zmienią się płynnie, czyli dźwięk z prawego głośnika będzie coraz cichszy, a z lewej strony coraz głośniejszy, to doznamy efektu, że źródło przemieszcza się z prawej do lewej. Taki obraz malowany dźwiękiem nazwano „sceną obrazu pozornego”. Obecnie specyfika dźwięku stereo może zostać podzielona na *true stereo* oraz *pan-pot stereo*. W **true stereo** dźwięk jest odtwarzany tak jak został nagrany- w stereofonicznej technice mikrofonowej, z naturalnym pogłosem pomieszczenia. Technika **pan-pot stereo** to dźwięk zarejestrowany monofonicznie i rozmieszczony w przestrzeni audio przez panoramowanie i dodanie efektów pogłosowych w programie DAW.

Obraz stereo daje nam możliwość poczucia przestrzeni, ale jest to jedynie płaszczyzna przed nami. Nie oddamy w ten sposób głębi i nie odtworzymy pola akustycznego, np. sali koncertowej, gdzie rolę grają również złożone zjawiska jak chociażby odbicia od ścian. Pomysł zastosowania więcej niż dwóch kanałów pojawił się w 1952 roku, również w kinie. System filmowy Cinema-Scope zawierał trzy kanały i jeden dodatkowy do efektów specjalnych. Pierwszym podejściem były układy kwadrofoniczne.

Kwadrofonia wykorzystuje cztery głośniki umieszone w równej odległości dookoła słuchacza. Największą trudnością jest tutaj rozdzielenie sygnałów. Przy ówczesnych możliwościach jakość i ilość danych możliwych do zapisania była za słaba. Nagrania wykonane metodą MS było przekształcane na kanały X i Y i podawane na przednie głośniki. Tylne głośniki były zasilane tylko sygnałem przekształconym, wyodrębnionym sygnałem S podawanym w przeciwfazie. System dopracowany przez Petera Scheibera w 1967 roku nie tylko wykorzystywał cztery głośniki w odpowiednim ustawieniu, ale też kodował 4 kanały audio wykorzystując 2 przesyłowe. Sygnał był dekodowany i odtwarzany na czterech głośnikach. Dzięki temu funkcjonujące ówcześnie narzędzi systemu stereo można było użyć do transmisji kwadrofonii. Takie kodowanie większej ilości kanałów do stereo nazwano **systemem matrycowym**. Od tej pory był on powszechnie używany.

Na tamtym etapie, odtworzenie źródeł pozornych po bokach lub za słuchaczem nie było możliwe. Dodatkowo, dla odpowiedniego efektu, słuchacz musiał znajdować się w ściśle określonym miejscu. Do dalszego rozwoju potrzebna była zatem większa ilość nie tylko głośników, ale i niezależnych kanałów. Układy kompresji danych (o których więcej w rozdziale *Formaty dźwięku*) umożliwiły opracowanie technologii nośników mieszczących wiele danych i zapisanie ich w systemie wielokanałowym.

4.2 Systemy Dolby

W tym miejscu historii na scenę wkracza laboratorium Dolby Inc., przełomowy, wielki gracz w świecie audio. W roku 1950 w jego laboratorium powstaje system

Dolby Surround. Jest to system transmisji dźwięku, który stał się standardem w technice kinowej, a chwilę później płynnie wszedł również na szerszy rynek jako „kino domowe". Widowiska kinowe weszły w nową erę, dzięki jeszcze lepszemu, wierniejszemu połączeniu dźwięku z obrazem. Początkowo systemy surround składały się nadal z czterech głośników, jednak niezależność kanałów dawała możliwość zaskoczenia widza spektakularnymi efektami kroków zbliżających się z tyłu czy samolotu przelatującego w poprzek sali. Otwierało to nowe drzwi dla twórców filmowych. Efekty specjalne stały się nową, kluczową niemalże, formą wyrazu artystycznego i uzyskały nową płaszczyznę realizacji.

Od tamtej pory systemy otaczania dźwiękiem, z nurtem zawrotnie szybko udoskonalanej technologii, rozwijają się w niesamowitym tempie. System surround stał się standardem, realizowanym przez wielu producentów na wiele sposobów. Poza możliwościami rozwoju przy okazji techniki kinowej rozwijało się również „kino domowe", które technologicznie wiele się od prawdziwego kina nie różniło, ale stworzyło koleją gałąź rynku muzycznego oferującego nagrania typu surround, nową jakość muzyki w odsłuchu domowym. To domknęło samo nakręcające się koło świata audio.

W pierwszym systemie surround od Dolby Inc. transmisja wciąż była dwukanałowa. W większości systemów Dolby stosuje się system matrycowy. Jednak to już ukazywało ogromne możliwości tego systemu. Kanały główne (R, L) są zasilane przez bezpośrednio przeniesiony sygnał, w którym można regulować balans i poziom wyjściowy. Daje to możliwość dostosowania zestawu do warunków pomieszczenia i preferencji słuchacza. Na głośniki surround (S) sygnał jest wydzielany na samym początku i przekazywany do układu różnicowego. Następnie przechodzi przez linię opóźniającą i odpowiednie filtry (antyaliasingowy i dolnoprzepustowy), po czym trafia z powrotem „na linię" do regulacji poziomu. Do umieszczenia źródła pozornego w obrazie nadal wykorzystuje się tylko głośniki przednie i stosunek sygnału między nimi. W związku z tym, poprawny odbiór obrazu pozornego wymagał umiejscowienia słuchacza dokładnie w osi zestawu oraz dość bliskiego umieszczenia głośników. Pojawia się tutaj pojęcie separacji- udział

sygnału z innych kanałów w mierzonym kanale, czyli tzw. przesłuchy. W najprostszy sposób można ją mierzyć podając sygnał na jeden z kanałów i mierząc wartość sygnału w pozostałych. Ponieważ sygnał surround wykorzystuje tylko różnicę kanałów wejściowych i ich przekształcenia to jego separacja, pomimo rozbudowy filtrów i kompresorów, wypadła dość słabo.

Tak powstał jego następca- **Dolby Surround Pro Logic**. Ulepszenia wprowadzone przez nową wersję systemu surround poskutkowały szybkim jego rozpowszechnieniem i powstaniem wielu nagrań muzycznych i kinowych przeznaczonych na dekodery Dolby Pro Logic. System kina domowego i kinowy różnił się tutaj już w zasadzie tylko ilością głośników, dostosowaną do wielkości sali. Największą rewolucją w stosunku do poprzednika było wykorzystanie aktywnego dekodera i dodatkowego kanału. Dolby Surround z dekoderem pasywnym przekształcał jedynie sygnał w ustalony sposób. W Dolby Pro Logic dekoder aktywny zmienia swoje parametry w zależności od sygnału wejściowego, a dodatkowy, niezależny kanał jest dedykowany nowemu głośnikowi centralnemu.

Zastosowano tutaj dodatkowe układy dostrajania, generator szumu oraz macierz adaptacyjną. W ten sposób z dwóch ścieżek uzyskujemy cztery kanały wyjściowe R, L, C, S. Wewnątrz macierzy adaptacyjnej następuje rozdzielenie sygnałów w układzie uwydatniania kierunkowego, tak aby uzyskać jak najlepszą separację i najdokładniejsze wyznaczenie kierunku dźwięku dominującego. Tak wydzielone sygnały znowu wracają „na linię" do regulacji poziomu i balansu.

Dodatkowy głośnik centralny daje przede wszystkim ściślejsze związanie dialogów filmowych ze środkiem sceny i łatwiejsze w odbiorze, szersze „centrum". Nie ma już potrzeby aby źródła pozorne umiejscowione centralnie realizować przez interpolację sygnału prawego i lewego, ani tego żeby widz siedział idealnie na osi. Teraz odpowiada za do odrębny kanał. Ponadto dekoder aktywny lepiej separuje kanał surround. Co prawda spada nieco poziom separacji między kanałami R i L, ale w znacznym stopniu podnosi się jakość wydzielenia kanału S, zdecydowanie zmniejsza poziom przesłuchów.

W swojej następnej odsłonie Dolby wyszedł naprzeciw audiofilom, w trosce o ich dotychczasowe zbiory. System **Dolby Pro Logic II** skupia się przede wszystkim na tym jak odtworzyć w formie dźwięku dookólnego nagrania stereofoniczne, będące wciąż w zdecydowanej większości. Nowy dekoder umożliwiał zaprogramowanie pola akustycznego na podstawie nagrania stereofonicznego.

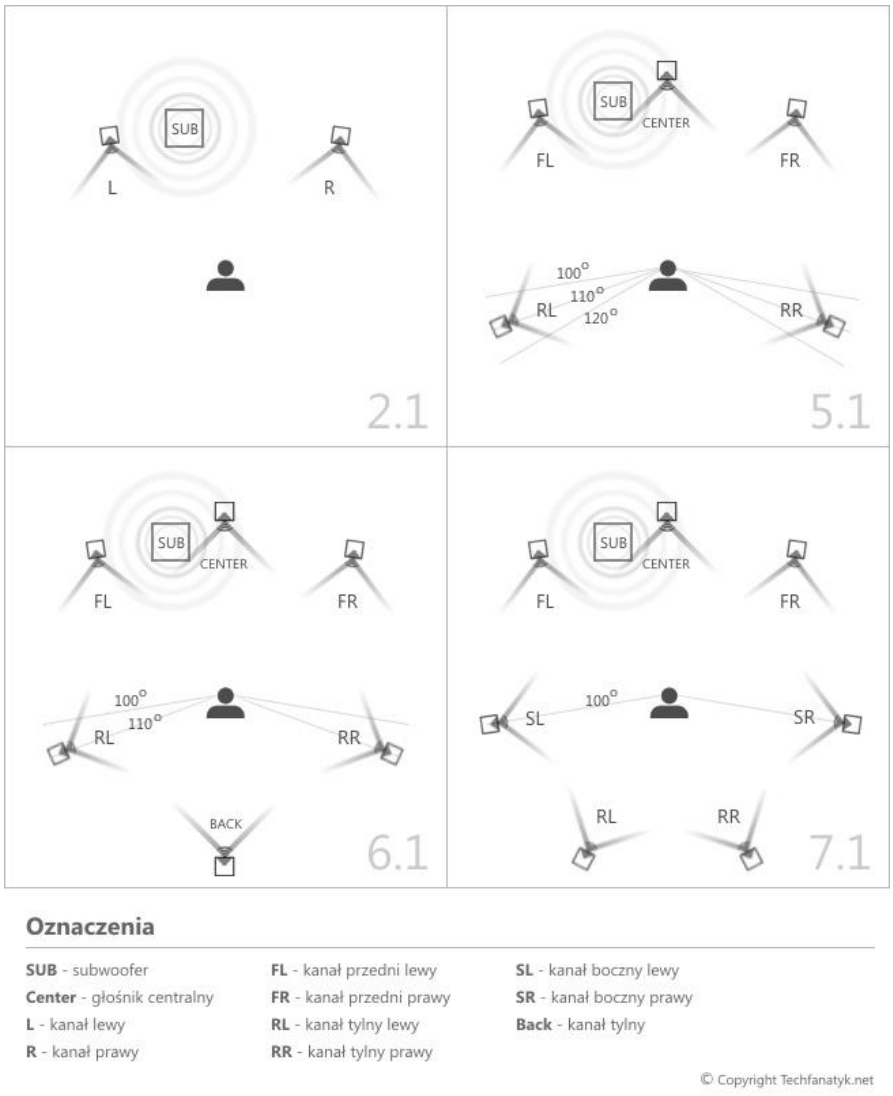
W obszarze zainteresowań twórców znalazły się głównie analogowe, stereofoniczne układy odtwarzania dźwięku czyli odtwarzacze klasycznych płyt LP czy CD, komputery czy domowe układy multimedialne. Poza tym osiągnięto jeszcze lepsze rezultaty w prawidłowej lokalizacji kierunkowości dominującego dźwięku oraz stabilne pole stereofoniczne między kanałami głównymi- R i L. W nowej wersji systemu wprowadzono dodatkowe funkcje:

-kontrola wymiarów- precyzyjna regulacja punktu centralnego na osi (przód-tył) daje możliwość dostosowania systemu do punktu odsłuchowego.

-kontrola szerokości kanału centralnego- umożliwia rozłożenia sygnału z kanału centralnego na lewy i prawa, co umożliwia jeszcze dokładniejszą regulację do warunków odsłuchowych oraz umieszczenie w stereofonii efektów dźwiękowych i dialogów.

-tryb panoramiczny- ta opcja w połączeniu z trybem „stereo” wykorzystuje głośniki surround, poszerzając tym samym pole odsłuchu.

Zatem Dolby Pro Logic postarał się o kompatybilność wstecz, dzięki algorytmom uprzestrzeniającym odświeżył odsłuch nagrań stereo oraz poprawił jakość samego odsłuchu przestrzennego. W międzyczasie pojawiły się konkurencyjne systemy takie jak DTS Digital Surround czy Circle Surround od RSP Technologies przeznaczone do wykorzystania w domu. O ile Circle Sourround nadal pracował na dźwięku analogowym, to DTS weszło już w erę cyfrową.



Rys 4.1 Układy głośników w systemach Dolby [29]

Chwilę później system **Cinema Digital Sound** wkroczył do kin z pierwszym cyfrowym, komercyjnym formatem i filmem „Dick Tracy". Konfiguracja głośników była typowa, ale dźwięk zapisywano na gruboziarnistej taśmie filmowej 35mm lub 70mm. Format CDS został opracowany we współpracy firm Kodak i Optical Radiation Corporation. Wykorzystywał kodowanie PCM z wykorzystaniem Delta Modulation. Niestety, cyfrowy zapis zastąpił optyczną ścieżkę analogową 35mm i magnetyczny zapis 70mm, co oznaczało brak zapasowej ścieżki audio. Ponieważ technologia LED i CCD nie była wtedy tak biegła, a do odczytania było milion pikseli na sekundę, awarie zdarzały się często, więc na wypadek zniszczenia ścieżki

dźwiękowej operatorzy kinowi musieli mieć dwie kopie filmu. A kopiowanie również było uciążliwe przy ówczesnej technologii. Brak zapasowej ścieżki audio był przeważającą wadą systemu. Film „Terminator 2" z 1991 roku był szczytem kariery CDS. Problemy z zawodnością systemu uwieńczyła przedwczesna premiera nowego systemu od Dolby. Po dwóch latach CDS wyszło z użycia. [11,17,18]

Tutaj laboratorium Dolby, z cennym doświadczeniem kosztownych pomyłek CDS, przedstawił nam „współczesność". **Dolby Digital (5.1)**- sześć niezależnych kanałów, nowy format kompresji, nowy nośnik, nowy standard telewizji cyfrowej. To wszystko w nowej odsłonie zestawu odsłuchowego, którego premiera kinowa miała miejsce w roku 1992, a dla kina domowego był to rok 1994. Forma Zoran Corporation wspomogła twórców produkując pierwszy układ scalony realizujący cały proces dekodowania. System łączył w sobie nowy standard z poprzednim (Pro Logic), dzięki czemu zebrał jeszcze większą popularność, również wśród kolekcjonerów poprzednich formatów. Początkowo system nazwano Dolby AC-3 od algorytmu, który zastosowano do kompresji danych cyfrowych. Potrzeba opracowania systemu kompresji wynikła z pomysłu umieszczenia ścieżki dźwiękowej pomiędzy perforacją taśmy filmowej i umieszczeniu na niej sześciu niezależnych kanałów. Tak AC-3 weszło na rynek, a okazawszy się bardzo uniwersalnym systemem, łatwym do adaptacji dla różnych zastosowań, wybrano go na standard na płytach DVD oraz wielokanałowego dźwięku w telewizji cyfrowej w USA. Więcej o tym formacie w rozdziale *Formaty dźwięku*. Wspomnianych sześć niezależnych kanałów w Dolby Digital dawało pełną separację, a co za tym idzie lepszą przestrzenność, precyzyjną lokalizację źródła i większy realizm. Daje to również jeszcze lepsze efekty rozbudowanych w tym systemie kanałów surround. Poza kanałami L, R i C kanał S został rozłożony na lewy i prawy- Sl, Sr oraz dodano kanał efektów niskoczęstotliwościowych LFE (Low Frequency Effects). Zakres częstotliwości realizowanych przez pierwszych pięć kanałów ma obejmować pełne pasmo tzn. od 3 Hz do 20 kHz. Kanał LFE skupia się na zakresie 3 Hz – 20 Hz czyli głownie na zwiększeniu ekspresji dynamicznych scen jak wybuchy, zderzenia itp.

Do współpracy dołączyła firma Yamaha, produkując zestawy kina domowego przeznaczone na Dolby Digital, wraz z nowymi procesorami opracowującymi dodatkowo dźwięk przesyłany do głośników surround. Efektem było jeszcze ściślejsze otoczenie dźwiękiem widza, dokładniejsze odtworzenie atmosfery. Sprzęt wyposażony w to ulepszenie nazwano „Tri-Field CINEMA DSP" i umożliwiał zaprogramowanie go na kilkanaście sposobów, tak aby można było wybrać najbardziej odpowiadający odbiorcy. System rozbudowywano o kolejne głośniki i głośno reklamowano nowe układy 7.1, 9.1 i kolejne. Pojawił się przesyt. Okazało się, że nawet w kinach niektóre głośniki są niemalże nieużywane. Rozbudowywanie systemu o kolejne przetworniki w osi azymutalnej nie dawało lepszych efektów. Dolby wystartował z nowym projektem. [18,4]

Dolby Atmos pojawił się w 2012 roku jako opracowany od początku w nowej technice system. Podstawą było skupienie się na większej przestrzenności dźwięku przez dodanie głośników sufitowych, wprowadzających w końcu nowy efekt, jakże użyteczny w kinie. Ponadto montaż systemu przewidywał użycie mikrofonów kalibracyjnych do prawidłowego rozmieszczenia głośników względem słuchacza i dostosowania np. barwy dźwięku do pomieszczenia. System dobrze się przyjął i szybko pojawił się w wersji kina domowego. W wersji kinowej przewidziane są 34 kanały (w tym 10 na suficie), a dla kina domowego jest to 7-11 głośników. Wykorzystana została technologia dźwięku obiektowego. Procesor zaprzężony jest do obliczania pożądanej lokalizacji dźwięku i rozkłada go na dostępne głośniki. Dzięki temu filmy z Dolby Atmos z dźwiękiem w dobrej jakości zmieściły się na płycie Blu-ray. [28,29, 18]

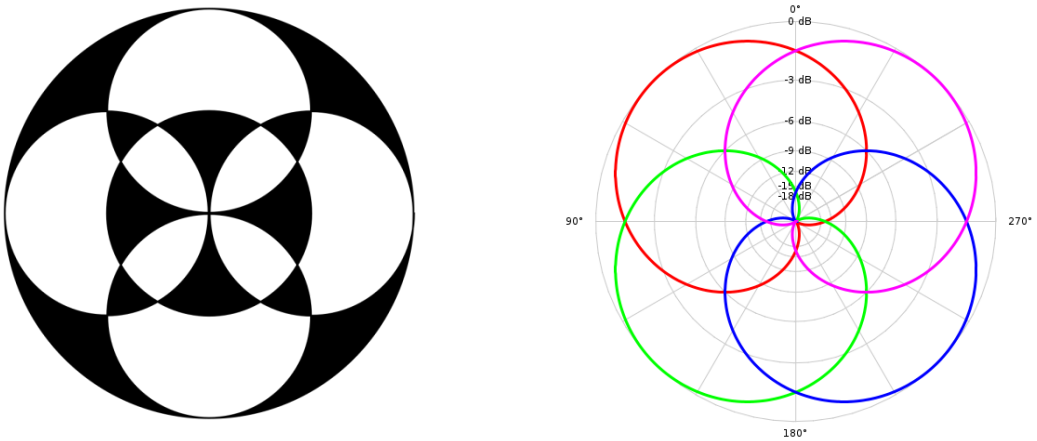
Obecnie system stereofoniczny jest już domyślnym sposobem odsłuchu i stał się na tyle oczywistym rozwiązaniem, że nie często nie zwraca się nawet uwagi na właściwe ustawienie głośników. Zgodnie z normą ustaloną przez European Broadcasting Union, ustanowioną w 1998 roku, głośniki powinny tworzyć w przybliżeniu trójkąt równoboczny z punktem odsłuchu i znajdować się na tej samej wysokości.

Aby właściwie odwzorować przestrzeń akustyczną, potrzebne jest więc odpowiednie pomieszczenie i ustawienie sprzętu. Dla systemu 5.1 wymagania są jeszcze bardziej restrykcyjne. Głośniki powinny być ustawione na okręgu o promieniu 2-4m. Nagranie możemy odebrać przestrzennie w rozmaitych kombinacjach i ustawieniach sprzętu, ale nie będą one odtwarzały wrażeń zgodne z zamysłem autora. Im bardziej rozbudowany układ tym większe znaczenie będzie miała odpowiednia konfiguracja odsłuchu. [16]

4.3 Ambisonia i najnowsze technologie

Absolutnie szczególna wśród systemów dźwięku przestrzennego jest ambisonia. Do reprodukcji na poziomie pierwszego rzędu ambisoni potrzebne jest 16 głośników, precyzyjnie ustawionych i skalibrowanych. Wykorzystuje się tutaj wektorowe funkcje panoramujące takie jak:

- VBAP- *Vector-Based Amplitude Panning*
- VBIP- *Vector-Based Intensity Panning*
- MDAP- *Multiple Direction Amplitude Panning*
- AIIRAD- *All-Round Ambisonic Decoding*



Rys 4.5 Logo Ambisoni (po lewej)- nawiązuje do kierunkowości przetworników w mikrofonie ambisonicznym (po prawej) [30]

Funkcje **VBAP** i **VPIB** wyznaczają pozycję wirtualnego źródła. Jeśli jeden z głośników znajduje się dokładnie w tej pozycji to tylko on będzie aktywny. W innym przypadku algorytm wykalibruje poziomy między sąsiednimi głośnikami. Daje to wyraźne odchylenie w postrzeganej szerokości i zabarwieniu. W przypadku ruchu źródła, przy przejściu pomiędzy zakresami (trójkątami na które podzielona jest sfera) mogę pojawić się intensywne, wyraźne skoki.

W metodzie **MDAP,** opracowanej na Alto University w Finlandii, użyte jest kilka kierunków wirtualnych źródeł w stałej konstelacji wokół kierunku zdarzenia dźwiękowego. Stabilizuje to szerokość i zabarwienie.

Najbardziej wszechstronną jest technika **AIIRAD**. W prosty sposób dekoduje sygnał w zwiększą płynnością niż VBAP i dobrze zachowuje treści ambisoniczne na dowolnych przestrzennych układach głośnikowych. Wykorzystuje kodowanie próbek do wirtualnego układu o bardzo wielu głośnikach (ponad 5000) i zbiera optymalne parametry wektorów kierunku źródła. Odtwarza nagranie w rzeczywistym układzie głośnikowym, dzięki zastosowaniu statystycznej matrycy VBAP.

Rys. 4.7 Zdjęcie z programu mikrofonu ZOOM przy odsłuchu formatu ambisonicznego

Ambisonia wymaga drogiego sprzętu, skomplikowanego dekodowania i tworzy duże rozmiary plików. Z drugiej strony jest elastyczna, ponieważ jedno nagranie może zostać przetworzone na dowolną liczbę kanałów i nie ma problemów z fazą. Poza tym jako jedyna reprezentuje pełnię dźwięku w 360 stopniach.

Idąc o krok dalej, chcąc jeszcze lepiej oddać przestrzeń, musimy przede wszystkim uzbroić się w… więcej głośników. Aby zagęścić siatkę źródeł dźwięku, stworzyć więcej kanałów i rozbudować przestrzeń odbiorcy. To podejście jest wykorzystane w metodzie **Wave Field Synthesis**, gdzie celem jest stworzenie całego wirtualnego środowiska akustycznego, z wykorzystaniem kilkuset głośników. W przypadku tego systemu, w odróżnieniu od pozostałych, lokalizacja słuchacza nie wpływa na odbiór dźwięku. Założenia opierają się na *zasadzie Huygensa*, wedle której każdy punkt do którego dotarła fala jest źródłem nowej fali kulistej. Dlatego w systemie WFS wyzbywamy się naturalnych odbić i ugięć fali, a wytwarzamy je sztucznie w (docelowo, w zamyśle, przy nieskończonej liczbie głośników) każdym punkcie przestrzeni nas otaczającej. Im więcej głośników tym dokładniejsza jest reprodukcja.

A może nie potrzeba jednak aż tylu głośników? **Soundbar** to zintegrowany system audio zawarty w jednej podłużnej odbudowie. W najprostszych przypadkach są to dwa głośniki, a w konstrukcjach projektorów dźwięku może ich być nawet kilkanaście. Projektorów dźwięku- bo w tym rozwiązaniu wykorzystuje się imitacje wirtualnych źródeł przez odbicia od ścian pomieszczenia. Wykorzystane są do tego modele psychoakustyczne i procesory DSP. Procesory **DSP** (Digital Signal Procesor) to układy do cyfrowej obróbki sygnału z przetwornikami o dużej mocy obliczeniowej i bardzo małych opóźnieniach. Są stosowane w zewnętrznych korektorach dźwięku. Najbardziej zaawansowane soundbary dają realistyczne efekty przestrzenne, ale tylko w rozszerzonej przedniej półsferze. Jedynie Yamaha opracowała technologię cyfrowej projekcji dźwięku (YSP), która dzięki odbiciom

ukierunkowanych wiązek daje efekt *full surround* i bardzo realistyczny dźwięk przestrzenny ze wszystkich stron. Efekt zostaje osiągnięty dzięki *InteliBeam*, czyli systemowi skanowania pomieszczenia. Do zestawu można podłączyć suboofer.

Rys 4.8 Soundbar AMBEO firmy Sennheiser

4.5 Stereo a binaural

Z drugiej strony mamy miniaturyzację i zbliżenie się do słuchacza. Dosłownie. Nagrania binauralne są przeznaczone wyłącznie na słuchawki. Z oczywistych powodów pierwsze zabawy z dźwiękiem przestrzennym przerabiano w technice stereofonicznej. Wraz z jej rozwojem pojawiała się coraz większa potrzeba odtworzenia dźwięku w formie jak najbliższej naturalnemu. W tym celu wykorzystywano coraz to więcej kanałów, głośników i opracowano wiele ulepszeń. Jednak anatomią człowieka i biologią słuchu, uchwyceniem tematu ze strony odbiorcy, zainteresowano się szczególnie przy konstrukcji słuchawek. Różnica pomiędzy tymi dwoma podejściami polega na całkowitym odseparowaniu kanałów w przypadku nagrania binauralnego. W prawym i lewym kanale reprodukowane są odmienne treści dźwiękowe.[15]

Można powiedzieć, że już paryska prezentacja z 1881 roku była wstępem do nagrań binauralnych, ponieważ mikrofony były ustawione na szerokość głowy, a odtwarzanie odbywało się na niezależnych kanałach, do każdego ucha osobno. Pierwsze słuchawki natomiast zupełnie nie były powiązane z branżą muzyczną.

Używano ich w centralach telefonicznych w latach 70. i 80. XIX wieku. W roku 1881 we Francji zaprezentowano system Théâtrophone stworzony przez Clementa Adera. Parę lat później, w roku 1890, francuzi mogli już posłuchać muzyki i przedstawień teatralnych prosto z Opery Garnier. Słuchawki były podpięte bezpośrednio do linii telefonicznej. Podobnym patentem, również wykorzystującym połączenie przez linię telefoniczną, mogła się pochwalić Wielka Brytania. Od roku 1895, dzięki systemowi Elektrophone, we własnym domu można było wysłuchać muzyki granej na żywo w lokalnej operze, a nawet mszy (oczywiście przy wykupionej subskrypcji).

Tak słuchawki zawitały w domach. Początkowo nadawane nagrania były stereofoniczne. Rozróżnienie nie było wtedy potrzebne. Jako że rozwiązanie nie było szczególnie wygodne, rozwój tej technologii odszedł w cień. Wskrzesiła go natomiast stacja radiowa z Connecticut ponad 50-siąt lat później. Odsłuch radia w technice stereofonicznej nie był jeszcze popularny więc sprawa nie była ani prosta, ani tania. Kanał lewy i prawy były nadawane na dwóch różnych częstotliwościach więc potrzebne były dwa takie same odbiorniki ustawione na te dwie częstotliwości. Mimo wszystko udało im się, maszyna ruszyła. Zainteresowano się techniką nagrań ukierunkowanych na słuchawki. Sprawa odpowiedniego oddania przestrzeni i umiejscowienia w niej instrumentów była ambicją głownie wielbicieli muzyki klasycznej. Pierwsze podejście tego typu w muzyce popularnej zaproponował Manfred Shunke. Nagrany przez niego album Lou Redda „Street Hassle" w 1978 roku określa się jako pierwsze komercyjne wykorzystanie systemu binauralnego w muzyce polularnej. Tym i kilkoma innymi albumami nagranymi w tej technice rozpowszechnili oni koncept Kunsthopf-Stereophonie (ang. dummy head recording), wykorzystujący bardziej czynnie wiedzę o przestrzenności ludzkiego słuchu.

System sztucznej głowy to zupełnie inne podejście do sprawy idealnego odsłuchu. Nie umieszczamy słuchacza w środowisku dźwięków, a obliczamy jak powinien to środowisko odebrać. Dzięki badaniom w oparciu o system sztucznej głowy dokładniej wyznacza się funkcje HRTF. Wszystko to po to aby później nagrania przeznaczone do odsłuchu na słuchawkach spleść z tymi funkcjami. Słuchawki powinny być możliwie jak najlepiej izolujące, ponieważ wszystkie efekty odbicia od

ciała, małżowiny i głowy, a nawet odpowiedź pomieszczenia w jakim mamy się wirtualnie znaleźć, zapewnia specjalnie opracowany splot. Celem jest „przeniesienie" słuchacza w wirtualną rzeczywistość- na koncert do sali koncertowej czy odległej planety w grze.

Zachowaniem otaczającej nas dźwiękowej przestrzeni zajmuje się **auralizacja**. Jest to technika kreacji przestrzeni nagrania, w której do sytuacji fonicznej nagranej bez żadnego pogłosu, niezależnie dodaje się odpowiedź pomieszczenia. Oznacza to, że możemy usłyszeć utwór na widowni filharmonii, w niewielkim pokoju lub na hali dworca. Zebrana uprzednio odpowiedź impulsowa pomieszczeń jest splatana z nagraniem i funkcją HRTF. Dzięki temu teoretycznie już odfiltrowany i „obliczony" dźwięk trafia bezpośrednio do naszych uszu. W praktyce jednak nie działa to tak doskonale, głownie ze względu na różnorodność funkcji HRTF. W tym przypadku różnice w ludzkiej anatomii są szczególnie zauważalne. Opracowany splot może dawać dobry efekt w odbiorze twórcy, ale HRTF odbiorcy może różnić się na tyle, że zupełnie nie poczuje zamysłu autora, inaczej odbierze tą przestrzeń. W związku z powyższym, techniki te są głównie wykorzystywane w grach, gdzie przestrzeń nie musi być jednoznaczna bo nie znamy jej empirycznie. Dźwięk przestrzenny w przypadku gier najlepiej dostosować jest do odsłuchu binauralnego, ponieważ pełni od rolę nie tylko w aspekcie atmosfery gry, ale także odgrywa ważną rolę w samej rozgrywce- kiedy na przykład dźwięk pomaga nam zlokalizować przeciwnika.

W nagraniach przeznaczonych do odsłuchu binauralnego odpowiedź pomieszczenia jest zawarta w nagraniu. Chcemy tylko jak najdokładniej odtworzyć rzeczywistą atmosferę. Dla odpowiedniego wrażenia przestrzenności, nagranie binauralne należy odtwarzać wyłącznie na słuchawkach, ponieważ odsłuch na głośnikach powoduje powstawanie przesłuchów międzykanałowych. Zmienia się barwa dźwięku, dźwięk przefiltrowany przez HRTF docierałby do nas podwójnie przefiltrowany, co uniemożliwia lokalizację. W odsłuchu binauralnym (w przeciwieństwie do stereofonicznego) kanały muszą być całkowicie odseparowane.

Słuchawki wykorzystane do odsłuchu binauralnego powinny mieć możliwie jak najbardziej płaską charakterystykę częstotliwościową. Można to osiągnąć przez

dodatkową korekcję lub kalibrację sprzętu. Dla gorszej jakości słuchawek charakterystykę wyrównuje się cyfrowo przetwarzając dźwięk. Jest to o tyle istotne, że składowe częstotliwości reprodukcji przekładają się na widmo i tym samym na lokalizację wirtualnych źródeł.

Rozdział 5: Formaty zapisu audio

Ponieważ w obszarze zainteresowań tej pracy znajdują się cyfrowe formy przekazu, warto poruszyć temat formatów zapisu oraz kompresji dźwięku i obrazu, niezbędnego do umieszczenia efektów w sieci. Przedstawię zatem kilka rozwiązań najbardziej popularnych lub historycznie istotnych.

Całą drogę jaką przechodzi dźwięk od mikrofonu do głośnika nazywa się **torem fonicznym**. Jest to zatem ciąg spiętych ze sobą urządzeń przetwarzających sygnał dźwiękowy. W poprzednich rozdziałach omówiono oba końce tego toru- elementy rejestrujące oraz odtwarzające dźwięk. Takie najprostsze tory foniczne stosowano do lat 30. XX wieku. Był to mikrofon, urządzenie rejestrujące i „z drugiej strony" urządzenie odtwarzające i słuchawki czy też głośniki. Do tej pory układ ten bardzo się rozbudował. W latach 50. kreacja nowego świata dźwiękowego, wykorzystującego modyfikacje i deformacje oryginalnych dźwięków wymusiła dodanie kolejnych elementów. Góruje też idea jak najlepszej wierności przekazywanego słuchaczowi dźwięku. Pomocne w tym okazały się wzmacniacze sygnału, regulacja poziomów nagrania, mierniki wysterowania czy regulacja poziomu odsłuchu. Przede wszystkim jednak cały proces znajdujący się pomiędzy omówionymi wcześniej skrajnymi ogniwami jest obecnie zawarty w cyfrowej formie dźwięku.

Zapis dźwięku zaczął się od fonografu i po raz kolejny…. Thomasa Edisona. Jego prezentacja miała miejsce 29 listopada 1877 roku, a patent otrzymał w 1878 roku. W fonografie membrana pobudzona dźwiękiem przekazywała drgania do igły żłobiącej w cynowej folii, nałożonej na obracający się wałek. Gdy igła przechodziła ponownie po folii, poruszała membraną w takt nagranego dźwięku. Była to pierwsza próba rejestracji umożliwiająca ponowne odtworzenie. Długość nagrania osiągała zaledwie 2 minuty i nie było możliwości zrobienia kopii. Zastąpiono więc folię woskowym wałkiem, a później celuloidem. Dzięki temu pomieszczono w zapisie już 4 minuty nagrania. Nie mniej jest to oczywiście zamierzchła historia. W roku 1929

wyprodukowano ostatni fonograf. Rynek podbiły już wtedy płyty winylowe. Zasada działania podobna- odtwarzanie dźwięku przez igłę w wyżłobionych na płycie rowkach. Później nastała era magnetofonów i kaset kompaktowych. Było to kolejne innowacyjne udoskonalenie ówczesnej technologii- tym razem z wykorzystaniem stalowej taśmy. Zwięźle przechodząc dalej, wszystko znów stało się historią gdy na scenę wkroczyły płyty CD. Był 17 sierpnia 1982 roku, gdy kooperacja firm Philips i Sony zaprezentowała premierową płytę kompaktową w fabryce w Langenhagen w Niemczech. Odtworzono walce Chopina w wykonaniu chilijskiego pianisty Claudio Arrau, który osobiście uruchomił nagranie w tym przełomowym wydaniu. Tym razem był to poliwęglanowy krążek z muzyką zapisaną cyfrowo, do którego odczytu wykorzystano laser. Spektakularna technologia osiągająca nieporównywalnie większą pojemność danych szybko osiągnęła niskie koszty i stała się standardem. Pierwsza płyta została wydana w 1982 roku (ABBA „The Visitors"), a trzy lata później można było już wybierać z pośród 6000 albumów. Można powiedzieć, że płyta CD w niezmienionej formie trwa już prawie 40 lat. I tu docieramy do punktu, chodź obecnie także niemal historycznego, to interesującego w kontekście dźwięku cyfrowego.

5.1 Dźwięk cyfrowy

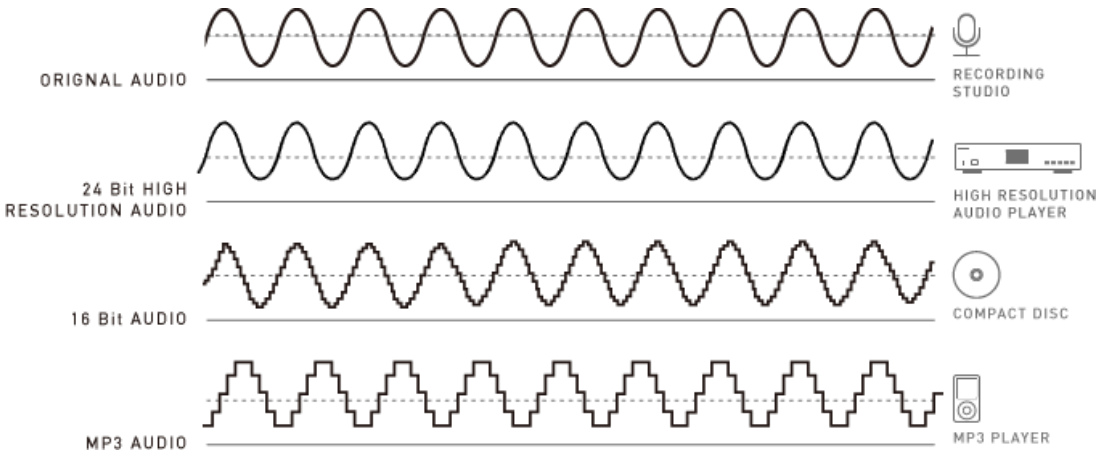
Zapis analogowy (od analogiczny, „taki sam") swojego czasu był uznawany z góry za lepszy, oryginalny. Obecnie, chociażby dzięki wspomnianym płytom CD, takie podejście jest zabawnym archaizmem. Typowy dla niej zapis tzn. 44,100 kHz w 16 - bitowym stereo już oddaje nagranie wystarczająco dobrze. Matematyczne wyliczenia pokazują, że taka rozdzielczość pokrywa w całości pasmo częstotliwości i niemalże cały zakres rozpiętości dynamicznej jaki obejmuje ludzki słuch.

Tłumaczeniem dźwięku „z naszego na komputerowy" i na odwrót zajmują się **kodeki** czyli **kodery** i **dekodery**. Kodery przetwarzają strumień danych (w naszym przypadku muzyki), na odpowiednią, wybraną formę cyfrową o określonej strukturze. Dekodery odwracają ten proces, dekodują ową formę z powrotem na strumień danych. Natomiast przechowywaniem tych „cyfr" zajmują się

kontenery. Zatem o tym jaką uzyskamy jakość decyduje wybrany kodek. Kontener odpowiada za sposób organizacji tych informacji. Możemy przechowywać w nim różne typy danych, nawet różnie zakodowanych.

Pierwszym, wartym omówienia przykładem, jest **system kodowania PCM** (ang. Pulse Code Modulation). Składa się on z kilku elementów, takich jak: filtr dolnoprzepustowy, układ próbkująco-zapamiętujący, przetwornik analogowo-cyfrowy i generator taktujący. Na wejściu kodera, filtr dolnoprzepustowy odcina częstotliwości nieużyteczne i zapobiega powstaniu efektu aliasingu. **Aliasing** jest zniekształceniem nagrania przez częstotliwości niepożądane, dający efekt brzęczenia, metalicznego brzmienia. Zbierając częstotliwości wyższe niż połowa częstotliwości próbkowania (tzn. „rozdzielczości z jaką oceniamy próbkę"), nie dostajemy jednoznacznej informacji czy była to częstotliwość f_0 czy dwa razy „gęstsza" czyli wyższa. Ta niejednoznaczność prowadzi do deformacji nagrania, która jest już nieodwracalna- dźwięk został błędnie zapisany. Aby jej zapobiec należy dopilnować aby nie pojawiły się w sygnale częstotliwości wyższe niż wspomniana połowa częstotliwości próbkowania, nazywana częstotliwością Nyquista. **Częstotliwość Nyquista** jest najwyższą dopuszczalną częstotliwością analogowego sygnału, która może być próbkowana poprawnie. W przypadku standardu CD- 44,1 kHz, jest to 22,5 kHz. Składowe harmoniczne powyżej częstotliwości Nyquista powinny zostać odcięte przez filtr. Częstotliwość Niquista ustanawia górną granicę częstotliwości sygnału w stosunku do częstotliwości z jaką próbkujemy ten sygnał. **Częstotliwość próbkowana** to ilość próbek zmierzonych w ciągu sekundy. Im wyższą częstotliwość przyjmiemy tym gęściej i dokładniej zbadamy sygnał. W przypadku audio jednak, w pewnym momencie dokładność ta nie wpływa już znacząco na jakość, ponieważ różnice nie będą dla nas słyszalne. Ze względów historycznych i technicznych, próbkowanie 44,1 kHz z płyty CD zostało przyjęte jako standard i punkt odniesienia. Wyższe częstotliwości (które ucinamy filtrem) i tak nie były by przez nas odebrane zgodnie z psychoakustycznym modelem. Jednak dla zachowania pewnej barwy dźwięku stosuje się również wyższe parametry- próbkowanie 48 kHz, a w studiach nagraniowych również 96 kHz i wyższe.

Proces próbkowania odbywa się w układzie próbkująco-zapamiętującym. W równych odstępach czasu (mierzonych przez taktowanie generatora- w ustalonej częstotliwości próbkowania) mierzona jest wartość napięcia danej próbki i zgodnie z wynikiem przypisywana odpowiednia liczba dziesiętna. Liczba jest zapamiętywana do momentu pojawienia się następnej próbki. Te informacje złożone w całość dają obraz amplitudy sygnału. W takiej formie można przystąpić do konwersji na system binarny. Przetwornik analogowo-cyfrowy przystępuje do kwantowania próbek czyli zaokrąglania odebranych liczb (określających napięcia) do swojej miary- zależnej od wybranej przez nas rozdzielczości tzn. 8, 16 czy 24 bity. Im więcej bitów na to przeznaczymy tym mniejszy popełnimy błąd przy zaokrąglaniu, tym dokładniejsze odwzorowanie oryginału.



Rys 5.1 Rozdzielczość bitowa i jakość dźwięku, dokładność zapisu fali dźwiękowej [31]

Kształt fali dźwiękowej jest dopasowywany do „komputerowych standardów". Jak dobrze to dopasowanie odda wersję oryginalną zależy od wspomnianej rozdzielczości, ponieważ decyduje ona o ilości poziomów do których musimy zrównać nasz sygnał. Dla 8 bitów jest to 2^8 czyli 256 poziomów, a dla 16 bitów jest to już 65 536 poziomów. Łatwo sobie wyobrazić, że przy takiej różnicy sygnał staje dużo mniej „kanciasty" w wersji 16-bitowej. To natomiast oddala nas w skali wykładniczej od problemu **szumu kwantyzacji** czyli zawieszenia informacji o sygnale w momentach pomiędzy następnymi próbkami (kiedy to układ

próbkująco- zapamiętujący właśnie zapamiętał liczbę i oczekuje na następną). Dodatkowo w tym celu stosuje się **nadpróbkowanie** (ang. *oversampling*) czyli uśrednianie poziomów między kolejnymi próbkami.

W pierwszych komputerowych czytnikach płyt CD ustanowiono za standard szybkość przesyłania danych 150 kB na sekundę (kBps). Szybko pojawiła się potrzeba poszerzenia tego przesyłu- powstały napędy CD-ROM o prędkości transferu danych 300 kBps, 600 kBps i szybsze. Początkowe parametry były wystarczające dla przepływności audio w formacie PCM. Pojawiły się jednak pewne ograniczenia ze strony pojemności pamięci, a później możliwości transferu w internecie. Potrzebna prędkość przesyłania danych czy też wykorzystana pamięć jest uzależniona od przepływności formatu jakim się posługujemy. **Przepływność** (bit rate) jest cechą sygnału cyfrowego, mówiącą o szybkości przesyłu informacji, ilości bitów (rzadko innych jednostek) przesłanych w ciągu sekundy- jednostka bps, b/s lub kBps, kB/s. Dla standardu płyty CD i kodowania PCM obliczenia są proste: dwa kanały zawierające 44 100 próbek na sekundę, po 16 bitów każda. Dostajemy przepływność 1 411 200 bps- 1,4112 Mbps. W przypadku formatów kompresowanych nie ma takiej prostej zależności. Przepływność zależy od stopnia kompresji. Razem ze stopniem kompresji spada jakość dźwięku, ale to również nie jest bezpośrednia zależność ponieważ zależy ona również od sposobu kompresji danych, od jakości samego kodeka. Przykładowo dla formatu FLAC z rozdzielczością płyty CD przepływność typowo osiąga 700-840 kbps. Dla formatu MP3 maksymalna przepływność to 320 kbps.

A zatem to tutaj zaczyna się temat jakości. System kodowania PCM jest najprostszym, popularnym rozwiązaniem. Nie stosuje się tu żadnej kompresji. Inne formy kodowania dźwięku są znacznie bardziej złożone i zwykle wykorzystują jakąś formę kompresji. Może ona być stratna lub bezstratna. W przypadku bezstratnej algorytm jest stworzony w taki sposób, aby plik zawierał mniejszą liczbę bitów, ale możliwe jest odtworzenie zakodowanej informacji w formie niezmienionej, identycznej z oryginałem. Kompresja stratna skupia się bardziej na zmniejszeniu

pliku niż jakości. Wykorzystuje bardziej restrykcyjne algorytmy w celu zmniejszenia liczby bitów, ale w niektórych przypadkach nie wpływa to znacząco na jakość i odtwarzana informacja jest bardzo blisko oryginału. Odpowiedni kodek przetwarza plik dźwiękowy na określony format cyfrowy, zakodowany w właściwy dla siebie sposób.

Płyta CD, która wprowadziła nas w erę cyfrowego zapisu dźwięku wykorzystuje standard CD-DA (Compact Disc Digital Audio). Zawiera się w nim liniowe kodowanie PCM o parametrach: 16 bitów na próbkę, przepustowość audio 1411,2 kbit/s, częstotliwość próbkowania 44,1 kHz. Przy takich założeniach możemy nagrać na płytę około 74 minuty muzyki. Ponieważ wzrosła popularność komputerów jako odtwarzaczy muzycznych, a pojemność dysków nie dawała możliwości przechowywania swojej dyskografii w pełnym formacie, rozpoczęto prace nad optymalnym sposobem kompresji, przechowywania i przesyłu danych.

Kontener **WAV** jest przykładem jednej z najpopularniejszych form przechowywania nieskompresowanych plików audio. Pod takim rozszerzeniem możemy zapisać zarówno dźwięk jaki i obraz wideo w dowolnie zakodowanej formie. Najczęściej jednak, ze względu na możliwości tego kontenera, wykorzystuje się format PCM i zapisuje muzykę w pełnej rozdzielczości. Generuje to oczywiście ogromne wykorzystanie pamięci. Jedna płyta CD zgrana do WAV zajmuje około 700 MB. Na początku ich kariery, dla dysków twardych komputerów było to o wiele za dużo. Dlatego pomimo koncepcji, która pojawiła się w latach 80., format WAV został wprowadzony dopiero w 1991 roku przez Microsoft i IBM przy okazji Windows 3.1. Odpowiednikiem dla Macintoshach jest format **AIFF**, a ponieważ komputery Apple rozprzestrzeniły się wśród dźwiękowców to stał się on standardem w produkcji. Rozdzielczość WAV utożsamia się z rozdzielczością płyty CD, a maksymalna jaką można obecnie spotkać to 32 bity przy 384 kHz. Ze względu na stosowanie 32-bitowych zmiennych maksymalny rozmiar pliku wynosi 4 GB. Podjęte zostały próby zmniejszenia plików dźwiękowych przy zachowaniu jakości. Tym charakteryzuje się kompresja bezstratna.

Wartym wspomnienia przykładem jest **Delta Modulation**, którą CDS zaproponowało w pierwszym kinowym systemie kinowym. Podczas gdy PCM zapisuje intensywność (poziom) każdej próbki w stosunku do poziomu 0 dB (co wymaga 16 bitów na każdą próbkę), Delta Modulation zapisuje tylko różnicę między kolejnymi próbkami, co wymaga mniej danych i może dać pozwolić na zmniejszenie pliku nawet 4:1. Działa ona najlepiej przy nagraniach głosu, ponieważ amplitudy nie są tu zbyt duże.

Zostało przyjęte, że kompresję uznajemy za bezstratną jeśli przepustowość jest nie mniejsza niż ta z płyty CD czyli 1411,2 kbps. Wydajność kompresji zależy od złożoności utworu. Takie kodeki wykorzystują liczne sprytne rozwiązania nie obniżające jakości nagrania. Na przykład zamiast zapisywać ciszę jako brak muzyki, próbka po próbce, można zapisać tylko czas jej trwania. Gdy kanał lewy i prawy są bardzo podobne, zamiast zapisywać je oddzielnie można zapisać tylko różnice między nimi. To i wiele innych sprytnych rozwiązań pozwoliło na ograniczenie rozmiarów plików zachowując najwyższą jakość.

Jednym z najpopularniejszych kodeków tego typu jest **FLAC** (Free Lossless Audio Codec). Jest często wybierany ze względu na swoją szybkość i małe użycie zasobów sprzętowych, a ponieważ bazuje na dokumentacji Open Source i nie jest objęty patentami umożliwia szeroki dostęp. Radzi sobie z nagraniami od mono, przez stereo do ośmiokanałowych i redukuje wielkość pliku do 50-70% rozmiaru oryginału. Został wprowadzony w 2001 roku i jest ciągle rozwijany oraz powszechnie używany jako format muzyki sprzedawanej w sieci. Należy do rodziny kodeków Ogg stworzonych przez organizacje non-profit xiph.org.

Analogicznie Apple wydało **ALAC** (Apple Lossless Audio Format) o rozszerzeniu .m4a (lub .mp4). Wprowadzony w 2004 roku format bazuje na podobnej zasadzie co FLAC. Kompresuje plik o średnio 52%. Nie generuje dużych obciążeń obliczeniowych więc może być zastosowany na słabszych urządzeniach jak iPody, ale poza środowiskami Apple format nie jest w zasadzie wspierany.[5,8,24]

5.2 Kompresja stratna

Chcąc ograniczyć wykorzystywaną pamięć jeszcze bardziej musimy już liczyć się z pewną utratą jakości. Mowa o stratnej kompresji danych. Kompresję wykonuje się tak, aby odczuwalna różnica jakości była jak najmniejsza. W przypadku dźwięku wykorzystuje się tutaj wiedzę z zakresu psychoakustyki- modele psychoakustyczne. Na ich podstawie wiemy gdzie można „przyciąć" trochę dźwięku, którego i tak byśmy nie usłyszeli. Pierwszym praktycznym przykładem może być ograniczenie zapisu do pasma słyszalności. W modelu przyjmuje się, że słyszymy dźwięki w zakresie 20 Hz- 20 kHz. Dlatego też większość cyfrowych odtwarzaczy muzycznych ma właśnie takie pasmo przenoszenia. Innym przykładem jest wykorzystanie faktu maskowania dźwięków. Z pośród dźwięków bliskich częstotliwościowo warto zapisać tylko ten najgłośniejszy- pozostałych i tak nie usłyszymy. Ten model jest zastosowano we wszystkich standardach MPEG.

W świecie audio historyczną karierę w dziedzinie kompresji dźwięku zrobił format **MP3 (MPEG 1, warstwa 3).** Powstał głównie z myślą o obrazie w 1982 roku w Instytucie Fraunhofer. W roku 1987 przy pracy nad radiofonią cyfrową opracowano algorytmy, które stały się podstawą kompresji dźwięku. Prace nad MPEG-1 Layer-3 zakończyły się w 1991 roku i format zaczął się reklamować jako jakość „prawie CD". Algorytm okazał się najbardziej optymalnym z rodziny MPEG. Kompresuje plik dźwiękowy nawet 11-krotnie względem WAV więc z nadejściem integracji ISDN i DSL stał się standardem muzyki w sieci. Przy uzyskanej przepływności 112-128 kbps dźwięk był zaskakująco dobry. Dlatego warto przyjrzeć się bliżej temu formatowi.

Przede wszystkim, w formacie MP3 szeroko wykorzystane zostały modele psychoakustyczne oraz kompresja stratna jak i bezstratna. W pierwszym etapie, z sygnału eliminowane są składowe niesłyszalne (lub słabo słyszalne)- kompresja stratna. W etapie drugim, eliminuje się nadmiarowe dane- kompresja bezstratna.

Algorytm operuje sygnałem próbkowanym częstotliwością do 48 kHz i optymalizuje go pod wyjściową przepustowość (dostępne są przepustowości od 32 do 320 kbps). Może operować na zarówno na dźwięku mono (głównie do kompresji głosu,

używany przy niskich przepływnościach- poniżej 32 kbps) jak i stereo (przepływność jest dzielona płynnie pomiędzy kanałami; używany do wysokich przepływności- 192 kbps i więcej) lub *dual channel* (z dwoma niezależnymi kanałami; wykorzystywany np. przy kilku wersjach językowych filmu). Możliwa jest również praca na dźwięku *joint stereo*, gdzie badane jest podobieństwo pomiędzy kanałami i w przypadku identycznego sygnału jest on zapisywany jako mono (co umożliwia kodowanie z większą dokładnością). W algorytmie *joint stereo* zawiera się też **stereofonia różnicowa**, która dzięki kodowaniu ścieżek jako sumy i różnice sygnałów pozwala zredukować dane nawet o 50%.

W procesie kodowania MP3 wykorzystane są procesy, które ze względu na uniwersalność warto omówić. Jednym z nich jest **DCT (Dyskretna Transformata Kosinusowa)**, która pomaga rozdzielić sygnał określony w dziedzinie czasu na sygnał z określony w dziedzinie częstotliwości. Na podstawie sygnału otrzymujemy odpowiadające mu współczynniki transformaty. Jest to proces odwracalny, na tym etapie, ze współczynników transformaty można odtworzyć dokładny sygnał bez strat. Większość współczynników jest zwykle bliska zeru, dzięki czemu w procesie kwantyzacji można je pominąć (co daje lepszą kompresję danych). W procesie **kwantyzacji** ogranicza się zbiór wartości sygnału do skończonej, określonej liczby bitów. Współczynniki DCT zostają przeskalowane, przez podzielenie ich przez właściwy współczynnik (znajdujący się w tabeli kwantyzacji, które są dobierane doświadczalnie) i zaokrąglane do najbliższej liczby całkowitej. Na tym etapie tracimy więc już pewne informacje.[7, 24, 25]

W 1952 roku David Huffman przedstawił inną metodę kodowania, która w algorytmie MP3 również została wykorzystana. **Kodowanie Huffmana** jest bezstratną metodą kompresji, łatwą w implementacji i prostą w założeniu. Wykorzystuje fakt, że pewne wartości w sygnale występują częściej niż inne. Zapisuje się ciąg znaków, którego długość jest zależna od częstotliwości wystąpienia symbolu. Jeżeli do ich zakodowania wykorzystamy krótsze słowa kodowe, a dla rzadszych wartości przeznaczymy dłuższe to w efekcie sumarycznie zmniejszymy wielkość zakodowanych danych.

Proces kodowania MP3 można przedstawić w kilku etapach. Najpierw sygnał dzieli się na mniejsze fragmenty tzw. *ramki* (o długości ułamka sekundy). Każda poprzedzona jest nagłówkiem zawierającym metadane o jej parametrach. Następnie dla sygnału określonego w czasie wyliczana jest jego reprezentacja w dziedzinie częstotliwości (widmo sygnału dźwiękowego). Ramki są dzielone najpierw na 32 pasma w wielofazowym banku filtrów, a później podpasma są przekształcane przez DCT, która generuje 18 współczynników dla każdego z podpasm. Daje to lepszą kontrolę przy analizie sygnału pod kątem progów maskowania i usuwa zbędne informacje. Widmo sygnału jest porównywane z modelem psychoakustycznym i koder ocenia które składowe są „najważniejsze"- najlepiej słyszalne, wymagające najwierniejszego odwzorowania. Składowe mało istotne mogą zostać zakodowane w pewnym przybliżeniu, a te najmniej istotne lub w ogólnie nie słyszalne zostają pominięte. Następuje optymalne przydzielenie bitów dla poszczególnych częstotliwości, tak by z uwzględnieniem decyzji modelu jak najwierniej zakodować sygnał. Proces kwantyzacji realizowany jest w formie dwóch pętli. Wewnętrzna pętla kontroluje współczynnik kompresji. Stosuje kwantyzacje, a następnie symuluje kodowanie współczynników i sprawdza czy „mieści się" w limicie przepływności. Jeśli nie, to z dopasowanym wskaźnikiem przyrostu pętla wykonuje się jeszcze raz. W pętli zewnętrznej zachodzi kontrola zniekształceń. Indywidualne współczynniki kwantyzacji zostają ustawione na 1. Obliczany jest błąd kwantyzacji i porównywany z oszacowanym przez model psychoakustyczny progiem percepcji. Jeśli błąd przekracza próg to współczynnik kwantyzacji zostaje odpowiednio zmieniony i pętla wykonuje się ponownie. Jeśli nie udaje się uzyskać żądanej przepływności i równocześnie spełnić wymagania modelu psychoakustycznego to dźwięk zostaje zakodowany bez spełnionych wymagań. W tym miejscu jest wykorzystywane kodowanie Huffmana, czyli dodatkowa, bezstratna kompresja usuwająca nadmiarowość danych. Aby możliwie najlepiej dopasować kompresję do danych źródłowych, wybierana jest najlepiej pasująca tablica kodów (z pośród zestawu kodów Huffmana). Dla różnych części widma wybierane są różne tablice, aby najlepiej skoordynować sygnał. Algorytm zaczyna od stworzenia histogramu i tworzy listę drzew binarnych, aby potem w pętli je

kompresować. W etapie końcowym formatowane są ramki wyjściowe i kolejno składane w pojedynczy ciąg bitów. W niektórych plikach MP3 dodatkowo wykonywana jest suma kontrolna- w każdej ramce zapisywana jest 16-bitowa liczba służąca weryfikacji poprawności strumienia. [1,2, 7, 24]

Współczynnik kompresji sygnału przez algorytm MP3 jest określony przez gęstość strumienia danych, *bitrate*. Jest to kompromis między rozmiarem pliku a jakością.

Rys. 5.2 Porównanie przepływności, rozmiaru i rozdzielczości wybranych formatów [31]

Gęstość strumienia może być stała- wtedy mówimy o **CBR** (Constatn Bit Rate), lub zmienna- **VBR** (Variable Bit Rate). Przy CBR na każdą sekundę nagrania przypada ta sama ilość bitów, co może spowodować nierównomierną jakość- spokojne fragment solowy będzie brzmiał lepiej niż podwójne forte całej orkiestry. W trybie kodowania VBR uwzględniona jest dynamika sygnału i przydziela więcej bitów na fragmenty bardziej złożone, a mniej na prostsze. Dzięki temu utwór ma stałą jakość. W przypadku VBR wymagane jest podanie przedziału tolerancji w jakim ma się mieścić gęstość strumienia bitowego. Ustalona z góry gęstość strumienia jest przydzielona do każdej ramki więc może się zdarzyć, że dla bardzo złożonych fragmentów przydzielony zakres tolerancji nie będzie wystarczający. W takiej sytuacji wykorzystywana jest **rezerwa bitowa**, zapewniona przez standard MP3. Rezerwa powstaje w miejscach pustych fragmentów ramek, w których zostało trochę miejsca.

Co ciekawe, specyfikacja formatu MP3 w dokumentach ISO/IEC 11172-3 nie określa dokładnie sposobu kodowania, a jedynie ogólny zarys wykorzystanej techniki

i wymagany poziom zgodności z normą. Ustalone są jedynie kryteria jakie ma spełniać struktura pliku by zaklasyfikować ją jako MP3. Taki zapis daje możliwość różnych realizacji koderów i dekoderów, a bazowe reguły zapewniają tylko funkcjonowanie standardu aby był odtwarzany przez dowolny dekoder.

Podsumowując, format MPEG-1 Layer-3 ma wiele zalet- od zmniejszenia rozmiarów pliku nawet 10-krotnie i stosunkowo dobrą jakość dźwięku, poprzez niewielką moc obliczeniową potrzebną do dekompresji nagrania, po darmowy dostęp do kodu źródłowego programów kodujących i dekodujących. Należy jednak pamiętać, że jest o format stratny, oparty na modelu psychoakustycznym, który może nie odpowiadać dokładnie wszystkim słuchaczom. Niektórzy usłyszą brakujące, wycięte składowe w dźwięku. Model jest dostosowany do uśrednionego HRTF-u. I tu jednak jest pewne pole manewru ponieważ przepływność formatu MP3 można regulować.

Wśród innych popularnych, stratnych algorytmów kompresji znajdzie się format **AAC (Advancent Audio Coding)** o rozszerzeniu .mp4 lub .m4a (które ponieważ wynikło z rozszerzenia standardu MPEG, pokrywą się z rozszerzeniem ALAC). Format jest kodekiem, bez kontenera multimedialnego. Pojawił się w 1997 roku jako owoc współpracy firm Fraunhofer IIS, AT&T Bell Labolatories, Sony, Solby i Nokia. Miał on na celu zlikwidowanie wad MP3, w prowadzając szereg ulepszeń dotyczących inteligencji i możliwości kodeka. Obsługiwał szerszy zakres częstotliwości próbkowania i teoretycznie nawet 48 kanałów. Był wydajniejszy i lepiej przenosił wyższe częstotliwości przy zachowaniu podobnych rozmiarów co MP3. Z biegiem lat, format MP3 został znacznie ulepszony i obecnie te różnice są coraz mniejsze. O lepszej jakości kodeka AAC decyduje lepsze zachowania charakteru nagrania, bliższe oryginału. Nie jest pojedynczym algorytmem, ale rodziną algorytmów w kilku profilach.

Wykorzystuje ulepszone metody filtrowania, kształtowania szumu (*noise shaping*) i zmniejsza błędy kwantyzacji. Predykcyjnie analizuje współczynniki nie tylko w dziedzinie czasu, ale i częstotliwości. W miarę możliwości, sposób kwantyzacji jest tak kalibrowany, aby szum kwantyzacji skupiał się wokół silnych dźwięków- wykorzystując TNS (*temporal noise shaping*), lub koduje się szum parametrycznie-

PNS (*perceptual noise substitution*). Przy cyfryzacji sygnału wykorzystuje się **kwantyzację wektorową**- kodowanie wektorów danych, nie pojedynczych liczb. W końcowym etapie, również kompresuje się sygnał kodowaniem Huffmana. Aby podnieść jakość względem formatu MP3, w algorytmie AAC+, definiowanym dla przepływności 128 kbps i dwóch kanałów, nie pozbywa się ostatecznie wysokich częstotliwości, a wykorzystuje **SBR** (*spectral band replication*). Ponieważ najwyższe częstotliwości są skorelowane z niższymi, są ich harmonicznymi, to można odtwarzać górą część spektrum na podstawie dolnej części widma. Zapis ogranicza się wtedy do parametrów je wiążących, dzięki czemu ograniczamy ilość danych. Jest to jedno z narzędzi wykorzystujące opis parametryczny. W podobny sposób opisuje się stereofonię. Zakodowany jest jeden sumaryczny kanał i zbiór parametrów (zależnych od częstotliwości). Wzajemna korelację kanałów opisuje IC (*interchannel coherence*), która reprezentuje zgodność statystycznych parametrów między kanałami. Im mniejsza jest korelacja tylko szerzej odbierzemy źródło dźwięku. Przy dekodowaniu sygnał zostaje przeskalowany i przesunięty w fazie oraz dodaje się splot sygnału z szumem w odpowiednich proporcjach, tak aby uzyskać różnice międzykanałowe.

AAC ogranicza wykorzystanie pamięci o 77-92% względem oryginału. Jest standardem wykorzystywanym między innymi dla dźwięku YouTube, radiofonii DAB+ i niektórych systemach telewizji cyfrowej. Jest także podstawowym kodekiem w urządzeniach Sony i Apple, co w dużej mierze zbudowało jego popularność.

Formatem zaproponowanym przez Microsoft jest **WMA** (Windows Media Audio), o rozszerzeniu .wma. Wprowadzony z Windows 98, stanowi konkurencję dla MP3 będąc częścią platformy Windows Media we wszystkich systemach napędowych Microsoft. Utrzymuje jakość 48 kHz z 24-bitową rozdzielczością i przepływności maksymalnej 192 kbps. Kompresuje plik od 86 do 95%, a więc jeszcze bardziej niż w przypadku MP3. Porównując z konkurentem, WMA brzmi lepiej tylko przy niskim *bitrate* tzn. do 96 kbps. Format nie jest też tak szeroko wspierany. Funkcjonuje zarówno jako kodek jak i kontener.

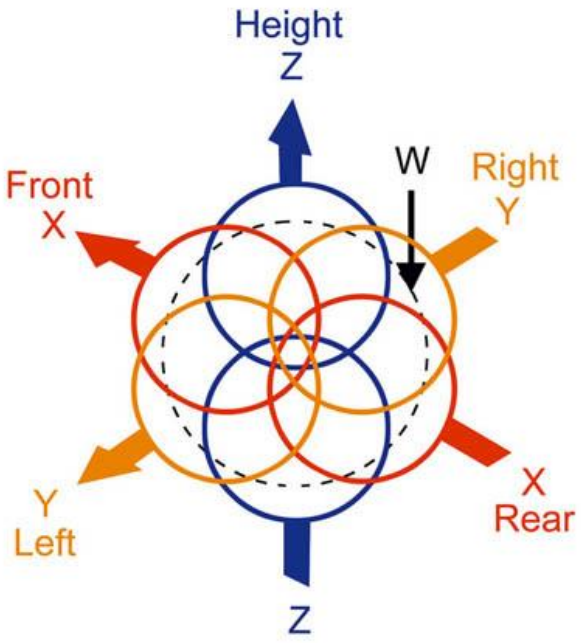
Odpowiedzią ze strony *open source* i fundacji Xiph.org jest **OOG Vobis** (z rozszerzeniem .ogg). Wydany w 1998 roku, nie ograniczony patentami, podobnie jak jest bezstratny odpowiednik. Obsługuje teoretycznie 255 kanałów i pond 16-bitową rozdzielczość w zakresie 6-48 kHz. Kompresuje plik o 74-90% względem oryginału. Nie jest jednak zbyt popularny, choć w 2005 roku wyróżniono go w magazynie Stereo i umieszczono na pierwszym miejscu w porównaniu formatów stratnych.

5.3 Formaty przestrzenne

Aby lepiej oddać przestrzenność nagrania, odwzorować nie tylko płaszczyznę horyzontalną, ale całą sferę, powstały formaty przestrzennego kodowania dźwięku. Takie kodowanie to tzw. **B-format**. Jest on przetwarzany z **A-formatu**, który odpowiada bezpośredniemu sygnałowi z mikrofonów. Następnie po liniowych przekształceniach, otrzymujemy B-format, zawierający kanały W, X, Y i Z, które są harmonicznymi sferycznymi kolejnych rzędów. Kanał W jest komponentem zerowego rzędu, a X, Y, Z to komponenty rzędu pierwszego. Dla poprawnej lokalizacji wysokich częstotliwości bardzo ważna jest poprawna matematyczna korekcja przy konwersji formatu A na B. Z dźwięku zarejestrowanego systemie Double MS na trzech kanałach, możemy przekształcić sygnał na poziomy B format według wzorów:

$$L = M_{front} + M_{rear}$$
$$X = M_{front} - M_{rear}$$
$$Y = S$$

Dla uzyskania czterech kanałów, musimy wykorzystać dedykowany mikrofon ambisoniczny, dzięki któremu dostaniemy czwarty kanał operujący płaszczyzną pionową.



Rys 5.4 Komponenty ambisoniczne [32]

Ideą jest przetwarzanie sygnałów A-formatu na B-format przy zastosowaniu liniowych kombinacji komponentów. Daje to możliwość odtworzenia dowolnej charakterystyki wirtualnego mikrofonu skierowanego w dowolnym kierunku. Dzięki temu na etapie postprodukcji, możemy zakodować dźwięk na dowolną ilość kanałów, w dowolnym systemie. Aby otrzymać lepsza lokalizację dźwięku, niezbędne jest zastosowanie wyższych harmonicznych sferycznych- większej liczby komponentów i ilości kanałów. Istnieje także możliwość zmatrycowania nagrania ambisonicznego na dwa kanały, z pominięciem Z (a więc informacji o lokalizacji góra- dół). Mówimy wtedy o **C-formacie** (czy też UHJ), który przydaje się gdy chcemy ograniczyć przepływność np. do transmisji w internecie. Obecnie jest to raczej format historyczny. Wszystkie dekodery w wersji hardware'owych i w formie wtyczek oraz potrzebne oprogramowanie są ogólnodostępne.

U podstaw ambisoni leżą matematyczne sformułowania harmonicznych sferycznych jako:

$$Y_k^m(\theta, \emptyset) = N_k^{|m|} P_k^{|m|} \sin(\emptyset) \cdot \begin{cases} \sin -m\theta, & gdy\ \text{m<0} \\ \cos m\theta, & gdy\ m \leq 0 \end{cases}$$

Y_k^m- sferyczna harmoniczna k-go rzędu z indeksem m (z zakresu $-k \leq m \leq +k$)- nazewnictwo z działu ambisoni nie jest spójne z matematycznym

$N_k^{|m|}$- współczynnik normalizacji

$P_k^{|m|}$- stowarzyszone wielomiany Legendre'a stopnia k rzędu m (rozwiązania kanoniczne równania różniczkowego Legendre'a zmiennej rzeczywistej z zakresu [-1,1]; rekurencyjne i ortogonalne)

θ- kąt azymutalny, początkowo zerowy i wzrastający lewoskrętnie

$\emptyset$- kąt elewacji, zerowy w płaszczyźnie azymutalnej , dodatni w górnej półsferze

Sferyczne harmoniczne są powiązane z sygnałem B-formatu $[B_k^m]$ poprzez równanie:

$$B_k^m = Y_k^m(\theta, \emptyset) \cdot S$$

gdzie S jest sygnałem źródłowym ukierunkowanym przez $(\theta, \emptyset)$. [30,32]

Pojawiło się kilka pomysłów redakcji formatów ambisonicznych- nie było jednomyślności, czy inspirować się fizyką kwantową, chemią czy grafiką komputerową? W związku z powyższym, powstały różne konwencje odnoście kolejności komponentów, normalizacji(czy też wagi współczynników) i polaryzacji.

Jeśli chodzi o kolejność komponentów (która jest w silnej korelacji z kierunkowością mikrofonów) to w tradycyjnym B-formacie wychodzi się od osi układu prawoskrętnego. Funkcjonuje tylko dla ambisoni rzędu zerowego i pierwszego. Dla spójności z wyższymi rzędami, opracowano konwencje bardziej uniwersalne, bazujące na symetrii względem osi Z. Wyróżnia się trzy z nich.

Furse-Malham- jest rozszerzeniem tradycyjnego B-formatu do rzędu trzeciego. Kolejność komponentów rozpoczyna oś Z, a następnie względem niej po obu stronach rozłożone są horyzontalne.

			W_0			
		Y_2	Z_3	X_1		
	V_8	T_6	R_4	S_5	U_7	
Q_{15}	O_{13}	M_{11}	K_9	L_{10}	N_{12}	P_{14}

Rys. 5.5 Kolejność komponentów w systemie Furse-Malham [30]

SID (Single Index Designation)- stosuje nomenklaturę trój-indeksową dla sferycznych harmonicznych. Jest kompatybilny z B-formatem pierwszego rzędu i kontynuuje zliczanie kolejnych sferycznych symetrycznie do osi Z, zaczynając od komponentów horyzontalnych. W związku z tym jest niekompatybilny z *Furse-Malham*. Nie jest szeroko wykorzystywany.

			0			
		2	3	1		
	5	7	8	6	4	
10	12	14	15	13	11	9

Rys. 5.6 Kolejność komponentów w systemie SID [30]

ANC (Ambisonic Channel Number)- jest najszerzej wykorzystywaną metodą zgodną z normalizacją SN3D i N3D, zdefiniowaną matematycznie jako $ACN = k^2 + k + m$

			0			
		1	2	3		
	4	5	6	7	8	
9	10	11	12	13	14	15

Rys. 5.7 Kolejnośc komponentów w systemie ACN [30]

Aby właściwie zrekonstruować przestrzeń dźwiękową ważne jest również zdefiniowanie normalizacji dla komponentów harmonicznych sferycznych, określenie ich wagi. Wyróżnia się cztery podejścia.

SN3D (Schmidt Semi-Normalizaion)- wykorzystuje normalizację w podejściu Schmidta (powszechnie używaną w innych dziedzinach). Współczynniki są zdefiniowane:

$$N_{k,m}^{SN3D} = \sqrt{(2-\delta_m)\frac{(k-|m|)!}{(k+|m|)!}} \; gdzie \; \delta_m = \begin{cases} 1, & gdy \;\; \text{m=0} \\ 0, & gdy \;\; m \neq 0 \end{cases}$$

Jest to wygodny sposób, ponieważ współczynniki są obliczane rekurencyjnie i w pierwszym rzędzie są wektorami jednostkowymi. Taka normalizacja jest szeroko wykorzystywana w połączeniu z ACN.

N3D- najbardziej oczywiste podejście do normalizacji, standardowa w fizyce i matematyce. Wykorzystuje bazy ortogonalne dla trójwymiarowego rozkładu. Dla idealnego rozproszenia pola 3D dawałaby równomierne rozłożenie mocy komponentów. Łatwa w dekodowaniu, ale dla za małej ilości głośników mogą pojawić się błędy.

Pomimo kilku możliwych kombinacji tych konwencji, korzysta się głównie z dwóch wersji B-formatów: AMB lub AmbiX. [30,32]

AMB (Microsoft Wave Format Extensible; WAVE-EX) to zapis B-formatu uwzględniający współczynniki wagowe FuMa (Furse Malham)- w zasadzie są one obecnie tożsame. Wprowadzony w 2001 roku przez Richarda Dobsona i oparty na formacie Microsoft'owym. Plik jest ograniczony parametrami kontenera WAV i może osiągać do 4 GB wielkości.

AmbiX (Ambisonic Exchange; .caf od Apple) omija problem limitu pliku 4GB. Normalizuje według SN3D. Numeruje kanały zgodnie z ANC, wnioskując z samej informacji o ilości kanałów. Nie zawiera dodatkowych metadanych- wystarczająca jest informacja o nagłówku .caf aby określić specyfikację. Koduje dźwięk liniowym PCM-em. Jest kompatybilny z formatem przyjętym przez Google dla YouYube 360.

Rozdział 6: Realizacja projektu

6.1 Specyfikacja wykorzystanego sprzętu

6.1.1 Kamera Garmin Virb 360

Kamera wykorzystana do rejestracji części wizualnej rejestruje obraz i dźwięk w 360 stopniach. Jest przystosowana do nagrań terenowych- ma klasę wodoszczelności do 10 m i temperaturę roboczą w zakresie 0-40^{o}C. Wymaga karty pamięci microSD USH-I o klasie U3 lub wyższej i pojemności do 128 GB. Nagrywa filmy w kilku trybach rozdzielczości:

- 5K: 2496×2496 (2 pliki) / szybkość 30fps; przepływność 100Mbps (50Mbps per file) : obraz 360^{o} bez łączenia
- 4K: 3840×2160 / szybkość 30fps; przepływność 80Mbps : obraz 360^{o} łączony
- 3.5K: 1760×1760 (2 pliki) / szybkość 60fps; przepływność 100Mbps (50Mbps na plik): obraz 360^{o} bez łączenia
- FHD: 1920×1080 / szybkość 120fps; przepływność 50Mbps, w formacie 16:9
- FHD: 1920×1080 / szybkość 60fps; przepływność 40Mbps, w formacie 16:9
- W trybie poklatkowym: w rozdzielczości 4K i 5K, w podziałką 2/5/10/30/60 sekund

W formacie 4K dostępna jest stabilizacja sferyczna. Maksymalny *bit rate* to 120 Mb/s. Wideo zapisane jest w rozszerzeniu .MP4. Dodatkowo zapisywane są metadane G-Metrix dla rzeczywistości rozszerzonej- kamera zawiera akcelerometr, barometr, żyroskop i kompas. Dostępna jest aplikacja na telefon, za pomocą której można sterować kamerą. W aplikacji przeznaczonej na komputer dostępna jest robocza, skompresowana wersja pliku i obróbka przy niższej rozdzielczości.

Dźwięk nagrywany jest przez kamerę z próbkowaniem 48 kHz na 4 kanałowy format ambisoni pierwszego rzędu z konwencją sortowania ACN i normalizacją SN3D. Kompresowany w AAC.

Rys 6.1 Kamera Garmin Virb 360

6.1.2 Mikrofon ZOOM H3-VR

Aby osiągnąć wyższą jakość dźwięku, do nagrań wykorzystano dodatkowo rejestrator ZOOM H3- VR, będący kompleksowym rozwiązaniem przeznaczonym do nagrań ambisonicznych, inspirowanym pierwowzorem Soundfield. Rejestrator zawiera 4 ukierunkowane kapsuły i rejestruje dźwięk przestrzennie, w 360 stopniach. Jednokierunkowe, kardioidalne mikrofony pojemnościowe przyjmują SPL do 120 dB.

Wbudowany potencjometr GAIN daje możliwość regulacji w zakresie od +18 do +48 dB. Do jego korekcji przewidziane jest wyjście słuchawkowe PHONE OUT na mini Jack. Odsłuch słuchawkowy jest podawany w trybie binaural. Maksymalne parametry rejestracji to 24 bity przy 96 kHz. Dostępne są trzy główne formaty nagrywania:

- Stereo: 2 kanałowy plik WAV przy rozdzielczościach: 44.1 kHz/16-bit, 44.1 kHz/24-bit, 48 kHz/16-bit, 48 kHz/24-bit, 96 kHz/16-bit, 96 kHz/24-bit

- Binaural: 2 kanałowy plik WAV przy rozdzielczościach: 44.1 kHz/16-bit, 44.1 kHz/24-bit, 48 kHz/16-bit, 48 kHz/24-bit (przy ustawieniach 96kHz/16bit lub 96kHz/24bit tryb binaural nie jest dostępny)
- Ambisonic: 4 kanałowy WAV przy rozdzielczościach: 44.1 kHz/16-bit, 44.1 kHz/24-bit, 48 kHz/16-bit, 48 kHz/24-bit, 96 kHz/16-bit, 96 kHz/24-bit. W zakresie nagrań ambisonicznych mamy do wyboru A-format oraz B-format (Fu/Ma lub AmbiX)

Do pamięci mikrofonu wymagana jest karta microSD/microSDHC/microSDXC klasy 4 lub wyższej. Przed zapisem urządzenie wymaga sformatowania karty. Możliwe jest podłączenie darty do 512 GB. Maksymalny rozmiar pliku to 2 GB (po przekroczeniu tej wielkości, automatycznie tworzy się następny plik z nagraniem, bez przerwania). Pliki WAV są wspierane przez BWF i iXML. Zasilanie zapewniają dwie baterie AA (koniecznie 1,5V- producent ostrzega, że litowe baterie 14500 3,7V mogą uszkodzić sprzęt). Kapsuły są dodatkowo otoczone konstrukcją chroniącą mikrofony przed mechanicznym zniszczeniem. Zestaw zawiera gąbkę mikrofonową (która została użyta we wszystkich nagraniach). Rejestrator jest wyposażony w wyświetlacz LCD 1,25”, dzięki któremu przede wszystkim możliwa jest konfiguracja urządzenia. Z najważniejszych opcji, na wyświetlaczu pojawia się: nazwa aktualnie nagrywanego pliku, długość obecnego nagrania i dostępna pozostała pamięć (w minutach), informacje o sferycznej orientacji mikrofonu oraz poziom GAIN. Pozycja mikrofonu jest odczytywana przez czujniki położenia przez 6-osiowy żyroskop. Możliwe jest podłączenie mikrofonu przez adapter Bluetooth BTA-1. Producent udostępnia aplikację na telefon, dzięki której można skalibrować sprzęt i sterować nim z odległości. Do ściągnięcia jest również dedykowana aplikacja na PC, która umożliwia odsłuch i konwersję nagrań. [33]

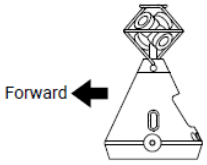
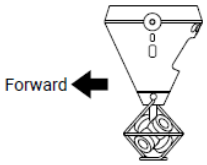
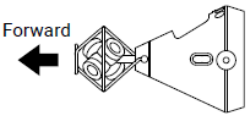
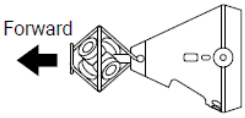
Rys. 6.2 Mikrofon ZOOM H3- VR

6.2 Klucz trasy, mapa i konfiguracja sprzętu

Kluczem w doborze rejestrowanych punktów były przede wszystkim topowe miejsca proponowane przez Google z subiektywną oceną przy doborze perspektywy oraz dodatkowo miejsca szczególnie atrakcyjne turystycznie lub charakterystyczne pod względem obrazu i dźwięku.

Kamera i mikrofon zostały zamontowane na statywie własnej konstrukcji, wykonanym tak, aby zminimalizować widoczność sprzętu na filmie.

Bardzo istotne jest umieszczenie mikrofonu i kamery w jednej osi i w zgodnym kierunku, aby obraz pokrywał się z kierunkowym dźwiękiem. W przypadku mikrofonu ZOOM H3 kierunek „do przodu" wskazuje czerwona dioda, a wyświetlacz LCD jest skierowany „do tyłu". W przypadku ustawień innych niż wybrany przeze mnie, mikrofon dostosowuje osi nagrania na podstawie danych z żyroskopu. Możliwe ustawienia są pokazane w instrukcji sprzętu:

Setting	Mic/recorder orientation	Explanation
Auto	–	The **H3-VR** automatically sets the mic position according to its orientation at the start of recording.
Upright		Use this setting to record with the **H3-VR** upright.
Upside Down		Use this setting to record with the **H3-VR** upside down.
Endfire		Use this setting to record with the **H3-VR** oriented horizontally with its display up.
Endfire Invert		Use this setting to record with the **H3-VR** oriented horizontally with its display down.

Rys 6.4 Pozycje mikrofonu ZOOM H3 [33]

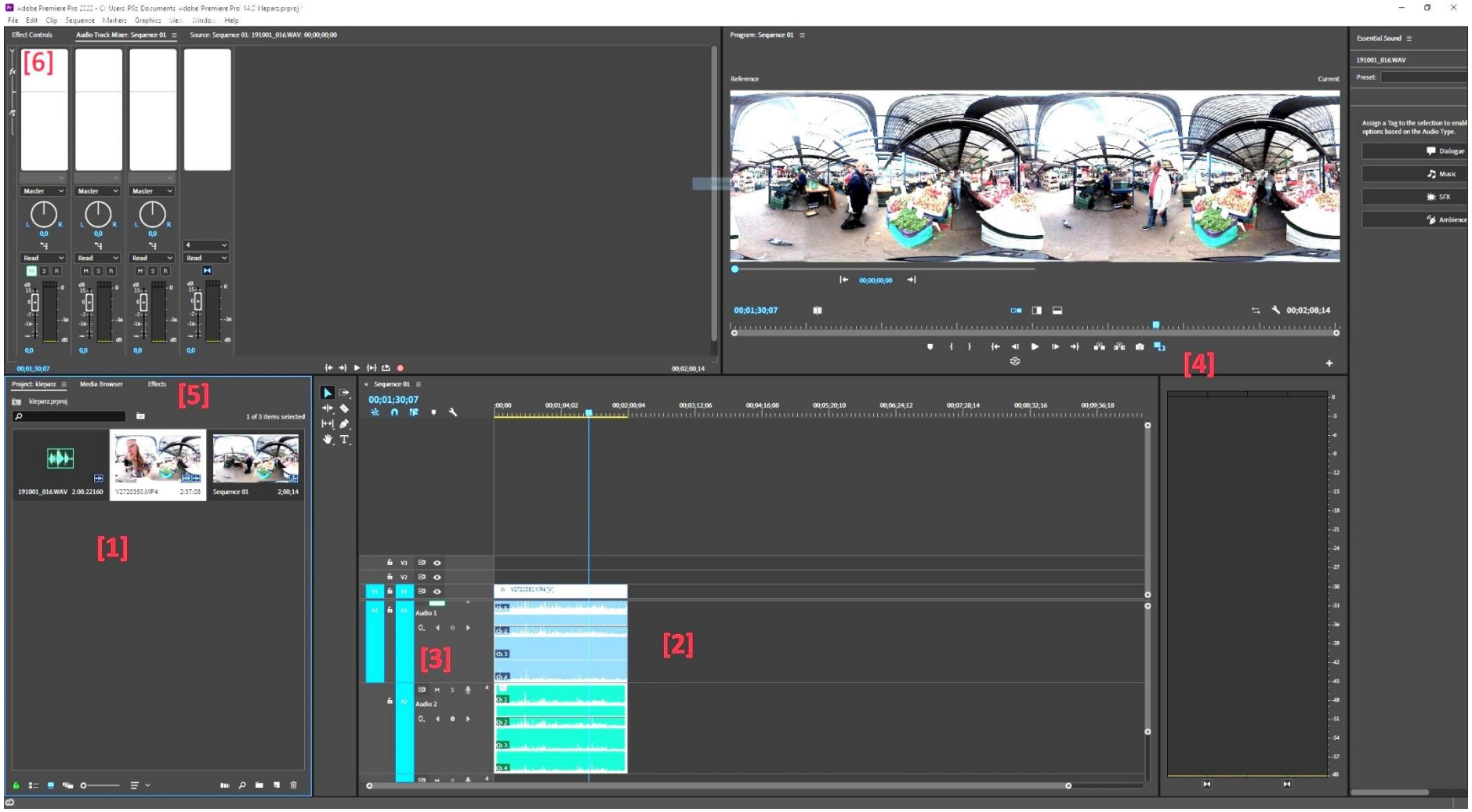
Mikrofon ma dodatkowy potencjometr do regulacji czułości (GAIN), które najlepiej ustawić przed samym nagraniem- przy nagraniach w terenie każdy punkt może mieć inne natężenie dźwięków otoczenia.

Aby dźwięk mógł być połączony z filmem w formie interaktywnego „poruszania się" razem z odbiorcą, konieczne było wykonanie nagrań dźwięku w konfiguracji mikrofonu dla B-formatu. Platforma YouTube wspiera wersję tego formatu w AmbiX, dlatego taką też wersję B-formatu wybrałam.

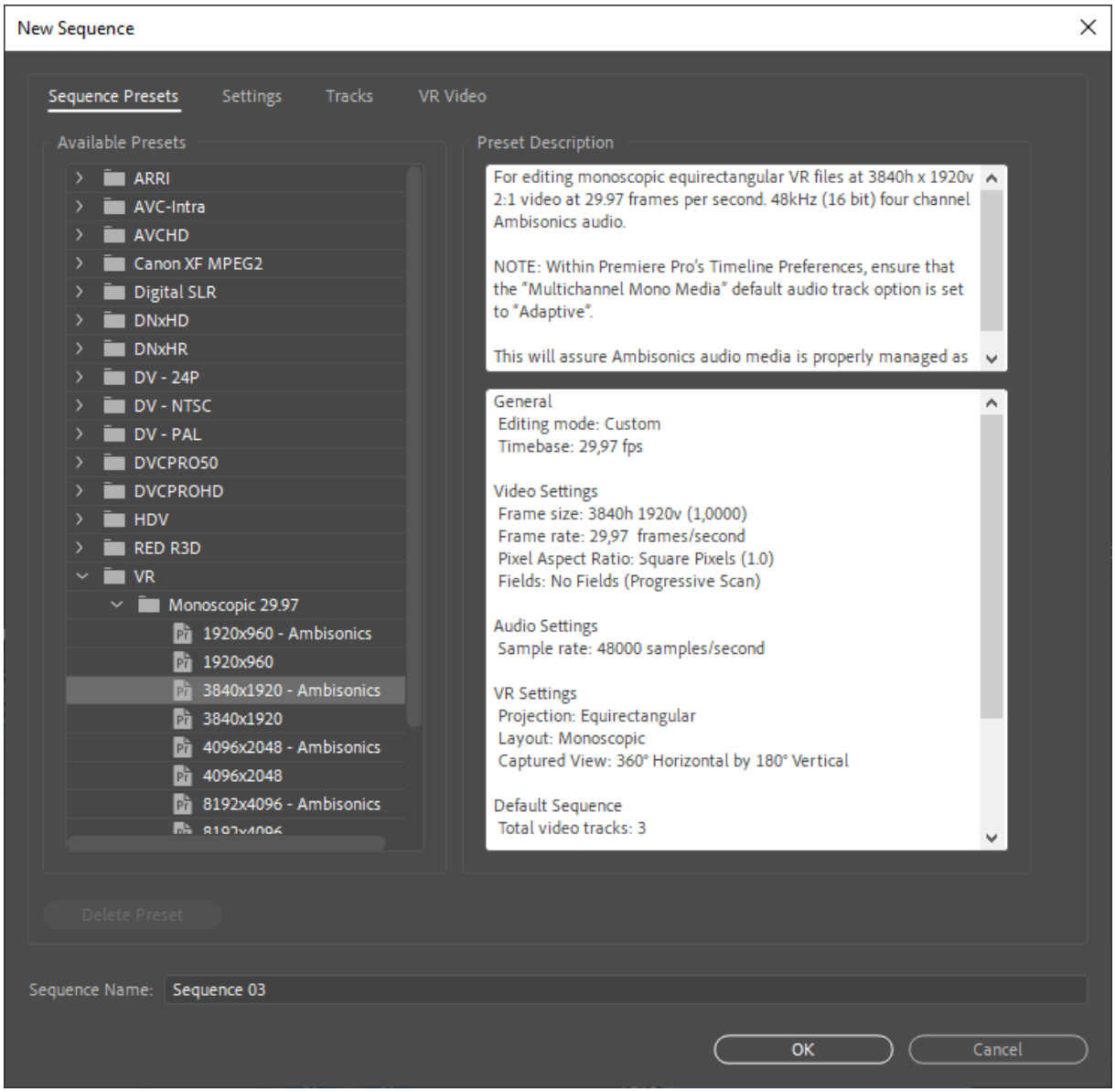
Nagrania filmów w formacie 4K w 360º, poza odpowiednim standardem karty, wymagały bardzo dużej pamięci. Można przyjąć że na jedną minutę nagrania potrzebne jest mniej więcej 1 GB pamięci.

Na podstawie kilku filmów instruktarzowych oraz instrukcji Google opisujących wymogi dotyczące nagrań filmów 360/VR z dźwiękiem ambisonicznym,

wykonałam montaż filmu i dźwięku w programie Premiere Pro w 7 dniowej wersji testowej. Alternatywą jest obróbka filmu i obrazu w niezależnych dedykowanych do tego programach (YouTube odsyła do strony instruktażowej jak w programie DAW Reaper wykorzystać polecane przez nich wtyczki do obróbki dźwięku ambisonicznego), a następnie splecenie ich bez podglądu. Wybrałam pracę w programie Premiere Pro, ponieważ jest to bardziej kompleksowe rozwiązanie zapewniające łatwą synchronizację filmu z dźwiękiem i podgląd efektów. Przedstawiam w punktach etapy obróbki pojedynczego filmu, wraz z wykonanymi przeze mnie kadrami kluczowych ustawień.

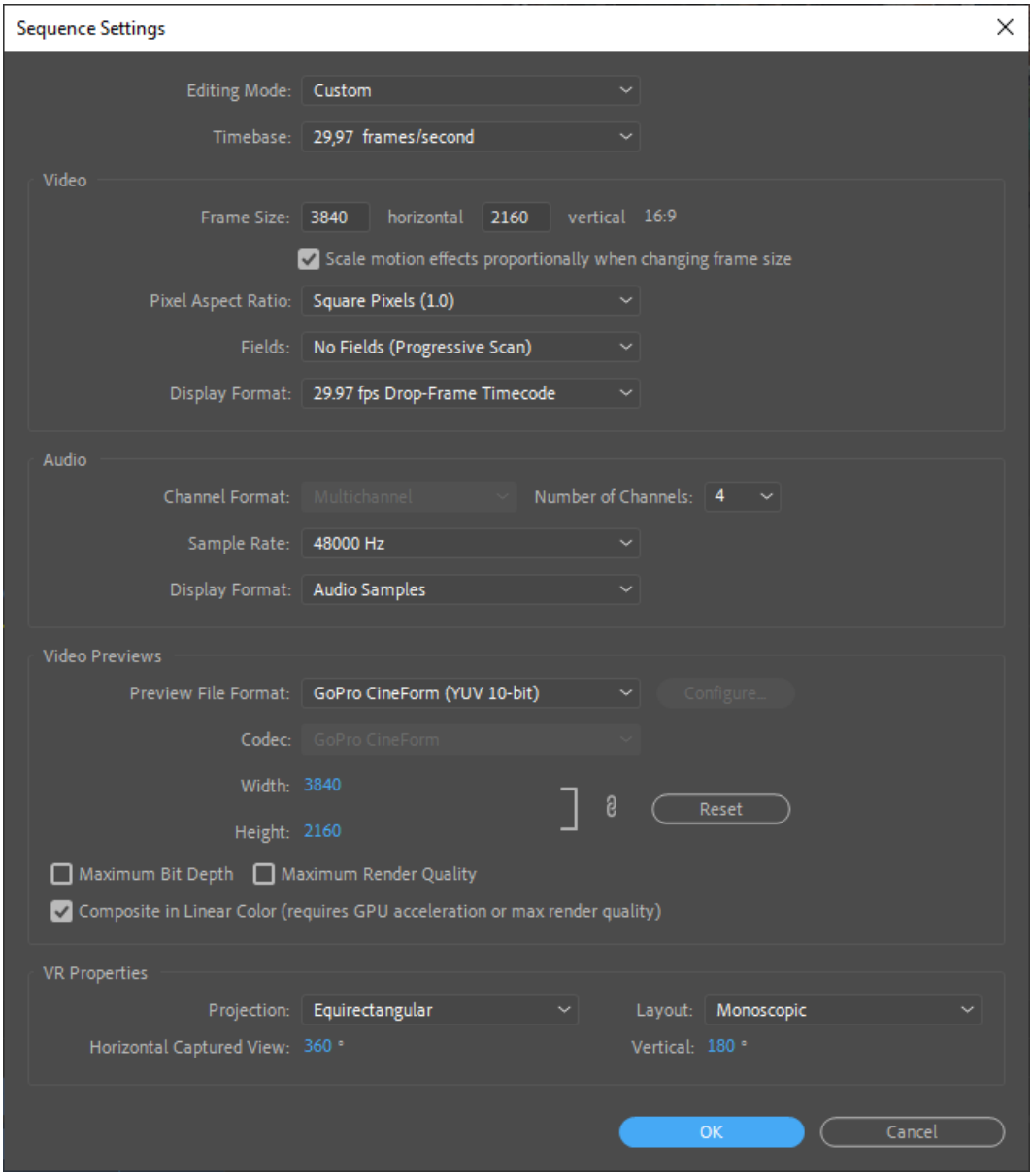

[6]
[5]
[4]
[1]
[2]
[3]
Effect Controls
Audio Track Mixer: Sequence 01
Source: Sequence 01: 191001_016.WAV: 00;00;00;00
Program: Sequence 01
Reference
Current
Essential Sound
Project: kleparz
Media Browser
Effects
Sequence 01
Audio 1
Audio 2
Dialogue
Music
SFX
Ambience

1) Zaimportowanie plików filmu AVI i nagrania WAVE z mikrofonu przeciągając je do bibioteki [1]
2) Tworzymy nową sekwencję wybierając: *File* > *New* > *Sequence*
3) Sekwencja powinna być przygotowana do pracy z obrazem 360 i dźwiękiem abisonicznym. Dlatego w jej ustawieniach [*Sequence Presets*] należy wybrać: *VR* > *Monoscopic* > *3840x 1920- Ambisonic* (dla kamery i formatu w którym wykonałam nagrania; należy wybrać rozdzielczość najbliższą naszemu nagraniu). Nazywamy sekwencję i wybieramy *Ok* aby ją stworzyć.

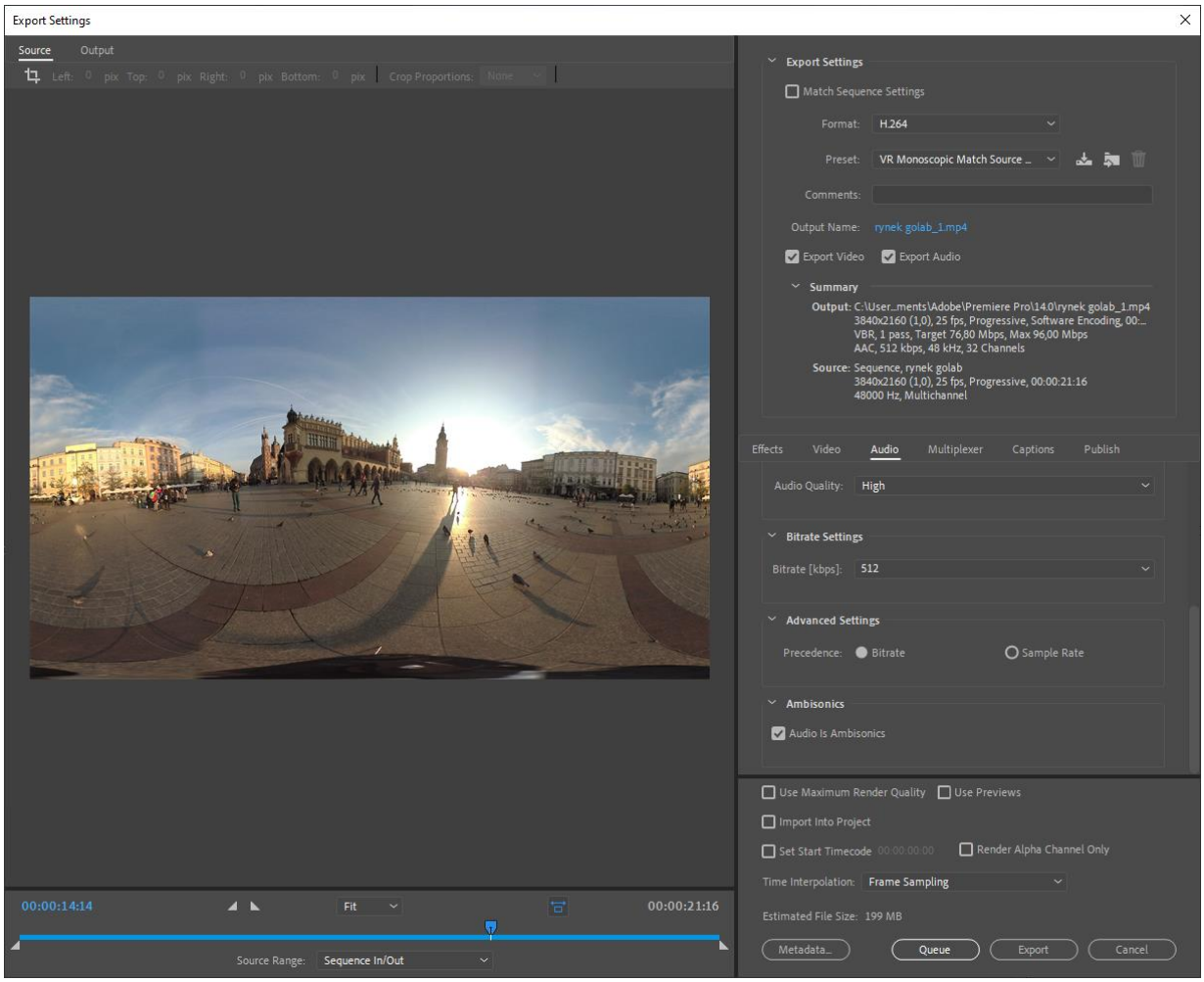


4) Program uprzedza, że przy takim wyborze należy dostosować *Multichanel* w ustawieniach audio. W przypadku najnowszej wersji programu ta opcja nie znajduje się w zakładce ustawień audio lecz w *Timeline*. Należy więc wybrać: *Edit* >*Preference* > *Timeline* > *Default Audio Tracks* > *Multichanel Mono Media* >*Adaptive*

5) Dopasowanie formatu wideo dokładnie do mojego filmu wykonuję w: *Sequence* > *Sequence Settings* > *Video* > *Frame Size* > …. w moim przypadku jest to 3840 x 2160. W tym samym miejscu można sprawdzić ustawienia audio. Zgodność częstotliwości próbkowania i dla pewności liczbę kanałów- których dla ambisoni powinno być 4.



6) Teraz możemy wczytać film i dźwięk do sekwencji, przeciągając je na odpowiednie ścieżki [2]. W programie Premiere Pro synchronizacja materiałów jest bardzo ułatwiona. Należy Zaznaczyć wszystkie tracki, kliknąć na nie prawym klawiszem myszy i wybrać: *Synchronize* > *Audio* > *Track Channel* > *Mix Down* >*Ok*
7) Wyciszam ścieżkę dźwiękową nagraną przez kamerę włączając na niej *Mute* [3].
8) Dla podglądu filmu w formie 360^{o} należy zaznaczyć opcję *Toggle VR Video Display* [4].
9) Aby dźwięk podążał za obrazem w zakładce konieczne jest założenie dodatkowego filtra na ścieżkę dźwiękową. W tym celu w oknie *Effect* [5] wybieram z folderu *Audio Effects* opcję *Panner- Ambisonics* i przeciągam ją na ścieżkę dźwiękową na tracku [2].
10) Na tym etapie najlepiej sprawdzić czy dźwięk jest w zgodnej osi z obrazem. Jeśli nie to w zakładce *Effect Controls* [6] trzeba obrócić w założonym efekcie *Panner- Ambisonics* obrócić odpowiednio kąt panoramy *Pan*. Dla właściwego odsłuchu przestrzennego, na track *Master* należy założyć filtr *Spacial* > *Binauralizer -Ambisonics*. Ważne żeby usunąć go przed eksportem pliku.
11) Gotowy film eksportujemy przez: *File* > *Export* > *Media…* co otwiera kolejne okno opcji.



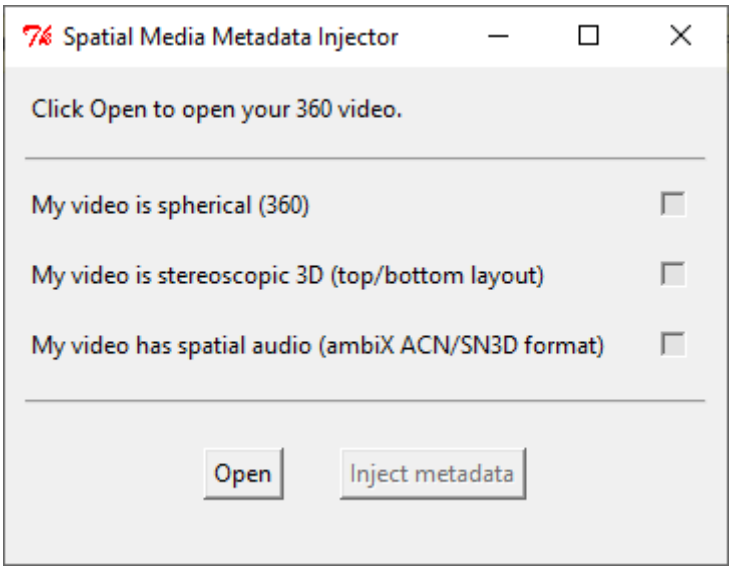
12) W tym miejscu format powinien być ustawiony na *H.264*, a preset na *Vr Monoscopic Maych Source- Ambisonics*. Przy takim wyborze reszta ustawień powinna sama się dostosować. Dla pewności- w zakładkach *Video* i *Audio* można sprawdzić wszystkie wybrane wcześniej konfiguracje. Na końcu tych zakładek zaznaczone powinny być opcje:

    Video> *Video is VR*, *Audio* > *Audio is Ambisonics*.

    Na samym dole okna wyświetla się szacowany rozmiar pliku po eksporcie.

13) Esportowany plik zapisuje się automatycznie w folderze dedykowanym programowi w dokumentqach użytkowanika. Zawiera audio w formacie AAC o przepływności 512 kbps, rozdzielczości 48 kHz.

14) Wyeksportowany plik trzeba zaopatrzyć w metadane zgodne ze standardami YouTube. Program jest dostępny na stronie instruktażowej za darmo- *Spacial Media Metadata Injector*.

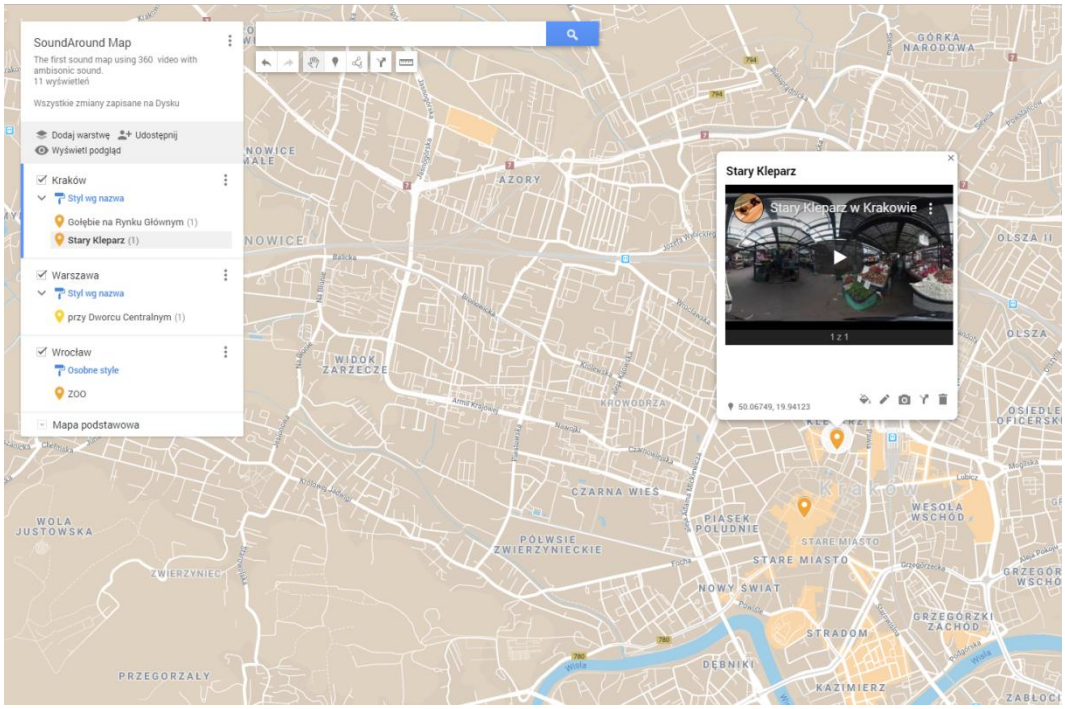


15) Wybierając *Open* wybieram plik z filmem i klikam *Inject Metadata*. Plik zapisuje się z końcówką *_injected* w tej samej lokalizacji co plik oryginalny. W tej formie film jest gotowy do publikacji.

6.3 Publikacja projektu

Zebrane filmy mają zostać umieszczone na mapie, a najwygodniejszym rozwiązaniem w pierwszym etapie projektu okazały się być Mapy Google. Jest to optymalne rozwiązanie umożliwiające wyznaczenia dokładnego punktu na mapie i załączenia do niego opisu oraz filmu z YouTube. Aby stworzyć taką mapę wymagane jest konto Google, do którego będzie ona przypisana. W tym przypadku jest to sprawa bardzo wygodna i intuicyjna, jedyna zaskakująca rzecz to fakt, ze Mapy Google to nie ta sama strona co Moje Mapy Google- z poziomu Map Google nie tworzymy swojej mapy, możemy jedynie zapisać swoje miejsca. Strona Moje Mapy Google to oddzielna platforma, w której tworzy się swój projekt od nanoszenia punktów i odnośników do wyboru swojego stylu mapy.

1) Aby najpierw opublikować film, na swoim koniecie YouTube wybieram opcje *Prześlij film*. Poza wyborem własnej nazwy i dodaniem opisu oraz tagów strona nie wymaga już dodatkowych ustawień.
2) Film zostaje najpierw przesłany, a później (o czym uprzednio informuje Youtube) dodatkową chwilę zajmuje mu jego publikacja wformie 360. W efekcie dostajemy link do udostepnienia filmu.
3) Na stronie *Moje Mapy Google* tworzę nową mapę [1].
4) W punkcie nagrania wstawiam pinezkę klikając *Dodaj znacznik* [2].
5) W tym miejscu mogę dodać swoją nazwę i opis oraz załączyć zdjęcie z dysku, z adresu URL lub adres URL do filmu na YouTube. Wybieram tą opcję i wczytuję film na YouTube.
6) Tak opublikowany punkt na mapie otwiera się jako miniaturka filmu- o tym czy już w formie 360 decyduje przeglądarka lub ustawienia telefonu. Klikając w logo YouTube na filmie możemy otworzyć go w pełnym rozmiarze na stronie. W przypadku przeglądarek na telefonie wyświetla się ostrzeżenie, że forma filmu dookólnego nie zostanie zachowana i zaleca się przejście do aplikacji YouTube. Tam możemy zobaczyć film w ostatecznej formie. Aplikacja wyświetla powiadomienie o wykorzystaniu dźwięku dookólnego i zaleca podłączenie słuchawek.



W związku z tym, że nagrania są wykonywane w miejscach bardzo popularnych, powstało pytanie czy nie ma przeciwskazań publikacji filmów zawierających wizerunki osób na nagraniu. Zgodnie z polskim prawem- art. 81 ust. 2 pkt 2, dopuszczone jest rozpowszechnianie wizerunku jeżeli stanowi on jedynie szczegół całości. Rozstrzygające jest tutaj ustalenie relacji między wizerunkiem, a pozostałymi elementami. Nagrania były wyselekcjonowane przeze mnie dodatkowo pod tym kątem- tak aby zawierały jak najmniej absorbujących postaci na filmie, a skupiały się na miejscu jako takim. Sprawy sądowe rozpatrywane pod tym kątem, opierają się na tymże artykule prawnym, w którym do opisu dozwolonych okoliczności nagrania użyte jest sformułowanie „takie jak…"- zgromadzenie , publiczna impreza czy krajobraz. Moja praca zdecydowanie wpisuje się w ten ostatni przykład. Przykładem takiej sprawy sądowej i jej rozstrzygnięcia jest wyrok Sądu Apelacyjnego w Warszawie z dnia 10.02.2005 r., I ACa 509/04, LEX nr 535042. [21]

Na ten moment (25.11.2019r.) mapa jest w dostępna w formie projektu na stronie Moje Mapy Google z publicznym dostępem do wyświetlania pod linkiem:

https://www.google.com/maps/d/viewer?mid=1ZC476y2Qn2vPUTYAnk8v1uDInNsHeteM&ll=51.1585952026431%2C19.017803649999905&z=7

Rozdział 7: Wnioski i perspektywy rozwoju projektu

Podsumowując, wykonana przeze mnie mapa dźwiękowa jest pierwszym zastosowaniem najnowszych, coraz szerzej wykorzystywanych technologii, w tej tematyce. Nagrania opierające się na dźwięku ambisonicznym w połączeniu z obrazem w 360 stopniach dają lepszą lokalizację źródeł dźwięku i wyraźniejszy obraz otoczenia, atmosfery, charakteru danego miejsca. Pomimo dostępności instrukcji i coraz łatwiejszego dostępu do odpowiedniego sprzętu, połączenie tych możliwości nie jest jeszcze tak popularne. Publikowane są niezależnie instrukcje dotyczące zarówno dźwięku przestrzennego jak i obrazu dookólnego. Tematem pierwszoplanowym jest w nich albo dźwięk, albo obraz. Tymczasem dopiero dobra jakość dźwięku w połączeniu z właściwym, odpowiadającym mu obrazem, daje wyraźne, realne wyobrażenie o przestrzeni dźwiękowej. Przy realizacji tego projektu udało mi się zebrać zalety najnowszych możliwości technicznych w jedno rozwiązanie i przedstawić optymalną drogę do jego osiągnięcia. Wiele z wyborów zostało podyktowane popularnością, która wiąże się w tym przypadku z optymalizacją. Wybór formatu dźwięku ambisonicznego, z technicznego punktu widzenia, nie określa jakości, a jedynie konwencję, więc został dostosowany do wymogów publikacji. Rodzaj kompresji wideo, przy obecnie wysokich możliwościach, również nie wpłyną znacząco na jakość, a jedynie wpisywał się w konwencję. Wybrany sposób publikacji daje możliwość wygodnego rozpowszechniania projektu, dzięki sprzężeniu kilku najpopularniejszych form wyszukiwania (Google, YouTube, Google Maps). Efekty pracy dały zadowalające wyniki, już pełniące swoją funkcję.

Następnym krokiem dla rozwoju projektu będzie stworzenie własnej strony internetowej poświęconej tematyce map dźwiękowych. Równocześnie, należałoby rozwijać bazę nagrań zarówno w Polsce jak i za granicą. Na podstawie przeglądu istniejących map dźwiękowych można zauważyć jak ważna jest przejrzystość i spójność strony, ponieważ dla osób z poza branży, tematyka ta może nie być intuicyjna. Przy tym, często do takich ludzi kierowana będzie owa prezentacja. Docelowo warto byłoby umieścić efekty pracy na stronie miasta, w formie reklamy,

wizytówki, ciekawostki, archiwizacji. W tym miejscu ergonomia i atrakcyjność oprawy graficznej nabierze jeszcze większego znaczenia. Na ten moment celem było uskutecznienie aktualnych technik z zakresu dźwięku i obrazu na rzecz map dźwiękowych. Cel ten został wykonany. Kolejnym wyzwaniem jest wdrożenie projektu do szerszego użytku, a zatem- projekt strony www, uwzględniający jej funkcjonalność z zastosowanymi tu technikami, rozszerzenie możliwości odbioru projektu (podróż wirtualna z Oculus Rift lub interaktywna instalacja w muzeum) oraz marketing.

Bibliografia

1. Marek Korbecki- „Komputerowe przetwarzanie dźwięku” wyd. Mikom; lipiec 1999
2. Wojciech Butryn- Dźwięk cyfrowy wyd WKŁ Warszawa 2002
3. https://depot.ceon.pl/bitstream/handle/123456789/8863/lokalizacja_pozornych_zrodel_dzwieku_w_nagraniach_binauralnych.pdf?sequence=1&isAllowed=y
4. Sztuka dźwięku technika i realizacja – Małgorzata Przedpełska-Bieniek wyd. Wojciech marzec Wawa 2017
5. Livesound.pl
6. http://www.highfidelity.pl/@numer--91&lang=
7. https://itszkola.edu.pl/materialy/ebibl/multimedia_technologie/Multimedia_technologie_internetowe_bazy%20danych_i_sieci_komputerowe_Metody_kodowania_i_przechowywania_sygnalow_dz%20wiekowych.pdf [październik 2019]
8. http://hifistudio.pl/artykuly-poradnikowe/formaty-plikow-audio-zestawienie/
9. https://encyklopedia.pwn.pl/haslo/Webera-Fechnera-prawo;3994481.html
10. Ryszard Gubrynowicz- „Narząd słuchu jako analizator akustyczny” wykład PJWSTK
11. Zeszyt dydaktyczny „Techniki nagrywania, kształtowania i odtwarzania dźwięku”- Andrzej Majkowski; Warszawa 2009
12. Seminarium „Head-Related Transfer Function”- Tilen Potisk; styczeń 2015
13. R. Murray Schafer- „The New Soundscape”
14. https://3dsound.info/Jak-ludzie-lokalizuja-dzwieki.html
15. Artykuł “Lokalizacja pozornych źródeł dźwięku w nagraniach binauralnych” – Adam Rosiński; Akademia Sztuk Pięknych w Gdańsku
16. European Broadcasting Union, Listening conditions for the assessment of soundprogramme material: monophonic and two–channel stereophonic; Genewa 1998
17. https://www.soundonsound.com/techniques/surround-sound-explained-part-3
18. Praca magisterska “Dopasowanie toru transmisji dźwięku do warunków odsłuchowych”- M. Kucharski, T. Stankiewicz; AGH Kraków 2013
19. http://www.krajobraz.kulturowy.us.edu.pl/publikacje.artykuly/niematerialne/kolacki.pdf
20. http://www.dsp.agh.edu.pl/_media/pl:dydaktyka:pkluska_praca_inzynierska.pdf
21. https://gdpr.pl/zdjecia-z-eventow-rozpowszechnianie-wizerunku

22. „Podstawy nagłośnienia i realizacji nagrań”- K. Sztekmiller; WKŁ 2008
23. Wykład „Percepcja i wartościowanie głośności” –Janusz Renowski
24. Wykład „Metody kodowania i przechowywania sygnałów dźwiękowych” –Andrzej Majkowski
25. Wykład nr. 11 „Dźwięk w multimediach”- Ryszard Gubrynowicz
26. https://www.comsol.com/blogs/analyzing-the-acoustics-of-a-head-and-torso-simulator/
27. https://www.x-kom.pl/poradniki/4839-proba-mikrofonu-doradzamy-jaki-mikrofon-do-komputera-wybrac.html
28. Prezentacja „Techniki mikrofonowe i cyfrowy stół mikserski”- Marcin Lewandowski
29. https://www.audiofanatyk.pl/podstawy-akustyki-i-rozmieszczania-glosnikow-stereo/
30. https://en.wikipedia.org/wiki/Ambisonics
31. https://www.technics.com/pl/high-res-audio/co-to-jest-high-resolution-audio.html
32. Wykład „Pantophonic Recording” – David Pickett
33. „Quick guide” do mikrofonu ZOOM H3- VR

Wszystkie rysunki zostały zaczerpnięte z podanych w bibliografii tytuów oraz stron internetowych dostępnych na dzień 20.11.2019r. Zdjęcia z rys. 3.8, 3.9, 4.7, 6.1 i 6.2 są zaczerpnięte ze stron firm dystrybuujących dany sprzęt.