AMBISONICZNA MAPA WYBRANYCH MIEJSC W TRÓJMIEŚCIE

Cezary PIETRZAK[1]
Piotr ODYA[2]
Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki,
Katedra Systemów Multimedialnych, ul. Gabriela Narutowicza 11-12, 80-233 Gdańsk
[1]cezarypiotrpietrzak@gmail.com
[2]pioodya@pg.edu.pl

Projekt miał na celu stworzenie ambisonicznej mapy Trójmiasta w formie aplikacji internetowej. Materiały wideo w technologii 360 z dźwiękiem w postaci sygnału ambisonicznego zostały zarejestrowane w lokalizacjach Trójmiasta, które uznano za charakterystyczne dla tej aglomeracji. Celem badawczym projektu było porównanie dostępnych algorytmów miksowania sygnałów ambisonicznych poprzez przeprowadzenie testów odsłuchowych. Przeprowadzono test porównań parami, aby uzyskać informacje o preferencjach odnośnie występowania w nagraniach ambisonicznych dodatkowego podkładu stereo oraz jego poziomu głośności. W drugim z testów zbadano jaki wpływ na odbiór nagrania ambisonicznego ma sposób oraz stopień jego przetwarzania. Wnioski otrzymane poprzez analizę wyników obu testów posłużyły jako wskazówki przy postprodukcji nagrań. Otrzymane materiały zostały zamieszczone na interaktywnej mapie w aplikacji internetowej.

1. WPROWADZENIE

Pojęcie systemu ambisonicznego po raz pierwszy pojawiło się w latach siedemdziesiątych ubiegłego wieku. W roku 1974, w magazynie *Studio Sound* ukazał się artykuł zatytułowany *What's wrong with quadrophonics?*, w którym Michael Gerzon wskazał wady popularnego wówczas systemu kwadrofonicznego. Wśród nich znalazły się między innymi problemy z odtwarzaniem lokalizacji poszczególnych źródeł dźwięku, stosunkowo mała powierzchnia dobrego odsłuchu czy podatność na zmiany tego obszaru, który był silnie uzależniony od pozycji słuchacza. W tym samym artykule Gerzon przedstawił również alternatywną koncepcję dla kwadrofonii, która dziś znana jest jako system ambisoniczny. Pomimo pracy włożonej przez Gerzona i jego współpracowników w rozwój ambisonii oraz jej zalet względem systemów kwadrofonicznych, nie przyjęła się ona wówczas w przemyśle nagraniowym [6] [11].

Jednak postępujący rozwój technologiczny oraz powszechna dostępność materiałów audiowizualnych w Internecie sprawiły, że nagrania ambisoniczne znalazły swoje miejsce w wydawałoby się dostatecznie złożonym i różnorodnym globalnym zbiorze multimediów. Szczególnie w ostatniej dekadzie nastąpił wzrost zainteresowania tym formatem dźwięku przestrzennego o czym świadczą jego implementacje w filmach 360 oraz rzeczywistości wirtualnej w serwisie społecznościowym *Facebook* oraz na platformie *YouTube*. W obu przypadkach określone zostały również wytyczne dotyczące publikowania materiałów zawierających dźwięk w postaci ambisonicznej odnośnie wspieranych formatów oraz wymaganych dodatkowych metadanych [3] [16].

W niniejszym artykule opisany został projekt, w którym za cel postawiono stworzenie ambisonicznej mapy Trójmiasta w formie aplikacji internetowej. Część badawcza pracy polegała na porównaniu dostępnych algorytmów miksowania sygnałów ambisonicznych. W tym celu przeprowadzono dwa testy odsłuchowe, których wyniki zostały następnie przeanalizowane i posłużyły jako wskazówki w dalszym

procesie obróbki nagrań. Otrzymane materiały zostały zamieszczone w aplikacji internetowej z interaktywną mapą, dostępną online.

2. REALIZACJA NAGRAŃ

Wybór urządzeń do przeprowadzenia nagrań podyktowany był mobilnością sprzętu, przy zachowaniu dostatecznie wysokiej jakości materiałów audio i wideo. Podczas realizacji skorzystano z mikrofonu ambisonicznego pierwszego rzędu *ZOOM H3-VR* z wbudowanym rejestratorem oraz z kamery *Insta360 ONE X*, które przedstawiono na rysunku 1. Oba urządzenia w czasie nagrań mocowano w jednej osi na dwustronnym uchwycie, kamera była skierowana do góry, a mikrofon do podłoża. Natomiast sam uchwyt montowany był na statywie.

Rys. 1. Mikrofon ambisoniczny pierwszego rzędu - ZOOM H3-VR oraz kamera 360 – Insta360 ONE X [9] [21]

W przypadku kamery zdecydowano się na ustawienie rozdzielczości 5.7K przy 30 klatkach na sekundę. Dodatkowo włączony został również tryb HDR, aby zwiększyć zakres dynamiki obrazu w nagraniu i uniknąć potencjalnych prześwietleń np. podczas rejestracji na pełnym słońcu. Natomiast dla sygnałów audio jako format wyjściowy ustawiono B-format przy częstotliwości próbkowania 96kHz i rozdzielczości 24 bity. Automatyczna konwersja rejestrowanego nagrania do odpowiedniej postaci, uwzględniającej orientację mikrofonu, była możliwa dzięki wbudowanemu w mikrofon żyroskopowi. Podczas nagrań korzystano również z możliwości monitorowania sygnału w czasie rzeczywistym poprzez wyjście słuchawkowe urządzenia, co pozwoliło na kalibrowanie sygnału [9] [21].

Materiały zarejestrowano w dwunastu wybranych miejscach, które uznano za charakterystyczne dla krajobrazu Trójmiasta:

- Gdynia: Port Gdynia (przy Muzeum Emigracji), Skwer Kościuszki, Bulwar Nadmorski im. Feliksa Nowowiejskiego, Gdynia Orłowo,
- Sopot: Park Północny, Molo, ul. Bohaterów Monte Cassino,
- Gdańsk: Park Oliwski, Nowy Port - Kapitanat Portu Gdańsk, Westerplatte, Długi Targ¸ Stogi - Plaża.

Aby w pełni oddać klimat danego miejsca przy jak najmniejszych zakłóceniach spowodowanych potencjalną obecnością ludzi i wzmożonym ruchem ulicznym, nagrania realizowano w godzinach porannych. W efekcie otrzymano szesnaście zestawów nagrań dźwiękowych oraz wideo zarejestrowanych we wszystkich wybranych wcześniej lokalizacjach.

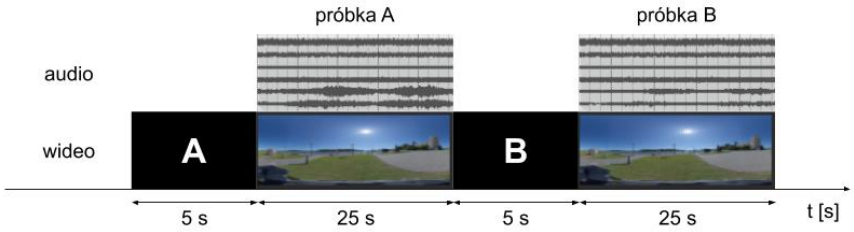
3. TEST PORÓWNAŃ PARAMI

Serwisem, w którym umieszczone miały zostać docelowe nagrania była platforma YouTube. Poza standardową opcją przesyłania plików wideo ze ścieżką audio mono lub stereo, serwis ten umożliwia

również publikowanie materiałów 360 z dźwiękiem przestrzennym. Obsługiwane są dwa następujące formaty: ambisonia pierwszego rzędu oraz ambisonia pierwszego rzędu z dźwiękiem stereo. Przy czym drugi z wymienionych formatów charakteryzuje się stałą pozycją ścieżki stereo, bez względu na ruch głowy słuchacza w danej chwili. Z uwagi na tę możliwość zdecydowano się na przeprowadzenie testu odsłuchowego, mającego na celu sprawdzenie preferencji użytkowników odnośnie występowania podkładu stereo oraz poziomu jego głośności w materiale 360 z dźwiękiem przestrzennym. Hipotezą postawioną w teście było założenie, że potencjalny odbiorca nagrania ambisonicznego, zarejestrowanego w przestrzeni publicznej, preferuje wersję pozbawioną dodatkowego podkładu muzycznego [16].

Jako formułę testu wybrano metodę porównań parami, jednak zamiast porównywania jakości próbek przed uczestnikami postawiono pytanie, która z prezentowanych próbek jest lepszej jakości. Dodatkowo zastosowano również powtórną rundę z zamienioną kolejnością próbek w celu zweryfikowania ocen testujących. Było to podyktowane założeniem, że uczestnikami testu będą osoby niebędące ekspertami. Również z tego względu liczność grupy testowej określono na dwadzieścia osób [2].

Zdecydowano się na przygotowanie sześciu próbek audiowizualnych, z których utworzono osiem par, w tym cztery z zamienioną kolejnością. Każda para prezentowana była w układzie przedstawionym na rysunku 2. Z zarejestrowanych wcześniej nagrań utworzono po cztery próbki testowe dla dwóch wybranych miejsc o różnych poziomach głośności podkładu muzycznego, które zostały oznaczone literami ABCD oraz numerem miejsca (1 lub 2). Informacje o poszczególnych próbkach oraz dobranych parach przedstawiono w tabelach 1 oraz 2.



Rys. 2. Struktura prezentacji próbek w przeprowadzonym teście preferencji dwójkowych

Tab. 1. Próbki użyte w teście porównań parami

Próbka	Podkład stereo	Miejsce nagrania
A1	brak podkładu	Gdynia – Bulwar Nadmorski im. Feliksa Nowowiejskiego
B1	z podkładem	
C1	z podkładem -6 dB	
A2	brak podkładu	Gdańsk – Kapitanat Portu Gdańsk
B2	z podkładem -3 dB	
C2	z podkładem -6 dB	

Tab. 2. Pary próbek testowych dla testu porównań parami

Nr pary	1	2	3	4	5	6	7	8
Próbka A	A1	A2	B1	B2	C1	A2	B1	C2
Próbka B	B1	C2	C1	A2	B1	B2	A1	A2

Ze względu na brak dedykowanej platformy do przeprowadzania oceny subiektywnej materiałów 360, test został utworzony przy pomocy aplikacji *Google Forms*, pozwalającej na stworzenie formularza, który może zawierać materiały wideo zamieszczone wcześniej w serwisie *YouTube*. Test podzielono na dziesięć sekcji. Pierwsza z nich, traktowana jako sekcja wstępna, zawierała opis przebiegu testu oraz dodatkowe informacje dotyczące konieczności odsłuchu za pośrednictwem słuchawek, obracania obrazem podczas prezentacji próbek oraz ograniczeń *Google Forms* związanych z brakiem możliwości odtwarzania próbek testowych w trybie pełnego ekranu. Kolejne osiem sekcji przeznaczono do prezentacji par próbek testowych, z czego w każdej z nich znajdował się odtwarzacz YouTube oraz pytanie

jednokrotnego wyboru dotyczące preferowanej przez słuchacza wersji. Na końcu formularza zamieszczono również ankietę z pytaniami dotyczącymi doświadczeń uczestników testu w zakresie realizacji i odbioru nagrań 360. Łączny czas trwania testu dla jednej osoby wynosił 8 minut. Między obiema częściami (rundami) nie stosowano przerwy.

W teście udział wzięło łącznie dwadzieścia jeden osób. Zgodnie z informacjami udzielonymi przez respondentów w ankiecie umieszczonej na końcu testu, jedynie sześciu z nich miało wcześniej doświadczenia w realizacji materiałów audiowizualnych, w tym tylko jedna osoba przy rejestracji wideo w technologii 360. Można więc uznać, że grupa testowa składała się z osób niebędących ekspertami. W tabeli 3 przedstawiono zsumowane odpowiedzi wszystkich uczestników w każdej prezentowanej parze, przy czym każda z próbek mogła otrzymać maksymalnie 21 głosów w kontekście jednej pary.

Tab. 3. Wyniki przeprowadzonego testu porównań parami

I część testu								II część testu							
numer pary															
1		2		3		4		5		6		7		8	
A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
A1	B1	A2	C2	B1	C1	B2	A2	C1	B1	A2	B2	B1	A1	C2	A2
13/21	8/21	14/21	7/21	7/21	14/21	7/21	14/21	13/21	8/21	16/21	5/21	7/21	14/21	9/21	12/21

Na podstawie oceny stabilności odpowiedzi każdego z uczestników z dalszej analizy wykluczono wyniki osób, które udzieliły odpowiedzi zgodnych w mniej niż trzech przypadkach. Następnie dla tak ograniczonego zbioru obserwacji wyznaczono liczbę odpowiedzi oddanych na każdą z próbek testowych, zarówno osobno dla obu części testu, jak i dla całego testu.

Z zaprezentowanych w teście par można wyodrębnić ich dwa rodzaje: pary, w których należało dokonać wyboru pomiędzy próbką bez podkładu muzycznego a próbką, w której podkład występuje oraz pary zawierające obie próbki z podkładem muzycznym o różnym poziomie głośności. W pierwszym z wymienionych rodzajów par uczestnicy wybierali wersję bez podkładu średnio w 14 na 21 przypadkach. W takim samym stosunku wybór ekspertów padał na próbki zawierające cichszą wersję podkładu dla par, w których nie występowała próbka bez podkładu.

Korzystając z otrzymanych danych wyznaczono statystykę chi kwadrat, porównującą wyniki obu części testu. Na jej podstawie można określić, czy istnieje istotna różnica między odpowiedziami uczestników w obu częściach testu. Wartość chi kwadrat obliczono według wzoru [10]:

$$X^2 = n \cdot \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}\right)^2}{n_{i\cdot} \cdot n_{\cdot j}} = 0{,}735 \tag{1}$$

gdzie:
r – liczba części testu,
s – liczba badanych obiektów,
n_{ij} – liczba wyborów j-tego obiektu w i-tej części testu,
$n_{i\cdot}$ – suma wyborów w i-tej części testu,
$n_{\cdot j}$ – suma wyborów j-tego obiektu we wszystkich częściach testu,
n – suma wszystkich wyborów podczas testu.

Uzyskana wartość statystyki chi kwadrat udowadnia, że przyjęto odpowiedni minimalny poziom zgodnych odpowiedzi udzielonych przez osoby predystynowane do bycia ekspertem równy 75%, a tym samym potwierdza właściwy dobór ekspertów do dalszej analizy wyników.

Statystyczną zgodność wyników określono poprzez porównanie obliczonej wartości chi kwadrat z wartością krytyczną, odczytywaną z tablic rozkładu tej statystyki dla wyznaczonej liczby stopni swobody oraz obranego poziomu istotności. Dla przyjętego poziomu istotności równego 0,05 oraz wyznaczonej liczby stopni swobody odczytano wartość krytyczną statystyki chi kwadrat równą 11,0705.

Obliczona wartość statystyki chi kwadrat nie przekracza wartości krytycznej, zatem można potwierdzić hipotezę o zgodności wyników. Następnie określono istotności różnic w wynikach dla poszczególnych par w obu turach, które zostały wyliczone według wzoru [10]:

$$z_{ij} = \frac{|p_i - p_j|}{\sqrt{\frac{(p_i + p_j) \cdot (2 - p_i - p_j)}{2 \cdot N}}} \tag{2}$$

gdzie:
z_{ij} – określone prawdopodobieństwo,
p_i – względna liczba głosów na i-te obiekty,
p_j – względna liczba głosów na j-te obiekty,
N – maksymalna liczba głosów, którą może otrzymać jeden obiekt,
n – liczba porównywanych obiektów,
m – liczebność grupy ekspertów.

Otrzymane wartości porównano z wartością graniczną z tablic rozkładu normalnego, która dla przyjętego poziomu istotności równego $\alpha = 0{,}05$ wynosi $z(\alpha) = 1{,}96$. Jeżeli wartość z_{ij} jest mniejsza od $z(\alpha)$, to nie istnieją podstawy do odrzucenia hipotezy dotyczącej braku istotnej różnicy między próbkami porównywanymi w danej parze (znak „–”). Natomiast w sytuacji odwrotnej należy uznać, że różnica ta jest statystycznie istotna (znak „+”). W tabeli 4 przedstawiono wartości prawdopodobieństwa z_{ij} oraz wynik porównania z wartością $z(\alpha)$ [10].

Tab. 4. Istotności różnic pomiędzy próbkami w danych parach

	pary próbek			
	1, 7	2, 8	3, 5	4, 6
z_{ij}	0,128	0,107	0,149	0,149
$z(\alpha) > z_{ij}$	–	–	–	–

Z poczynionych obliczeń jednoznacznie wynika, że nie ma podstaw do odrzucenia hipotezy o braku istotnej różnicy między porównywanymi obiektami w danej parze. Uzyskane wyniki można więc uznać za istotne i stabilne. Przeprowadzona analiza potwierdziła założoną wcześniej hipotezę, zatem można uznać, że docelowy odbiorca nagrania ambisonicznego, zrealizowanego w przestrzeni publicznej, będzie preferował wersję „naturalną”, bez podkładu muzycznego.

4. TEST PARAMETRYCZNY

Ze względu na wielość dostępnych wtyczek przeznaczonych do postprodukcji sygnałów ambisonicznych, zdecydowano się na przeprowadzenie testu parametrycznego, który miał na celu uzyskanie wytycznych dotyczących wartości parametrów poszczególnych wtyczek oraz ich potencjalnej przydatności podczas obróbki materiału w procesie tworzenia docelowej bazy nagrań, co miało zostać sprawdzone poprzez przetestowanie wpływu różnych ustawień na odbiór materiału przez słuchaczy.

Podczas wyboru testowanych wtyczek kierowano się charakterem zarejestrowanych nagrań. Z uwagi na to, że materiały były realizowane w różnych warunkach i pojawiły się w nich niepożądane składowe o różnych zakresach częstotliwości, zdecydowano się na przetestowanie działania korektora barw w zakresie niskich i wysokich częstotliwości, a także kompresora wielopasmowego w zakresie niskich częstotliwości, celem sprawdzenia wpływu na odbiór materiału przy zminimalizowaniu występowania częstotliwości odpowiadających powiewom wiatru oraz powstających niepożądanych składowych o wysokich częstotliwościach. Przetestowano również działanie obu powyższych wtyczek przy wzmacnianiu charakterystycznych dźwięków otoczenia, znajdujących się w zakresie średnich częstotliwości. Postanowiono także w ramach eksperymentu sprawdzić jak na odbiór nagrania wpłynie wyrównanie poziomów dla kanałów w sygnale ambisonicznym. W tym celu przetestowano dwie wtyczki

wyrównujące wzmocnienia kanałów ortogonalnych względem kanału dookólnego. Materiały realizowano również w miejscach, w których znajdował się bodziec dźwiękowy, charakterystyczny dla danego otoczenia. Dlatego też przetestowano wpływ działania wtyczek wzmacniających dany kierunek lub przestrzeń w nagraniu.

Analogicznie jak w przypadku testu porównań parami, z uwagi na docelowe przeznaczenie materiałów, założono, że w teście będą uczestniczyć osoby nie będące ekspertami. Minimalną liczebność grupy testowej określono na 35 osób. Formułę testu zdefiniowano kierując się wymaganiami stosowanymi w podejściu typu MUSHRA [14]. Na ich podstawie przyjęto długość próbek testowych równą 10 sekund. Zadaniem uczestników testu w przypadku każdej z próbek była jej subiektywna ocena w skali od 1 do 10, gdzie wartość 10 traktowana była jako najwyższa możliwa ocena.

Do stworzenia formularza testowego ponownie posłużyła platforma *Google Forms*. Dla każdego z testowanych parametrów wyodrębniono osobną sekcję prezentującą przypisane do niej próbki. Ponieważ serwis ten nie służy docelowo do przeprowadzania odsłuchowych testów parametrycznych, nie było możliwe spełnienie warunków sposobu prezentacji próbek w formie dowolnego przełączania się między nimi oraz oceny próbek w skali ciągłej, co wymagane jest w przypadku testów typu MUSHRA. Z tego powodu zdecydowano, aby dla każdej próbki przeznaczyć osobną instancję odtwarzacza *YouTube* oraz pytanie o jej ocenę. Podobnie jak w formularzu przygotowanego dla testu porównań parami, na jego początku zamieszczono sekcję wstępną zawierającą informacje dotyczące przebiegu testu [14].

Na potrzeby testu stworzono łącznie 41 próbek testowych. Mając na uwadze wyniki testu porównań parami, zrezygnowano z dodatkowego podkładu muzycznego. Rolę kotwicy w każdym z prezentowanych zestawów próbek stanowił sygnał niepoddany działaniu wtyczki w badanym aspekcie.

W teście udział wzięło łącznie trzydzieści pięć osób. Spełnione zostało zatem założenie dotyczące liczebności grupy testowej. W pierwszej kolejności wyłoniono grupę ekspertów, czego dokonano na podstawie wariancji obliczonej dla każdego uczestnika w kontekście całego testu. Z dalszej analizy wykluczono wyniki sześciu osób, dla których wariancja wyniosła poniżej założonego poziomu równego 0,75. Wyodrębniono w ten sposób grupę ekspertów liczącą 29 osób.

Następnie dla otrzymanej grupy ekspertów wygenerowano ponownie histogramy rozkładu ocen poszczególnych próbek testowych oraz obliczono dla nich statystyki opisowe. W każdym z badanych przypadków wartości średniej arytmetycznej oraz mediany dla poszczególnych próbek były zbliżone do siebie. Największe różnice między tymi statystykami zauważono dla próbek bez przetwarzania w sekcjach nr 4, 5, 6 oraz 9, które oscylowały w granicach od 0,7 do 0,9, co świadczy o symetryczności rozkładu ocen względem jego środka [8].

Natomiast analizując poziomy wariancji, będącej miarą rozproszenia wyników względem wartości średniej, najwyższe wartości zaobserwowano dla sekcji, w których testowane było działanie wtyczek uwydatniających wybrane kierunki lub przestrzenie w nagraniu. Ponadto porównano wartości wariancji ocen ekspertów z wartościami uzyskanymi dla wszystkich uczestników testu. Po wyodrębnieniu grupy ekspertów dla ocen próbek testowych znajdujących się w pierwszych siedmiu sekcji testowych wartość wariancji wzrosła. Jednak w tych sekcjach, w których badano wtyczki uwydatniające wybrane kierunki w nagraniu, miara rozproszenia nieznacznie spadła w ośmiu na dwanaście przypadków badanych w tych sekcjach. Takie zależności pomiędzy wartościami statystyki wariancji otrzymanymi dla obu grup potwierdzają prawidłowy dobór ekspertów [8].

Z uwagi na występowanie miar porządkowych oraz fakt, że otrzymane rozkłady odpowiedzi nie były rozkładami normalnymi, zdecydowano się na przeprowadzenie testu Kruskala-Wallisa dla wyników ekspertów. Hipotezą zerową w tym teście jest założenie, że średnie rangi lub mediany dwóch serii danych są takie same. Przeciwny przypadek stanowi zaś hipotezę alternatywną. Jako hipotezę do przeprowadzanego testu przyjęto domyślną hipotezę zerową. W pierwszej kolejności obliczono wartość statystyki T dla całego zbioru odpowiedzi według następującego wzoru [15]:

$$T = \frac{(N-1)(\sum_{i=1}^{k} \frac{{s_i}^2}{n_i} - C)}{S_r - C} = 49{,}872 \tag{3}$$

gdzie:
N – liczba elementów zbioru,
k – liczba grup, na które został podzielony zbiór,
s_i – suma rang elementów grupy i,
n_i – liczebność grupy i,
S_r – suma kwadratów rang wszystkich elementów zbioru,
C – wartość korekty dla średniej.

Hipotezę zerową testu sprawdzono poprzez porównanie obliczonej statystyki T z wartością krytyczną chi kwadrat, odczytaną z tablic rozkładu chi kwadrat dla wyznaczonej liczby stopni swobody równej 40 oraz obranego poziomu istotności – 0,05. Dla tych wartości odczytano poziom krytyczny statystyki chi kwadrat równy 55,7585, zatem nie można odrzucić hipotezy zerowej, ponieważ wyznaczona wartość statystyki T nie przekroczyła wartości krytycznej [15].

W dalszej analizie sekcje testu zostały podzielone na grupy według przeznaczenia przetwarzania danych wtyczek. W celu sprawdzenia hipotezy zerowej dla każdej z grup wyznaczono wartość statystyki T oraz liczbę stopni swobody, dla których odczytano wartość krytyczną statystyki chi kwadrat. Według otrzymanych obliczeń nie istnieją podstawy do odrzucenia hipotezy zerowej, ponieważ dla żadnej z wydzielonych grup wartość statystyki T nie przekroczyła wartości krytycznej chi kwadrat odczytanej dla założonego poziomu istotności równego 0,05 oraz określonej liczby stopni swobody.

Kolejne kroki analizy przeprowadzono przy użyciu skryptu w języku R oraz pakietu *fBasics* zawierającego zbiór funkcji do analizy danych, w tym również funkcję dedykowaną dla testu Kruskala-Wallisa. Serie odpowiedzi wyłonionych ekspertów, stanowiące dane wejściowe, zapisano w pliku CSV. Podając kolejne wektory odpowiedzi do funkcji w kombinacji „każdy z każdym", otrzymano macierz statystyki chi kwadrat oraz macierz p-wartości dla całego testu [20].

Następnie z otrzymanych macierzy wyodrębniono zbiory wartości statystyk dla każdej z grup. Uzyskane wartości statystyki chi kwadrat dla wszystkich badanych próbek w kolejnych grupach są mniejsze od wyznaczonej dla każdej z nich wartości krytycznej, co świadczy o istotności różnic między udzielanymi odpowiedziami. Analizując obliczone p-wartości można uznać, że nie ma podstaw do odrzucenia hipotezy zerowej, ponieważ wszystkie są większe od założonego poziomu istotności [8].

Dodatkowo porównano sumy ocen oddanych na próbki testowe w celu uzyskania informacji o preferencjach wyznaczonej grupy ekspertów dotyczących charakteru oraz stopnia przetworzenia dźwięku ambisonicznego w prezentowanych przykładach. Przygotowane zestawienie nie wskazało jednoznacznie faworytów wśród nagrań ocenianych przez ankietowanych. Z tego powodu sumy ocen porównano ponownie, jednak tym razem pod uwagę wzięto jedynie oceny powyżej średniej arytmetycznej dla danego obiektu testowego ocenianego przez grupę wyznaczonych ekspertów. Na podstawie wyników dla każdej sekcji wyznaczono najwyżej oceniane próbki oraz odpowiadające im wartości testowanych parametrów, które zamieszczono w tabeli 5.

Tab. 5. Preferowane próbki oraz przypisane do nich wartości testowanych parametrów

nr sekcji	nr próbki	testowany parametr	wartość parametru
1	1 lub 3	częstotliwość odcięcia filtru górnoprzepustowego	bez lub 120Hz
2	3	poziom tłumienia górnozakresowego korektora półkowego lub filtru pasmowo-zaporowego ustawionego w górnym paśmie	korektor półkowy: -12dB
3	1 lub 4	poziom wzmocnienia filtru pasmowo-zaporowego ustawionego w środkowym paśmie	bez lub 4dB
4	1	stopień kompresji	bez
5	1 lub 4	stopień kompresji	bez lub 8:1
6	1 lub 4	stopień kompresji	bez lub 8:1
7	3	parametr Envelopment	0,27
8	3	stopień kompresji	4:1
9	2	wzmocnienie	2dB
10	1	stopień przekształcenia	bez

Wyniki analizy przeprowadzonego testu są w wielu aspektach wyraźnie niejednoznaczne. Przyczyn takiego stanu rzeczy należałoby szukać w kilku czynnikach. Pierwszym z nich jest niewątpliwie nieodpowiedni dobór uczestników, o czym świadczą m.in. wartości wariancji uzyskane przez uczestników w poszczególnych sekcjach. W przypadku przeprowadzania testu, w którym oceniane próbki przetworzone zostały w stopniu średnim lub znacznym, rekomenduje się dobranie grupy testowej składającej się z osób, które można określić mianem ekspertów. Jednak z uwagi na ograniczenia występujące podczas realizacji niniejszej pracy dyplomowej możliwe było zorganizowanie testu wyłącznie w formie zdalnej, co niosło za sobą dodatkowe komplikacje w postaci braku kontroli nad przebiegiem sesji odsłuchowych oraz dysproporcji w odbiorze próbek testowych ze względu na zróżnicowany sprzęt, z którego korzystali uczestnicy. W związku z tym wnioski otrzymane podczas przeprowadzonej analizy wyników należy traktować raczej jako punkt wyjścia do dalszych rozważań na temat przetwarzania nagrań ambisonicznych niż jako konkretne wskazówki dotyczące tego zagadnienia [14].

5. STWORZENIE BAZY NAGRAŃ ORAZ APLIKACJI

Dla każdego z miejsc, w którym zarejestrowano materiały otrzymano zestaw dwóch nagrań: wideo oraz audio, odpowiednio w formatach INSV i WAV. Należy jednak zaznaczyć, że pliki z rozszerzeniem INSV nie są domyślnie wspierane przez programy służące do postprodukcji materiałów wideo. Rozwiązaniem tego problemu okazała się być wtyczka dostępna w pakiecie z darmowym oprogramowaniem *Insta360 Studio* udostępnianym przez producenta kamery, która jest przeznaczona do użycia wyłącznie we współpracy z oprogramowaniem *Adobe Premiere Pro* w wersjach 2018, 2019 lub 2020. Dzięki niej możliwa jest obsługa oraz edycja plików w formacie INSV. W związku z ograniczoną możliwością wyboru programu do edycji materiałów wideo, zdecydowano się na użycie *Adobe Premiere Pro* w wersji 2019 [9] [12].

Po zapoznaniu się z zarejestrowanymi materiałami wideo uznano, że poza standardową obróbką w postaci przycięcia ścieżek do docelowej długości, konieczna będzie ich dodatkowa edycja ze względu na przechylenia obrazu względem podłoża oraz na widoczny w kadrze statyw oraz mikrofon. W celu wypoziomowania obrazu użyto wtyczki *VR Sphere Rotation* dostępnej w ramach oprogramowania *Adobe Premiere Pro*. Przykład porównania przed i po wprowadzonej korekcie zaprezentowano na rysunku 3. Natomiast znajdujące się w kadrze statyw oraz mikrofon zasłonięto, nakładając na obraz maskę z logiem Katedry Systemów Multimedialnych. Materiały wideo wyeksportowano bez ścieżki dźwiękowej w formacie MPEG-4 dla parametrów jednakowych z parametrami zarejestrowanego nagrania [12].

Rys. 3. Efekt wypoziomowania obrazu za pomocą wtyczki *VR Sphere Rotation*

Do postprodukcji nagrań audio wybrano cyfrową stację roboczą *Reaper* w wersji 6.02, która pozwala na miksowanie oraz eksportowanie sygnałów wielokanałowych. Dla każdego z projektów, w którym przeprowadzono obróbkę dźwięku, ustawione zostały parametry tożsame z parametrami zarejestrowanych sygnałów [13].

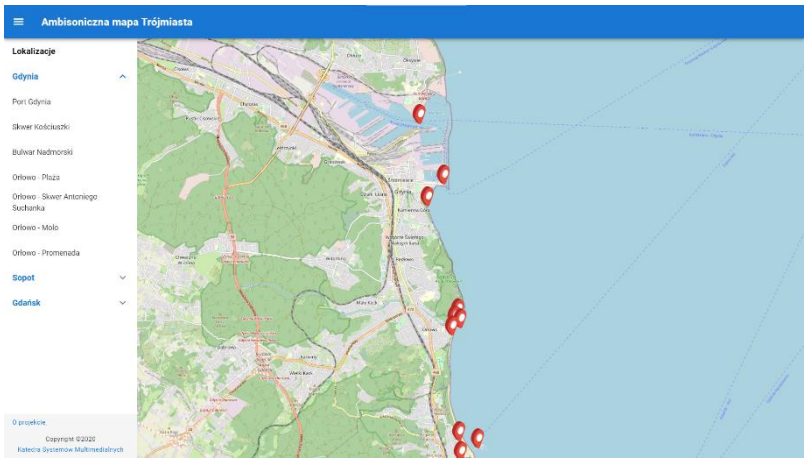
Uwzględniając wyniki testu porównań parami zrezygnowano z dodatkowego podkładu stereo, więc routing kanałów został ustawiony tak, aby zarówno dla kanału głównego jak i pozostałych sygnał wyjściowy był w postaci czterokanałowej. Z uwagi na fakt, że nagrania zostały zarejestrowane w

ambisonicznym B-formacie, nie było konieczne użycie enkodera. Monitorowanie sygnału poprzez słuchawki umożliwiło ustawienie na kanale głównym wtyczki dekodującej sygnał do postaci binauralnej [1] [5].

Ze względu na niejednoznaczne wyniki testu parametrycznego, zdecydowano się jedynie na korekcję częstotliwości w dolnym oraz górnym paśmie, sugerując się wariantami, które były najlepiej oceniane przez ekspertów. Pliki dźwiękowe zostały wyeksportowane w formacie WAV dla ustawień projektu.

Przygotowane w ten sposób pliki audio oraz wideo połączono w kontenerze MPEG-4 za pomocą kodeka FFMPEG. Następnie przy użyciu narzędzia *Spatial Media Metadata Injector* dodano również wymagane metadane, zgodnie z wymaganiami platformy *YouTube* [4] [7] [16].

Ostatnim etapem projektu było stworzenie aplikacji internetowej z interaktywną mapą, zawierającą przygotowane wcześniej nagrania. W tym celu wykorzystano pakiet *Vue.js* i kompatybilne z nim biblioteki komponentów: *Vuetify* oraz *VueLayers*. Finalny wygląd aplikacji, która została umieszczona na serwerze Katedry Systemów Multimedialnych, przedstawiono na rysunku 4 [17] [18] [19].



Rys. 4. Aplikacja internetowa z interaktywną mapą

6. WNIOSKI

W ramach niniejszego projektu zrealizowano serię materiałów audiowizualnych dla wybranych lokalizacji Trójmiasta. Nagrania zostały poddane procesowi postprodukcji i zamieszczone na interaktywnej mapie w aplikacji internetowej. Istnieje przy tym możliwość łatwej rozbudowy mapy o nowe nagrania, wykonywane w kolejnych miejscach w Trójmieście.

Przeprowadzony test porównań parami wykazał, że docelowi odbiorcy materiałów zdecydowanie preferują nagrania pozbawione dodatkowego podkładu muzycznego. Natomiast w przypadku drugiego testu wnioski pochodzące z analizy jego wyników są niejednoznaczne. Prawdopodobnym tego powodem były ograniczenia związane ze zdalną formą testu, co wpłynęło bezpośrednio na brak kontroli nad przebiegiem sesji odsłuchowych oraz na dysproporcje w odbiorze próbek testowych przez poszczególnych uczestników.

LITERATURA

[1] 5 Tips for Mixing Ambisonics | 360° | VR | Spatial Audio | Part 5/7 | Berklee Online, https://www.youtube.com/watch?v=1dvK4ojh2b0, (data dostępu: 05.10.2020).
[2] ANSTENDIG M. B., AB Testing, *A misapplication of visual criteria in audio*, 2006. http://www.anstendig.org/ABTesting.html, (data dostępu: 05.10.2020).

[3] Facebook 360 Video, https://facebook360.fb.com/, (data dostępu: 05.10.2020).
[4] FFmpeg Codecs Documentation, https://www.ffmpeg.org/ffmpeg-codecs.html, (data dostępu: 05.10.2020).
[5] FRANK M., ZOTTER F.: *Ambisonics – A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality*, Springer Topics in Signal Processing 2019.
[6] GERZON M.: *What's wrong with quadrophonics*, Studio Sound, May 1974, https://www.audiosignal.co.uk/Resources/What_is_wrong_with_quadraphonics_A4.pdf, (data dostępu: 05.10.2020).
[7] google/spatial-media: Specifications and tools for 360º video and spatial audio., https://github.com/google/spatial-media, (data dostępu: 05.10.2020).
[8] HEUMANN C., SCHOMAKER M., SHALABH, *Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R*, Springer International Publishing, Switzerland, 2016.
[9] Insta360 ONE X - Own the moment., https://www.insta360.com/product/insta360-onex, (data dostępu 05.10.2020).
[10] KOSTEK B., Wykład z przedmiotu Technologia nagrań – Testy subiektywne, Politechnika Gdańska – Wydział Elektroniki, Telekomunikacji i Informatyki – Katedra Systemów Multimedialnych, https://multimed.org/student/tn/testy_subiektywne.pdf, (data dostępu: 05.10.2020).
[11] Michael Gerzon Audio Pioneer – Ambisonics, https://www.michaelgerzonphotos.org.uk/ambisonics.html, (data dostępu: 05.10.2020).
[12] Profesjonalny edytor wideo | Adobe Premiere Pro, https://www.adobe.com/pl/products/premiere.html, (data dostępu: 05.10.2020).
[13] REAPER | Audio Production Without Limits, https://www.reaper.fm/, (data dostępu: 05.10.2020).
[14] Recommendation ITU-R BS.1534-1 – Method for the subjective assessment of intermediate quality level of coding systems, https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-1-200301-S!!PDF-E.pdf, (data dostępu: 05.10.2020).
[15] SPRENT P., SMEETON N.C., *Applied Nonparametric Statistical Methods, Third Edition*, Champan & Hall/CRC, 2001.
[16] Używanie dźwięku przestrzennego w filmach 360° i rzeczywistości wirtualnej (VR), https://support.google.com/youtube/answer/6395969, (data dostępu: 05.10.2020).
[17] Vue.js, https://vuejs.org/, (data dostępu: 05.10.2020).
[18] Vuetify, https://vuetifyjs.com/en/, (data dostępu: 05.10.2020).
[19] VueLayers :: Homepage, https://vuelayers.github.io/, (data dostępu: 05.10.2020).
[20] WUERTZ D., SETZ T, CHALABI Y., MAECHLER M., Package fBasics, Rmetrics – Markets and Basic Statistics, 2020.
[21] ZOOM H3-VR Handy Recorder | Zoom, https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h3-vr-handy-recorder, (data dostępu: 05.10.2020).

AN AMBISONIC MAP OF SPECIFIC PLACES IN THE TRI-CITY

The project aimed to create an ambisonic map of the Tri-City in the form of a web application. The video materials were recorded in 360 technology with sound in the form of an ambisonic signal in locations of the Tri-City, which were considered to be characteristic for this agglomeration. The research objective of the study was to compare the available algorithms for mixing ambisonic signals by conducting listening tests. The pairwise comparison test was performed to obtain information about the preferences regarding the presence of additional head-locked stereo background music and its level. In the second test the impact on the reception of the ambisonic recording was examined in terms of the method and degree of its processing. The conclusions obtained by analyzing the results of both tests served as guidelines for the post-production of the recordings. The prepared materials were placed on an interactive map in the web application.