Artykuły recenzowane

Studium badawcze młodych akustyków 2024

Redakcja: Bartłomiej Chojnacki

OSKA

Ogólnopolska Studencka
Konferencja Akustyków

7-10.03 2024

Kraków

Artykuły recenzowane

Studium badawcze młodych akustyków 2024

Wydawnictwo konferencyjne Ogólnopolska Studencka Konferencja Akustyków OSKA 2024

Redakcja: *dr inż. Bartłomiej Chojnacki*

Afiliacja redaktora:
AGH Akademia Górniczo-Hutnicza
Wydział Inżynierii Mechanicznej i Robotyki
Katedra Mechaniki i Wibroakustyki

Opracowanie i projekt okładki: Piotr Książek, Agnieszka Puzio, Jakub Werwiński

---

Wydawnictwo Polskie Towarzystwo Akustyczne Oddział w Krakowie,
ul. Uniwersytetu Poznańskiego 2
61-614 Poznań

shaping engineers

Nie jesteśmy tradycyjną firmą technologiczną ani instytutem badawczym.

Łączymy najlepsze elementy obu, bez ograniczania żadnego z nich, a nasz ośrodek **ARIC (Acoustic Research and Innovation Center)** jest sercem naszej działalności.

Jesteśmy społecznością inżynierów i naukowców, kreatywnych umysłów skupionych na rozwiązywaniu problemów technologicznych i eksplorowaniu nowych kierunków badań.

Zainteresowani?
Poznajmy się.
www.kfb-acoustics.com

HITACHI
Inspire the Next

Shape Tomorrow Today

Work With Purpose

What we do is always evolving.
We advance a sustainable energy future for all.

hitachienergy.com/careers

Apply now

Hitachi Energy

OSKA

Ogólnopolska Studencka Konferencja Akustyków

7-10.03 2024

Kraków

Spis treści:

Patroni

Patron Główny

Polskie Towarzystwo Akustyczne oddział Krakowski

Patron Honorowy

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Patroni medialni

Organizatorzy

oska-konferencja.pl

Miłosz **DERŻKO**[1]

POLITHEREMIN - KONCEPCJA I FAZA WSTĘPNA BUDOWY PROTOTYPU

POLYTHEREMIN - CONCEPT AND PRELIMINARY PHASE OF PROTOTYPE CONSTRUCTION

[1] AGH Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

milder@student.agh.edu.pl

Streszczenie

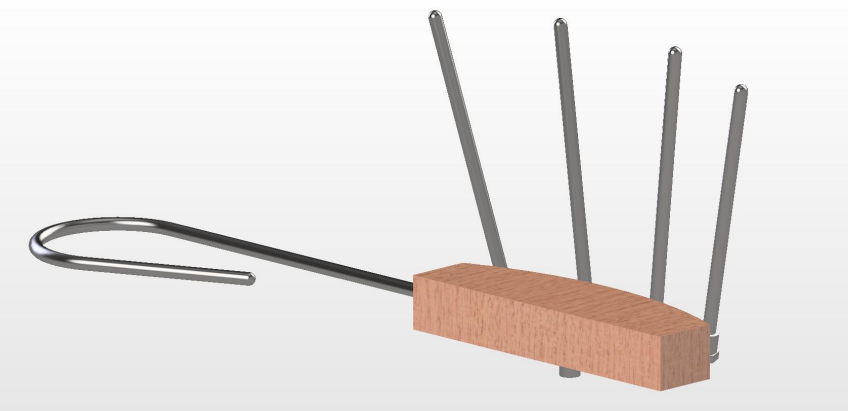
Projekt zakłada rozwinięcie innowacyjnego instrumentu, bazującego na konstrukcji Theremina. Instrument zostanie wyposażony w cztery anteny, z których każda będzie kontrolować jeden ton, sterowany przez odpowiedni palec ręki muzyka. Sygnał z anten będzie pochodził z modulowanego generatora przebiegu prostokątnego. Częstotliwość ta, zostanie odczytana za pomocą wewnętrznego procesora co uefektywni późniejszą analizę. Bliskie położenie anten spowoduje korelacje między poszczególnymi tonami, co utrudni grę na instrumencie.W związku z tym, zastosowany zostanie algorytm dekorelujący, mający na celu skuteczne odseparowanie sygnałów pochodzących od poszczególnych anten. Dźwięk będzie generowany w domenie analogowej za pomocą oscylatorów sterowanych cyfrowo, co pozwoli uzyskać brzmienie najbardziej zbliżone do pierwotnego instrumentu. Dodatkowo, urządzenie będzie posiadało możliwość udostępniania zdekorelowanych sygnałów z anten do zewnętrznych instrumentów za pomocą interfejsu MIDI. Takie podejście otwiera nowe perspektywy eksperymentalne i kreatywne w dziedzinie muzyki elektronicznej.

1 Wprowadzenie

We wczesnych latach XX wieku w instytucie Inżynierii Fizycznej Rosyjski fizyk Lev Sergeyevich Termen zaczął pracę nad wpływem obecności człowieka na sygnały radiowe. Radar miał wykrywać obecność człowieka poprzez zmianę fluktuacji sygnału radiowego nadawanego przez antenę [1]. Była to jedna z przyczyn powstania instrumentu. Theremin był jednym z pierwszych elektrofonów produkowanych masowo. Konstrukcja instrumentu w ogólności jest dosyć prosta. Posiada on dwie anteny, z czego jedna kontroluje wysokość tonu. Druga jest odpowiedzialna za głośność sygnału wyjściowego. Zwykle, obydwie anteny są położone w dwóch prostopadłych płaszczyznach, eliminuje to wzajemny wpływ na

siebie. Barwa dźwięku Thereminu podyktowana jest jego konstrukcją, natomiast współczesne wersje traktują ten instrument jako rodzaj kontrolera MIDI. Takie rozwiązanie pozwala na swobodną zmianę brzmienia instrumentu poprzez podłączenie do zewnętrznego instrumentu [2].

Przedstawiony instrument, jest swego rodzaju rozszerzeniem koncepcji Lwa Termena. Klasycznie do zmiany wysokości tonu grającego należy użyć całej ręki. W przypadku PoliTheremina mamy do czynienia z zestawem czterech anten które będą wykrywać odległości względem odpowiednich palców jednej ręki. Dodatkowo zostanie zamontowana antena głośności, kontrolowana przez drugą rękę muzyka podobnie jak w klasycznym instrumencie.



Rysunek 1: Wstępna wizualizacja instrumentu

Jak można zauważyć na rysunku 1 anteny są zamontowane stosunkowo blisko siebie. Niestety bliskość palców uniemożliwia odpowiedniego odseparowania anten. Ciągnie to za sobą konsekwencje w postaci skorelowanych sygnałów. Można spróbować zmniejszyć ten efekt poprzez umieszczenie anten nierównolegle. Oczywiście nie wyeliminuje to wszystkich problemów, ale może pomóc. Aby całkowicie wyeliminować korelacje między sygnałami z anten, zastosowany zostanie model uczenia maszynowego wytrenowany na odpowiednio dużym zbiorze danych. Proces ten zostanie opisany w dalszych etapach pracy nad instrumentem.

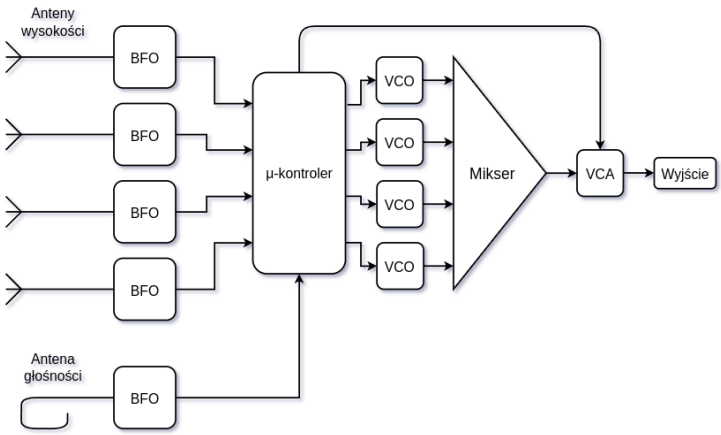
2 Działanie instrumentu

PoliTheremin do zmiany częstotliwości tonów, będzie wykorzystywał zestaw anten monopolowych. Jest to rodzaj anten emitujący fale elektromagnetyczną wokół przewodnika z taką samą mocą. Gdy analizujemy jedną antenę, można przyjąć, że posiada ona charakterystykę dookólną. Sytuacja się komplikuje gdy w polu bliskim znajdzie się kawa-

łek przewodnika, wówczas zgodnie z zasadą wzajemności staje się on również anteną. W związku z tym, wpływa on na charakterystykę monopola. Oczywiście ręka muzyka jest w tym przypadku również jest przewodnikiem, tworzącym zakłócenia w polu bliskim. Jest to jednak zjawisko wskazane dla działania instrumentu.

Sytuacja się komplikuje gdy oprócz ręki, w polu znajdzie się jeszcze kilka innych anten. Wówczas charakterystyka dookólna dla każdej z nich się zmieni [3]. Aby instrument mógł działać mimo naruszonej kierunkowości, należy odseparować od siebie sygnały każdej z anten. W procesie przetwarzania sygnałów zakładamy użycie sieci neuronowej, której zadaniem będzie wykrycie oraz usunięcie korelacji między sygnałami.

Projekt działania instrumentu w postaci schematu blokowego można zobaczyć na rysunku 2. Wejścia cyfrowe mikrokontrolera będą mierzyć częstotliwość wejściową pochodzącą z generatorów zdudnieniowych (BFO), które w połączeniu z antenami tworzą moduły detektorów odległości. Po poddaniu przez proces uczenia maszynowego, sygnały przepłyną do oscylatorów sterowanych napięciowo (VCO), aby na końcu zostały złączone w jeden wyjściowy sygnał. Jego amplituda wyjściowa zostanie regulowana poprzez antenę głośności. Zostanie ona położona prostopadle do pozostałych anten, co zniweluje wzajemny wpływ na pozostałe anteny.

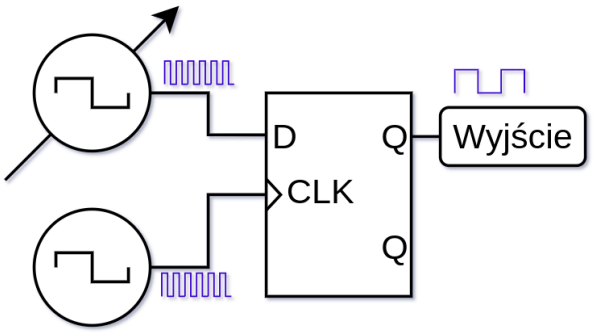


Rysunek 2: Schemat blokowy instrumentu

2.1 Detektor odległości

Działanie Theremina opiera się na uzależnieniu pojemności anteny z dowolnym parametrem układu. W tym przypadku, jednym ze sposobów jest stworzenie dwóch generatorów. Jeden z nich będzie wzorcowym, ze stałą częstotliwością, natomiast drugi będzie regulowany. Ponieważ jest to generator o sprzężeniu pojemnościowym, dołożenie do niego

dodatkowej pojemności skutkuje zmianą częstotliwości wyjściowej. Zmiany te są bardzo małe jednak gdy zmierzymy częstotliwości dudnień sumy dwóch generatorów możemy zaobserwować, o wiele większe zmiany częstotliwości w zakresie słyszalnym. Układ detektora położenia został też dobrze opisany w artykule [4]. W tym projekcie zostanie użyta jego cyfrowa wersja z pewną modyfikacją. Jego schemat mieści się na rysunku 3.



Rysunek 3: Cyfrowa implementacja generatora zdudnieniowego

Na powyższym schemacie można zauważyć, że zmodyfikowany został sposób uzyskania częstotliwości pośredniej. Ponieważ sygnały wykorzystane w układzie są prostokątne, zamiast sygnału dudniącego, który trzeba będzie filtrować można wykorzystać układ próbkujący oraz zjawisko aliasingu. Taką implementację zaproponował twórca projektu OpenTheremin [5] który w wersji czwartej instrumentu wykorzystał przerzutnik typu D jako układ próbkujący. W takim przypadku częstotliwość wyjściowa układu jest dana zależnością.

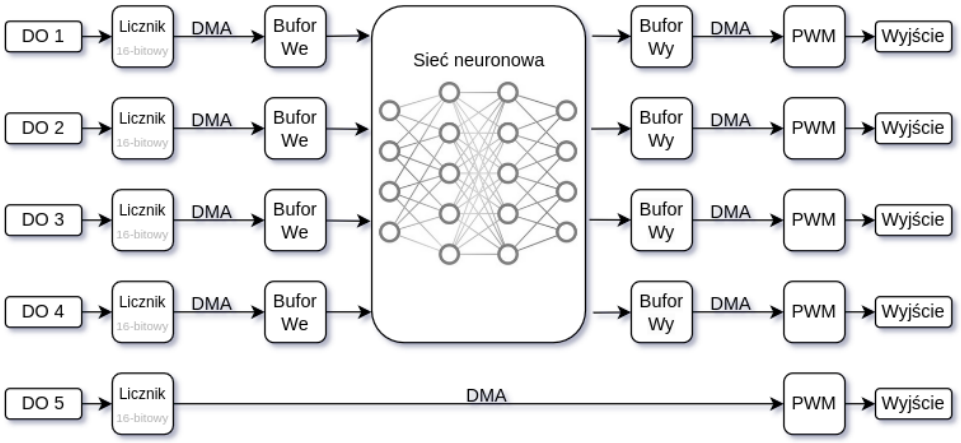
$$f = |f_{ref} - f_{reg}| \tag{1}$$

Gdzie f_{ref} oraz f_{reg} są wartościami częstotliwości z generatora referencyjnego oraz regulowanego. Taki sposób zmiany częstotliwości wymaga mniej elementów w części analogowej co przekłada się również na koszty związane z budową instrumentu.

2.2 Architektura programu

Mikrokontroler użyty w projekcie to RP2040 firmy Raspberry Pi Ltd [6]. Zawiera on szereg peryferiów które pozwolą odciążyć procesor. Architektura programu podzielona jest na trzy części - akwizycji danych, dekorelacji, oraz generacji. Z uwagi na fakt wykorzystania sieci neuronowej, dekorelacja sygnałów jest jedynym zadaniem które musi być wykonane przez rdzeń procesora. Mimo wszystko implementacja sieci, na mikrokontrolerze

mocno go obciąża, należy więc pozostałe zadania wykonać jak w sposób jak najbardziej odrębny od głównego procesora. Zapewni to zarówno płynność działania programu jak i niską latencję instrumentu. Wizualizację zaplanowanej architektury programu można zobaczyć na rysunku 4.

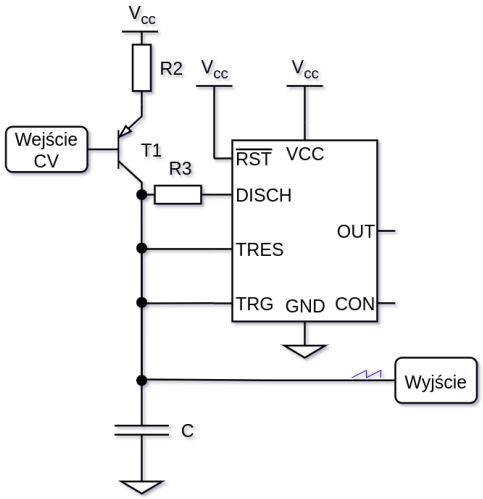


Rysunek 4: Plan przepływu sygnałów w mikrokontrolerze RP2040

Częstotliwość sygnału pochodzącego z detektorów odległości (DO) zostaje mierzona przez liczniki przyporządkowane do odpowiednich wejść cyfrowych. Każdy z nich zlicza impulsy dla każdego okresu sygnału wejściowego, wynik zostaje zapisany w odpowiednim rejestrze pamięci mikrokontrolera. Rejestr ten zostaje kopiowany do bufora wejściowego sieci neuronowej za pomocą kontrolera DMA. Ponieważ zliczanie dla każdego wejścia odbywa się asynchronicznie, wszystkie bufory należy ze sobą zsynchronizować. Zostanie to zrealizowane przez osobny licznik wysyłający sygnał do kontrolera DMA ze stałą częstotliwością. Po skopiowaniu wszystkich wartości z rejestrów zostanie wysłany sygnał przerwania do procesora, aby następnie przepuścił dane przez sieć neuronową. Sieć posiada cztery wyjścia które należy z powrotem przenieść do rejestrów generatorów PWM. Tak jak wcześniej, również zostanie do tego użyty kontroler DMA. Po wygenerowaniu sygnału PWM zostanie on wprowadzony do generatorów tonu grającego kreujących ostateczną barwę instrumentu.

2.3 Sekcja wyjściowa

Dźwięk instrumentu zależy od zastosowanych oscylatorów. Nie musimy tutaj implementować standardu $\frac{V}{oct}$ ponieważ można to osiągnąć w dziedzinie cyfrowej. Wybrano więc prosty układ VCO z wykorzystaniem popularnego układu scalonego NE555. Schemat układu został przedstawiony na rysunku 5.



Rysunek 5: Schemat generatora VCO

Jak widać, układ nie jest zbyt skomplikowany, jego działanie opiera się na ładowaniu oraz rozładowaniu kondensatora w odpowiednim momencie. Źródło prądowe R2, T1 ładuje kondensator C prądem stałym, w skutek czego napięcie na nim rośnie w sposób liniowy. Bo naładowaniu kondensatora do napięcia progowego, następuje jego nagłe rozładowanie przez R3 oraz pin DISCH układu NE555. Po rozładowaniu kondensatora, napięcie wyjściowe znów rośnie od zera do $\frac{2}{3}V_{cc}$ co zamyka pętlę generatora.

Częstotliwość generowanego przebiegu zależy od prądu ładowania kondensatora. Prąd ten pochodzi z regulowanego źródła prądowego R2, T1. Regulacja źródła prądowego polega na ustaleniu prądu płynącego przez rezystor R2. Jest on wprost zależny od napięcia wejściowego zgodnie z prawem Kirchoff'a. Zakładając, że $U_{BE} = const$, zależność między prądem ładowania a napięciem wejściowym wygląda następująco.

$$I_C = -\frac{U_{Wej}}{R_2} - \frac{V_{cc} - U_{BE}}{R_2} \tag{2}$$

Gdzie I_C to prąd ładowania kondensatora. Zakładając, że czas rozładowania kondensatora jest nieskończenie mały, częstotliwość generatora można zapisać w postaci.

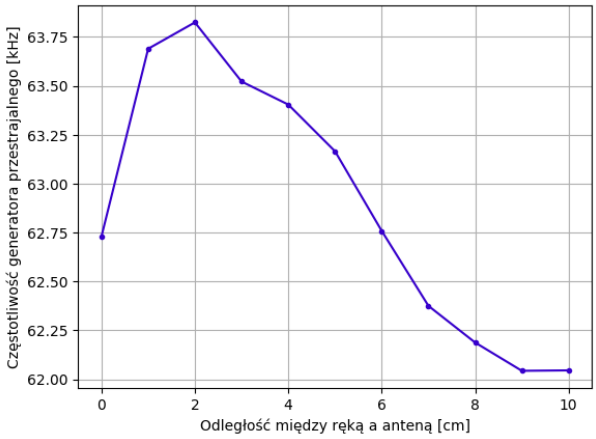
$$f = \frac{3I_C}{2U_{cc}C} \tag{3}$$

Gdzie C to pojemność kondensatora. Sekcja wyjściowa ma generować cztery przebiegi piłokształtne o częstotliwościach różnych względem siebie. Zostają one następnie zmiksowane oraz przepuszczone przez wzmacniacz sterowany napięciem. Instrument na wyjściu powinien wygenerować akord grany przez muzyka.

3 Pomiary wstępne

Eksperyment z użyciem prototypowej wersji układu został zrealizowany w celu zbadania zależności między odległością od anteny a częstością detektora odległości. Moduł detektora został skonstruowany przy użyciu układu 74HC14, zawierającego sześć negatorów z przerzutnikiem Schmitta. Schemat modułu jest tożsamy z przedstawionym w rozdziale 2.1. Pomiar częstotliwości poszczególnych generatorów przeprowadzono za pomocą analizatora stanów logicznych Saleae Logic z pasmem przenoszenia 4MHz. Częstotliwość sygnałów została określona metodą szczytową, uśredniając wyniki z dwudziestu okresów sygnału.

W celu zweryfikowania, czy detektor częstotliwości poprawnie odczytuje odległość ręki, przeprowadziliśmy pomiary wstępne. Zmierzono odpowiedzi detektora dla dziesięciu różnych położeń od zera do dziesięciu centymetrów. Wyniki pomiarów wskazują, że generator wzorcowy jest nastrojony na częstotliwość około 60 kHz, podczas gdy oscylator przestrajalny generuje zakres od około 62 kHz do 64 kHz. Wartości wyjściowe generatora przedstawione są na rysunku 6.



Rysunek 6: Zależność położenia ręki od częstotliwości wyjściowej generatora przestrajalnego

Analiza charakterystyki detektora ukazuje, że zależność jego odczytów od odległości nie jest liniowa, co potwierdzają badania opisane w artykule [4]. Ponadto, można zaobserwować spadek częstotliwości poniżej dwóch centymetrów od anteny. Zważywszy na fakt, że układ został skonstruowany na płytce prototypowej, mogą występować znaczne pojemności wzajemne w układzie. W miarę zbliżania się ręki do układu, wpływała ona na niego coraz bardziej, co mogło przyczynić się do spadku częstotliwości. Kolejna wersja prototypu będzie musiała być bardziej odporna na zakłócenia wynikające z ręki znajdującej

się blisko układu. Można to osiągnąć poprzez zwiększenie długości anteny oraz umieszczenie układu na płytce PCB. Ta modyfikacja powinna zapewnić bardziej stabilną pracę detektora odległości.

4 Podsumowanie

Wyniki przeprowadzonych pomiarów wskazują, że detekcja położenia stanowi obiecującą podstawę do dalszych badań i doskonalenia prototypu instrumentu. Kolejnym krokiem w projekcie jest zbudowanie układu akwizycji sygnału z detektora odległości. Konieczne będzie stworzenie programu do detekcji częstotliwości oraz przesłania wartości do komputera. Zebrane dane posłużą jako zbiór testowy oraz walidacyjny do uczenia sieci neuronowej. Następnie sieć tą należy zaimplementować w mikrokontrolerze RP2040 tak, aby nie blokował pozostałych funkcjonalności.

PoliTheremin jest jeszcze na wczesnym etapie rozwoju. Jego konstrukcja będzie prawdopodobnie ewoluować w trakcie pracy nad projektem, co może się przyczynić do rozwiązania istniejących problemów. Projekt prezentuje duże pole rozwoju w zakresie szeroko pojętej elektroniki audio oraz programowania systemów wbudowanych. Kolejne działania będą sukcesywnie dokumentowane oraz w miarę możliwości publikowane.

Literatura

[1] Anonymous MIT student, From Physical Law to Artistic Expression, 21M.380 Music and Technology, 2009

[2] `https://www.moogmusic.com/products/etherwave-theremin`, dostęp: 13.01.2024r.

[3] Carmen Bachiller Martin, Jorge Sastre Martinez, Amelia Ricchiuti, Hector Esteban Gonzalez, and Carlos Hernandez Franco, Study of the Interference Affecting the Performance of the Theremin, Hindawi Publishing Corporation International Journal of Antennas and Propagation

[4] Kenneth D. Skeldon, Lindsay M. Reid, Viviene McInally, Brendan Dougan, and Craig Fulton, Physics of the Theremin, American Association of Physics Teachers, 1998

[5] `https://www.gaudi.ch/OpenTheremin/`, dostęp: 13.01.2024r.

[6] `https://datasheets.raspberrypi.com/rp2040/rp2040-datasheet.pdf`, dostęp: 13.01.2024r.

Michał **KAMIŃSKI**[1], Kajetana Marta **SNOPEK**[1],

ANALIZA CECH CHARAKTERYSTYCZNYCH PRAWIDŁOWYCH I WADLIWYCH REALIZACJI FONEMU /R/

ANALYSIS OF THE CHARACTERISTICS OF CORRECT AND WRONG REALIZATION OF THE PHONEME /R/

[1] Politechnika Warszawska

michal-_kaminski@wp.pl

Streszczenie

Referat dotyczy problemu niejednoznacznej reprezentacji fonemu /r/ i prezentuje wyniki pracy inżynierskiej [1]. W literaturze sprzed 50 lat poświęconej wymowie Polaków przyjmowało się, iż wariantem podstawowym fonemu /r/ była spółgłoska drżąca dziąsłowa dźwięczna („voiced alveloar trill"). Ostatnie badania sprzed dekady wykazują jednak, iż większość Polaków uderza jeden raz koniuszkiem języka o wałek dziąsłowy (a niekiedy wyłącznie przybliża, a nie wibruje nim). Częstość każdej z realizacji zależy m.in. od otoczenia głoski czy tempa mowy. W ramach eksperymentu zdecydowano się na zbadanie artykulacji również osób posiadających różne rodzaje rotacyzmu. Ze względu na brak istnienia ogólnodostępnej bazy danych zawierającej nagrania osób z tą wadą wymowy, postanowiono stworzyć własną. Sygnały testowe zostały zarejestrowane w komorze bezechowej, a następnie posegregowane i poddane wstępnej obróbce w celu ułatwienia późniejszej analizy przebiegów czasowych i spektrogramów. Uzyskane wyniki mogą być punktem wyjścia do bardziej dogłębnych analiz wykorzystywanych np. w automatycznych algorytmach rozpoznawania wad wymowy.

1 Wprowadzenie teoretyczne o sygnale mowy

Proces generowania sygnału mowy można spróbować wytłumaczyć opierając się o teorię sygnału. W takim ujęciu sygnał powstały w krtani (źródło) przechodzi przez trakt głosowy modelowany jako układ liniowy [2]. Jeżeli oznaczy się widmo mocy sygnału wychodzącego z krtani jako $S_r(f,t)$, funkcję transmitancji traktu głosowego jako $H(f,t)$ oraz widmo mocy sygnału mowy jako $S_x(f,t)$, to uzyska się zależność:

$$S_x(f,t) = S_r(f,t)|H(f,t)|^2 \tag{1}$$

Wyróżnia się dwa główne typy źródeł ([3] [4]). Pierwszy z nich, ton krtaniowy, występuje dla mowy dźwięcznej oraz powstaje w czasie fonacji czyli naprzemiennego zamknięcia i otwarcia głośni. Wychodzący z tchawicy strumień powietrza jest blokowany z częstotliwością równą liczbie drgań fałd głosowych na sekundę odpowiadającą tonowi podstawowemu, która dla każdego mówcy może być inna. Na widmo tonu krtaniowego składają się też częstotliwości harmoniczne, które tłumione około 12dB/oktawę są wielokrotnością tonu podstawowego. Drugi z nich, czyli szum biały, ma miejsce dla mowy bezdźwięcznej i jest swobodnym przepływem mas powietrza przez krtań, gdyż nie następuje okresowe zamknięcie głośni.

Następnie pobudzenie jest poddane artykulacji, czyli filtracji, która jest uzależniona od silnie zmiennej w czasie charakterystyki czasowo-częstotliwościowej traktu głosowego, który jest złożony z kilku jam (gardłowej, ustnej, nosowej) czyli, inaczej, z układu rezonatorów. Spowoduje ona zwiększenie amplitudy wokół powstałych częstotliwości rezonansowych. Tak samo jak ton krtaniowy, będzie ona zależała od cech osobniczych danego mówcy.

Wychodzący z ust głos jest przetwarzany analogowo-cyfrowo w celu uzyskania przebiegu czasowego. Może on posłużyć do analizy mowy w dziedzinie czasu, wyznaczenia funkcji autokorelacji (nazywanej tutaj autokorelacją) bądź po zastosowaniu na nim transformacji Fouriera przedstawienia zmieniającego się w czasie widma w formie spektrogramu.

2 Modelowa realizacja fonemu /r/

Według wielu publikacji z dziedziny fonetyki i fonologii(m.in. [5],[6]), istnieje więcej niż jedna realizacja fonemu /r/. Oznacza to tyle, że kilka dźwięków (alofonów) odbieramy jako jeden element logiczny i zapisujemy jedną literą. Mając wspólny mianownik jakim jest dziąsłowość, będą się one różnić między sobą takimi cechami dystynktywnymi jak dźwięczność (dźwięczna/bezdźwięczna), występowaniem palatalizacji (twarda/zmiękczona) czy liczbą uderzeń (drżąca/uderzeniowa). Palatalizowane alofony będą się wyróżniały zbliżeniem środkowej części języka do podniebienia twardego.

Jako wariant podstawowy fonemu /r/ została wytypowana spółgłoska drżąca dziąsłowa dźwięczna ([5],[7]) Podczas realizacji czubek języka, czyli inaczej *apex*, naprzemiennie szybko uderza przednie dziąsła i cofa się. Od alofonu jednouderzeniowego różni się tym, że zetknięcie musi się pojawić co najmniej 2 razy.

Według między innymi [7] spółgłoska ma budowę trójfazową złożoną z segmentów: konsonantycznego-wokalicznego-konsonantycznego. Zarówno pierwszy, jak i trzeci, posiadający nadzwyczajnie niską energię i trwający około 20ms, pojawia się podczas uderzenia i posiada silne ugięcia w F_3. Gdyby tylko raz się pojawił, można by mówić o alofonie w wersji jednouderzeniowej. Element wokaliczny, który trwa już 1.5-krotnie dłużej, przypada na chwilę odsunięcia *apexa* i w brzmieniu przypomina samogłoskę zwaną "neutralnym e" bądź "szwa", występującą np. na początku w angielskim wyrazie ⟨about⟩. Według [8] częstotliwości formantowe takiego elementu można przyrównać do częstotliwości kolejnych fal stojących mających długość 0,25l, 0,75l, 1,25l..., gdzie l to długość typowego traktu głosowego (17cm) i wynoszą 500Hz, 1500Hz, 2500Hz.... F_1 faktycznie pokrywa się z tymi wyznaczonymi przez innych badaczy: 600Hz([5]) , 400-450Hz([6]), 500Hz([7]).

Wada wymowy, polegająca na deformacji głoski r, nazywa się rotacyzmem właściwym [9] i jej charakter może dotyczyć m.in. obecności (lub braku) dźwięczności/wibracji, miejsca artykułowania (np. w krtani, gardle czy między podniebieniem miękkim a tylną częścią języka) bądź używania nieprawidłowych artykulatorów (np. języczka czy warg).

3 Dotychczasowe badania nad fonemem /r/

Wieloraka natura fonemu /r/ zaciekawiła polskich uczonych, którzy postanowili zobaczyć jak obecnie wygląda typowa realizacja. W artykule z 2010 roku [10] S. Jaworski postanowił sprawdzić jakie rodzaje artykulacji, i jak często, występują w pozycji między samogłoskami dla różnych temp i występowania akcentu bądź jego braku.

Najczęściej pojawiła się realizacja "voiceless tap "bezdźwięczne jedno uderzenie", które pojawiało się przy niezwieraniu fałd głosowych. Przypomina ona element konsonantyczny, pojawiała się też częściej w sylabach akcentowanych i normalnym tempie mowy.

Drugą najpopularniejszą odmianą było "fricativisation", gdzie koniuszek języka zamiast uderzenia tak bardzo przybliża się do dziąsła, że tworzy małą szczelinę, przez którą przechodzący strumień powietrza tworzy szum o wysokiej amplitudzie w wyższych częstotliwościach, charakterystyczny dla spółgłosek szczelinowych. Odmiana pojawia się zdecydowanie częściej dla szybkiego tempa i braku akcentu.

Kolejną artykulacją jest "approximantisation of [r]", która różni się od poprzedniej rozmiarem szczeliny - w tym przypadku jest ona na tyle duża, aby nie powstał szum, ale na tyle mała, aby aproksymantu nie uznawać za samogłoskę. Amplituda sygnału nieznacznie się zmniejsza, widać też ugięcia częstotliwości formantowych, szczególnie czwartej.

Ostatnia, najrzadziej pojawiająca się artykulacja, to uznawany za podstawowy wariant "trill", czyli po prostu wielouderzeniowe [r]. Co ciekawe, pojawiło się ono zaledwie w 1.3% przypadków. Może się narodzić pytanie - czy naprawdę przez te kilkadziesiąt lat wymowa Polaków zmieniła się tak bardzo?

Warto też się zastanowić - co w przypadku innej pozycji niż VrV lub innych zasad klasyfikacji? Przykładowo Ł. Stolarski w swoich publikacjach [11][12] wyodrębnił w odmiennych otoczeniach różne typy artykulacji. Najczęściej pojawiała się "a tapped stop(s)", a następnie, w zależności od pozycji i uwzględniając odmiany występujące częściej niż 5%: "(weak) closure(s) with an immediate intersification in higher frequencies", "a trilled stop(s)", "a trilled fricative(s)/ a fricative(s)/ an approximant(s)"/.

4 Stworzenie bazy sygnałów testowych

W celu przeanalizowania zarówno prawidłowej wymowy /r/, jak i tej wadliwej, nazywanej rotacyzmem, postanowiono stworzyć własną bazę sygnałów testowych. W badaniu wzięło udział 18 osób, głównie studentów, z czego 5 osób miało poprawną wymowę (2 kobiety i 3 mężczyzn) oraz 13 osób "rerało"(5 kobiet i 8 mężczyzn). Większość badanych pochodziła z Warszawy bądź Lubelszczyzny. Każda z osób w swoim tempie czytała do mikrofonu wyrazy z kartki.

Na badania składały się trzy eksperymenty. W pierwszym z nich ustawiono /r/ w pozycjach: CrV, VrC oraz Vr# (w wygłosie, za każdą z samogłosek oprócz nosowych). W dwóch pierwszych pozycjach sprawdzono czy na realizację /r/ wpływa fonem spółgłoskowy i uwzględniono każdy z możliwych: ⟨p, b, t, d, k, g, cz, dż, ć, dź, c, dz, m, n, ń, l, f, w, s, z, sz, ż, ś, ź, h, ł, j⟩. Łącznie uzyskano 60 słów zdatnych do badania. Propozycje wyrazów zapożyczono z publikacji Łukasza Stolarskiego [11], [12] oraz wymieszano je z 60 słowami niezawierającymi /r/. Następnie poproszono, aby osoby badane przeczytały wszystkie trzyliterowe kombinacje, w których /r/ jest w pozycji interwokalicznej, co dało 36 wyrazów na osobę. W ostatnim eksperymencie postanowiono sprawdzić jak osoby badane będą artykułowany różną długość fonemu /r/. W tym celu poproszono uczestników o przeczytanie słów ⟨mira⟩, ⟨mirra⟩ i utrzymanie samego dźwięku na parę sekund ⟨rrrrrr⟩. Ostatecznie na powstałą bazę danych składało się dokładnie 99 przebiegów /r/ od jednej osoby, co daje ich łącznie 1 782.

Nagrania zrealizowano w komorze bezechowej Zakładu Elektroakustyki Wydziału Elektroniki i Technik Informacyjnych Politechniki Warszawskiej, której poziom dźwięku za-

kłóceń nie przekracza 20dB(A), czas pogłosu wynosi poniżej 50ms, a jej objętość to około 250m^3, dzięki czemu we wnętrzu powstają warunki pola fali swobodnej.

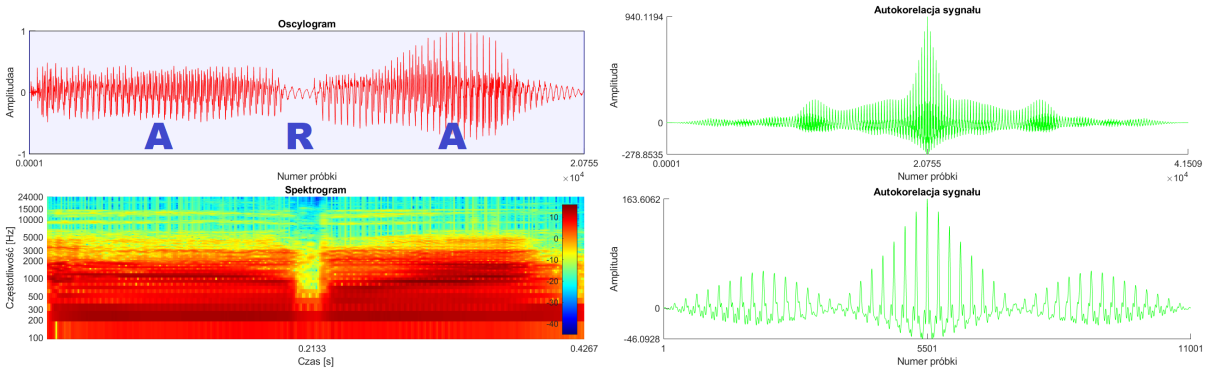
Wpierw wychodząca z ust fala akustyczna została zbierana przez oddalony o około 15 centymetrów od warg mikrofon pojemnościowy Neumann U 87 Ai, który, ustawiony na charakterystykę kardioidalną, może rejestrować dźwięk w zakresie 20-20kHz. Podłączono go do interfejsu audio Focusrite Scarlett 816 3rd Gen. Pomimo możliwości próbkowania aż do 192kHz, wybrano 48kHz w formacie mono .wav 24-bitowy PCM, aby pokryć cały zakres słyszalności ludzkiego ucha. Następnie interfejs podłączono do laptopa i zapisywano wcześniej zarejestrowane przebiegi w programie REAPER v6.82. Kolejno wyodrębniono z każdego nagrania fragment zawierający /r/ w sąsiedztwie jednego dźwięku przed i po (przykładowo - z wyrazu ⟨srebro⟩ zostaje ⟨sre⟩) i znormalizowano w sposób szczytowy.

Taki rodzaj segmentacji pozwala na zachowanie możliwie najkrótszego sygnału, gdyż wysublimowanie samego /r/ byłoby bardzo problematyczne. Co prawda wyodrębnienie wyłącznie elementu wokalicznego [r] byłoby możliwe z założeniem postawienia granic w elementach konsonantycznych, jednakże przecież to nie jest jedyna postać fonemu. W przypadku jednokrotnego uderzenia języka odseparowanie "ciszy" sprawia wrażenie słuchowe "buczenia", które nie przypomina /r/. Już w samej czynności ruchu czubka do wałka przydziąsłowego kryje się badany dźwięk. Jako, że ten etap nie ma wyłącznie jednej postaci, gdyż jest zależny od sąsiedniego dźwięku, to postanowiono zawrzeć "nawiązkę" otoczenia. Dodatkowo baza zawiera też nieprawidłowe realizacje, które warto zaobserwować w szerszym kontekście.

5 Główne rodzaje przebiegów czasowych poprawnej wymowy

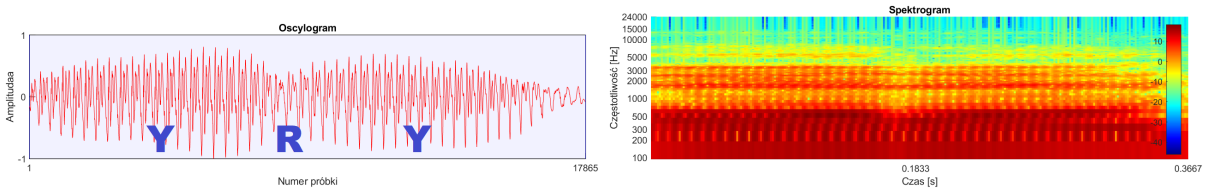
Zaobserwowano, że prawidłowe realizacje nie zawierają jednej ścisłej postaci, wobec czego trudnym zadaniem byłoby wyznaczenie konkretnych wartości numerycznych. Postanowiono zatem, że najlepszym będzie ogólny przegląd najbardziej reprezentatywnych sygnałów dla każdej z odmian możliwych do wyodrębnienia.

Pierwszą z nich jest nagły jednorazowy spadek amplitudy, który przypomina element konsonantyczny (rys. 1) i powstaje na skutek jednego uderzenia języka. Podczas około 20 milisekund trwania elementu można dostrzec przebieg przypominający sinusoidę złożoną z kilku okresów. Dzięki dużym skokom amplitudy na przebiegu autokorelacji dla całego sygnału można dostrzec "wcięcia", które będą bardzo widoczne, jeżeli odpowiednio ograniczy się sygnał do postaci 10101.



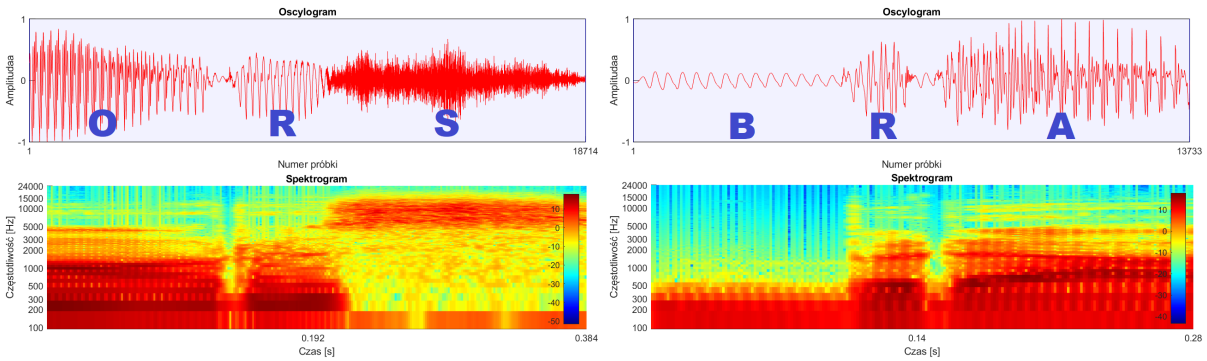
Rysunek 1: Oscylogramy i spektrogramy oraz autokorelacje realizacji /r/ jednouderzeniowe dla ⟨ara⟩, osoba W02

Można również wyróżnić przebieg mniej wcięty, podobny do "r uaproksymantowanego", jak w przypadku pracy J. Stolarskiego (rysunek 2). Co prawda ciągłość formantów zachowano, ale można zaobserwować minimalny spadek energii.



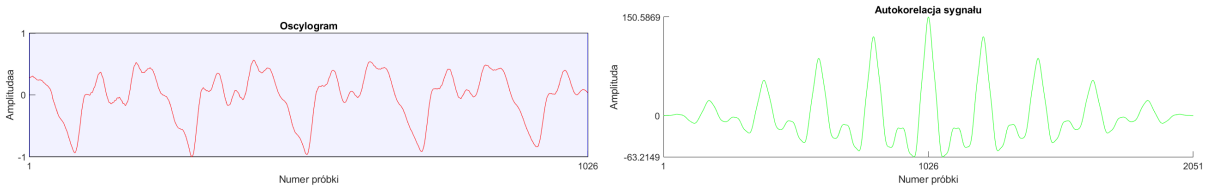
Rysunek 2: Oscylogram i spektrogram realizacji "r uaproksymantowane"⟨yry⟩, osoba M01

Kolejnym typem jest pojawienie się elementu wokalicznego przy klasycznej realizacji fonemu trwającego około 25-35ms i przypominającego na spektrogramie samogłoskę (rys. 3). Kształt przebiegu może się zwężać w niesymetryczny sposób przy którejś ze stron.



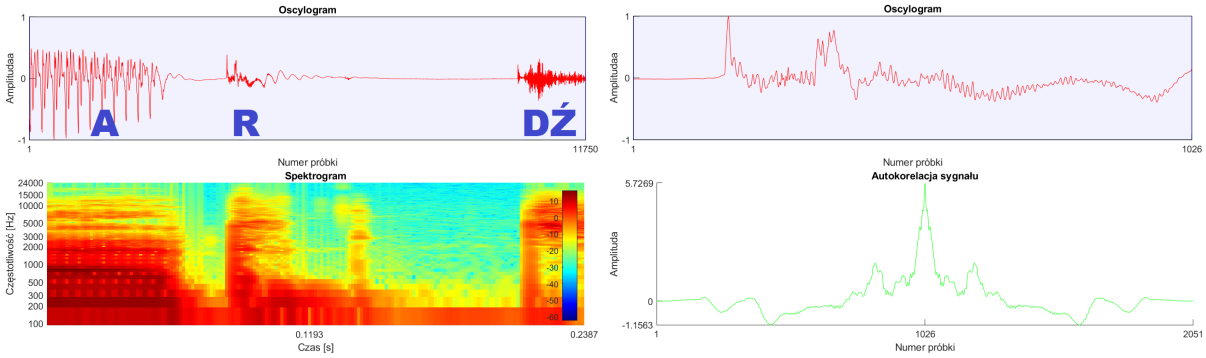
Rysunek 3: Oscylogramy i spektrogramy realizacji z elementem wokalicznym dla ⟨ors⟩, osoba W02 i ⟨bra⟩, osoba M01

Gdyby wyznaczyć autokorelację całego sygnału, można zauważyć mniej wyraźne "wcięcie", natomiast operacja ta dla wyłącznie elementu wokalicznego będzie przypominała charakterem samogłoskę (patrz. rys.4).



Rysunek 4: Oscylogramy fragmentu elementem wokalicznym i jego autokorelacje dla: ⟨aru⟩, osoba W02

Ostatnią wyróżnioną odmianą jest "r frykatywne", które sprawia wrażenie "szumu", jak gdyby masy powietrza się ścierały. Pojawia się ono najczęściej w ubezdźwięcznionym otoczeniu i przy wymowie szybkiej i niewyraźnej. Jego realizacja zaczyna się ciszą (rys 5.), gdy nagle amplituda wzrasta, a przebieg przypomina charakterystyką mowę bezdźwięczną ze względu na gęstość przejść przez zero. Ze spektrogramów można odczytać duży wpływ wysokich częstotliwości w chwili skoku nawet do 8kHz. W powiększonych fragmentach przebiegu czasowego widać, że sygnał nie jest okresowy. Ponadto pojawia się również szum addytywny o wyższych częstotliwościach, co potwierdza również kształt funkcji autokorelacji.



Rysunek 5: Oscylogram i spektrogram całego sygnału oraz oscylogram i autokorelacja dla fragmentu "r frykatywnego"dla ⟨ardź⟩, osoba W02

6 Przebiegi wymowy wadliwej

Jako, że każda osoba posiadająca rotacyzm realizuje /r/ w inny, indywidualny sposób, trudnym jest wyznaczenie typowych przebiegów jak w poprzednim punkcie. Co prawda czasem można wskazać segmenty przypominające element konsonantyczny, frykatywny, wokaliczny czy uaproksymantowany, jednakże budowa fonemu /r/ potrafi być o wiele bardziej złożona. Analiza porównawcza dla następnych punktów skupiających się na różnych długościach /r/ czy elemencie wokalicznym może prowadzić do ciekawych wniosków.

7 Porównanie elementów wokalicznych wymowy prawidłowej i z rotacyzmem w fragmencie ⟨sre⟩

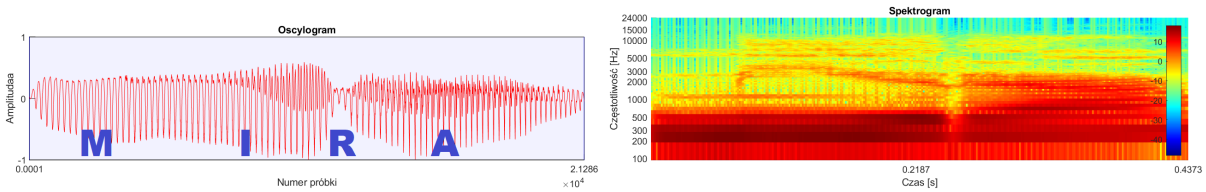
Podczas testów zaobserwowano, iż dla fragmentu ⟨sre⟩ pojawia się najczęściej segment podobny do elementu wokalicznego. Postanowiono sprawdzić takie parametry jak częstotliwość pierwszego formantu (wokół jakiej częstotliwości oscyluje drugie największe zgromadzenie energii) oraz, o ile to możliwe, czas trwania segmentu.

W przypadku F_1 uzyskano niższy dla kobiet (445-527Hz) niż dla mężczyzn (457-567Hz), jednak zawsze oscylował on wokół 500Hz, co potwierdza teorię na temat zależności częstotliwości formantowych od długości traktu głosowego. Warto pamiętać o tym, że kobiety mają krótszy kanał głosowy, co znalazło swoje odzwierciedlenie w wynikach. Ponadto dla osób z rotacyzmem, częstotliwości te były odrobinę wyższe, jednakże może to być cecha osobnicza mówców.

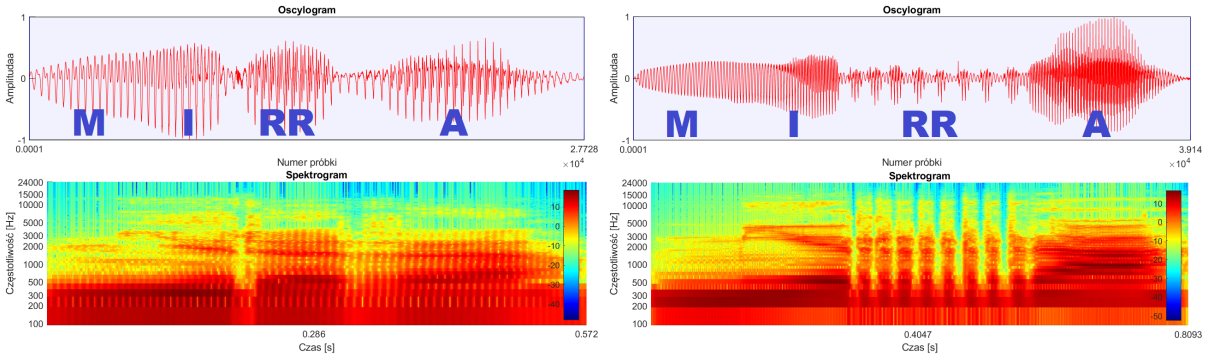
Jeżeli chodzi z kolei o czas trwania, to, pomimo braku reprezentatywności spowodowanej różnym tempem mowy, można zauważyć, że trwał on dosyć krótko, przeważnie 25-35ms z maksymalną odnotowaną wartością 56ms i minimalną 23ms.

8 Realizacja różnych długości fonemu dla wymowy prawidłowej na przykładzie wyrazu ⟨mira⟩ i ⟨mirra⟩

Postanowiono zbadać w jaki sposób różna długość fonemu /r/ wpływa na jego realizację. Dla wyrazu ⟨mira⟩ wszystkie osoby z prawidłową wymową realizowały go poprzez jedno uderzenie języka jako element konsonantyczny (rys 6.). W przypadku przedłużonego trwania dla wyrazu ⟨mirra⟩ (rys 7.) okazało się, że artykulacja może przebiegać w dwojaki sposób - albo z bardzo długim (do 80ms) elementem wokalicznym oddzielonym od reszty dwoma uderzeniami języka (bądź przybliżeniami), albo z faktycznymi wibracjami języka, które przypominają nałożenie wielu elementów konsonantycznych na długi segment wokaliczny bądź zwyczajne złożenie wielu krótszych.



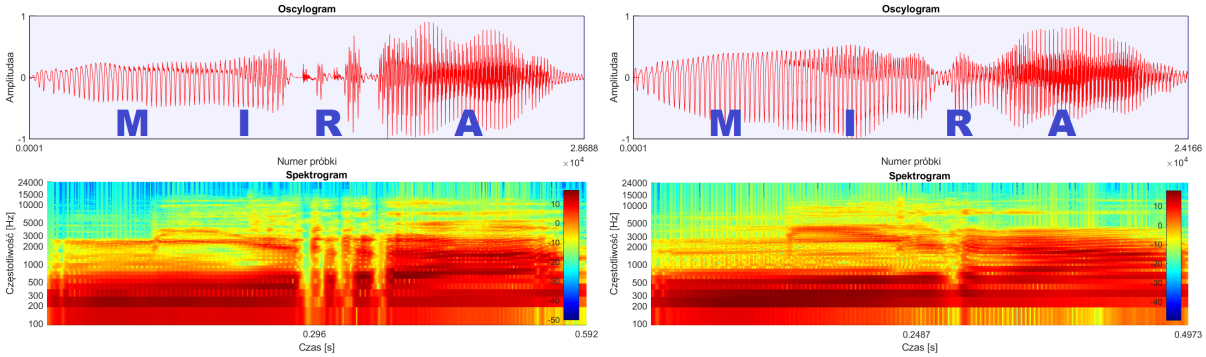
Rysunek 6: Oscylogram i spektrogram dla prawidłowej wymowy ⟨mira⟩, osoba W01



Rysunek 7: Oscylogramy i spektrogramy dla prawidłowej wymowy ⟨mirra⟩ poprzez: długi element wokaliczny (osoba M02) i wibracje języka (osoba W02)

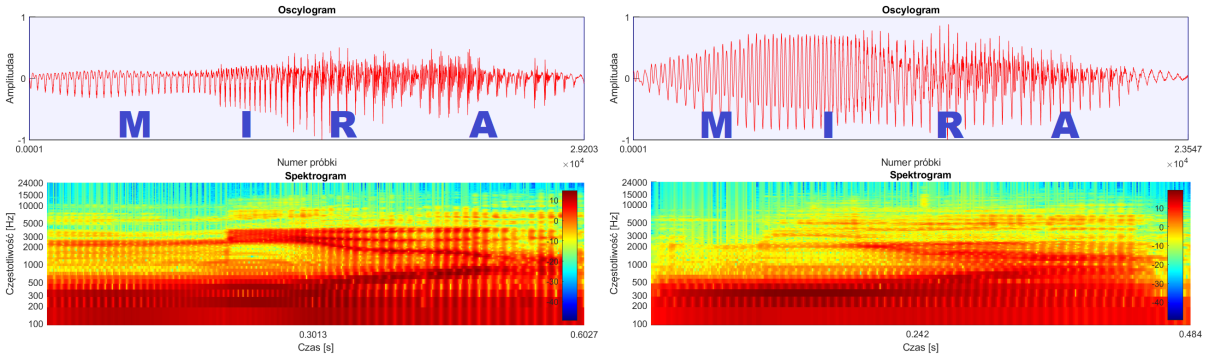
9 Postać spodziewanego elementu konsonantycznego dla wymowy z rotacyzmem na przykładzie wyrazu ⟨mira⟩

Skoro dla wyrazu ⟨mira⟩ wypowiadanego przez osobę z poprawną wymową pojawiał się element konsonantyczny, to postanowiono sprawdzić, jaki substytut będą się starały wyartykułować osoby z rotacyzmem. Otrzymano trzy główne kategorie przebiegów. Pierwszy z nich (dla np. W03 i W04, które generowały silnie brzmiące drżenia), polega na zjawisku wytworzenia wibracji zamiast po prostu jednokrotnego uderzenia języka, co skutkuje nietypową polisegmentalną budową (rys. 8.).



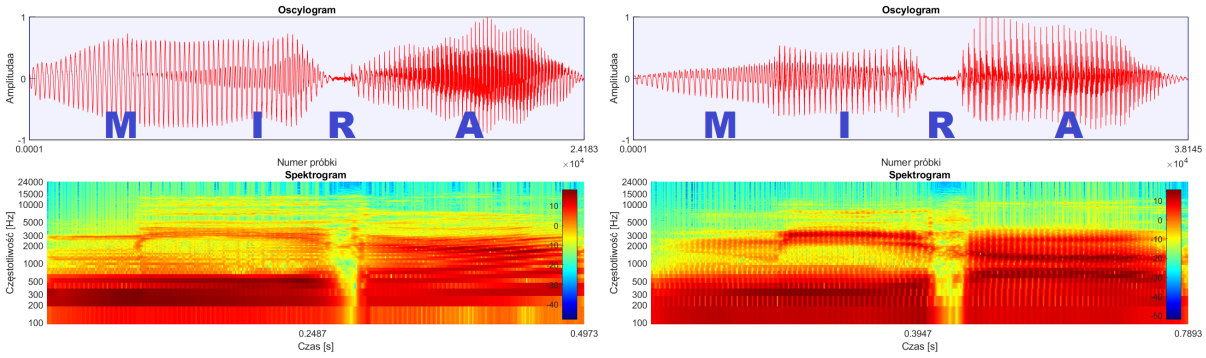
Rysunek 8: Oscylogramy i spektrogramy dla nieprawidłowej wymowy ⟨mira⟩, odmiana "zwibrowana"(osoby W03 i W04)

Zdecydowanie najczęściej zaobserwowano przebieg przypominający "r uaproksymantowane"(rysunek 9) ze względu na brak elementów konsonantycznych, a jedynie niewielkie zafalowania amplitudy obwiedni sygnału. Owe wahania mają sprawić wrażenie odrębnej głoski, aby zapewnić niepostrzeżone przejście częstotliwości formantowych z ⟨i⟩ do ⟨a⟩.



Rysunek 9: Oscylogramy i spektrogramy dla nieprawidłowej wymowy ⟨mira⟩, odmiana "zafalowania amplitudy"(osoby M06 i M07)

Wyróżniono także rodzaj realizacji, której ze słuchu przypomina "przydech", stłumione ⟨h⟩ (rys. 10). Na oscylogramie można zauważyć wcięcie, jednakże swoją zawartością widmową przypomina ono mniej element konsonantyczny, a bardziej "r frykatywne".



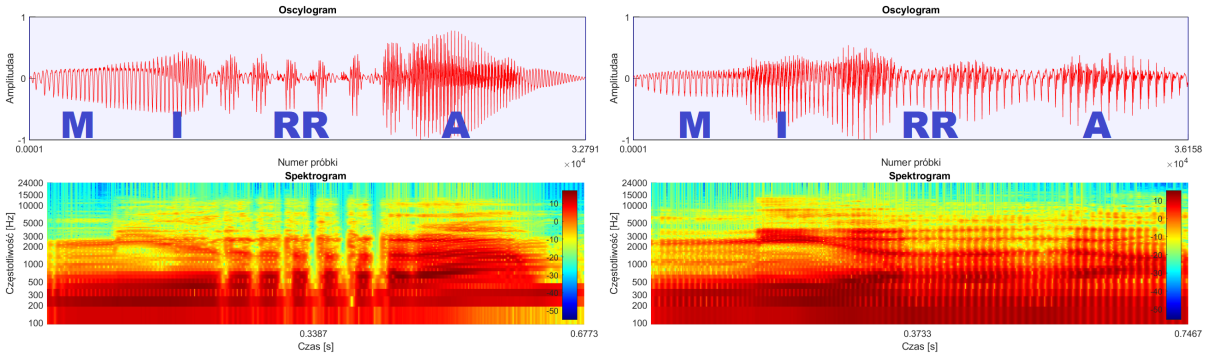
Rysunek 10: Oscylogramy i spektrogramy dla nieprawidłowej wymowy ⟨mira⟩, odmiana "przydech"(osoby W05 i M10)

10 Postać spodziewanych wibracji języka dla wymowy z rotacyzmem na przykładzie wyrazu ⟨mirra⟩

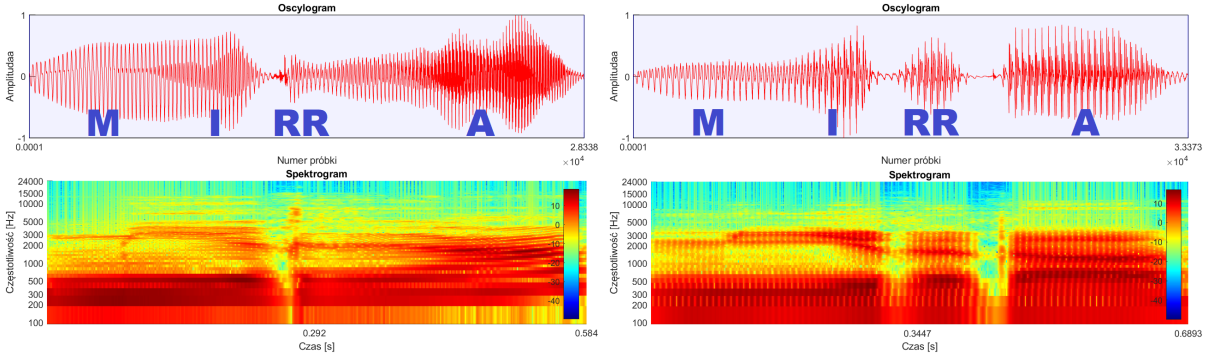
Analogiczny tok rozumowania przyjęto dla wyrazu ⟨mirra⟩ - tym razem sprawdzono, jak osoby z rotacyzmem postarają się wyartykułować drżenie języka, którego nie używają. Co prawda z trudnością wyróżniono główne rodzaje, jednakże wspólnym mianownikiem ich wszystkich jest dłuższe trwanie dźwięku.

Osoby realizujące odmianę "zwibrowaną" brzmią bardziej naturalnie, ponadto przebiegi czasowe przypominają już te dla wibracji przy prawidłowej artykulacji (rys. 11). Dla osób z "zafalowaniem amplitudy"widać większą liczbę zafalowań obwiedni, a z kolei badani robiący "przydech" realizują silniejsze ⟨h⟩, dzięki czemu jest łatwiejsze do wyodrębnienia. Zauważono również, że niekiedy ktoś był w stanie wytworzyć jeden duży

element wokaliczny, jednakże jego brzmienie było nienaturalne (rys. 12).



Rysunek 11: Oscylogramy i spektrogramy dla nieprawidłowej wymowy ⟨mirra⟩, odmiany: "zwibrowana"i "zafalowania amplitudy"(osoby W03 i M07)



Rysunek 12: Oscylogramy i spektrogramy dla nieprawidłowej wymowy ⟨mirra⟩, odmiany: "przydech"i "względnie poprawna"(osoby W05 i M10)

11 Podsumowanie

W niniejszym referacie zaprezentowano jak nieuchwytną naturę w wymowie prawidłowej ma fonem /r/, co zwiększa trudność w ustaleniu czym właściwie on jest. Interesującym jest też to, że pomimo małej liczebności grupy otrzymano szerokie spektrum obserwacji różnych postaci rotacyzmu. Dzięki stworzonej bazie nagrań można przeprowadzić dalsze badania w np. dziedzinie cepstrum. Używane w systemach rozpoznawania mowy współczynniki mel-cepstralne wykorzystują skalę melową, która polega na psychoakustycznym odbieraniu dźwięków. Jeżeli ludzie rozpoznają wady wymowy ze słuchu, to takie parametry mogłyby być dobrymi wskaźnikami rotacyzmu.

12 Podziękowanie

Publikacja została przygotowana dzięki wsparciu Fundacji Wspierania Rozwoju Radiokomunikacji i Technik Multimedialnych.

Literatura

[1] M. Kamiński, "Analiza cech charakterystycznych języka mówionego z wadami wymowy", prac. inż., Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska, 2024 (opiekun pracy dr hab. inż. Kajetana Snopek)

[2] R. Makowski, "Automatyczne Rozpoznawanie Mowy - wybrane zagadnienia", Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2011

[3] T. P. Zieliński, Cyfrowe przetwarzanie sygnałów - Od teorii do zastosowań", Warszawa: Wydawnictwa Komunikacji i Łączności, 2007

[4] V.Dellwo, M.Huckvalei, M.Ashby, "Speaker Classification I"w *How Is Individuality Expressed in Voice? An Introduction to Speech Production and Description for Speaker Classification*, Springer Berlin, Heidelberg, 2007

[5] B. Wierzchowska, "Wymowa polska", Warszawa: Państwowe Zakłady Wydawnictw Szkolnych, 1971

[6] L. Dukiewicz, I. Sawicka, "Fonetyka i fonologia", Kraków: Wydaw. IJP PAN, 1995

[7] W. Jassem, "Podstawy fonetyki akustycznej", Warszawa: Państwowe Wydawnictwo Naukowe, 1973

[8] R. Gubrynowicz, "Komputerowe modelowanie artykulacji głosek języka polskiego", Instytut Podstawowych Problemów Techniki PAN, 2000

[9] A. Sołtys-Chmielowicz, "Zaburzenia artykulacji. Teoria i praktyka", Oficyna Wydawnicza IMPULS, 2013

[10] S. Jaworski, "Phonetic Realisations of the Polish Rhetoric Intervocalic Position : A Pilot Study", *Annales Neophilologiarum*, Wydział Filologiczny Uniwersytetu Szczecińskiego", 2010

[11] Ł. Stolarski, "Tap as the basic allophone of the Polish rhotic", *Linguistica Silesiana*, Polska Akademia Nauk, Oddział w Katowicach, 2013

[12] Ł. Stolarski, "Further Analysis of the Articulation of /r/ in Polish – the Postconsonantal Position", *SKY Journal of Linguistics*, 2015

Tomasz **KOPCIŃSKI**[1], Maurycy **KIN**[1]

PROJEKT I WYKONANIE LAMPOWEGO WZMACNIACZA HI-FI Z LAMPAMI TYPU NUVISTOR W OBWODACH PRZEDWZMACNIACZA.

DESIGN AND REALIZATION OF A HI-FI TUBE AMPLIFIER WITH NUVISTOR TUBES IN THE PREAMPLIFIER.

[1]Politechnika Wrocławska

253179@student.pwr.edu.pl

Streszczenie

Niniejszy referat ma na celu pokazanie sposobu budowy, wykonania oraz wykonania pomiarów wzmacniacza elektroakustycznego zbudowanego w oparciu o lampy elektronowe, szczególnie nuvistory wykorzystane w obwodzie przedwzmacniacza. Lampy nuvistorowe 6N52S wykorzystano w stopniu wejściowym oraz układzie odwracającym fazę w układzie samosymetryzyującym (tzw. kołyski). W pojedynczej końcówce mocy wykorzystano parę pentod mocy EL84 pracujących w układzie Push – Pull. We wzmacniaczu zastosowano również pasywny układ regulacji barwy dźwięku w układzie Baxhandalla. W dokumencie przedstawiono kolejne etapy projektowania, budowy oraz pomiarów. Wykonane urządzenie oraz przeprowadzone pomiary i testy subiektywne potwierdzają możliwość zastosowania lamp nuvistorowych w urządzeniach pracujących z sygnałem fonicznym. Zastosowanie nuvistorów pozwoliło na obniżenie napięć anodowych względem tradycyjnie stosowanych triod ECC83, zmniejszenie wymiarów urządzenia oraz uzyskanie pasma przenoszenia od 20 Hz do 20 kHz z nierównością -2,6 dB/+0,0 dB w stosunku do 1kHz.

1. Wprowadzenie

Niniejszy artykuł ma przedstawić sposób wykonania wzmacniacza zintegrowanego bazującego na technologii lampowej. Szczególnym aspektem pracy są wykorzystane w niej lampy w stopniu przedwzmacniacza. Lampy miniaturowe, nuvistory zostały wykorzystane w stopniu przedwzmacniacza. Lampy nuvistorowe znalazły szerokie zastosowanie w technice analogowej w układach wysokiej częstotliwości. W układach pracujących z sygnałem fonicznym są stosowane przez firmę „Nuvista”, a niegdyś znalazły zastosowanie w mikrofonach oraz magnetofonach studyjnych.

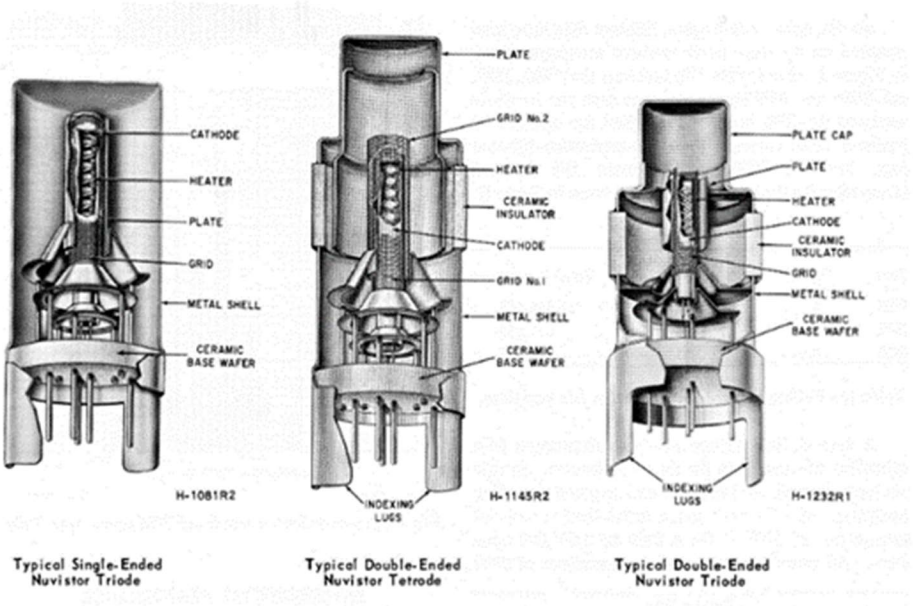
2. Cel

Projekt oraz wykonanie wzmacniacza mocy z wykorzystaniem lamp nuvistorowych w stopniu przedwzmacniacza ma pokazać, że wykorzystanie lamp miniaturowych, pierwotnie dedykowanych dla zakresu w.cz., jest możliwe także w obwodach elektronicznych małej częstotliwości. nuvistory pozwalają na zmniejszenie wymiarów urządzenia oraz uzyskanie parametrów wzmacniacza pozwalających na spełnienie założeń normy Hi-Fi.

3. Nuvistor

Lampy nuvistorowe wynaleziono w 1959 r. i stosowano w wielu konstrukcjach militarnych, przemysłowych i konsumenckich, takich jak głowice telewizyjne czy promy kosmiczne. Konstrukcja nuvistora składa się z lekkich wsporników i cylindrycznych elektrod zamkniętych w metalowo ceramicznej obudowie. Budowa cylindryczna umożliwia uzyskanie porównywalnych parametrów elektrycznych i termicznych. Przekroje podstawowych typów lamp nuvistorowych przedstawiono na rysunku 1.

Ponadto znaczącą przewagą lamp nuvistorowych jest ich niezawodność wynikająca ze specyficznej konstrukcji lamp tego typu oraz wykazują dużą odporność na promieniowanie jonizujące i niskie mikrofonowanie, co przekładało się na zastosowanie ich w przemyśle militarnym.[1]



Rysunek 1. Przekroje nuvistorów [1]

4. Założenia projektowe

Projekt zakłada wykonanie wzmacniacza lampowego z końcówką mocy typu Push-Pull o parametrach U_{wej} =150 mV i mocy wyjściowej minimum 10 W. Do tego celu wykorzystano końcówkę zrealizowaną na dwóch lampach EL84 oraz przedwzmacniacz zbudowany w oparciu o nuvistory 6S52N. Projektowane urządzenie będzie posiadać dwa niezależne układy przedwzmacniacza z korekcją barwy dźwięku oraz końcówką mocy. Zasilanie układu będzie realizowane przez wspólny zasilacz z mostkiem Gretza, filtracją napięcia oraz opóźnionym załączaniem napięcia anodowego.

5. Konstrukcja przedwzmacniacza

Przedwzmacniacz wykonano przy wykorzystaniu 3 lamp 6N52S z czego para lamp pracuje jako odwracacz fazy w układzie tzw. kołyski. Każdy z dwóch stopni posiada wzmocnienie około 40 V/V. Konstrukcję stopni oporowych oparto na wiedzy zawartej w książce „Wzmacniacze elektroniczne" G. S. Cykin [2]. Wejście wzmacniacza może pracować z sygnałem o amplitudzie 150 mV. Wejścia dla sygnałów o większym napięciu są realizowane poprzez rezystorowe dzielniki napięcia, które nie powodują obciążenia wyjść urządzeń podłączanych do wzmacniacza. Pierwszy stopień jest tradycyjnym układem wzmacniacza oporowego z automatyczną polaryzacją siatki, w której ujemne napięcie siatki sterującej względem katody lampy uzyskuje się poprzez spadek napięcia na rezystorze wpiętym w szereg pomiędzy katodą, a masą układu. Wzmocniony sygnał wyjściowy ze stopnia oporowego jest pobierany z anody lampy poprzez kondensator foliowy 100 nF o maksymalnym napięciu pracy 450 V, przez co sygnał podawany na układ korekcji częstotliwościowej jest pozbawiony składowej stałej równej napięciu na anodzie lampy poprzedniego stopnia wzmacniającego. Do obliczeń przyjęto napięcie zasilające równe 200 V przy napięciu na anodzie lampy równym 100 V. Wysokie napięcie na anodzie lampy powoduje uzyskanie niższych zniekształceń sygnału na stopniu oporowym. Przyjęte napięcie siatki sterującej wynosi -1 V. Zastosowane elementy według obliczeń pozwalają na równomierne wzmocnienie sygnału w pełnym paśmie akustycznym.

6. Wykorzystane elementy

Lampa EL84

Lampa EL84 jest pentodą mocy wykorzystywaną w stopniach końcowych wzmacniaczy w klasie Single Ended oraz Push-Pull. Ze względu na dostępność i przystępne ceny jest chętnie wykorzystywana zarówno w amatorskich konstrukcjach jak i urządzeniach wysokiej klasy.

Lampa nuvistorowa 7895/6S52N

Lampy nuvistorowe zaliczają się do lamp niezawodnych, które charakteryzują się dużą niezawodnością i trwałością, niewielkim rozrzutem parametrów, odpornością na wibracje i wstrząsy i małym przydźwiękiem i szumami. Ponadto są to lampy o niezwykle małych wymiarach i mają możliwość pracy w dużym zakresie temperatur (od -100 °C do 350 °C). Lampy 7895/6S52N są triodami o dużym wzmocnieniu K_a=64. W opisywanej pracy lampy tego typu będą wykorzystywane w stopniu przedwzmacniacza oraz stopniu odwracającym fazę w układzie samosymetryzującym [5].

7. Stopień oporowy

Obliczenia

Obliczenia zostały wykonane w oparciu o rozwiązania przedstawione w literaturze [2], [3]. Wzory od 1 do 43 przedstawiają poszczególne etapy obliczeń stopnia oporowego prezentowanego wzmacniacza.

Początkowym etapem obliczania stopnia oporowego na triodzie jest założenie warunków, w jakich ma on pracować. Przyjęto wartości:

$$f_d = 20{,}00\ Hz \quad (1)$$

$$f_g = 20{,}00\ kHz \quad (2)$$

$$M_d \leq 1{,}06 \quad (3)$$

$$M_g \leq 1{,}03 \quad (4)$$

Amplituda napięcia na wyjściu stopnia powinna wynosić 6 V przy doprowadzeniu na wejście napięcia o amplitudzie 150 mV. Napięcie anodowe podawane z zasilacza do stopnia oporowego wynosi 200 V. Zatem wzmocnienie jakie musi posiadać wybrana przez nas lampa można określić ze wzoru:

$$K_u = \frac{U_{wyj}/U_{wej}}{0{,}75} = \frac{6/0{,}15}{0{,}75} = 53{,}33 \quad (5)$$

Dane katalogowe lampy nuvistorowej 6S52N [4], [5]:

$$C_{wej} = 4{,}75\ \mu F \quad (6)$$

$$C_{wyj} = 2{,}40\ \mu F \quad (7)$$

$$C_{s_1/a} = 0{,}85\ \mu F \quad (8)$$

$$K_a = 64{,}00 \qquad (9)$$

$$\rho_{as} = 6{,}80\ k\Omega \qquad (10)$$

Wartość opornika anodowego dobiera się zgodnie z zasadą:

$$R_a = 7\rho_{as} = 47{,}00\ k\Omega \qquad (11)$$

Wartość rezystora siatkowego przyjmuje się jako dziesięciokrotność wartości rezystora anodowego:

$$R_s = 10R_a = 470{,}00\ k\Omega \qquad (12)$$

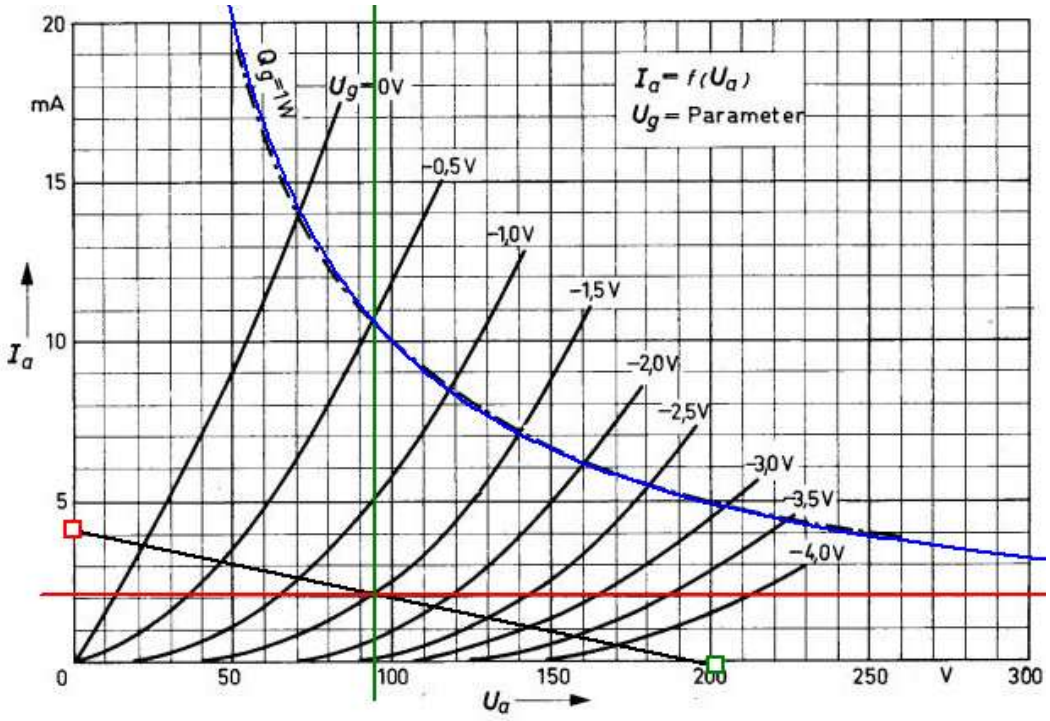
Następnie bazując na charakterystykach anodowych wybranej lampy określa się punkt pracy lampy. Założone parametry:

$$E_a = 200{,}00\ V \qquad (13)$$

$$R_a = 47{,}00\ k\Omega \qquad (14)$$

$$I_a = \frac{E_a}{R_a} = \frac{200}{47000} \approx 4{,}26\ mA \qquad (15)$$

$$U_{s_0 min} = \frac{U_{wyj}}{0{,}75K_a} + 0{,}7 = \frac{6}{0{,}75 \times 64} + 0{,}7 = 0{,}83\ V \qquad (16)$$



Rysunek 2. Wyznaczenie punktu pracy lampy z wykorzystaniem programu ECCLab [6]

Rysunek 2 przedstawia wyznaczanie punktu pracy z wykorzystaniem programu ECCLab oraz charakterystyk anodowych nuvistora. Charakterystyki lampy na podstawie noty katalogowej nuvistora 7895 będącego odpowiednikiem rosyjskiej triody 6S52N.[5]

Aby dobrać punkt pracy leżący w połowie prostej obciążenia wybrano napięcie siatki równe -1,5 V względem katody lampy. Dla takiego założenia:

$$U_a = 95{,}00\ V \qquad (17)$$

$$I_a = 2{,}40\ mA \qquad (18)$$

$$U_{s_0} = 1{,}50\ V \qquad (19)$$

Nachylenie stycznej do charakterystyki siatkowej w tym punkcie wynosi:

$$\rho_a = \frac{\Delta U}{\Delta I} = \frac{30}{0{,}002} = 15{,}00\ k\Omega \qquad (20)$$

Z tego:

$$R_{a\sim} = \frac{R_a R_s}{R_a + R_s} = \frac{47 \times 10^3 \times 470 \times 10^3}{47 \times 10^3 + 470 \times 10^3} \approx 22{,}10\ k\Omega \qquad (30)$$

oraz

$$k_{uśr} = K_a \frac{R_{a\sim}}{\rho_a + R_a} = 64 \frac{22{,}1 \times 10^3}{15 \times 10^3 + 22{,}1 \times 10^3} \approx 39{,}90 \qquad (31)$$

Taki średni współczynnik wzmocnienia stopnia jest wystarczający, ponieważ:

$$6 : 39{,}9 \approx 0{,}15\ V \leq 0{,}15\ V \qquad (32)$$

Jako pojemność montażu elementu przyjęto wartość jak do lamp z cokołem palcowym równą:

$$C_m = 6{,}00\ pF \qquad (33)$$

Dla której całkowita pojemność obciążenia stopnia jest równa:

$$C_0 = C_{wyj1} + C_m + C_{wej2} + C_{prz2}\left(1 + \frac{U_{amax2}}{U_{smax2}}\right)$$
$$= 1\ pF + 6\ pF + 10{,}8\ pF + 0{,}5\ pF(1 + 55) = 45{,}80\ pF \qquad (34)$$

Aby wyliczyć współczynnik zniekształceń dla górnej częstotliwości pracy należy obliczyć:

$$R_{rg} = \frac{\rho_a R_a}{\rho_a + R_a} = \frac{15 \times 10^3 \times 47 \times 10^3}{15 \times 10^3 + 47 \times 10^3} \approx 11{,}37\ k\Omega \qquad (35)$$

A następnie przejść do wyliczenia współczynnika zniekształceń dla górnych częstotliwości:

$$M_g = \sqrt{1 + \left(\omega_g \times C_0 \times R_{rg}\right)^2} = \sqrt{1 + (6{,}28 \times 20 \times 10^3 \times 45{,}8 \times 10^{-12} \times 11{,}37 \times 10^3)^2}$$
$$= 1{,}00 \qquad (36)$$

Uzyskanie ujemnego napięcia na siatce sterującej będzie pochodzić ze spadku napięcia na rezystorze katodowym o oporności równej:

$$R_k = \frac{U_{s0}}{I_{k0}} = \frac{U_{s0}}{I_{a0}} = \frac{1{,}5}{0{,}0024} = 625\ \Omega \approx 680{,}00\ \Omega \qquad (37)$$

Aby obliczyć moc wydzielaną na poszczególnych rezystorach korzysta się ze wzorów:

$$P_{R_s} = \frac{U_{s_max2}^2}{2R_s} = \frac{10^2}{2 \times 470 \times 10^3} = 1{,}06\ \mu W \qquad (38)$$

$$P_{R_k} = I_{k0}^2 R_k = I_{a0}^2 R_k = 2{,}4^2 \times 10^{-6} \times 680 = 3{,}90\ mW \qquad (39)$$

$$P_{R_a} = I_{a_0^2} \times R_a = 2{,}4^2 \times 10^{-6} \times 47 \times 10^3 = 0{,}27\ W \qquad (40)$$

Pojemności kondensatorów włączanych w układ dla założenia, że zniekształcenia dla dolnych częstotliwości mają być mniejsze od M_d=1,06 ($Md=M_{ds} \times M_{dk}$, $M_{ds}=M_{dk}$=1,03) wylicza się ze wzorów:

$$C_s = \frac{0{,}159}{f_d R - s\sqrt{M_{ds}^2 - 1}} = \frac{0{,}159}{16 \times 470 \times 10^3 \times \sqrt{1{,}03^2 - 1}} = 86{,}00\ nF \qquad (41)$$

$$C_k = \frac{0{,}159}{f_d R_k} \sqrt{\frac{(1 + S_{kd} R_k)^2 - M_{dk}^2}{M_{dk}^2 - 1}} = \frac{0{,}159}{16 \times 680} \sqrt{\frac{(1 + 1{,}73 \times 10^{(-3)} \times 680)^2 - (1{,}03)^2}{(1{,}03)^2 - 1}}$$
$$\approx 120{,}00\ \mu F \qquad (42)$$

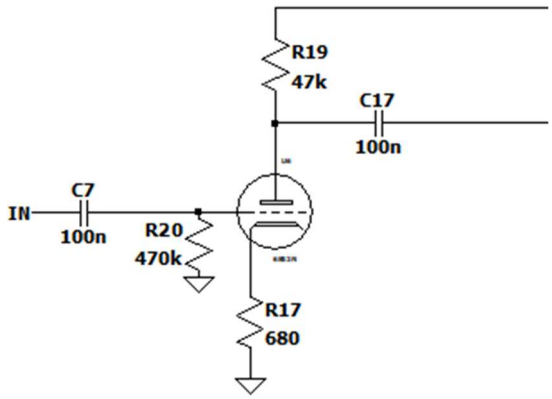
Gdzie:

$$S_{kd} = S_{ad} = \frac{K_a}{\rho_a + R_{a\sim}} = \frac{64}{15 \times 10^3 + 22{,}1 \times 10^3} = 1{,}73 \times 10^{-3} \qquad (43)$$

Schemat pierwszego stopnia wzmacniającego

Wejście wzmacniacza jest chronione przed pojawieniem się stałego napięcia na wejściu przez kondensator C7. Rezystor siatkowy R20 odpowiada za rezystancję układu widzianą przez urządzenie podłączone do wejścia wzmacniacza. Ponadto ustala on napięcie na siatce, zapobiegając przepływaniu przez lampę zbyt dużego prądu anodowego. Rezystor R17 odpowiada za ustalenie napięcia na katodzie lampy względem siatki lampy będącej na potencjale 0 V. Rezystor R19 odpowiada za ustalenie napięcia na anodzie lampy. Spadek napięcia na tym rezystorze odpowiada wzmocnionemu i odwróconemu w fazie sygnałowi

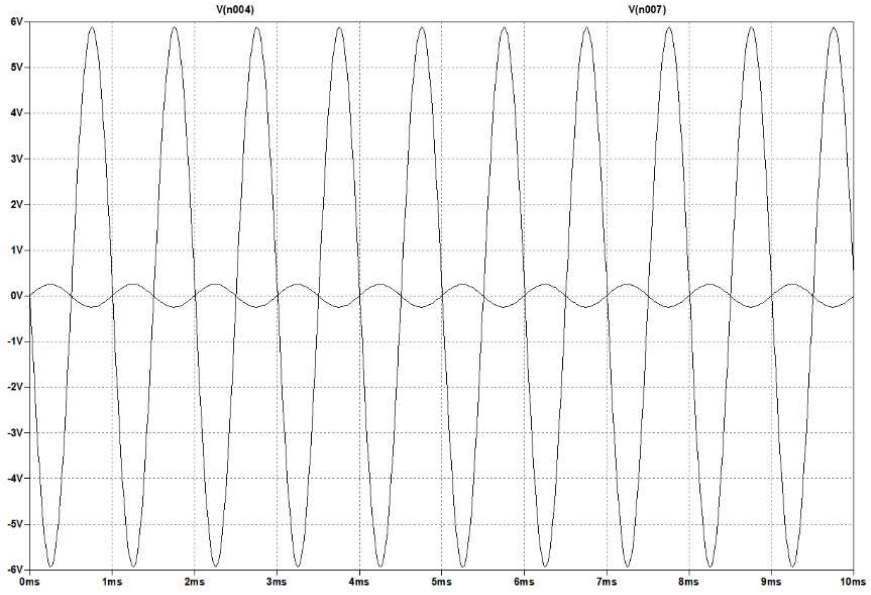
doprowadzonemu na siatkę lampy. Kondensator C17 niweluje składową stałą napięcia ze wzmocnionego sygnału, który doprowadzany jest do kolejnych stopni wzmacniających.



Rysunek 3. Schemat pierwszego stopnia wzmacniającego na triodzie 6S52N

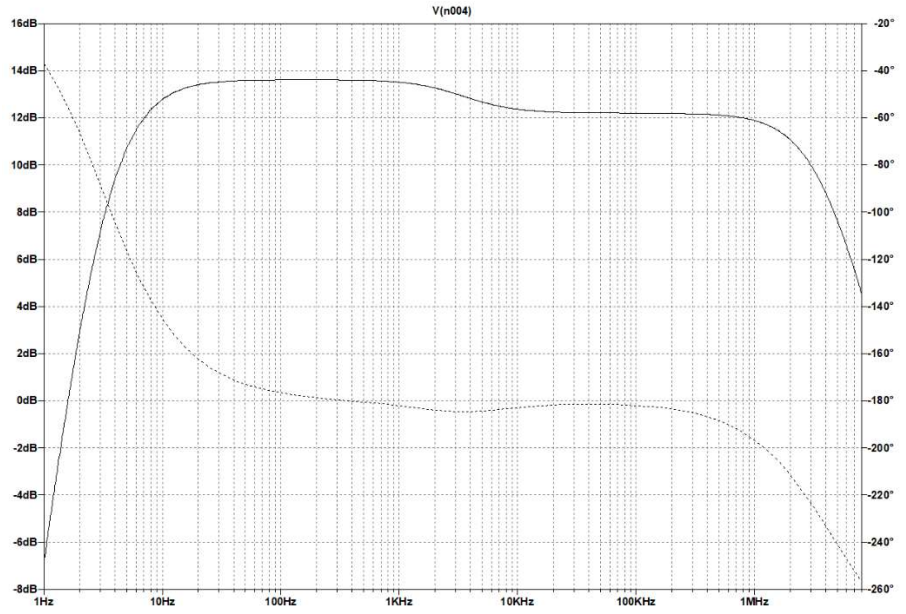
Symulacje

Do wykonania symulacji wykorzystano program LTSpice [7] firmy Analog Device's wraz z dodatkowymi bibliotekami elementów [8]. Wykonano symulacje czasowe oraz częstotliwościowe układu przedstawionego na rysunku 3.



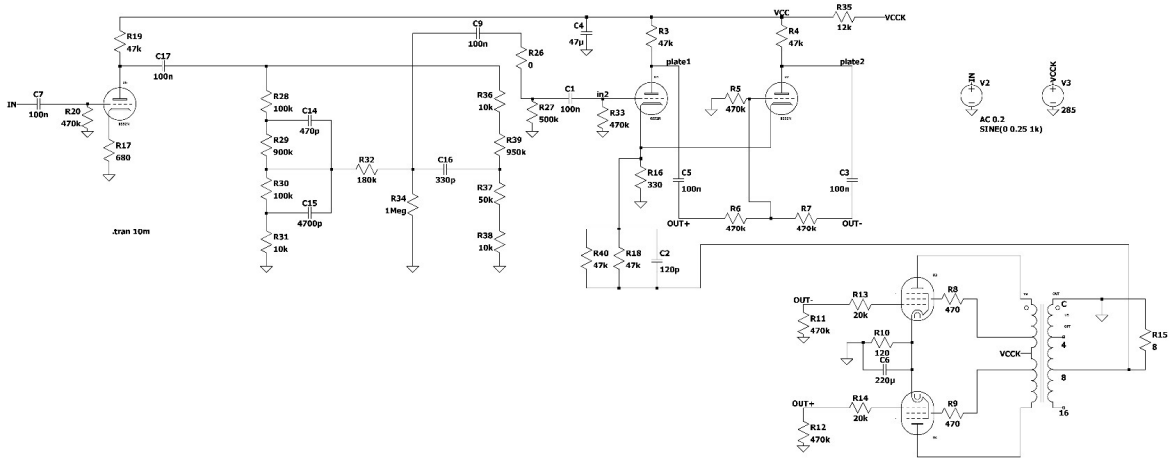
Rysunek 4. Symulacja przebiegu czasowego dla stopnia oporowego przy f = 1 kHz

Jak wynika z przedstawionej na rysunku 4. symulacji można stwierdzić, iż zaprojektowany układ wzmacniacza oporowego na triodzie nuvistorowej 6S52N działa poprawnie dla badanej częstotliwości. Sygnał na wyjściu badanego stopnia jest odwrócony w fazie względem sygnału doprowadzanego na siatkę i wzmocniony około 38,5 razy.



Rysunek 5. Charakterystyka amplitudowa oraz fazowa w funkcji częstotliwości badanego wzmacniacza oporowego w zakresie od 1 Hz do 8 MHz

Z wykonanej symulacji przedstawionej na rysunku 5 można odczytać, iż od 20 Hz do 20 kHz pasmo przenoszenia stopnia wynosi +0,1/-1,2 dB w stosunku do 1 kHz. Użyteczny zakres częstotliwości pracy badanego wzmacniacza przy spadku wzmocnienia o 3 dB w stosunku do wzmocnienia dla częstotliwości 1 kHz zawiera się w paśmie od 4,5 Hz do 3 MHz. Wykonane pomiary pozwalają stwierdzić, iż badany stopień przedwzmacniacza spełnia założenia normy DIN45500, dla której pasmo przenoszenia wzmacniacza przy nierównomierności wzmocnienia równej ± 1,5 dB zawiera się w przedziale od 40 Hz do 16 kHz. Schemat urządzenia został przedstawiony na rysunku 6.



Rysunek 6. Schemat ideowy pojedynczego kanału wzmacniacza

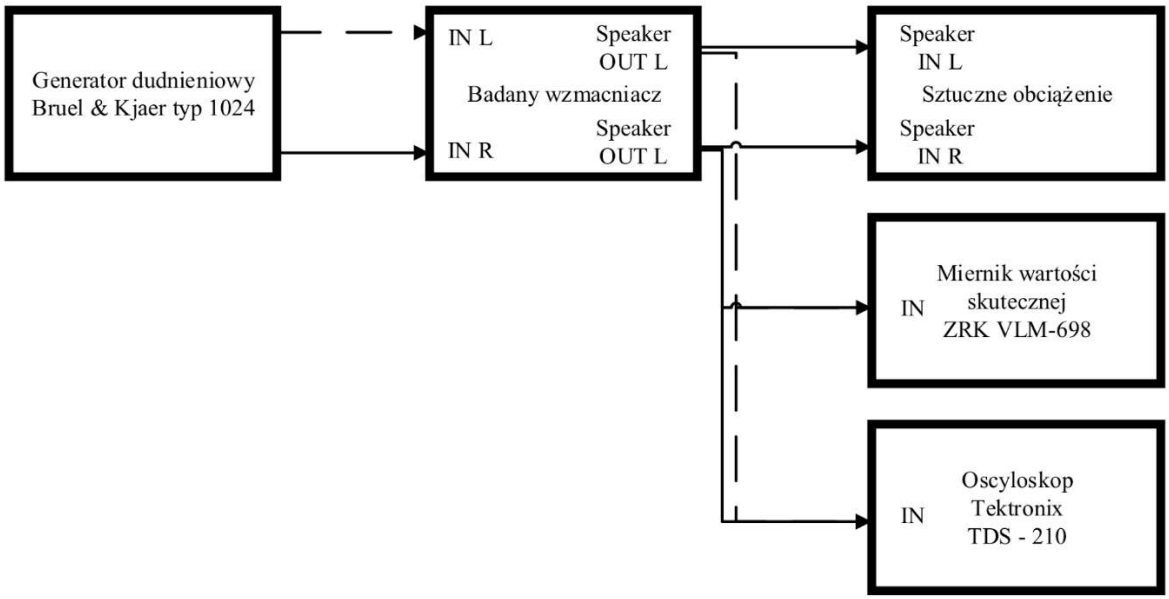
8. Pomiary urządzenia

Pomiary wykonano przy wykorzystaniu generatora dudnieniowego, miernika wartości skutecznej, oscyloskopu oraz sztucznego obciążenia o rezystancji 8 Ω.

Wykaz wykorzystanych urządzeń:

- Generator dudnieniowy Bruel & Kjaer typ 1024
- Miernik wartości skutecznej ZRK VLM-698
- Oscyloskop Tektronix TDS 210

Schemat układu pomiarowego został przedstawiony na rysunku 7.



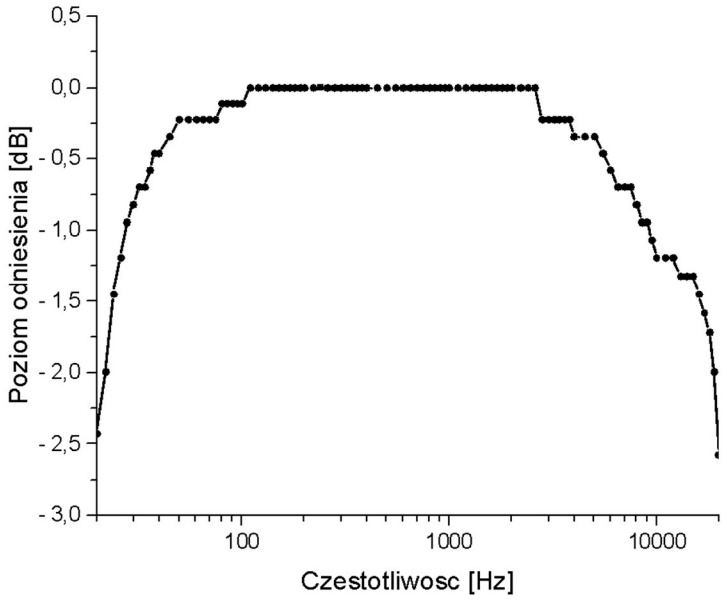
Rysunek 7 Schemat układu pomiarowego

Pomiary przeprowadzono dostarczając do wejścia wzmacniacza sygnału sinusoidalnego o napięciu 430 mV RMS. Dla pomiaru mocy maksymalnej zwiększano napięcie aż do uzyskania przesterowania sygnału widocznego na oscyloskopie (zastosowanie takiej metody pomiaru wynikało z ograniczonych możliwości urządzeń wykorzystanych w pomiarach). Widoczne zniekształcenia uzyskano dla napięcia wejściowego 750 mV RMS (zgodnie z normą opisującą pomiary wzmacniaczy dla pozostałych pomiarów napięcie na wejściu wzmacniacza powinno być o 10 dBu [około 320 mV] niższe niż maksymalne napięcie wejściowe, przy którym napięcie wyjściowe wzmacniacza jest ograniczone zniekształceniami). Celem przeprowadzenia pomiarów było sprawdzenie czy zbudowany wzmacniacz spełnia założenia normy Hi-Fi (DIN 45500).

Uzyskane wyniki:

Tabela 1. Uzyskane parametry

Pasmo przenoszenia	-2,6 dB/+0,0 dB od 20 Hz do 20 kHz w stosunku do 1 kHz
Przesłuchy międzykanałowe	-52 dB
Różnice międzykanałowe	0,1 dB
SNR	-54 dB
Moc wyjściowa	12 W



Rysunek 8. Pasmo przenoszenia badanego wzmacniacza w odniesieniu do 1 kHz

Zaprojektowany i wykonany wzmacniacz spełnia założenia normy Hi-Fi (DIN 45500) [9] w zakresie przeprowadzonych testów. Pasmo przenoszenia wzmacniacza zawiera się w przedziale od 20 Hz do 20 kHz przy nierównomierności -2,6 dB w stosunku do 1 kHz przy częstotliwości 20 kHz oraz -2,4 dB dla częstotliwości 20 Hz. Charakterystyka częstotliwościowa badanego wzmacniacza została przedstawiona na rysunku 8. Zgodnie z założeniami normy DIN 45500 w paśmie do 40 Hz do 16 kHz nierównomierności w paśmie przenoszenia nie przekraczają ±1,5dB. Przesłuchy międzykanałowe dla częstotliwości 1 kHz wynoszą -52 dB, co również przekracza założenia normy, które mówią o minimum -50 dB dla 1 kHz. Różnica między kanałami stereo przy doprowadzeniu sygnału sinusoidalnego o częstotliwości 1 kHz do obu wejść wzmacniacza i pomiarze napięcia na jego wyjściach wynosi 0,1 dB co również przekracza założenia normy mówiące o różnicy mniejszej niż 3 dB. SNR badanego wzmacniacza wynosi -54 dB, a moc wyjściowa około 12 W. Szacowana wartość THD przy mocy wyjściowej wynosi 1 %. Szacunki oparto na pomiarach kształtu sygnału

wyjściowego wykonanych oscyloskopem z pomiarem THD. Tabela 1 przedstawia uzyskane wyniki.

9. Testy subiektywne

Badania subiektywne polegały na odsłuchaniu trzech utworów, („Pirates of the Caribbean" – Hans Zimmer (album „Live In Prague"), „Sunrise, sunset" – Zdzisława Sośnicka (album „Musicals"), „Get Lucky" – Daft Punk and Pharrell Williams (album „Random Access Memories"). Źródłem sygnału fonicznego był odtwarzacz CD Yamaha CD750, źródłem dźwięku były zestawy głośnikowe Tonsil Bolero 200. Badanie polegało na odsłuchaniu trzech utworów z różnych gatunków muzycznych i opisanie różnic w odbiorze pomiędzy dźwiękiem z wzmacniacza tranzystorowego względem zbudowanego wzmacniacza lampowego. Uzyskane odpowiedzi od wszystkich badanych pozwoliły na określenie subiektywnych parametrów dźwięku uzyskiwanego przy użyciu zbudowanego wzmacniacza.

Wszystkie osoby określiły wzmacniacz jako bardzo dobry, przewyższający model referencyjny. Średnica pasma była określana jako dobrze brzmiąca (klarowne wokale), basy i soprany były wyraźnie słyszalne, bez zbędnych podbić. Dźwięk był detaliczny, a całość określana była jako bardzo dobra.

10. Podsumowanie

Nuvistory mogą z powodzeniem być wykorzystywane w układach elektronicznych małej częstotliwości pracujących z sygnałem fonicznym. Zastosowanie nuvistorów pozwala na zmniejszenie napięć anodowych oraz zaoszczędzenie miejsca względem tradycyjnych lamp typu ECC83 stosowanych w obwodach przedwzmacniaczy lampowych. Ponadto są bardziej wytrzymałe na uszkodzenia mechaniczne.

Wykonane pomiary pokazują, że zastosowanie lamp nuvistorowych przynosi zadowalające rezultaty. Moc wyjściowa wzmacniacza jest porównywalna z seryjnymi produktami w klasie AB wykorzystującymi lampy EL84. Zgodnie z założeniami normy PN-EN 61305 część 3 – Specyfikacja parametrów i metody ich pomiaru [10] – wzmacniacze pasmo przenoszenia wzmacniacza powinno zawierać się w granicach ± 2 dB dla wejść z korekcją. Dla badanego wzmacniacza pasmo przenoszenia w granicy +2,0 dB / -0,0 dB od 22 Hz do 18 kHz.

Dla mocy wyjściowej 12 W szacowana wartość THD wzmacniacza wynosi 1 %.

Literatura

[1] N. J. RCA Electronic Components and Devices Harrison, „RCA Nuvistors Industrial and Military”, 1967.

[2] G.S. Cykin, *Wzmacniacze elektroniczne*. Moskwa: Wydawnictwa Komunikacji i Łączności, 1965.

[3] Aleksander Zawada, *Lampy elektronowe w aplikacjach audio*. Legionowo: Wydawnictwo btc, 2011.

[4] Leonard Niemcewicz, *Lampy elektronowe i półprzewodniki*. Warszawa: Wydawnictwa Komunikacji i Łączności, 1975.

[5] RCA Electron Tube Division, „RCA 7895 datasheet”, Harrison.

[6] Wojciech Staszak, „ECCLab”. 2007.

[7] Analog Device’s, „LTSpice”.

[8] Jerzy Witkowski, „http://156.17.38.202/”.

[9] Deutsches Institut für Normung, „DIN 45500”. 1966.

[10] PKN, *PN-EN 61305 część 3 – Specyfikacja parametrów i metody ich pomiaru*. Polska: PKN, 2001.

Dominika **KUCZAK**[1], Maurycy **KIN**[1]

BADANIE WPŁYWU PRZETWARZANIA SYGNAŁU NA ZMIANY WRAŻEŃ SŁUCHOWYCH NAGRAŃ DŹWIĘKOWYCH

INVESTIGATING THE EFFECT OF SIGNAL PROCESSING ON CHANGES IN THE AUDITORY SENSATIONS OF SOUND RECORDINGS

[1]Politechnika Wrocławska
dominika.kuczak@outlook.com

Streszczenie

W niniejszej pracy przeprowadzono analizę nagrań, które poddano dwóm rodzajom przetwarzania: kompresji dynamicznej o różnych wartościach współczynnika kompresji oraz zmianie prędkości odtwarzania. Badanymi sygnałami były fragmenty utworów muzycznych oraz nagrania mowy. W ramach analizy przeprowadzono dwa eksperymenty: pierwszy wykonany za pomocą ankiety online z wykorzystaniem 5-stopniowej skali oceny ACR (ang. Absolute Category Rating), natomiast drugi przeprowadzono w jednakowych warunkach odsłuchowych przy użyciu metody porównawczej CCR (ang. Comparison Category Rating) o 7-stopniowej skali porównawczej. Wyniki pomiarów wykazały, że pomimo pogarszającej się oceny badanych atrybutów wrażenia słuchowego wraz ze zwiększeniem się prędkości odtwarzania, zrozumiałość wypowiedzi słownych była oceniana na poziomie akceptowalnym. Oznacza to, że można dokonać skrócenia czasu odtwarzania wypowiedzi, prowadząc do oszczędności czasu lub zmniejszenia objętości informacyjnej nagrania. W przypadku kompresji dynamiki można stwierdzić, że wysoki współczynnik kompresji powodował pogorszenie jakości ogólnej ocenianych sygnałów.

1. Wstęp

W warunkach powiększającej się ilości danych gromadzonych wraz ze zwiększaniem zawartości przetwarzanych i gromadzonych informacji istotnym zagadnieniem okazuje się czas poświęcony na przetwarzanie tych informacji. Szczególna sytuacja miała miejsce podczas pandemii SARS-CoV-2, kiedy to proces studiowania polegał na zdalnym uczestnictwie w zajęciach, a następnie ponownym odsłuchaniu i obejrzeniu prezentowanych nagrań. Zajmowało to znaczną ilość czasu, zdecydowanie większą w porównaniu do tradycyjnej metody nauki. Aby przyspieszyć ten proces, bardzo często powtórne odsłuchanie czy obejrzenie wykładu następowało ze zwiększoną prędkością odtwarzania.

W niniejszej pracy przeprowadzono analizę atrybutów dźwięku nagrań zmodyfikowanych w oparciu o różne parametry przetwarzania sygnału. Przedmiotem badań jest porównanie wrażeń słuchowych przy odsłuchu fragmentów tych samych nagrań dźwiękowych, poddanych różnym zabiegom przetwarzania sygnału. Niniejsza praca skupia się na dwóch rodzajach modyfikacji: zmianach prędkości oraz kompresji dynamicznej.

Bezpośrednią inspiracją tematu pracy był odczuwalny dyskomfort w trakcie odsłuchu nagranych wykładów o zwiększonej prędkości odtwarzania, a także wyznaczenie akceptowalnej redukcji dynamiki sygnału fonicznego. Zarówno skrócenie nagrania jak i kompresja dynamiki sygnału mogą posłużyć jako sposób na zmniejszenie objętości przesyłanych i magazynowanych danych.

2. Metodyka badań

W celach badawczych wykorzystano nagrania mowy oraz muzyki o różnorodnym charakterze dynamicznym, z których wyodrębniono fragmenty zawierające temat utworu o długości od 35 do 55 sekund w taki sposób, aby tworzyły zamknięty element struktury muzycznej [1]. Dokonano także nagrań mowy dwóch głosów: męskiego i żeńskiego, trwających około 15 sekund oraz 6-sekundowego dialogu przeprowadzonego przez te same osoby. Nagrania zrealizowano przy pomocy mikrofonu Superlux E205U oraz oprogramowania komputerowego REAPER v. 6.82.

Wszystkie pliki poddano różnym zabiegom przetwarzania sygnału. W przypadku kompresji dynamicznej, do badań został wybrany parametr współczynnika kompresji o wartościach 2:1, 3:1 oraz 4:1. Ustawienia pozostałych parametrów (próg zadziałania, czas zadziałania oraz regeneracji) [2] dobrano indywidualnie dla każdego nagrania. Zmiana prędkości była wyznaczana w postaci procentowej [3,4]. Testy przeprowadzono dla wartości: -5%, 30%, 50%, 100%. Operacje związane z modyfikacją kompresji dynamicznej zostały przeprowadzone z wykorzystaniem oprogramowania REAPER v. 6.82 oraz Samplitude v.8. Zmiany prędkości odtwarzania próbek dokonano przy pomocy programu Audacity 3.3.3. Po każdej operacji daną próbkę eksportowano do pliku *.wav, co zapewniło pełną kompatybilność stosowanych systemów.

Wykorzystywany algorytm do zmian skali czasu - WSOLA (ang. Waveform Similarity Overlap-Add) dzieli sygnał na ramki, które następnie na siebie nakłada w sposób nieregularny, aby zachować ciągłość fazową. Program Audacity wykorzystuje go do jednoczesnej zmiany prędkości jak i wysokości dźwięku w nagraniu.

Badanymi atrybutami wrażenia słuchowego były: naturalność brzmienia, ogólna jakość nagrania oraz wyrazistość mowy (w przypadku nagrań słownych) lub przejrzystość sceny dźwiękowej (w przypadku nagrań muzycznych).

2.1 Eksperyment I

Pierwszy eksperyment został przeprowadzony on-line z wykorzystaniem internetowego oprogramowania Formularz Google. Mając na względzie różnorodny sprzęt, wykorzystywany do odsłuchu oraz warunki odsłuchowe, w ankiecie zapytano o rodzaj urządzenia jak również o otoczenie, w jakim przeprowadzono ocenę. Badaniu podlegał cały materiał dźwiękowy z uwzględnieniem oryginalnych fragmentów utworów. Celem testu była ocena indywidualna za pomocą 5-stopniowej skali metody oceny punktowej ACR (ang. Absolute Category Rating) [5], którą zaprezentowano w Tabeli 1.

Tabela 1. Skala oceny ACR dla poszczególnych cech wrażeniowych

cechy wrażeniowe / skala ocen	**jakość ogólna**	**naturalność brzmienia**	**wyrazistość mowy / przejrzystość utworu**
1	zła	sztuczne	niewyraźny/nieprzejrzysty
2	słaba	mało naturalne	słabo wyraźny/przejrzysty
3	dostateczna	średnio naturalne	średnio wyraźny/przejrzysty
4	dobra	dość naturalne	dobrze wyraźny/przejrzysty
5	doskonała	naturalne	bardzo wyraźny/przejrzysty

Materiał testowy podzielono na pięć sekcji, a w każdej z nich nagrania ułożono w przypadkowej kolejności. Po dokonanej ocenie, w celu dodatkowej analizy, badanym zadano pytanie dotyczące znajomości utworów muzycznych wykorzystanych w testach.

Eksperyment przeprowadzono na grupie 13 osób w wieku: od 11 do 60 lat. Znaczna większość wykonała badanie w cichym pomieszczeniu. Słuchacze mogli wykonać badanie zarówno z wykorzystaniem głośników, jak i słuchawek. Ponadto połowa słuchaczy nie rozpoznała utworów muzycznych prezentowanych podczas badania.

Dokonana analiza statystyczna otrzymanych rezultatów wykazała, iż grupa odsłuchowa nie odpowiadała w sposób jednorodny, co potwierdzono za pomocą testu jednorodności wariancji udzielonych ocen.

2.2 Eksperyment II

Ze względu na fakt, iż w poprzednim eksperymencie nie otrzymano homogenicznych wariancji ocen słuchaczy poszczególnych atrybutów wrażeń słuchowych, postanowiono wykonać kolejne badanie z wykorzystaniem oceny komparatywnej CCR (ang. Comparison Category Rating) [5], która pozwoliła na uzyskanie dokładniejszych wyników subiektywnej oceny jakości. Skala wykorzystanej metody została zaprezentowana w Tabeli 2.

Tabela 2. Skala oceny CCR dla poszczególnej cechy wrażeniowej

cechy wrażeniowe / **skala ocen**	**jakość ogólna**	**naturalność brzmienia**	**wyrazistość mowy / przejrzystość utworu**
+3	dużo lepsza	dużo bardziej naturalna	dużo bardziej wyrazisty/przejrzysty
+2	lepsza	bardziej naturalna	bardziej wyrazisty/przejrzysty
+1	nieznacznie lepsza	nieznacznie bardziej	nieznacznie bardziej wyrazisty/przejrzysty
0	taka sama	taka sama	taka sama
-1	nieznacznie gorsza	nieznacznie mniej naturalna	nieznacznie mniej wyrazisty/przejrzysty
-2	gorsza	mniej naturalna	mniej wyrazisty/przejrzysty
-3	dużo gorsza	dużo mniej naturalna	dużo mniej wyrazisty/przejrzysty

Drugie badanie zostało przeprowadzone na terenie Politechniki Wrocławskiej w odpowiedniej sali odsłuchowej, która spełnia wymagania norm [6,7]. Podczas tego badania warunki odsłuchowe, a także wykorzystywany sprzęt, były jednakowe dla każdego słuchacza. Osoby badane zapytano o wiek i o znajomości utworów muzycznych wykorzystanych w testach.

Badaniu poddano wyłącznie próbki wyodrębnione z całości materiału dźwiękowego o wartości wariancji, otrzymanej z pierwszego eksperymentu, nieprzekraczającej 1 dla oceny jakości ogólnej. W Tabeli 3 zestawiono wykorzystany materiał dźwiękowy.

Tabela 3. Zestawienie wykonanych stopni danego rodzaju przetwarzania dla poszczególnych próbek

próbki mowy		
rodzaj wypowiedzi	**wykonane stopnie kompresji**	**wykonane stopnie zmiany prędkości**
dialog	3:1, 4:1	-5%
głos męski	2:1, 3:1, 4:1	-5%, 30%
głos żeński	2:1, 3:1, 4:1	-5%, 100%
próbki muzyczne		
wykonawca i tytuł utworu	**wykonane stopnie kompresji**	**wykonane stopnie zmiany prędkości**
David Bowie „Never Let Me Down”	---	-5%, 100%
Ralph Vaughan Williams „The Wasps”	2:1, 3:1, 4:1	-5%
David S. Ware Quartet „Bliss Theme”	2:1	-5%

Ze względu na brak doświadczenia słuchaczy, na początku badania przedstawiono uczestnikom proces testowy oraz wykorzystaną metodę oceny. Następnie przeprowadzono eksperyment, który trwał 20 minut. Po każdej parze ocenianych sygnałów słuchacze udzielali odpowiedzi na pytania związane z ocenianymi atrybutami.

Badanie zostało przeprowadzone na grupie 10 słuchaczy w wieku od 20 do 30 lat. Znaczna większość rozpoznała utwór wykonany przez D. Bowiego, jedna osoba znała również utwór „Bliss Theme”. Pozostałe nagrania nie zostały rozpoznane przez słuchaczy.

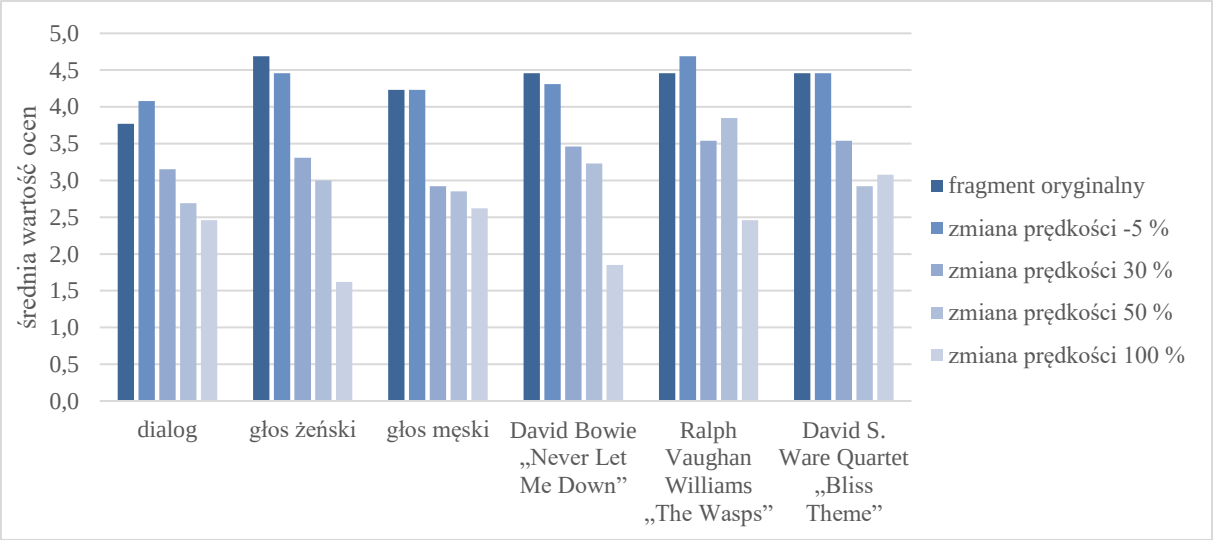
3. Wyniki

3.1 Eksperyment I

W celu weryfikacji statystycznej otrzymanych wyników przeprowadzono analizę statystyczną uzyskanych ocen. Zastosowano do tego test Bartletta [8] oparty o statystykę $\chi 2$ przy poziomie istotności $\alpha = 0{,}05$.

Dla pierwszego badania, we wszystkich trzech badanych atrybutów dźwięku uzyskano wartości $p < 0{,}05$ co oznacza, że wariancje odpowiedzi słuchaczy są niejednorodne. Pomimo tego faktu postanowiono przeanalizować otrzymane rezultaty w celu znalezienia istniejących prawidłowości oraz trendów w ocenie słuchaczy.

Na Rysunku 1 przedstawiono uśrednione wartości oceny zmian prędkości dla atrybutu „ogólna jakość dźwięku”, natomiast w Tabelach 4 oraz 5 przedstawiono wyniki oceny dla dwóch wybranych atrybutów.



Rysunek 1. Wyniki wartości średniej dla oceny jakości ogólnej dźwięku przy zmianach prędkości odtwarzania zmierzone metodą ACR

Tabela 4. Wyniki wartości średniej pierwszego badania dla naturalności brzmienia

rodzaj wypowiedzi lub utworu	**fragment oryginalny**	**stopień kompresji**			**stopień zmiany prędkości**			
		2:1	**3:1**	**4:1**	**-5 %**	**30 %**	**50 %**	**100 %**
dialog	3,62	3,92	4,15	4,08	4,00	2,23	1,69	1,31
głos żeński	4,77	4,15	4,54	4,23	4,62	1,77	1,85	1,15
głos męski	4,23	4,00	4,46	4,23	4,00	1,85	1,77	1,54
David Bowie „Never Let Me Down”	4,46	4,15	4,15	4,23	4,15	2,38	2,38	1,15
Ralph Vaughan Williams „The Wasps”	4,46	4,31	4,46	4,31	4,54	3,23	3,23	1,46
David S. Ware Quartet „Bliss Theme”	4,23	4,23	3,69	3,69	4,23	3,08	2,00	2,00

Tabela 5. Wyniki wartości średniej pierwszego badania dla zrozumiałości mowy dla nagrań mowy, bądź przejrzystości utworu dla utworów muzycznych

rodzaj wypowiedzi lub utworu	**fragment oryginalny**	**stopień kompresji**			**stopień zmian prędkości**			
		2:1	**3:1**	**4:1**	**-5 %**	**30 %**	**50 %**	**100 %**
dialog	4,23	4,23	4,23	4,46	4,46	3,31	2,69	1,62
głos żeński	4,54	4,69	4,85	4,62	4,62	2,92	3,15	1,15
głos męski	4,38	4,38	4,54	4,38	4,46	3,23	2,85	2,15
David Bowie „Never Let Me Down”	4,46	3,92	3,92	3,77	4,38	3,08	2,54	1,54
Ralph Vaughan Williams „The Wasps”	4,46	4,15	4,31	4,23	4,77	3,62	3,62	2,31
David S. Ware Quartet „Bliss Theme”	4,23	3,92	3,69	3,92	4,38	3,46	2,62	2,69

Analizując otrzymane wyniki można stwierdzić, że kompresja dynamiczna testowanych nagrań nie wpływa znacząco na wartość oceny poszczególnych atrybutów. Niekiedy zaobserwowano jednak poprawę jakości ogólnej oraz zrozumiałości, zwłaszcza w przypadku dialogu i stopnia kompresji 4:1. Może to być spowodowane wyrównaniem składowych dźwięków mowy w zakresie wyższych częstotliwości, wpływających na wrażenie wyrazistości, a tym samym zrozumiałości mowy [9,10]. Z tego też względu poniższe zjawisko znalazło zastosowanie w dźwiękowych systemach ostrzegawczych, dla których powinno zapewnić się dobrą zrozumiałość nadawanych komunikatów [2,11]. W przypadku próbek zawierających sygnały muzyczne nie odnotowano żadnych prawidłowości uzyskiwanych ocen w zależności od stopnia kompresji.

W przypadku zmiany prędkości odtwarzania, a zwłaszcza przyspieszania nagrań, można zauważyć mocną degradację jakości sygnałów muzycznych we wszystkich ocenianych aspektach. Spadek oceny jakości odnotowano także dla sygnałów mowy, niemniej jednak dla przyspieszenia odtwarzania o 30% w przypadku zrozumiałości nie jest on taki drastyczny, jak w przypadku muzyki. Może to być spowodowane sposobem percepcji treści słuchanych wypowiedzi [12,13,14]. Można zauważyć, że degradacja naturalności brzmienia, jak i jakości ogólnej, jest znaczna.

3.2 Eksperyment II

Dokonana analiza statystyczna otrzymanych rezultatów z wykorzystaniem testu Bartletta jednorodności wariancji [8] (podobnie jak poprzednio - statystyka $\chi 2$ przy poziomie istotności $\alpha = 0,05$) wykazała, że wariancje ocen grupy odsłuchowej dla wszystkich badanych atrybutów wrażenia słuchowego są homogeniczne.

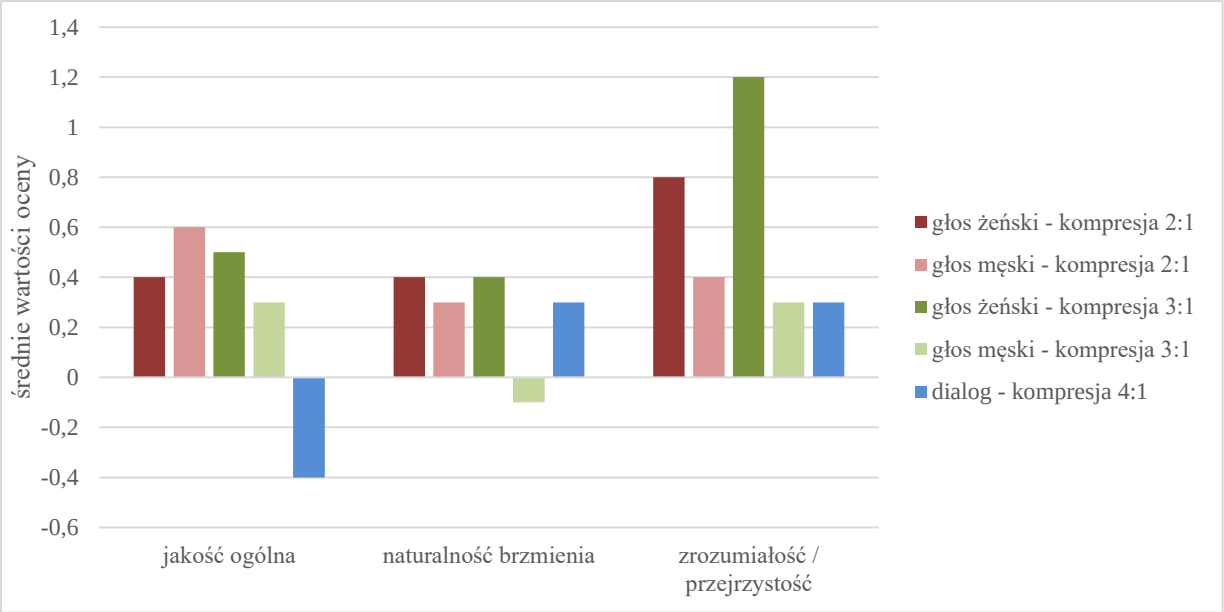
W Tabeli 6 przedstawiono wyniki oceny ogólnej jakości, naturalności brzmienia oraz zrozumiałości mowy. Rysunek 2 przedstawia średnie wartości oceny dla sygnałów poddanych kompresji dynamicznej. Ze względu na oczywisty wpływ przetwarzania na atrybuty wrażeniowe sygnałów muzycznych, zaprezentowano jedynie wyniki dla sygnałów mowy.

Na podstawie wyników można stwierdzić, że zastosowana kompresja dynamiki w zdecydowanej większości przypadków poprawiła nieco ocenę ogólną jakości nagrania, jak i zrozumiałości oraz naturalności brzmienia. Dla potwierdzenia statystycznej istotności uzyskanych wyników zastosowano test t-Studenta [15] na poziomie istotności $\alpha = 0,05$, uzyskując wartości p większe od 0,05 dla stopni kompresji 2:1 oraz 3:1. Oznacza to, że zastosowanie kompresji dynamiki o wyżej wymienionych stopniach kompresji nie wpływa istotnie na ocenę badanych atrybutów wrażenia. Jedynie dla stopnia kompresji 4:1 oraz oceny jakości ogólnej uzyskano wartość $p = 0,01$, co oznacza, że zastosowanie kompresji dynamiki ze stopniem 4:1 wpływa na wartość oceny jakości ogólnej.

Nieco inaczej przedstawia się sytuacja dla zmiany prędkości odtwarzania: dla zwolnienia o 5% odnotowano statystyczną istotność zmiany oceny jedynie w przypadku jakości ogólnej ($p = 0,03$). Dla pozostałych atrybutów dźwięku uzyskano wartości $p > 0,05$, co oznacza, że zmiana prędkości odtwarzania polegająca na zwolnieniu nagrania o 5% nie powoduje zmiany oceny badanych atrybutów. Dla przyspieszenia odtwarzania o 30% oraz 100% zmiany wszystkich badanych atrybutów są istotne statystycznie: dla obu wartości zmiany prędkości odtwarzania odnotowano pogorszenie oceny subiektywnej, co potwierdzają wyniki uzyskane w eksperymencie 1. Należy jednak zauważyć, że degradacja zrozumiałości jest znacznie mniejsza niż uzyskana w poprzednim eksperymencie. Przyjmując za dopuszczalną jakość rezultaty oceny CCR przyjmujące wartości większe od -2 [16,17] to zastosowane przyspieszenie nagrań spełnia kryterium dopuszczalnej zrozumiałości oraz jakości ogólnej dźwięku mowy, co umożliwiałoby zastosowanie tego zabiegu do celów transmisyjnych.

Tabela 6. Wyniki wartości średniej oraz wariancji drugiego badania dla jakości ogólnej, naturalności brzmienia oraz zrozumiałości dźwięków mowy

rodzaj wypowiedzi lub utworu	**rodzaj przetwarzania**	**stopień**	**jakość ogólna**	**naturalność brzmienia**	**zrozumiałość mowy/przejrzystość utworu**
dialog	kompresja	3:1	0,60	0,00	0,40
	kompresja	4:1	-0,40	0,30	0,30
	zmiana prędkości	-5 %	0,00	-0,20	0,20
głos żeński	kompresja	2:1	0,40	0,40	0,80
	kompresja	3:1	0,50	0,40	0,80
	kompresja	4:1	1,00	0,70	1,20
	zmiana prędkości	-5 %	0,00	-0,50	0,60
	zmiana prędkości	100 %	-1,90	-1,70	-1,70
głos męski	kompresja	2:1	0,60	0,30	0,40
	kompresja	3:1	0,30	-0,10	0,30
	kompresja	4:1	0,20	0,40	0,50
	zmiana prędkości	-5 %	0,10	0,80	0,30
	zmiana prędkości	30 %	-1,30	-1,60	-1,40



Rysunek 2 Wyniki wartości średniej sygnałów mowy poddanych kompresji zmierzone metodą CCR, gdzie próbkami wzorcowymi były sygnały przetworzone

4. Dyskusja wyników

Dla większości próbek przetworzonych pod względem kompresji o stopniu 2:1, w pierwszym eksperymencie otrzymano jednakowe bądź niższe wartości oceny niż dla stopnia 4:1 lub fragmentu oryginalnego. Wykonane testy statystyczne wykazały, iż wartości oceny otrzymane dla wszystkich nagrań poddanych kompresji o stopniu 2:1 nie są istotne statystycznie w przypadku wszystkich badanych cech wrażeniowych. Oznacza to, iż dany stopień opisywanego rodzaju przetwarzania nie wpływa na zmiany wrażeń słuchowych, a zatem i ocenę jakości badanych próbek dźwiękowych w stosunku do próbek nieprzetworzonych [18]. Analiza statystyczna przy ocenie metodą ACR dla kompresji o stopniu 3:1 wykazała zbliżone bądź niższe wartości oceny, co do ocen fragmentów oryginalnych. Testy statystyczne przeprowadzone w badaniu porównawczym CCR pozwalają na wyciągnięcie jednakowych wniosków jak w przypadku kompresji o niższym współczynniku. Oznacza to, iż nie wpływają one na zmiany wrażeń słuchowych.

Najgłębsza kompresja oceniana była w sposób znacznie odmienny w zależności od nagrania bądź danej cechy wrażeniowej. Wyniki badania metodą CCR wykazały, że słuchacze oceniali fragment oryginalny lepiej w porównaniu do próbek zmodyfikowanych. Analiza statystyczna wykazała, iż wartości są istotne statystycznie dla ogólnej oceny jakości oraz ogólnych wrażeń. Z tego wynika, iż dokonanie kompresji o stopniu 4:1 pogarsza odczucia

wrażeniowe dla wspomnianych cech wrażeniowych. Dla pozostałych - nie występują zmiany wrażeń odsłuchowych.

Nieznaczne spowolnienie prędkości odtwarzania było oceniane różnorodnie w zależności od rodzaju próbki oraz cech wrażeniowych. Nagranie przeprowadzonego dialogu było jednak pozytywnie odbierane w każdej z ocenianych cech wrażeniowych, co okazało się zbieżne z danymi literaturowymi [3,17]. Wszystkie próbki poddano ponownie testom, które wykazały pozytywne odczucia: analiza statystyczna wykazała jednakowe wnioski wyłącznie dla ogólnej oceny jakości.

Ocena nagrań o zwiększonej prędkości odtwarzania, wykonanych metodą ACR, wykazała, iż wraz ze zwiększeniem stopnia przetwarzania, ocena poszczególnych cech wrażeniowych jest niższa. Ponadto wyniki badania komparatywnego CCR charakteryzują się zbieżnymi wnioskami do wartości otrzymanych w pierwszym badaniu. Analiza statystyczna dla przyspieszenia próbek o 100% wykazała, iż we wszystkich ocenianych cechach wrażeniowych zmiana jest istotna statycznie. Oznacza to, iż zwiększenie prędkości odtwarzania nagrań mowy pogarsza estetyczne wrażenie słuchowe, ale nie na tyle, aby były one niezrozumiałe i nieprzydatne w odbiorze.

Biorąc pod uwagę sposób percepcji sygnału mowy uzyskane rezultaty można wykorzystać do ograniczenia objętości informacji [4]: przyspieszając odtwarzany plik o 30% skraca się nagranie także o 30%, co niezależnie od formatu zapisu pozwala na redukcję rozmiaru pliku także o minimum 30%. Nie powoduje to drastycznego pogorszenia zrozumiałości przekazywanych treści [14], co można wykorzystać choćby do odsłuchiwania wykładów lub prelekcji.

Zastosowanie kompresji dynamiki sygnałów, jakkolwiek nie wykazało istotnych zmian w ocenie poszczególnych atrybutów, także pozwala na wprawdzie niewielką redukcję objętości plików, zwłaszcza mowy. Może to być wskazówką do opracowania algorytmów kompresji stratnej danych, które mogłyby być bardziej efektywne niż obecnie istniejące.

5. Podsumowanie

Przeprowadzone badania wykazały, że wraz ze zwiększeniem prędkości odtwarzania nagrań, pogarsza się ocena wrażeniowa słuchacza. Analiza statystyczna wyników wykazała zbieżność wysuwanych wniosków.

Próbki poddane kompresji były oceniane przez słuchaczy w sposób indywidualny. Metoda oceny skalowania absolutnego ACR nie wykazała istotnego wpływu rodzaju przetwarzania na uzyskane wyniki. Dodatkowo, trend zachowania wartości średnich ocen w zależności od stopnia kompresji, otrzymanych w badaniu z zastosowaniem paradygmatu oceny porównawczej CCR, nieznacznie odbiegał od trendu zmian ocen słuchaczy uzyskanych w przypadku metody ACR bez porównania z wzorcem.

Uzyskane wyniki oceny sygnałów mowy pozwalają prognozować, że można dokonywać wydatnej zmiany prędkości odtwarzania plików bez drastycznego pogorszenia jakości i zrozumiałości, co pozwala na oszczędności czasowe. Powyższe wnioski można wykorzystać do redukcji objętości informacyjnej plików zawierających sygnały mowy poprzez skrócenie ich za pomocą przyspieszenia w fazie bezpośrednio po dokonaniu rejestracji sygnałów. Z analizy wynika, że wartościami parametrów, pozwalającymi na maksymalną oszczędność czasową i wielkości plików, były 30 % w przypadku zmiany prędkości nagrań oraz współczynnik kompresji 3:1.

Literatura

[1] Łętowski T., *Słuchowa ocena urządzeń elektroakustycznych*, Centralny ośrodek badawczo-rozwojowy RiTV, 1976.

[2] PN-IEC 268-8 „Urządzenia i systemy elektroakustyczne - Układy automatycznej regulacji wzmocnienia", 1994.

[3] Chu E., Cheng J.-T., Chen C.-P., *Audio Time-Scale Modification with Temporal Compressing Networks,* ACM Multimedia Asia (MMAsia '23), Tainan, Tajwan, 06–08 grudzień 2023.

[4] Kupryjanow A., Czyżewski A., *A non-uniform real-time speech time-scale stretching method,* International Conference on Signal Processing and Multimedia Applications (SIGMAP 2011), Seville, Hiszpania, s. 1-7, 2011.

[5] Brachmański S., *Wybrane zagadnienia oceny jakości transmisji sygnału mowy,* Oficyna Wydawnicza Politechniki Wrocławskiej, 2015.

[6] ITU-T P.800 „Methods of subjective determination of transmission quality", 1996.

[7] EBU Tech. 3276 „Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic", 1998.

[8] Greń J., *Statystyka matematyczna. Modele i zadania*, PWN, 1977.

[9] Moore B. C. J., *Wprowadzenie do psychologii słyszenia*. PWN, 1997.

[10] Boike K. T., Souza P. E., *Effect of Compression Ratio on Speech Recognition and Speech-Quality Ratings With Wide Dynamic Range Compression Amplification*, J. Speech, Language and Hearing Research, 43 no.2, s.456-468, 2000.

[11] PN-EN 54-24 „Systemy sygnalizacji pożarowej – cz. 24: Dźwiękowe systemy ostrzegawcze – Głośniki", 2008.

[12] Butler G., McManus F., *Psychologia - bardzo krótkie wprowadzenie*, Prószyński i S-ka, 1999.

[13] Flanagan J. L., *Speech analysis, Synthesis and Perception*, Springer Verlag, 1965.

[14] Włodarczyk M., Sekalski P., *Evaluation of Time-Scale Modification Methods for Audio Signals on Mobile Devices with Android OS*, Mat. 21st Int. Conference Mixed Design of Integrated Circuits & Systems MIXDES, IEEE, s. 451-454, 2014.

[15] Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M., *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach - część II*. PWN, 1999.

[16] Oh W., *Review of Standard Sound Quality Assessment Methods for the Transmitted and Processed Sounds*, The Journal of the Acoustical Society of Korea, 32 no. 3, s. 214-226, 2013.

[17] Driedger J., Müller M., *A review of time-scale modification of music signals*, Applied Sciences (Switzerland), 6(2):57, 2016.

[18] Ohlmann K., Kollmeier B., Denk F., *Factors Affecting Sound Quality in Acoustically Transparent Hearing Devices*, J. Audio Eng. Soc., 72 no. 1/2, s. 16-32, 2024.

Tomasz **PIWOWARSKI**[1], Bartłomiej **CHOJNACKI**[1]

METODY KONTROLI KIERUNKOWOŚCI DŹWIĘKU ZA POMOCĄ ZWROTNICY CYFROWEJ W ZESTAWIE GŁOŚNIKOWYM

THE METHODS OF SOUND DIRECTIVITY CONTROL WITH THE DIGITAL CROSSOVER IN THE SPEAKER SET

[1]Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, al. Mickiewicza 30, 30-059 Kraków

piwowarski@student.agh.edu.pl

Streszczenie

Kierunkowość zestawu głośnikowego jest parametrem zmiennym i zależnym od wielu czynników. Przyczyną jego złożoności jest obecność różnych przetworników elektroakustycznych rozmieszczonych najczęściej wzdłuż obudowy głośnikowej. Dla pasm częstotliwości, w których sygnał odtwarzany jest przez więcej niż jeden głośnik, istnieje prawdopodobieństwo występowania interferencji fal wygaszających, zwłaszcza poza osią główną głośnika. Na przestrzeni lat, wielu badaczy prowadziło prace nad zmniejszeniem występowania wspomnianych zniekształceń, poprzez wykorzystanie odpowiednich technik projektowania zwrotnic cyfrowych. Jednak różnorodny sposób prowadzenia dotychczas badań uniemożliwia porównanie proponowanych metod.

Celem pracy jest implementacja wybranych rozwiązań, zachowując podczas wykonywania prac jednakowe warunki akustyczne, tę samą aparaturę pomiarową oraz wykorzystując jeden zestaw głośnikowy. W referacie przedstawiono założenia poszczególnych technik wraz z opisem ich realizacji oraz wyniki pomiarów odpowiedzi zestawu z zaimplementowanymi zwrotnicami. Rezultaty badań zostały zestawione z zaprojektowaną zwrotnicą referencyjną i omówione za pomocą zaproponowanych metod porównawczych, w celu doboru najskuteczniejszego rozwiązania.

1. Wprowadzenie

Stosowanie zwrotnicy głośnikowej w zestawach głośnikowych jest kluczowe, z uwagi na występowanie różnych przetworników. Układ ten jednak oddziałuje na cechy wielodrożnych systemów, takie jak przesunięcie fazowe sygnału, czy kierunkowość [1]. Udowodniono, że na drugą z wymienionych charakterystyk wpływ ma różnica w czasach dotarcia fal dźwiękowych z głośników zestawu do punktu odsłuchowego, która jest powiązana z różnicą faz sygnałów zależną od częstotliwości. Omawiane przesunięcie czasowe występujące w głównej mierze w okolicy częstotliwości podziału powodując wygaszanie się fal akustycznych generowanych przez poszczególne przetworniki [2], a jego przyczyną są różne pozycje pojedynczych źródeł dźwięku na obudowie zestawu [3].

Podjęte zostały próby kompensacji omawianego zjawiska mające prowadzić do redukcji różnic czasów dotarcia fal dźwiękowych z pojedynczych głośników do punktu odbioru i niwelacji zniekształceń powstających poza osią główną zestawu. Lipshitz i Vanderkooy zaproponowali montaż przetworników na odgrodzie w sposób „schodkowy" lub wykorzystanie pochyłej płaszczyzny frontowej obudowy [4]. Działania te miały na celu wyrównanie opóźnień czasowych poprzez kontrolę dystansu głośnika od punktu odbioru. Rozwiązania nie spotkały się jednak z pozytywnym odbiorem ze względu na walory estetyczne i ekonomiczne. Wyniki badań przedstawione przez Shaieka ukazały, że najlepsze zachowanie kierunkowości uzyskiwane jest poprzez wykorzystanie głośników koaksjalnych posiadających wspólną oś [5]. Stanowi to jednak znaczące ograniczenie w doborze przetworników dla zestawu. Największą popularność kontroli kierunkowości oraz najszersze spektrum możliwości oferują algorytmy cyfrowe, implementowane za pomocą zwrotnicy aktywnej [6].

W pracy wykonano implementację dwóch wybranych metod kontroli kierunkowości zestawu głośnikowego wykorzystujących filtry cyfrowe. Catalá Iborra zaproponował projekt zwrotnicy głośnikowej oparty na uśrednieniu charakterystyk amplitudowo-częstotliwościowych oraz wyrównaniu faz przetworników względem siebie, w wyznaczonym oknie odsłuchowym [7]. Murray opracował praktykę wyrównywania faz głośników o wspólnych częstotliwościach podziału w zadanych punktach pomiarowych, przy wykorzystaniu filtrów zwrotnicy zaprojektowanej w tradycyjny sposób [8]. Ze względu na niejednorodność wykorzystanej przez autorów aparatury, układów i zestawów, warunków akustycznych prowadzenia badań oraz metodologii przedstawiania wyników, niemożliwe jest porównanie działania omówionych metod. W tym celu ujednolicono metodologię prowadzenia badań poprzez wykorzystanie tego samego, trójdrożnego zestawu głośnikowego oraz

cyfrowego procesora sygnałowego. Za jego pomocą wdrożono opisane przez Catalá Iborrę i Murray'a praktyki wraz ze zwrotnicą zaprojektowaną w sposób tradycyjny, pełniącą funkcję zwrotnicy odniesienia. Charakterystyki kierunkowości zestawu z zaimplementowanymi metodami zbadano poprzez wykonanie pomiarów jego odpowiedzi impulsowych w warunkach bezechowych, na półsferze z zadaną rozdzielczością kątową. Na ich podstawie obliczono średnią arytmetyczną wraz z poziomem odniesienia i rozstępem oraz odchylenie standardowe. Parametry te umożliwiły ocenę skuteczności zrealizowanych praktyk, na podstawie której wyłoniono najskuteczniejszą metodę kontroli kierunkowości zestawu głośnikowego.

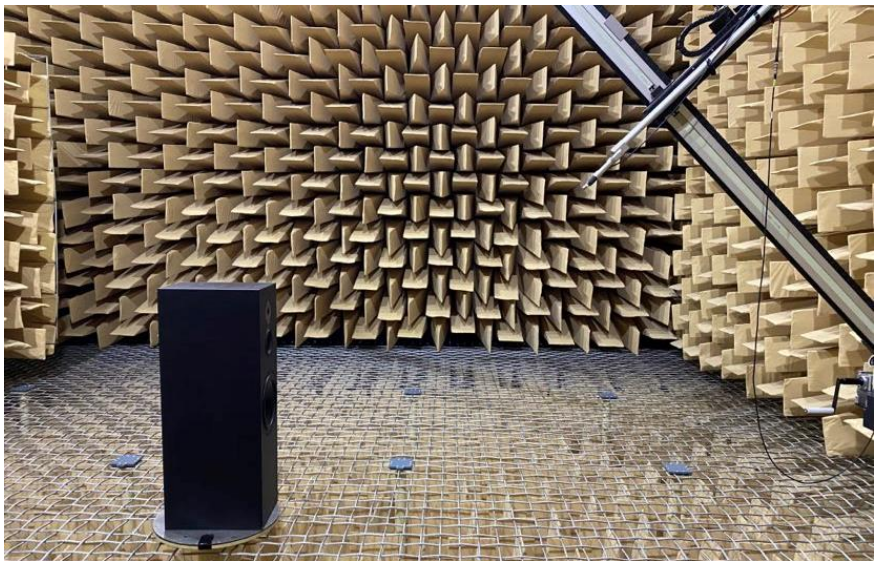
2. Metodologia i implementacja

2.1. Ujednolicenie metodologii badań

W celu umożliwienia porównania badanych metod, wykonano serię działań prowadzącą do ujednolicenia metodologii implementacji algorytmów oraz pomiarów. Wybrane praktyki kontroli kierunkowości realizowane były na trójdrożnym zestawie głośnikowym, za pomocą cyfrowego procesora sygnałowego miniDSP DDRC-24. Filtry wszystkich implementowanych zwrotnic były filtrami Butterwortha o różnych częstotliwościach podziału, zależnych od wykorzystywanej metody, oraz o tych samych rzędach dla poszczególnych rodzajów filtrów:

- filtr dolnoprzepustowy – II rząd,
- filtr pasmowoprzepustowy – III rząd,
- filtr górnoprzepustowy – IV rząd.

Pomiary kierunkowości zestawu głośnikowego wykonywane były w komorze bezechowej (pole swobodne), na półsferze o promieniu 2 m, z rozdzielczością kątową 5°. W tym celu wykorzystano manipulator pozycjonujący mikrofon w elewacji oraz stół obrotowy, na którym umieszczono zestaw głośnikowy (rysunek 1). Pomiary odpowiedzi impulsowych wykonano wykorzystując sygnał sinusoidalny przestrajany logarytmicznie z trzykrotnym uśrednieniem. Aby umożliwić porównanie badanych metod do przypadku referencyjnego, zgodnie z tradycyjnymi metodami zaprojektowana została zwrotnica odniesienia.



Rysunek 1. Zdjęcie wykonane w trakcie pomiarów, przedstawiające manipulator z zainstalowanym mikrofonem oraz zestaw głośnikowy zlokalizowany na stole obrotowym; komora bezechowa Laboratorium Akustyki Technicznej AGH w Krakowie

2.2. Implementacja I metody kontroli kierunkowości zestawu głośnikowego

Opisywana praktyka została zaproponowana przez Catalá Iborrę [7]. W pierwszej kolejności wybrany został zakres okna odsłuchowego, w obrębie którego realizowane będą techniki uśredniania i interpolacji charakterystyki amplitudowo-częstotliwościowej zwrotnicy oraz faz głośników względem siebie. Na podstawie charakterystyk kierunkowości przetworników elektroakustycznych wykorzystanych w zestawie głośnikowym, wybrany został zakres okna (-60°, +60°) w płaszczyźnie horyzontalnej. Ze względu na spadek wpływu głośnika niskotonowego na pasmo przenoszenia zestawu wraz ze wzrostem elewacji, zastosowany został zakres (0, +50°) w płaszczyźnie wertykalnej.

Na podstawie pomiarów odpowiedzi impulsowych przetworników zestawu głośnikowego wykonanych w zadanym oknie z rozdzielczością kątową 5°, w każdym z punktów pomiarowych zaprojektowane zostały filtry zwrotnicy głośnikowej. Łącznie wykonano ich 275. Następnie uśredniono ich charakterystyki amplitudowo-częstotliwościowe za pomocą operacji interpolacji geometrycznej, przedstawionej za pomocą wzoru (1):

$$Geometric\_Interpolation = Value_1 * (\frac{Value_2}{Value_1})^{ratio} \quad (1)$$

gdzie:

Geometric_Interpolation – wynik operacji interpolacji geometrycznej wg [9],

Value1 – pierwsza wartość poddawana operacji,

Value2 – druga wartość poddawana operacji,

ratio – wskaźnik stosunku interpolacji (jeżeli *ratio* = 0.0, wówczas funkcja zwraca *Value1*, jeżeli *ratio*=1.0, funkcja zwraca *Value2*).

W wyniku powyższych działań otrzymano zestaw filtrów o następujących częstotliwościach podziałów:

- filtr dolnoprzepustowy: 644 Hz,
- filtr pasmowoprzepustowy: 632 Hz, 2398 Hz,
- filtr górnoprzepustowy: 2512 Hz.

Kolejnym etapem implementacji metody było wyrównanie faz przetworników elektroakustycznych względem siebie. W tym celu wykonano serie pomiarów odpowiedzi impulsowych zestawu w zadanym oknie odsłuchowym. Z każdą kolejną serią wprowadzano przesunięcia fazowe z krokiem ±10° na przetworniki skrajne (niskotonowy i wysokotonowy). Wykonując operację średniej ważonej, gdzie wagi wyznaczono w oparciu o metodę Tylki i Chiueiri'ego [10], otrzymane zostały uśrednione w oknie odsłuchowym charakterystyki amplitudowo-częstotliwościowe dla każdego wprowadzonego przesunięcia fazy. Do zwrotnicy wprowadzono przesunięcia skutkujące maksymalną sumą uśrednionych amplitud w paśmie częstotliwości podziału:

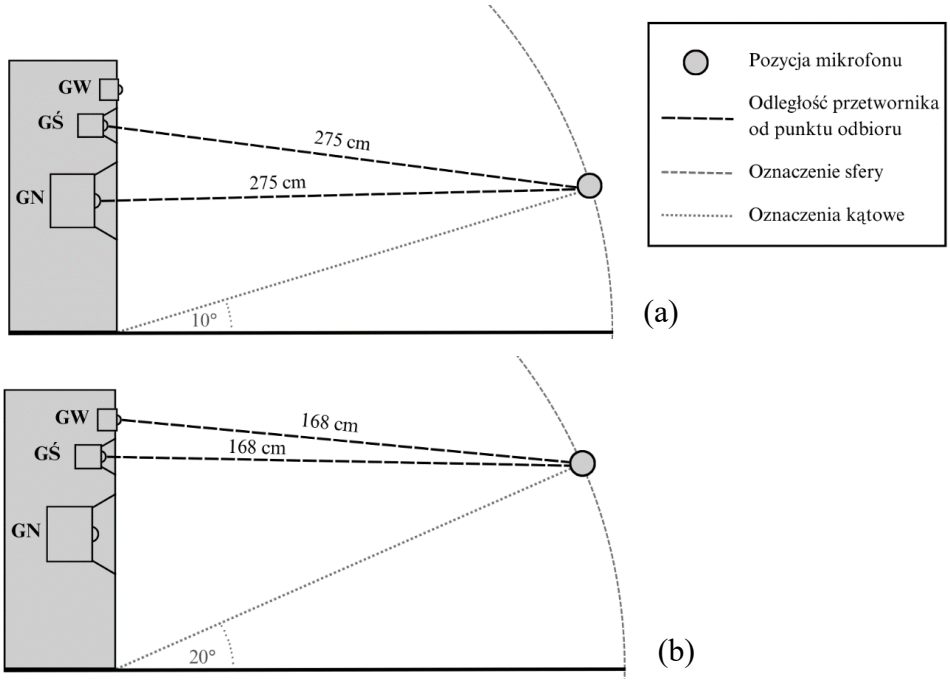
- głośnik niskotonowy: -10° (opóźnienie czasowe równe 1,25 ms),
- głośnik wysokotonowy: -40° (opóźnienie czasowe równe 0,35 ms).

2.3. Implementacja II metody kontroli kierunkowości zestawu głośnikowego

Technika opisana została przez Murray'a [8]. Ze względu na brak opisanej przez autora metodyki projektowania filtrów, podczas realizacji wykorzystywano bank filtrów zwrotnicy referencyjnej.

Pierwszym działaniem było określenie dwóch punktów pomiarowych, o równych odległościach dla przetworników z wspólnym zakresem odtwarzania sygnału w częstotliwościach podziału (tj. niskotonowy i średniotonowy oraz średniotonowy i wysokotonowy). W celu zachowania ciągłości metodologii badawczej, punktu rozmieszczono na sferze pomiarowej (rysunek 2):

- punkt wspólny dla głośnika niskotonowego i średniotonowego: 0° w osi poziomej, 10° w osi pionowej (odległość mikrofonu od głośników: 175 cm);
- punkt wspólny dla głośnika średniotonowego i wysokotonowego: 0° w osi poziomej, 20° w osi pionowej (odległość mikrofonu od głośników: 168 cm).



Rysunek 2. Pozycje mikrofonów podczas pomiarów do omawianej metody; (a) pozycja mikrofonu dla wyrównania faz GN i GŚ; (b) pozycja mikrofonu dla wyrównania faz GŚ i GW; GN – głośnik niskotonowy; GŚ – głośnik średniotonowy; GW – głośnik wysokotonowy

Następną procedurą było wyrównanie faz w wyznaczonych punktach odbioru głośników skrajnych do głośnika średniotonowego, przy finalnym zachowaniu zgodnej polaryzacji wszystkich urządzeń. Aby to osiągnąć odwrócono polaryzację przetwornika niskotonowego i wysokotonowego, a następnie wprowadzano na nie opóźnienia czasowe sygnałów, skutkujące największymi zniekształceniami charakterystyki amplitudowo-częstotliwościowej w pasmach podziału zwrotnicy. Ostatnim krokiem było ponowne odwrócenie polaryzacji omawianych głośników, aby były one zgodne z głośnikiem średniotonowym. W rezultacie sygnał

przetwornika niskotonowego został opóźniony o 0,60 ms, natomiast sygnału przetwornika wysokotonowego nie poddano opóźnieniom.

3. Wyniki

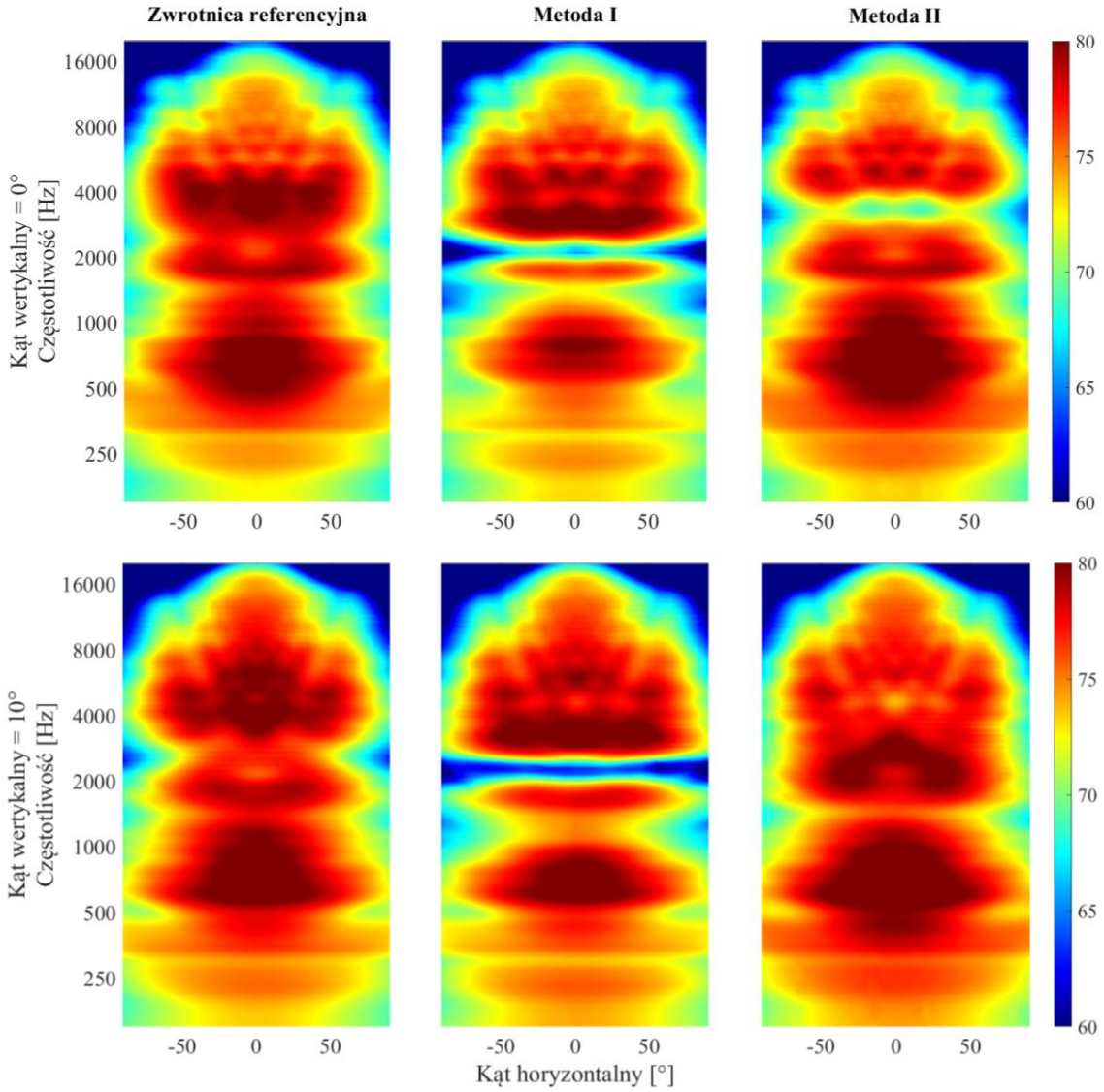
Charakterystyki zaprojektowanych zwrotnic cyfrowych, będącymi rezultatami realizacji metod kontroli kierunkowości dźwięku zestawu głośnikowego wraz z charakterystyką zwrotnicy referencyjnej przedstawione zostały w tabeli 1. Różnice w działaniu poszczególnych zestawów filtrów zobrazowane zostały w sposób graficzny, za pomocą średniej arytmetycznej oraz odchylenia standardowego. Za zbiór danych do operacji oceny skuteczności, przyjęto wyniki pomiarów odpowiedzi impulsowych zestawu w zakresie osi poziomej (-90°, +90°) oraz w zakresie osi pionowej (0°, +90°). W celu minimalizacji wpływu parametrów przetworników oraz ich zakresów odtwarzania na ocenę, analizy wykonywano w zakresie częstotliwości (100 Hz, 8000 Hz).

Tabela 1. Zestawienie charakterystyk zaprojektowanych na podstawie badanych metod zwrotnic cyfrowych wraz z charakterystyką zwrotnicy referencyjnej

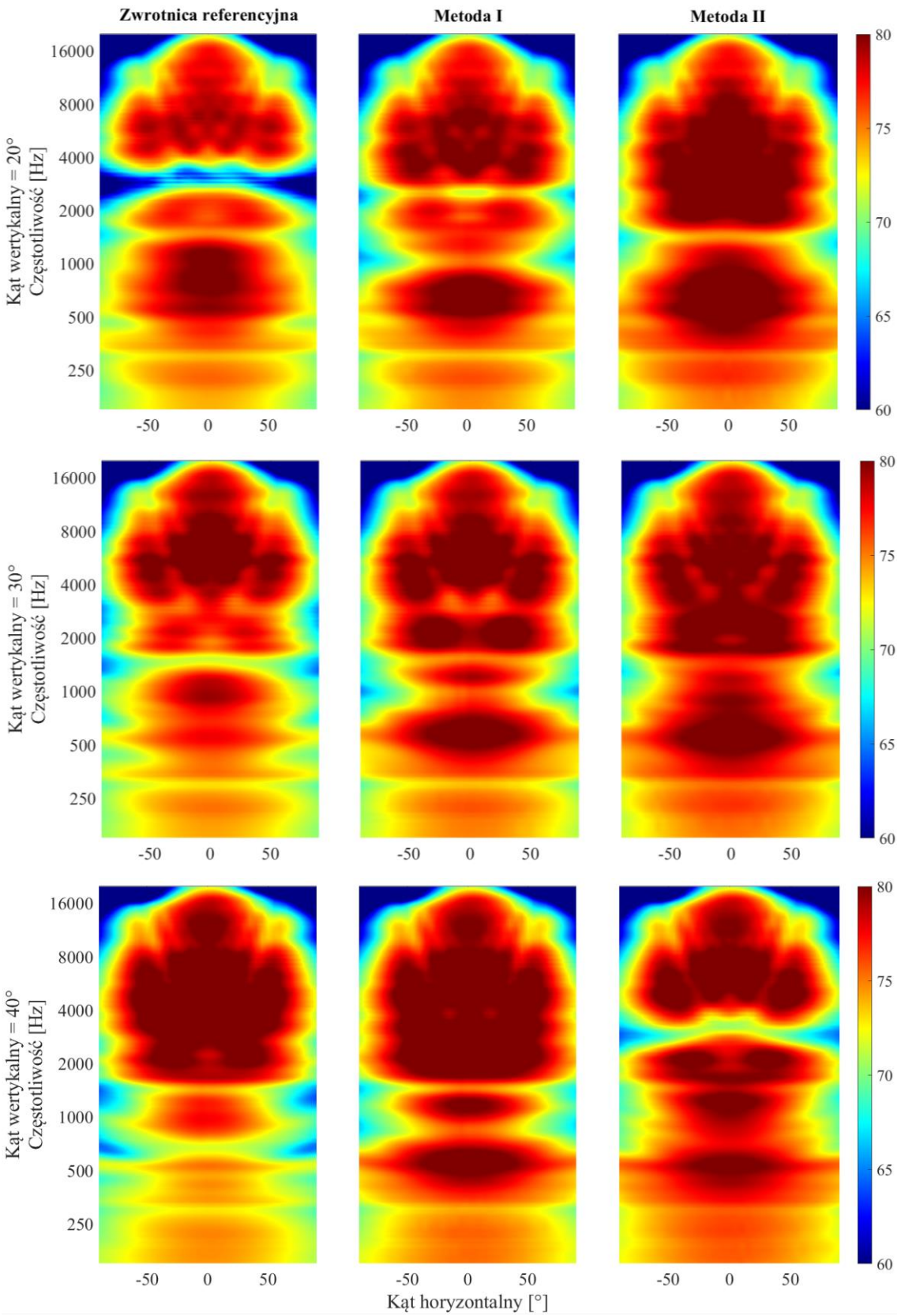
Metoda	**Filtr**	**Rodzaj filtru [-]**	**Rząd [-]**	**Częstotliwość podziału [Hz]**	**Polaryzacja przetwornika [-]**	**Opóźnienie sygnału [ms]**
zwrotnica referencyjna	LPF	Butterwortha	II	650	+	0,00
	BPF	Butterwortha	III/III	600/3110	-	0,00
	HPF	Butterwortha	IV	3390	+	0,00
metoda I	LPF	Butterwortha	II	644	+	1,25
	BPF	Butterwortha	III/III	632/2398	-	0,00
	HPF	Butterwortha	IV	2512	+	0,35
metoda II	LPF	Butterwortha	II	650	+	0,60
	BPF	Butterwortha	III/III	600/3110	+	0,00
	HPF	Butterwortha	IV	3390	+	0,00

3.1. Wykresy zależności poziomu ciśnienia akustycznego od częstotliwości i kąta horyzontalnego

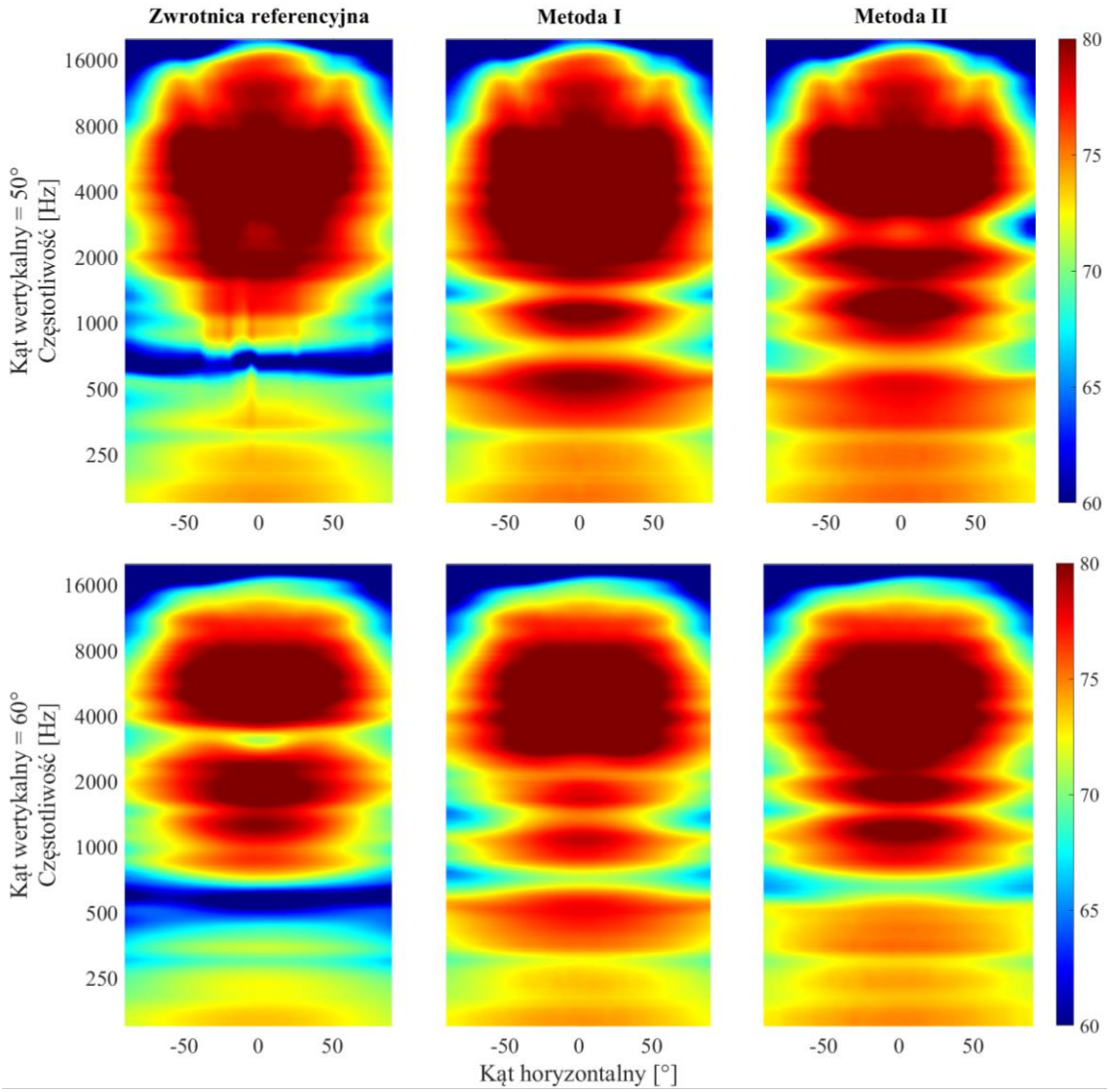
Wyniki pomiarów odpowiedzi impulsowych zestawu głośnikowego na sferze, przedstawione zostały na rysunkach 3-5 w postaci charakterystyk zależności poziomu ciśnienia akustycznego od częstotliwości i wychylenia na osi poziomej. Aby umożliwić ocenę zachowania zestawu dla omawianych wariantów w osi pionowej, zaprezentowane zostały wykresy w zależności od pozycji na osi pionowej, w zakresie (0°, 60°) z krokiem 10°.



Rysunek 3. Charakterystyki zależności znormalizowanego poziomu ciśnienia akustycznego od częstotliwości i kąta horyzontalnego; zakres wertykalny: (0°, 10°), krok: 10°



Rysunek 4. Charakterystyki zależności znormalizowanego poziomu ciśnienia akustycznego od częstotliwości i kąta horyzontalnego; zakres wertykalny: (20°, 40°), krok: 10°



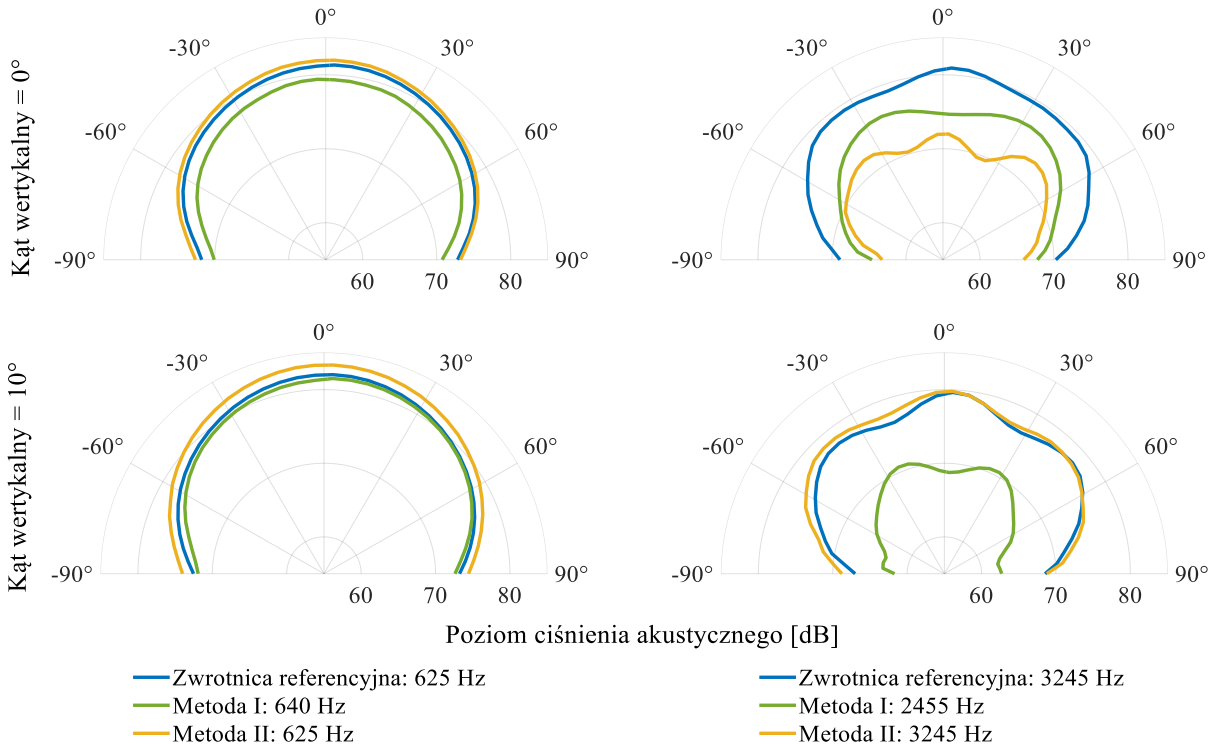
Rysunek 5. Charakterystyki zależności znormalizowanego poziomu ciśnienia akustycznego od częstotliwości i kąta horyzontalnego; zakres wertykalny: (50°, 60°), krok: 10°

Zjawisko zniekształcenia kierunkowości zestawu w częstotliwościach podziału zwrotnicy zauważalne jest zarówno ze wzrostem kąta horyzontalnego jak i ze wzrostem kąta wertykalnego. Występują one dla wszystkich badanych wariantów, jednakże dla praktyki wyrównywania fazy w punkcie (metoda II) są najmniejsze. Ze wzrostem elewacji, na osi nie występują wówczas wartości poziomu ciśnienia akustycznego mniejsze, niż 70 dB oraz szerokość charakterystyki względem kąta horyzontalnego jest najbardziej równomierna dla wszystkich elewacji. Metoda wyrównania poziomu i fazy w oknie (metoda I) pokazała istotną interferencją wygaszającą dla górnej częstotliwości podziału w elewacji 10°, która objawia się spadkiem poziomu ciśnienia akustycznego do 60 dB. Również równomierność poziomu ciśnienia akustycznego względem kąta horyzontalnego jest mniejsza, niż względem

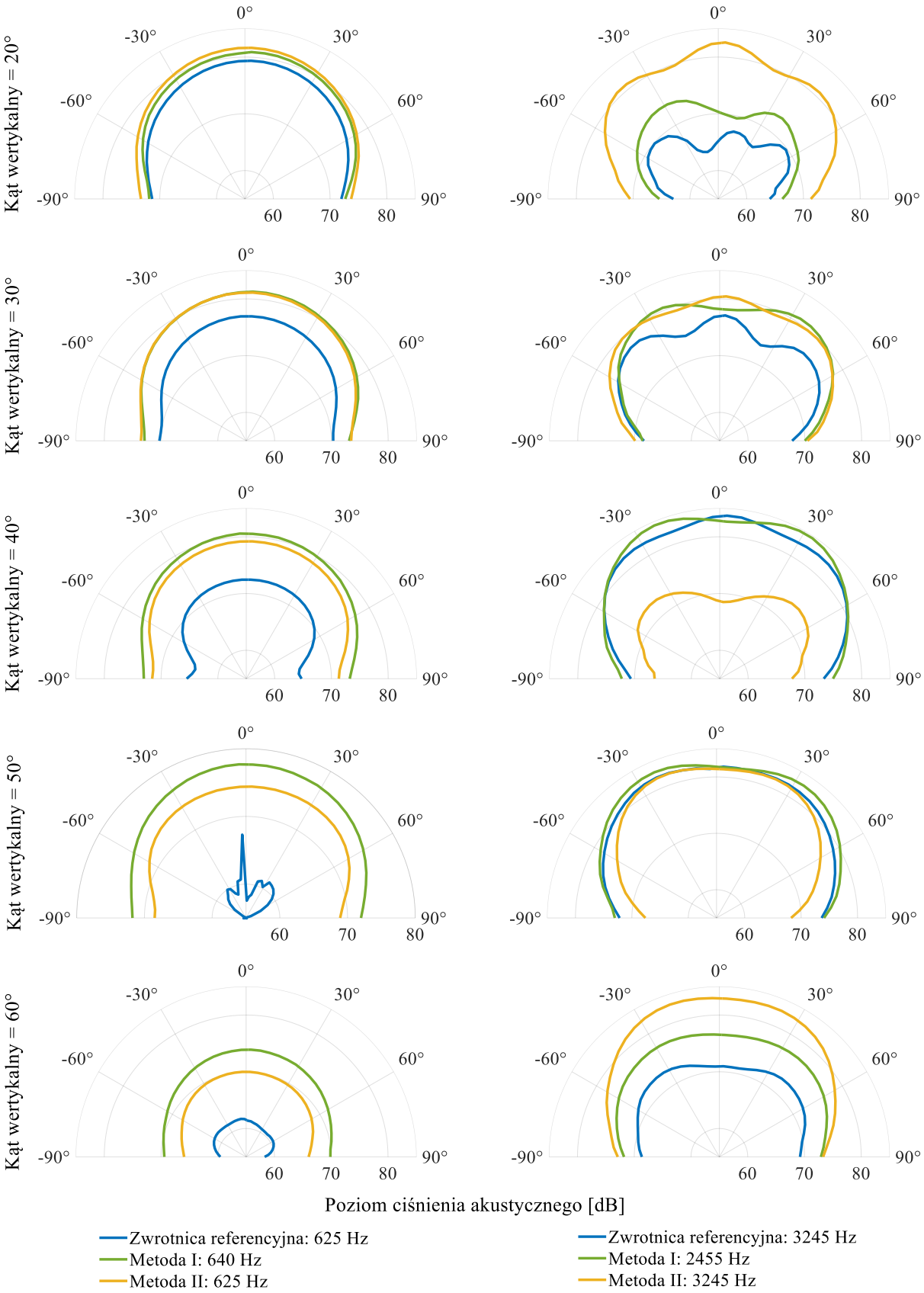
poprzedniej omawianej metody. Zwrotnica referencyjna posiada największą liczbę interferencji wygaszających.

3.2. Wykresy charakterystyki kierunkowości

Charakterystyki kierunkowości zestawu głośnikowego z zaimplementowanymi metodami jej kształtowania w zestawieniu ze zwrotnicą referencyjną, przedstawione zostały na rysunkach 6-7. Do porównania wybrano ponownie wyniki z zakresu osi pionowej (0°, +60°), z krokiem 10°. Wykresy ukazują charakterystyki dla uśrednionych częstotliwości podziału zwrotnic, wykorzystywanych w omawianych praktykach. Zabieg ten umożliwia analizę kierunkowości zestawu w najistotniejszych pod kątem zniekształceń częstotliwościach.



Rysunek 6. Charakterystyki kierunkowości zestawu głośnikowego dla uśrednionych częstotliwości podziału; zakres kąta wertykalnego: (0°, 10°), krok: 10°



Rysunek 7. Charakterystyki kierunkowości zestawu głośnikowego dla uśrednionych częstotliwości podziału; zakres kąta wertykalnego: (20°, 60°), krok: 10°

Powyższe wykresy dla dolnej częstotliwości podziału wskazują, że charakterystyka kierunkowa dla obu metod kontroli kierunkowości dźwięku, zmienia się nieznacznie względem elewacji w zakresie (0°, 40°). Dla kolejnych kątów wertykalnych poziom ciśnienia akustycznego zaczyna spadać przy zachowaniu charakterystyki zbliżonej do wszechkierunkowości. Większy spadek poziomu widoczny jest dla wyrównania fazy w punkcie (metoda II). W przypadku zwrotnicy referencyjnej jednak widoczne są zniekształcenia oraz istotnie różnie od charakterystyki wszechkierunkowej, od kąta wertykalnego 40° wzwyż. Największe anomalie zachodzą w przypadku elewacji równej 50°.

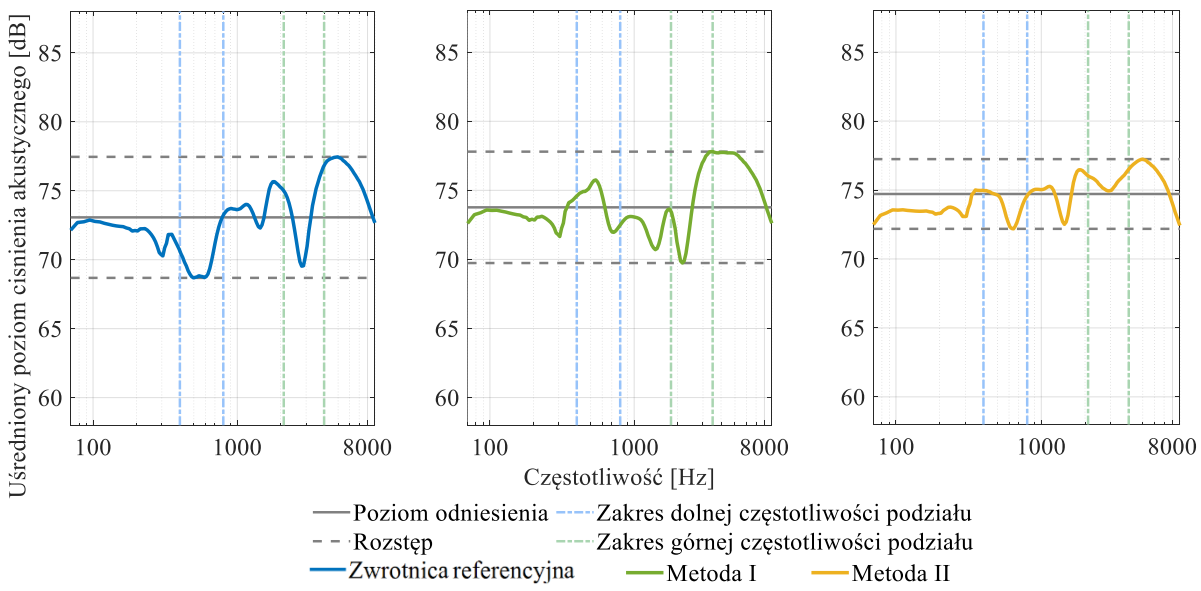
Górna częstotliwość podziału zwrotnicy ukazuje większe różnice pomiędzy badanymi wariantami. Najbardziej równomierna charakterystyka kierunkowości została uzyskana w przypadku wyrównania fazy w punkcie (metoda II). Występuje dla niej spadek poziomu ciśnienia akustycznego przy zachowaniu kształtu w kącie wertykalnym 0°. W elewacji 40° zauważalne jest zniekształcenie charakterystyki, natomiast inne elewacje wskazują na charakterystykę zbliżoną do wszechkierunkowości, przy zachowaniu wysokiej skuteczności. Zwrotnica oparta o metodę wyrównania poziomu i fazy w oknie (metoda I) powoduje widoczne zniekształcenia w kątach wertykalnych z zakresu (0°, 20°). Kolejne elewacje przedstawiają równomierny rozkład poziomu ciśnienia akustycznego w zależności od kąta horyzontalnego. Zwrotnica referencyjna powoduje widoczne zniekształcenia kształtu charakterystyki w kącie horyzontalnym 20° oraz widoczny spadek amplitudy dla elewacji 60°.

3.3. Średnia arytmetyczna, poziom odniesienia i rozstęp

Wykonanie obliczeń średniej arytmetycznej poziomów ciśnienia akustycznego w zadanych zakresach umożliwiło zróżnicowanie wybranych metod kontroli kierunkowości dźwięku oraz ocenę ich skuteczności. Wspomniany sposób uśrednienia w celu analizy charakterystyk amplitudowo-częstotliwościowych miał już miejsce w badaniach innych autorów [11]. Nie zdecydowano się na wykorzystanie średniej geometrycznej, ponieważ postać danych nie ma charakteru zespolonego, a co zatem idzie analizie podlega jedynie odpowiedź amplitudowa, bez odpowiedzi fazowej. Przypadki użycia operacji geometrycznych opisane zostały przez Sierę [12].

W pierwszej kolejności został obliczony omawiany parametr poprzez uśrednienie wszystkich pomiarów wykonanych we wspomnianym zakresie. Wyniki zależne są od częstotliwości. Następnie w celu wyznaczenia rozstępu charakterystyk w stosunku do

liniowego poziomu odniesienia, obliczone zostały wartości maksymalne oraz minimalne z uzyskanych średnich arytmetycznych. Za poziom odniesienia przyjęto wartość środkową rozstępu. Wyniki obliczeń dla zwrotnicy referencyjnej i badanych metod przedstawione zostały na rysunku 8 oraz w tabeli 2.



Rysunek 8. Średnia arytmetyczna charakterystyk amplitudowo-częstotliwościowych zestawu głośnikowego, badana w zakresie kąta horyzontalnego (-90°, +90°) oraz w zakresie kąta wertykalnego (0°, +90°), w odniesieniu do skuteczności średniej i tolerancji w pasmach (100 Hz, 8000 Hz)

Tabela 2. Poziom odniesienia oraz tolerancja badanych metod, określone na podstawie średniej arytmetycznej charakterystyk amplitudowo-częstotliwościowych w pasmach (100 Hz, 8000 Hz)

Metoda	**zwrotnica referencyjna**	**metoda I**	**metoda II**
Poziom odniesienia [dB]	73,1	73,8	74,7
Rozstęp [dB]	±4,4	±4,0	±2,5

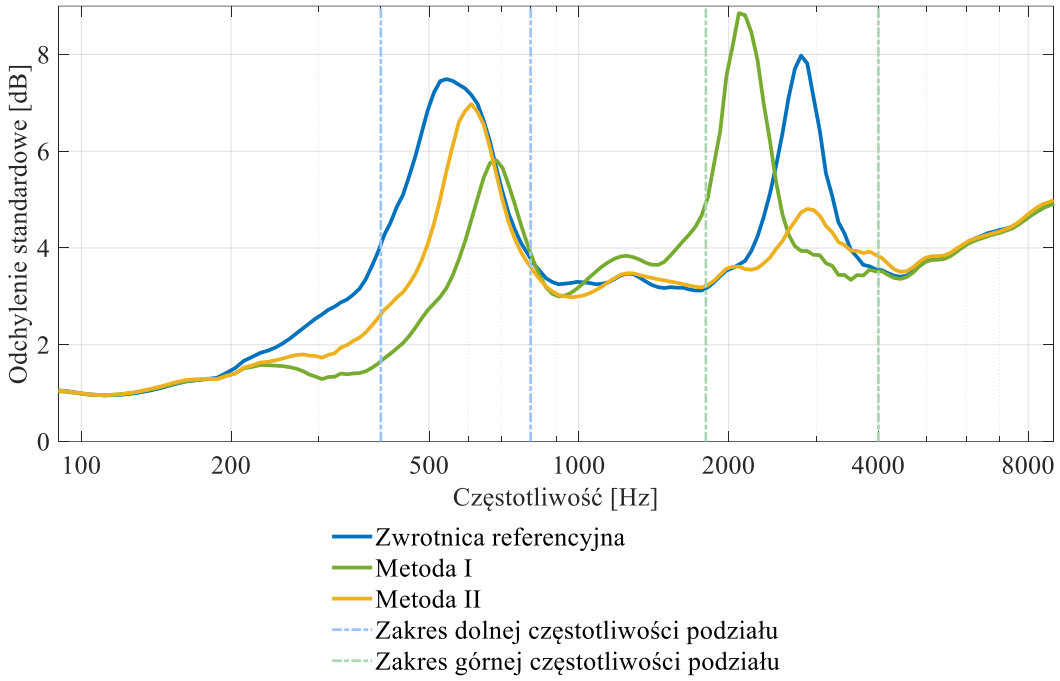
Analiza powyższych wyników wskazuje, że średnia o najmniejszych zniekształceniach i kształtem najbardziej zbliżonym do funkcji liniowej uzyskana została dla metody II. Rozstęp względem punktu odniesienia dla omawianej metody jest również najmniejsza co oznacza, że ze wszystkich analizowanych praktyk, jej charakterystyka kierunkowości jest najbardziej zbliżona do wszechkierunkowości. Różnica rozstępu pomiędzy metodą I a zwrotnicą referencyjną jest równa ±0,4 dB.

Na podstawie powyższej analizy stwierdza się, że najskuteczniejszą praktyką kontroli kierunkowości zestawu głośnikowego jest metoda II. Badane parametry wskazują, że metoda I

charakteryzuje się większą skutecznością niż zwrotnica referencyjna, jednakże różnice nie są istotne. Oznacza to jednakże, że wybrane metody kształtowania kierunkowości zestawu przyniosły zakładane efekty i zmniejszyły jego zniekształcenia poza osią główną zestawu.

3.4. Odchylenie standardowe

Odchylenie standardowe wykorzystywane było przez innych autorów w celu oceny kierunkowości dźwięku [13]. Im wyższe parametr przyjmuje wartości, tym większa jest kierunkowość, natomiast wraz ze spadkiem odchylenia standardowego, zwiększa się wszechkierunkowość. W badanym przypadku, mniejsze wartości omawianego parametru wskazują na większą skuteczność metody, natomiast sam parametr zależny jest od częstotliwości. Wyniki obliczeń dla każdej praktyki kontroli kierunkowości zestawu głośnikowego przedstawione zostały na rysunku 9 oraz w tabeli 3.



Rysunek 9. Odchylenie standardowe obliczone z charakterystyk amplitudowo-częstotliwościowych zestawu głośnikowego, badana w zakresie kąta horyzontalnego (-90°, +90°) oraz w zakresie kąta wertykalnego (0°, +90°)

Tabela 3. Wartości odchylenia standardowego obliczone z charakterystyk amplitudowo-częstotliwościowych zestawu głośnikowego, badana w zakresie kąta horyzontalnego (-90°, +90°) oraz w zakresie kąta wertykalnego (0°, +90°), podane w pasmach tercjowych

Metoda	zwrotnica referencyjna	metoda I	metoda II
Średnie odchylenie standardowe [dB]	3,86	3,44	3,38

Powyższy wykres wskazuje, że w zakresie dolnej częstotliwości podziału, największą skuteczność osiągnięto poprzez implementację zwrotnicy opartej o metodę I. Obie metody kontroli kierunkowości zestawu głośnikowego okazały się skuteczniejsze od wykorzystania zwrotnicy referencyjnej. Analiza parametru dla górnej częstotliwości podziału wskazuje, że najmniejsze odchylenie standardowe uzyskane zostało dla metody II. Najmniej skuteczną metodą w omawianym paśmie okazała się metoda dominująca dla dolnej częstotliwości podziału, czyli metoda I.

Uśrednione wartości odchylenia standardowego wskazują, że najskuteczniejszą metodą jest metoda II, dla której uśredniona wartość parametru wynosi 3,38 dB. Metoda I również okazała się skuteczniejsza od zwrotnicy odniesienia, uzyskując wartość 3,44 dB. Uśredniona wartość odchylenia standardowego dla zwrotnicy referencyjnej wynosi 3,86 dB. Oznacza to, że badane metody kontroli kierunkowości zestawu głośnikowego zwiększają jego wszechkierunkowość.

4. Podsumowanie

Podczas prowadzenia badań wykonano skuteczną implementację wybranych metod kontroli kierunkowości zestawu głośnikowego z wykorzystaniem zwrotnicy cyfrowej, w jednorodnych warunkach akustycznych. Wykorzystanie tej samej aparatury umożliwiło dalszą analizę, której celem było wyłonienie praktyki najskuteczniejszej.

Na podstawie przeprowadzonych analiz stwierdzono, że obie badane metody kształtowania kierunkowości zestawu zmniejszają zniekształcenia charakterystyki amplitudowo-częstotliwościowej poza osią główną zestawu, względem zwrotnicy odniesienia. Dowiedziono również, że wspomniane zadanie najskuteczniej realizuje technika zaproponowana przez Murray'a [8], czyli metoda II. Charakteryzuje się ona najmniejszym rozstępem względem poziomu odniesienia w analizie średniej arytmetycznej oraz najmniejszą

wartością średniego odchylenia standardowego. Dla górnej częstotliwości podziału przyjmuje ona wartości mniejsze od zwrotnicy referencyjnej o ponad 3 dB, natomiast od metody I mniejsze o ok. 4 dB. W okolicy dolnej częstotliwości (200 – 800 Hz) największą skutecznością na podstawie analizy odchylenia standardowego charakteryzuje się metoda I. Uzyskane wyniki wskazują, że technika oparta o wyrównanie faz przetworników przy zachowaniu zgodnej polaryzacji najskuteczniej minimalizuje zniekształcenia charakterystyki amplitudowo-częstotliwościowej poza osią zestawu.

Dodatkowo na podstawie otrzymanych wyników poziomu odniesienia założyć można, że wykorzystywanie metod kontroli kierunkowości zestawu głośnikowego zwiększa jego skuteczność w oknie odsłuchowym.

Literatura

[1] S. Cecchi, V. Bruschi, S. Nobili, A. Terenzi, and V. Välimäki, "Crossover Networks: A Review," *Journal of the Audio Engineering Society*, vol. 71, pp. 526–551, Jan. 2023, doi: 10.17743/jaes.2022.0100.

[2] D. G. Fink, "Time Offset and Crossover Design," *J. Audio Eng. Soc*, vol. 28, no. 9, pp. 601–611, 1980, [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=3963

[3] H. Shaiek and J.-M. Boucher, "Optimizing the Directivity of Multiway Loudspeaker Systems," *EURASIP J Audio Speech Music Process*, vol. 2010, Jan. 2010, doi: 10.1155/2010/928439.

[4] S. P. Lipshitz and J. Vanderkooy, "In-Phase Crossover Network Design," *J. Audio Eng. Soc*, vol. 34, no. 11, pp. 889–894, 1986, [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=5237

[5] H. Shaiek, "Optimizing wide band coaxial loudspeaker systems using digitalsignal processing techniques," Ph.D. dissertation, TELECOM, Bretagne, France, 2007.

[6] A. Rimell and M. O. Hawksford, "Reduction of Loudspeaker Polar Response Aberrations Using Psychoacoustically Motivated Adaptive Sub-Band Algorithms," 1994.

[7] V. M. C. Iborra, "Crossover design based on median level and phase correction within a listening window."

[8] J. A. Murray, "Microalignment of Driversvia Digital Technology*."

[9] D. W. Gunness, "Loudspeaker Transfer Function Averaging and Interpolation." [Online]. Available: www.aes.org.

[10] J. G. Tylka and E. Y. Choueiri, "On the Calculation of Full and Partial Directivity Indices." [Online]. Available: http://www.princeton.edu/3D3A/Directivity.html

[11] V. M. C. Iborra and F. F. Li, “DSP loudspeaker 3D complex correction.” [Online]. Available: http://www.aes.org/e-lib.

[12] J. Sierra, J. Kamrava, P. Espinosa, J. M. Arneson, and P. Kohut, “Statistical and Analytical Approach to System Alignment,” in *Audio Engineering Society Convention 145*, Oct. 2018. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=19783

[13] B. Chojnacki, “Numerical Directivity Simulations Of Speaker Arrays For Omnidirectional Sound Source Quality Assessment,” *Vibrations in Physical Systems*, vol. 33, no. 1, 2022, doi: 10.21008/j.0860-6897.2022.1.01.

Magdalena **PUCHALSKA**[1], Andrzej **WICHER**[1]

ANALIZA DZIAŁANIA UKŁADU SŁUCHOWEGO Z WYKORZYSTANIEM METOD OBIEKTYWNYCH U OSÓB WE WCZESNEJ FAZIE CHOROBY ALZHEIMERA

ANALYSIS OF THE FUNCTIONING OF THE AUDITORY SYSTEM USING OBJECTIVE METHODS IN PEOPLE IN THE EARLY PHASE OF ALZHEIMER'S DISEASE

[1] Katedra Akustyki, Wydział Fizyki, Uniwersytet im. Adama Mickiewicza w Poznaniu
ul. Uniwersytetu Poznańskiego 2, 61-614 Poznań

Adres e-mail autora korespondencyjnego: magda@puchalski.it

Streszczenie

Niniejsza praca opisuje badania słuchu wykonane osobom we wczesnej fazie lub z podejrzeniem choroby Alzheimera (AD). Przed objawami charakterystycznymi tej choroby występują tzw. objawy prodromalne, do których mogą należeć zaburzenia funkcji słuchowych. Referat analizuje wstępne wyniki badań słuchu u osób we wczesnej fazie lub z podejrzeniem AD w celu znalezienia potencjalnie wartościowych wyników, które mogłyby wskazywać na użyteczność testów w badaniach przesiewowych.

Pacjentom wykonano obiektywne badania słuchu (otoemisja produktu zniekształceń nieliniowych, badanie potencjałów wywołanych z pnia mózgu oraz badanie potencjału P300), a analiza tych wyników miała na celu weryfikację hipotezy o zmianach wartości mierzonych wielkości w tej grupie badanych względem osób zdrowych. W tym kontekście założono, że szczególnie miarodajne może być badanie potencjału P300, który opiera się na analizie morfologii fal mózgowych o długim czasie utajenia, czyli pochodzących z najwyższych pięter centralnego układu nerwowego, CNS (*Central Nervous System*), włącznie z procesem kognitywnym.

Zaobserwowano związek interlatencji w badaniu ABR (*Auditory Brainstem Response*) ze stopniem otępienia. Prócz tego w wielu przypadkach w uzyskanych wynikach badań występują odstępstwa od wartości normatywnych, które mogą sugerować niedobory w CNS. Mimo, że na obecnym etapie badań nie można jednoznacznie stwierdzić jaka jest przyczyna tych odchyleń od normy, jednak otrzymane wyniki sugerują, ze badania w tym zakresie powinny być kontynuowane.

1. Wprowadzenie

Narząd słuchu to zespół struktur w organizmie człowieka umożliwiający przetwarzanie fali akustycznej na wrażenie słuchowe. W jego skład wchodzi część obwodowa oraz część ośrodkowa. Część obwodowa zawiera ucho zewnętrzne, środkowe i wewnętrzne oraz nerw słuchowy. Nerw słuchowy łączy obie części, dostarczając informacje z części obwodowej do części ośrodkowej, centralnego układu nerwowego.

Centralny układ nerwowy jest niejako podzielony na obszary odpowiedzialne za poszczególne funkcje. Przykładem mogą być dwa ośrodki odpowiedzialne za rozumienie mowy - ośrodek Broki oraz ośrodek Wernickego. Działanie tych układów w korze mózgowej jest możliwe dzięki bardzo zróżnicowanym funkcjom neuronów. Są one niejako podzielone pod względem spełniających funkcji, toteż każde uszkodzenie powoduje upośledzenie konkretnych umiejętności, np. w wyniku choroby neurodegeneracyjnej. Przykładowo - niektóre z neuronów w korze słuchowej człowieka reagują tylko na wzrost częstotliwości, inne na zmniejszanie się częstotliwości, jeszcze inne tylko na wzrost częstotliwości sygnałów wysokoczęstotliwościowych, i zmniejszanie się częstotliwości sygnałów niskoczęstotliwościowych itd. (Ozimek, 2018). Udowodniono także, że w zależności od częstotliwości, uaktywniają się inne miejsca kory słuchowej mózgu - cała droga słuchowa zachowuje więc tonotopową strukturę (Humphries i in., 2010, Pruszewicz i Obrębowski, 2010).

Istnieje wiele schorzeń objawiających się neurodegeneracją, czyli utratą komórek nerwowych. Należą do nich np. choroba Parkinsona, stwardnienie zanikowe boczne, stwardnienie rozsiane, pląsawica Huntingtona, czy choroba Alzheimera.

1.1. Choroba Alzheimera

Choroba Alzheimera (AD - *Alzheimer's Disease*) to nieodwracalne, postępujące zaburzenie neurologiczne charakteryzujące się degeneracją komórek mózgowych. AD jest najczęstszą przyczyną otępienia. Zgodnie z definicją prezentowaną przez WHO - *World Health Organization* (ICD-10) otępienie to zespół objawów spowodowany przewlekłą lub postępującą chorobą mózgu, charakteryzujący się licznymi upośledzeniami funkcji poznawczych (pamięć, myślenie, orientacja, rozumienie, liczenie, język, uczenie się,

ocenianie). W zależności od źródła, choroba Alzheimera stanowi przyczynę otępienia u 60-70% z 50 milionów ludzi, u których je zdiagnozowano (WHO, 2020).

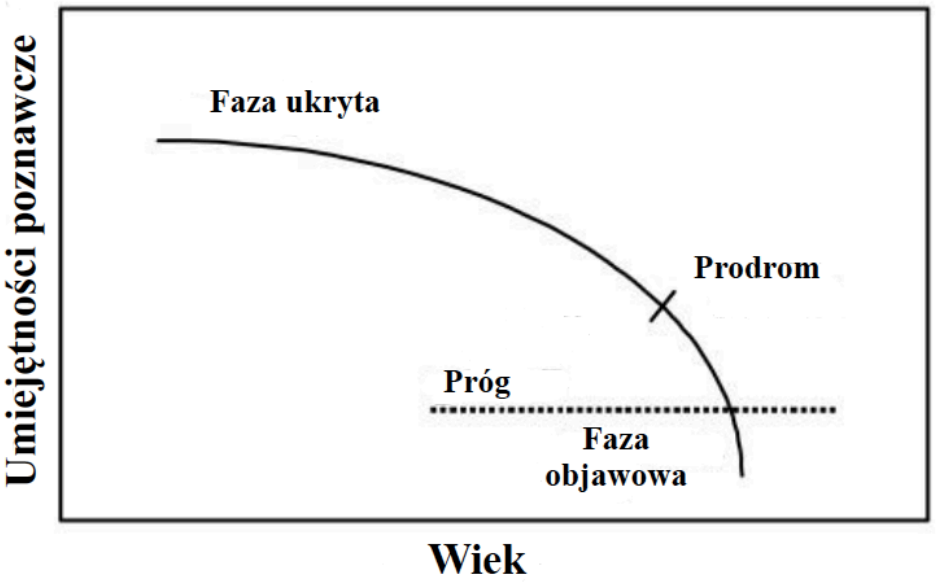
Przebieg choroby sprawia, że najbardziej zauważalne objawy występują przeważnie u osób powyżej 65. roku życia (Cichacz-Kwiatkowska i in., 2016). Istnieją jednak przesłanki, że najczęściej choroba rozwija się znacznie wcześniej, niż zostanie wykryta. Pierwsze objawy pojawiają się kiedy znacząca liczba neuronów ulegnie uszkodzeniu lub uszkodzenie dotyczy określonej części ośrodkowego układu nerwowego. W przypadku chorób neurodegeneracyjnych procedura wprowadzenia nowego leku na rynek jest dłuższa, droższa i bardziej ryzykowana niż w przypadku innych leków (Kaitin i Milne, 2011). AD jest chorobą nieuleczalną, a podejmowana terapia ma na celu złagodzenie objawów oraz zahamowanie i spowolnienie postępujących zmian. Z tych powodów kluczowe jest wczesne rozpoznanie choroby i wdrożenie odpowiedniego leczenia.

W związku z hipotezami dostępnymi na temat procesu powstawania AD, poza wywiadem psychiatryczno-neurologicznym, istnieje kilka sposobów na obiektywne wykrycie w organizmie chorego substancji wywołujących neurodegenerację. Markery biologiczne (w skrócie **biomarkery**) to mierzalne wskaźniki stanów normalnych lub patogennych, takich jak obecność lub nasilenie choroby (Strimbu & Tavel, 2010). W AD możliwym do wykrycia biomarkerem jest białko tau oraz β-amyloid. Są one możliwe do wykrycia na dwa sposoby - za pomocą pobrania i analizy płynu mózgowo-rdzeniowego lub pozytonowej tomografii emisyjnej (PET). Oba te sposoby są albo inwazyjne, albo kosztowne i trudno dostępne.

Istnieje pilna potrzeba prostych, niedrogich, nieinwazyjnych i łatwo dostępnych narzędzi diagnostycznych dla choroby Alzheimera. Nowe technologie testowania mogą wspierać rozwój leków na wiele sposobów. Istnieją także badania próbujące powiązać badania krwi z wczesnym wykryciem choroby Alzheimera (Alzheimer's Association, 2020). Krok ten byłby przełomowy, ze względu na dostępność i małą cenę takiego rozwiązania, jednak rozwiązanie to nie weszło jeszcze do powszechnego użytku.

1.1.2. Faza prodromalna choroby

Objawy prodromalne to wczesne, nieswoiste objawy, występujące przed pojawieniem się charakterystycznych objawów choroby (Rysunek 1.1.) W chorobach neurodegeneracyjnych takich jak AD, lub choroba Parkinsona, faza prodromalna (przedkliniczna) może trwać przez lata lub dekady. Objawy obejmują bardzo lekkie zmiany w zachowaniu czy sposobie mówienia (Marjorie i in., 1985, Pistono i in., 2020). Zmiany te są możliwe do zauważenia przez osoby bliskie i możliwe do wykrycia przez wyspecjalizowanego neurologa lub psychiatrę, posiadającego odpowiednie narzędzia do badań.



Rysunek. 1.12. Model choroby Alzheimera. (na podstawie: Welsh-Bohmer, 2008, tłumaczenie własne).

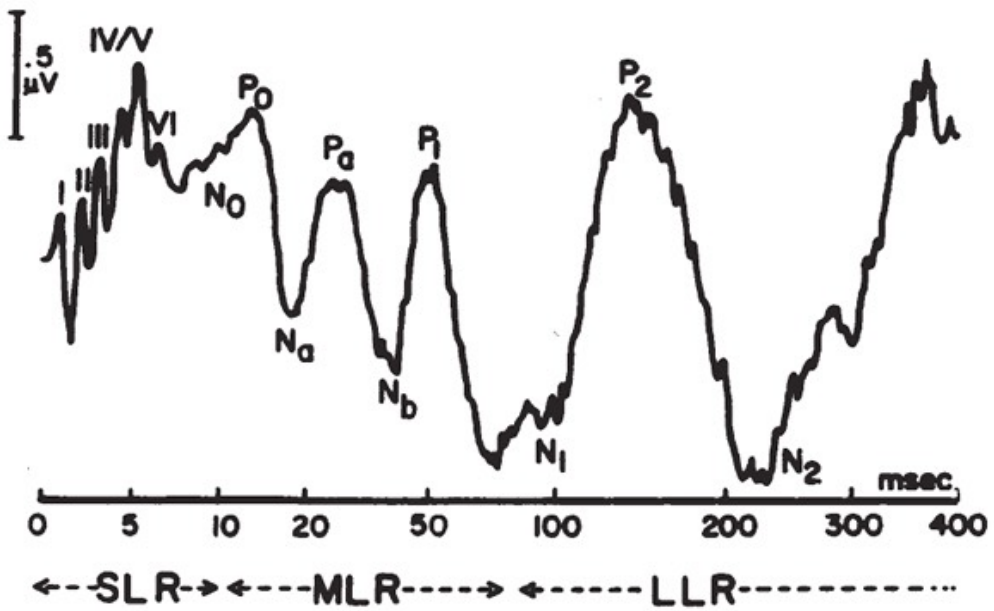
1.2. Badania słuchowe

W tej pracy skupiono się na związku odbiegających od normy badań słuchu z rozwojem choroby Alzheimera w fazie prodromalnej. Aby wykluczyć sytuacje, w których mylnie zinterpretowano niezwiązane z neurodegeneracją uszkodzenie słuchu jako nieswoisty objaw AD, osobom badanym wykonano rozszerzoną diagnostykę narządu słuchu. Należało wykluczyć uszkodzenia przewodzeniowe, uszkodzenia ucha środkowego (takie jak np. otoskleroza), czy nieprawidłowe funkcjonowanie ślimaka. W ramach badań diagnostycznych

słuchu wykonano więc wywiad audiologiczny, audiometrię tonalną, tympanometrię oraz pomiar otoemisji akustycznych.

1.2.1. Słuchowe potencjały wywołane pnia mózgu

Mianem ABR (*Auditory Brainstem Response*) określa się obiektywne badanie elektrofizjologiczne polegające na rejestracji **potencjałów słuchowych odbieranych z pnia mózgu**. Odpowiedzi pnia mózgu reprezentują aktywność elektrofizjologiczną na wyższych piętrach drogi słuchowej. Wywołuje się je za pomocą bardzo krótkich impulsów tonalnych lub trzasków, których czasy trwania są rzędu mikrosekund, a są one powtarzane ok. 30-40 razy na sekundę. Wywołują one wrażenie trzasku o maksimum energii przypadającym na 2-4 kHz i są podawane na słuchawki. Dzięki temu badaniu z elektroencefalogramu można odczytać informację o progu słyszenia badanego. Obrazują ją krzywe uzyskiwane dla kolejnych poziomów sygnału, w szczególności wyróżniany na tych krzywych fragment zapisu ABR z charakterystycznym załamkiem zwany **falą V** (Rysunek 1.2.). Badanie daje najbardziej wiarygodne wyniki, jeśli jest wykonywane podczas snu pacjenta. ABR bada potencjały o **krótkim** czasie utajenia. W tym badaniu uwaga pacjenta nie ma znaczenia, jednak, dla jak najmniejszych zakłóceń, powinien on nieruchomo leżeć lub spać, gdyż skurcze mięśni mogą powodować zakłócenia. Na Rysunku 1.2. przedstawiono podział na potencjały o krótkiej, średniej i długiej latencji.



Rysunek 1.2. Słuchowe potencjały wywołane na logarytmicznej osi czasu z powszechnym podziałem na potencjały o krótkiej (SLR - *Short Latency Response*), średniej (MLR - *Medium Latency Response*) i długiej (LLR - *Long Latency Response*) latencji. W lewej części rysunku widoczne są także fale ABR od I do V (ASHA, 1987).

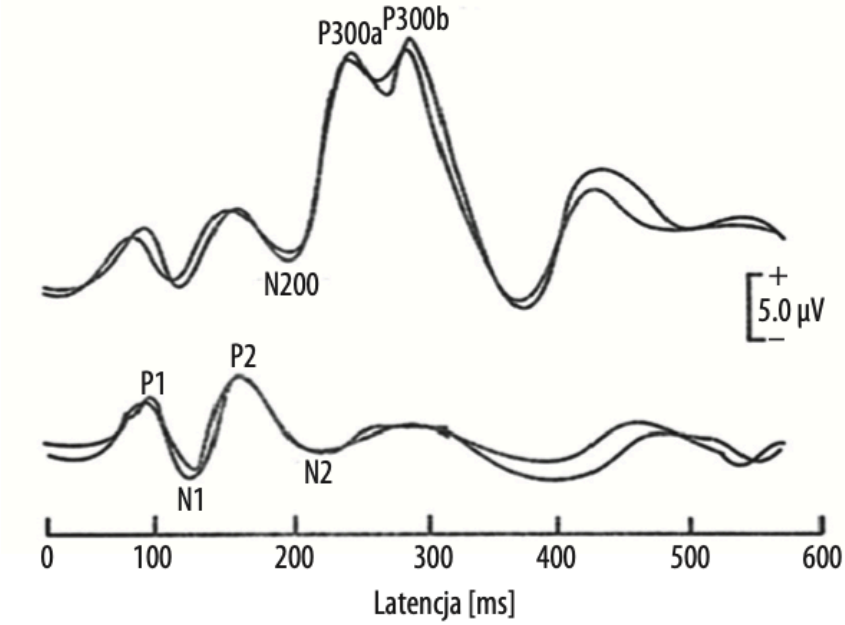
Badanie ABR, dzięki analizie interlatencji, umożliwia stwierdzenie zaburzeń neurologicznych (Szymańska i in., 2010). Należy także zauważyć, że bardzo często zapisy ABR mogą być zmienione w zaburzeniach pozaślimakowych pomimo faktu, że próg słyszenia może być prawidłowy (Kochanek, Śliwa, 2017). Ponadto, w przeciwieństwie do innych badań z zakresu AEP (*Auditory Evoked Potentials*), ABR jest badaniem powszechnie używanym, dzięki czemu wartości referencyjne są dobrze znane i możliwe jest dokładne stwierdzenie wartości odbiegających od normy. Innymi celami wykonywania badania odpowiedzi z pnia mózgu jest ocena otoneurologiczna, pozwalająca na opisanie funkcji całej drogi słuchowej. Umożliwia ono także diagnostykę centralnych zaburzeń słuchu - ocenę czynności kojarzeniowo-poznawczych w segmencie słuchowym ośrodkowego układu nerwowego. Jest także badaniem uzupełniającym w zaburzeniach procesu komunikatywnego, np. afazji (Pruszewicz i Obrębowski, 2010).

1.2.2. Potencjał poznawczy P300

Fala P300 to potencjał wywołany o długim czasie utajenia, który umożliwia badanie kory słuchowej w ośrodkowym układzie nerwowym. Liczba 300 w nazwie tego potencjału

odnosi się bezpośrednio do czasu, po jakim się on pojawia. Potencjał ten reprezentuje aktywność korową obejmującą umiejętności dyskryminacyjne, integracyjne i uwagę, będąc wskaźnikiem szybkości przetwarzania korowego. P300 wzbudził duże zainteresowanie społeczności badawczej, ponieważ często jest nieprawidłowy u pacjentów z zaburzeniami funkcji poznawczych (Kraus & McGee, 1994). Potencjał ten odzwierciedla proces decyzyjny w korze mózgowej.

Badanie potencjału P300 polega na podaniu badanemu dwóch rodzajów bodźców tzw. bodźca *standard* i *dewiant* (lub *oddball*). Zadaniem badanego jest skupianie uwagi na bodźcu *dewiant*. Na Rysunku 1.3. przedstawiono przykładowy przebieg elektroencefalogramu będący wynikiem badania potencjału P300.



Rysunek 1.3. Przykład słuchowych potencjałów korowych zarejestrowanych w odpowiedzi na bodźce *dewiant* (górna część rysunku) oraz bodźce *standard* (dolna część rysunku) (Milner, 2015).

Istnieją naukowe przesłanki, że wiele z tych potencjałów, z potencjałem P300 na czele, jest powiązanych ze schorzeniami powodującymi degenerację komórek nerwowych ośrodkowego układu nerwowego (Lee, 2021, Kopka & Truszczyński, 2013, Słotwiński & Zagrajek, 2012). Jednym z takich schorzeń jest choroba Alzheimera.

1.3. Choroby neurodegeneracyjne a badania słuchowe

Choroba Alzheimera jest schorzeniem rozwijającym się zazwyczaj na długo przed tym, kiedy zostaje postawiona diagnoza. Nie istnieją aktualnie farmaceutyki pozwalające na całkowite zatrzymanie lub cofnięcie rozwoju choroby (Papadaniil i in., 2016). Jednocześnie wszystkie powszechnie stosowane metody stawiania jednoznacznej diagnozy są inwazyjne, trudno dostępne lub kosztowne. Istnieje zatem potrzeba przyspieszenia procesu stawiania diagnozy, tak, by jak najwcześniej wdrożyć leczenie, które - hipotetycznie - w fazie prodromalnej lub wczesnym stadium choroby może przynieść o wiele lepsze skutki. Jednym ze sposobów może być wykonanie szeregu badań ukazujących funkcjonowanie zmysłów oraz umiejętności poznawczych osoby badanej.

Dowiedziono, że u osób z rozwiniętym AD występują istotne różnice w amplitudach i latencjach fali P300, które mają związek z zaburzeniami umiejętności kognitywnych. W związku z tym może być ona używana jako badanie potwierdzające upośledzenie funkcji poznawczych w związku z chorobą Alzheimera, ale także innych chorób otępiennych i zaburzeń psychiatrycznych, takich jak schizofrenia, zaburzenia obsesyjno-kompulsywne, depresja, choroba Parkinsona, czy pląsawica Huntingtona (Lee, 2021, Kopka & Truszczyński, 2013, Słotwiński & Zagrajek, 2012). W dostępnych badaniach nie ma jednak konsensusu co do wartości latencji i amplitud, które jednoznacznie świadczyłyby o chorobie Alzheimera - istnieje potrzeba ich standaryzacji (Pedroso i in., 2012). Ponadto, brakuje wiedzy na temat tego, jakie wyniki tych badań uzyskują osoby we wczesnym stadium i fazie prodromalnej AD.

Badania późnych potencjałów wywołanych (ERP - *Event Related Potentials*) można przeprowadzać w różnoraki sposób z użyciem nieco odmiennych metod. Jednym ze sposobów stymulacji jest np. używanie dodatkowego bodźca-dystraktora, np. w postaci szumu białego (Cecchi i in., 2015). W zależności od rodzaju bodźca oraz od miejsca umieszczenia elektrody, zauważono różnice dla fali P300 (wraz z podziałem na załamek P3a i P3b) oraz dla innych poprzedzających ją późnych potencjałów, takich jak N100, N200, P200 itd. W większości przypadków różnice polegały na opóźnieniu latencji lub na zmniejszeniu amplitudy fal.

Kolejnym badaniem mogącym mieć związek z neurodegeneracją jest badanie DPOAE (*Distortion Product Otoacoustic Emission*) - otoemisji produktu zniekształceń nieliniowych. Dzięki temu badaniu można ocenić funkcjonowanie takich struktur drogi słuchowej jak jądra ślimakowe, kompleks oliwkowy górny, wzgórek dolny itd. (Liu i in., 2021). DPOAE nie jest powszechnie wykorzystywane w diagnostyce zaburzeń kognitywnych, jednak istnieje związek między występowaniem blaszek Aβ w mózgu obciążonego chorobą Alzheimera, a znacznym (10 - 25 dB) obniżeniem progów uzyskanych w badaniu DPOAE.

1.4. Cel pracy oraz hipoteza badawcza

Celem niniejszej pracy było zbadanie wpływu choroby Alzheimera w fazie prodromalnej na słuchowe potencjały wywołane: odpowiedzi z pnia mózgu oraz falę P300 i stwierdzenie, czy mogą one być rzetelnym biomarkerem stwierdzającym wczesny rozwój AD.

W związku ze stosunkowo dużą liczbą publikacji potwierdzających referencyjne wartości latencji i amplitud w badaniu ABR, zdecydowano się na jego wykonanie. Spodziewano się związku pomiędzy odbiegającymi od norm wynikami tego badania oraz rozwojem choroby Alzheimera. Szczególny nacisk położono na analizę interlatencji i komponentów obwodowych zapisu (fala I-III, III-V, I-V), gdyż spodziewany jest ich związek z zaburzeniami neurologicznymi (Szymańska i in., 2010).

Hipoteza 1: Wyniki badań fali P300 (która ma związek z umiejętnościami kognitywnymi) u osób w fazie prodromalnej choroby Alzheimera będą statystycznie istotnie różne od wyników badań osób zdrowych.

Hipoteza 2: Wyniki badania ABR (odpowiedzi z pnia mózgu) u osób w fazie prodromalnej choroby Alzheimera będą statystycznie istotnie różne od wyników badań u osób zdrowych.

2. Metodyka badań

Decyzja o wyborze testów diagnostycznych została dokonana przez zespół badawczy złożony m. in. z akustyków i lekarzy specjalistów z zakresu neurologii, działający w ramach

Alzheimer Prediction Project[1]. Pacjentom w ramach diagnostyki wykonywano także m. in. badania wzroku, czy subiektywne badania słuchu, jednak w tej pracy skupiono się na obiektywnych badaniach słuchu. Zespół badawczy uzyskał zgodę na wykonywanie badań od Komisji Bioetycznej przy Wielkopolskiej Izbie Lekarskiej.

Każda osoba przed przystąpieniem do badań wyraziła pisemną zgodę na udział w eksperymentach oraz przetwarzanie danych osobowych. Zgoda zawierała informacje na temat przebiegu badań oraz omówienie problemu związanego z moralnym problemem ewentualnego postawienia diagnozy choroby Alzheimera, gdyż nie istnieje na nią lek.

Przed rozpoczęciem badań przesiewowych zadbano o komfort pacjenta - zapewniono mu coś do picia oraz poinstruowano na temat przebiegu badań, a w razie potrzeby odpowiadano na wszystkie pytania. Przeprowadzono także szczegółowy wywiad audiologiczny. W ramach skriningu wykonano otoskopię i audiometrię tonalną, badając przewodnictwo powietrzne i kostne. W tabelach nie uwzględniono wartości progów przewodnictwa kostnego, ponieważ nie wykazały one rezerwy ślimakowej. W części obiektywnej badań zawarto tympanometrię oraz badanie otoemisji produktu zniekształceń nieliniowych. Kolejnymi badaniami było, wykonanie z użyciem elektrod rejestracji odpowiedzi z pnia mózgu (ABR) oraz badanie potencjału P300. Podczas badania P300 pacjentowi odtwarzano dwa rodzaje sygnałów - sygnał tonalny *frequent* o częstotliwości 2 kHZ pojawiający się z częstością 80% oraz sygnał *rare* o częstotliwości 1,5 kHz pojawiający się z częstością 20%. Pacjent słyszał jeden ton na sekundę i był instruowany, aby ze skupieniem zwracać uwagę na sygnały typu *rare* i każdy z nich zliczać za pomocą klikacza, który znajdował się na jego palcu.

W badaniu udział wzięło 14 osób - 10 kobiet i 4 mężczyzn. Były to osoby z przedziału wiekowego od 40 do 83 lat. Średni wiek, a zarazem mediana wyniosły 60 lat.

Celem badania było sprawdzenie, czy istnieje korelacja między ubytkiem słuchu, a utratą pamięci. Ze względu na problematykę zagadnienia, nie udało się jednak uzyskać wyników obiektywnych badań neurologicznych pod kątem potwierdzenia wczesnej fazy choroby Alzheimera od wszystkich osób. Czynnik utraty pamięci ujęto więc w parametrze, który nazwano subiektywną oceną stopnia otępienia badanego. Z pacjentami

[1] strona internetowa Alzheimer Prediction Project - https://alz.put.poznan.pl (dostęp 21.03.2024 r.)

przeprowadzano wywiad, który zawierał pytania takie jak „Czy pogorszyło się u Pana/Pani tempo czytania?", „Czy zauważył/a Pan/Pani problem z przypominaniem sobie słów?" itp. Z tych odpowiedzi utworzono następnie wskaźnik, zgodnie ze wzorem, że każda odpowiedź twierdząca oznacza wzrost subiektywnej oceny stopnia otępienia badanego o 1 punkt. Wskaźnik ten wynosił, zależnie od osoby badanej, od 0 do 8 punktów.

3. Wyniki

Uzyskane wyniki badań poddano analizie w programie JASP, posłużono się klasycznym testem ANOVA. Dane wejściowe do analizy statystycznej umieszczono w Tabeli 4.1. We wszystkich analizach skupiono się tylko na uszach prawych osób badanych. Kolejne osoby badane zakodowano oznaczeniami Alz1, Alz2 itd.

W kolumnach 1-3 przedstawiono latencje poszczególnych fal zarejestrowanych podczas badania ABR, wyrażone w milisekundach. W kolejnych kolumnach znajdują się interlatencje, będące różnicami między poszczególnymi latencjami. Dalej znajdują się latencje poszczególnych fal zarejestrowanych w badaniu potencjału P300, także wyrażone w milisekundach. Jak można zauważyć - dla dwóch osób nie udało się wykonać badania, a u niektórych osób nie było możliwe wyróżnienie niektórych załamków na krzywej (np. Alz6 i Alz7).

Wartość graniczną słuchu normalnego ustalono na poziomie 25 dB HL. W celu wyznaczenia średniej, brano pod uwagę wartości progów dla 500, 1000, 2000 oraz 4000 Hz, obliczając z nich HTL (*Hearing Threshold Level)*. Wykonując analizy, posłużono się nazwą czynnika „czy_norma_sluch_wg_WHO", który klasyfikował wartości na binarny zbiór, na podstawie odpowiedzi tak/nie. Natomiast kryteriami do interpretacji wystąpienia lub braku DPOAE były średni poziom otoemisji ze wszystkich pasm częstotliwości na poziomie większym niż -10 dB SPL oraz stosunek sygnału do szumu (SNR - *Signal to Noise Ratio*) na poziomie większym niż 9 dB. W jednej z kolumn umieszczono także wartość SNR z badania DPOAE. Jako „normę" w badaniu tympanometrycznym kwalifikowano tympanogramy typu A. W badaniu DPOAE oraz tympanometrii posłużono się analogicznymi do tympanometrii nazwami czynników tj. „czy_DPOAE_wystapila" oraz „czy_tympano_norma". Pytania te klasyfikowały poszczególne osoby badane do zbiorów zgodnie z odpowiedzią tak/nie.

Podział na dwie kategorie kolumnie stopien_otepienia wyglądał następująco: osobom, które uzyskały od 0 do 4 punktów subiektywnej oceny otępienia przydzielono kategorię „brak/małe”, a osobom, które uzyskały 5-8 punktów, kategorię „średnie/duże”. Temu czynnikowi w programie JASP nadano nazwę „stopien_otepienia” i w zależności od liczby „uzyskanych” punktów przydzielał on każdą osobę do dwuelementowego zbioru „brak/małe” lub „średnie/duże”. Dane na temat subiektywnej oceny stopnia otępienia osoby badanej nie zostały zebrane od wszystkich osób - brakuje dwóch wyników, ponieważ nie wszystkie osoby przechodziły jednakową diagnostykę psychiatryczną. Aby zapewnić zanonimizowanie danych, nazwiska osób były kodowane przez lekarza kwalifikującego, a wszelkie wyniki badań były przypisywane bezpośrednio do kodu osoby tak, aby identyfikacja danych osobowych oraz powiązanie ich z wynikami badań nie były możliwe.

Na wszystkich wykresach przedstawiających wyniki badań ABR za pomocą czarnych słupków przedstawiono przedziały ufności 95%. Było to możliwe dzięki temu, że badanie ABR jako jedyne było powtarzane trzykrotnie dla każdej osoby.

Tabela 4.1. Dane wejściowe do analizy statystycznej

kod_pacjenta	1fala_ABR	3fala_ABR	5fala_ABR	interwal1-3_ ABR	interwal3-5_ ABR	interwal1-5_ ABR	P1_P300	N1_P300	P2_P300	N2_P300	P3_P300	wiek	czy_norma_sluch _wg_WHO	Czy_DPOAE_ wystapila	DPOAE_war tosci_SNR	Czy_tympano_ norma	typ_tympano gramu	stopien_otepienia
Alz1	1,81	4,18	6,02	2,37	1,85	4,21	32	84	212	276	350	73	nie	nie	6,58	tak	A	
Alz2	1,74	4,12	6,15	2,38	2,03	4,40	38	110	220	316	382	83	nie	nie	3,54	nie	As	
Alz3	1,78	3,81	5,80	2,03	1,99	4,02	32	78	128	166	244	78	nie	nie	4,02	tak	A	brak/małe
Alz4	1,73	3,63	5,45	1,90	1,81	3,71	40	82	154	280	344	40	tak	tak	28,4	tak	A	brak/małe
Alz5	1,62	3,83	5,60	2,21	1,77	3,98	282	320	368	414	452	64	tak	tak	21,38	tak	A	średnie/duże
Alz6	1,89	4,00	5,85	2,11	1,84	3,96	36	82	144		288	62	nie	nie	3,16	tak	A	średnie/duże
Alz7	1,24	3,81	5,69	2,57	1,88	4,45	44	86	140		388	51	tak	tak	25,2	tak	A	brak/małe
Alz8	1,63	3,89	5,77	2,26	1,88	4,13	50	90	180	264	326	46	tak	tak	14,4	tak	A	średnie/duże
Alz9	1,70	3,73	5,54	2,03	1,81	3,84						45	tak	tak	16,54	tak	A	brak/małe
Alz10	2,12	4,01	5,95	1,89	1,94	3,83						69	tak	nie	1,84	tak	A	brak/małe
Alz11	1,79	3,78	5,48	1,99	1,70	3,69	44	94	204	254	284	53	tak	tak	16	nie	C	średnie/duże
Alz12	1,53	3,79	5,87	2,26	2,08	4,33	38	76	196	300	352	54	nie	nie	7,44	nie	As	brak/małe
Alz13	1,83	4,07	5,74	2,23	1,68	3,91	52	90	286	348	456	72	tak	tak	11,36	tak	A	średnie/duże
Alz14	1,76	3,68	5,34	1,92	1,66	3,59	36	70	256	328	366	58	tak	tak	15,94	tak	A	średnie/duże

Latencja załamka P3 fali P300 nie miała istotnych statystycznie związków z żadnym z czynników. Wyniki testu zaprezentowano w Tabeli 4.1.

Tabela 4.1. Efekty główne dla zmiennej zależnej P3_P300

Czynnik	Suma kwadratów	F	p
czy_norma_sluch_wg_WHO	7442,438	2,018	0,186
czy_DPOAE_wystapila	7442,438	2,018	0,186
czy_tympano_norma	711,111	0,163	0,695
stopien_otepienia	2160	0,419	0,536

Na danej grupie badanej nie wykryto również związku między wiekiem a stopniem otępienia. Ten wynik zaznaczono w tabeli żółtym kolorem, tak jak inne, najbardziej interesujące według autorów, wyniki. Wystąpiła korelacja między wiekiem a średnim poziomem HTL (czy_norma_sluch_wg_WHO) oraz wystąpieniem DPOAE (Tabela 4.2.) - średni poziom HTL rósł wraz z wiekiem, a poziom otoemisji wraz z wartością SNR malał. To potwierdza, że wraz z wiekiem może rosnąć szansa na ubytek odbiorczy ślimakowy.

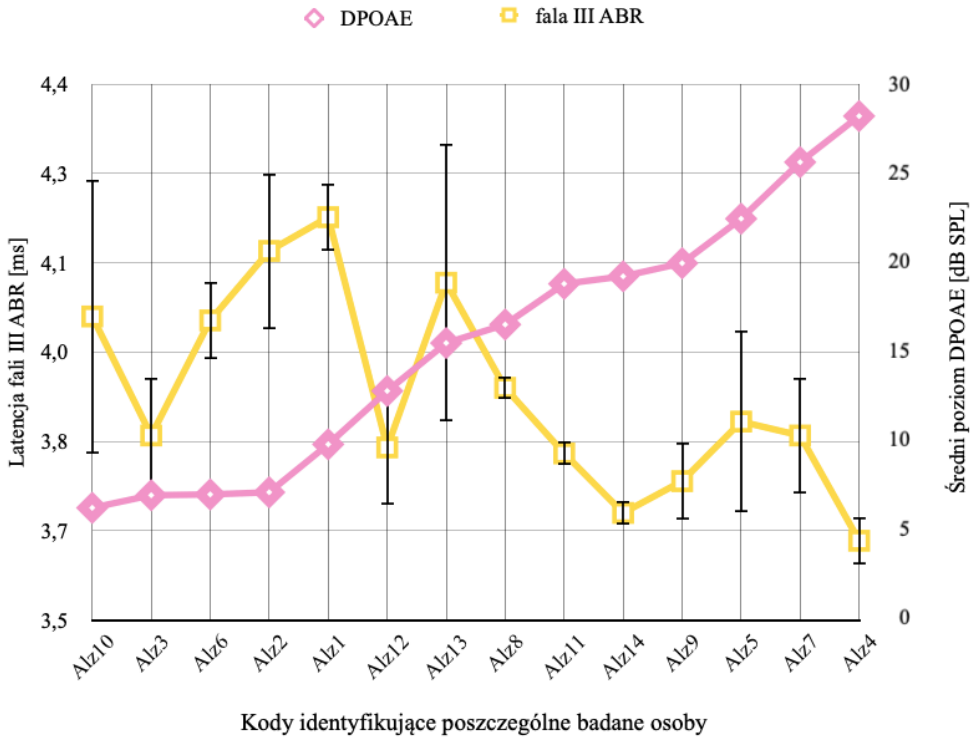
Tabela 4.2. Efekty główne dla zmiennej zależnej wiek

Czynnik	Suma kwadratów	F	p
czy_norma_sluch_wg_WHO	691,429	5,312	0,04
czy_DPOAE_wystapila	900,72	7,990	0,015
czy_tympano_norma	29,126	0,157	0,699
stopien_otepienia	27	0,184	0,677

Zaobserwowano związek między wystąpieniem DPOAE, a latencją fali III w badaniu ABR. Opóźnienie latencji tego załamka może mieć związek np. z zaburzeniami nerwu słuchowego (Kochanek, 2017). Wyniki tej analizy przedstawiono w Tabeli 4.3. Na Rysunku 4.1. zaprezentowano zależności, które wystąpiły dla tych wartości - wraz ze wzrostem poziomu otoemisji malała latencja fali III ABR. W celu ukazania zależności uszeregowano wartości latencji fali III w funkcji rosnącej.

Tabela 4.3. Efekty główne dla zmiennej zależnej fala_3_ABR

Czynnik	Suma kwadratów	F	p
czy_norma_sluch_wg_WHO	0,076	3,150	0,101
czy_DPOAE_wystapila	0,113	5,385	0,039
czy_tympano_norma	0,0009201	0,03	0,865
stopien_otepienia	0,018	1,015	0,337



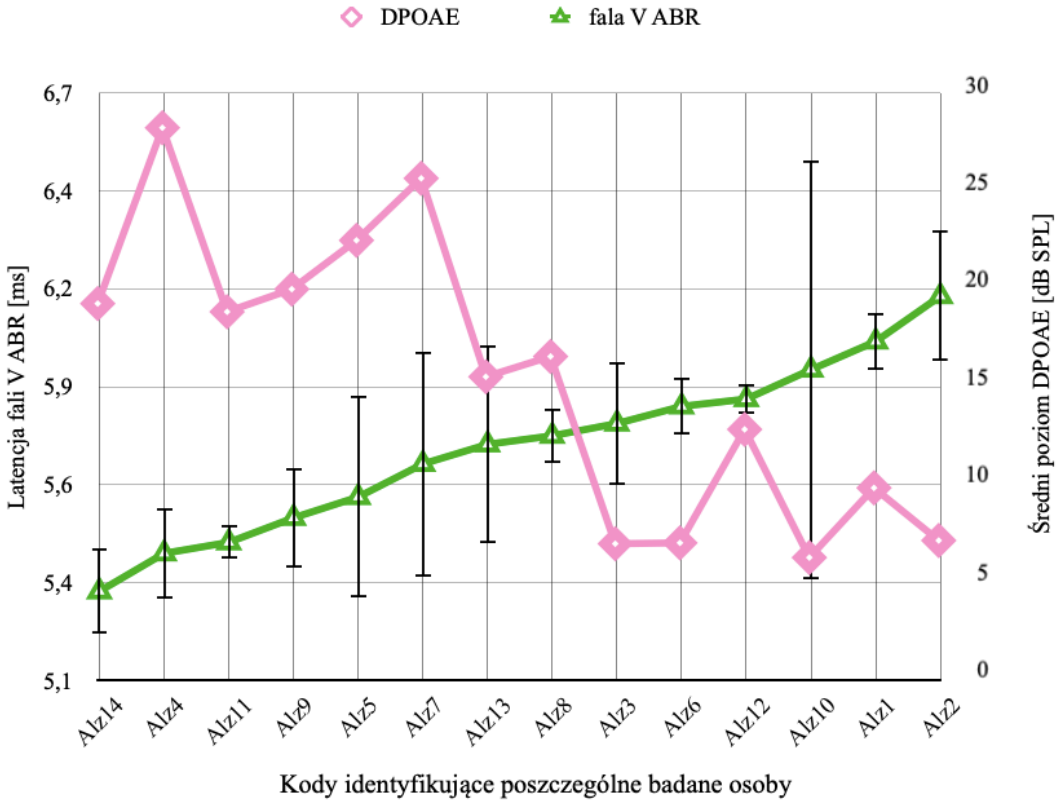
Rysunek 4.1. Porównanie wartości między poziomem DPOAE, a latencją fali III w badaniu ABR.

Zauważono istotny statystycznie związek latencji fali V z wystąpieniem otoemisji akustycznej oraz ze średnim poziomem słyszenia. Nie wykazano jednak związku fali V ABR z subiektywnym stopniem otępienia badanego. Wyniki znajdują się w Tabeli 4.4.

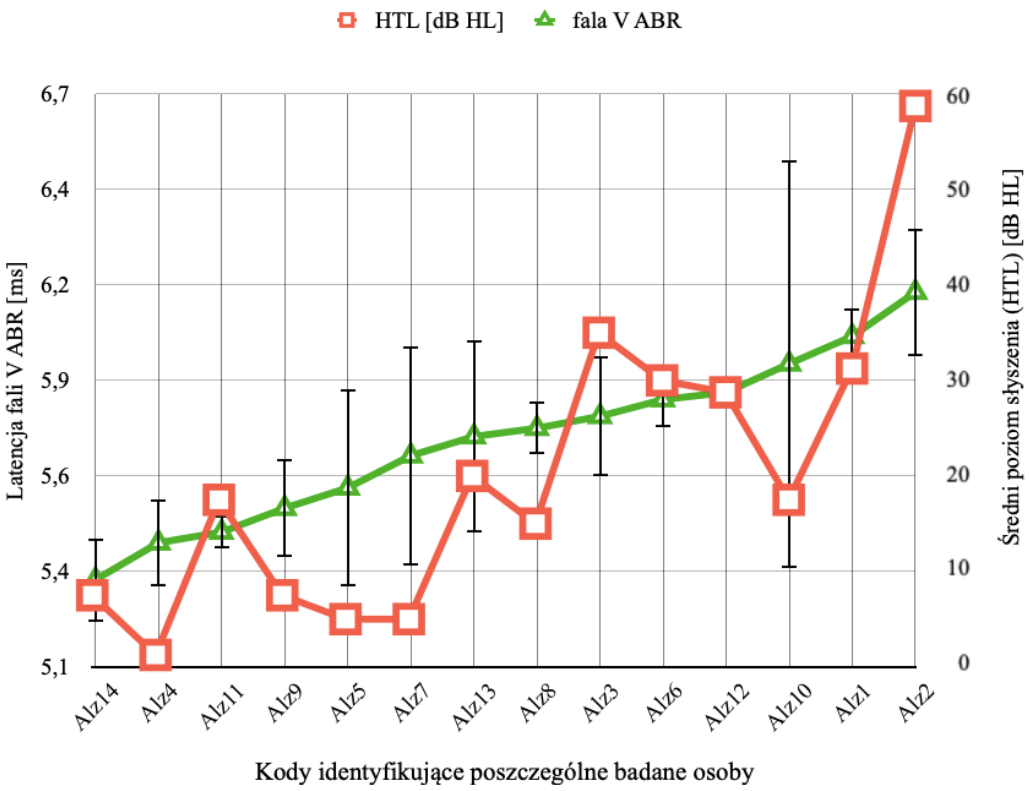
Tabela 4.4. Efekty główne dla zmiennej zależnej fala_5_ABR

Czynnik	Suma kwadratów	F	p
czy_norma_sluch_wg_WHO	0,328	10,771	0,007
czy_DPOAE_wystapila	0,450	22,208	<0,001
czy_tympano_norma	0,037	0,683	0,425
stopien_otepienia	0,022	0,6	0,457

Na Rysunkach 4.2. oraz 4.3. przedstawiono, jak ilościowo zmieniają się dane, dla których wykryto zależności - fala V ABR wraz z DPOAE oraz fala V ABR ze średnim poziomem słyszenia. Poziom DPOAE spadał wraz ze wzrostem latencji fali V, z kolei średni poziom słyszenia rósł wraz ze wzrostem latencji fali V. W celu ukazania zależności na obu rysunkach uszeregowano wartości latencji fali V w funkcji rosnącej.



Rysunek 4.2. Porównanie wartości między latencją fali V ABR, a poziomem DPOAE.

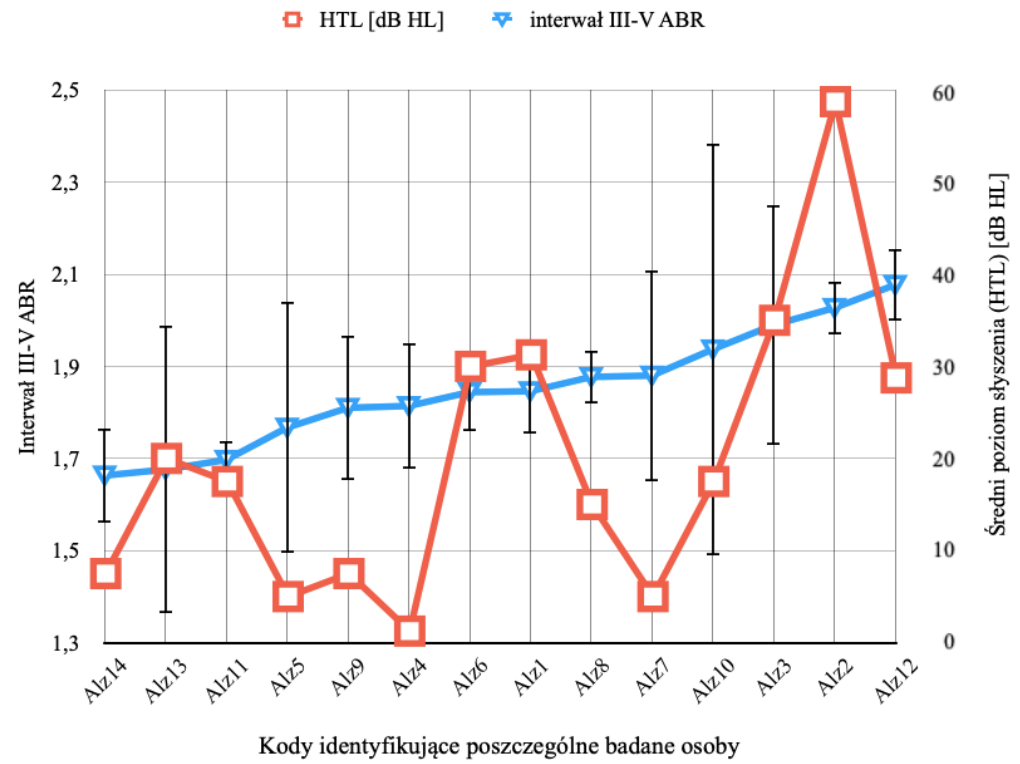


Rysunek 4.3. Porównanie wartości między latencją fali V, a średnim poziomem słyszenia (HTL).

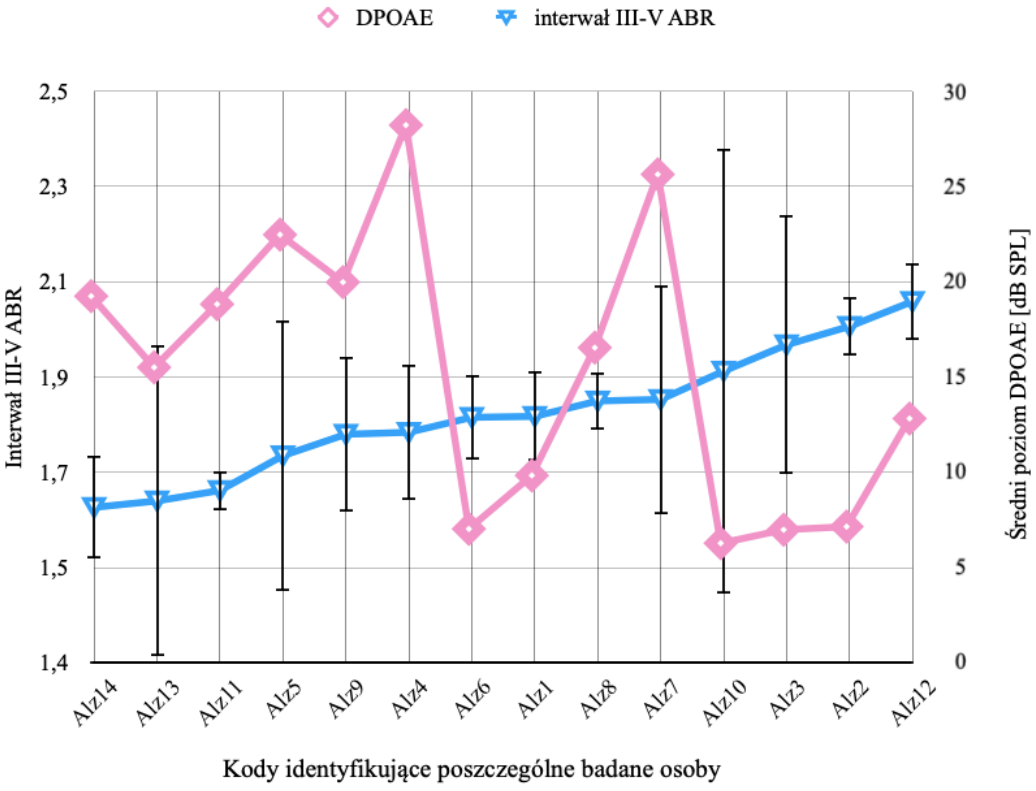
Analiza interwału pomiędzy falą III i V w badaniu ABR wykazała istotną korelację czasu trwania tegoż interwału ze średnim progiem słyszenia oraz z wystąpieniem DPOAE. Najbardziej interesującym wynikiem jest jednak związek tego interwału z subiektywnym stopniem otępienia badanego. Wyniki przedstawiono w Tabeli 4.5. Na Rysunkach 4.4. - 4.6. przedstawiono zależności poszczególnych parametrów od interwału III-V ABR.

Tabela 4.5. Efekty główne dla zmiennej zależnej interwał_3-5_ABR

Czynnik	Suma kwadratów	F	p
czy_norma_sluch_wg_WHO	0,088	8,666	0,012
czy_DPOAE_wystapila	0,112	13,617	0,003
czy_tympano_norma	0,027	1,732	0,213
stopien_otepienia	0,081	8,426	0,016



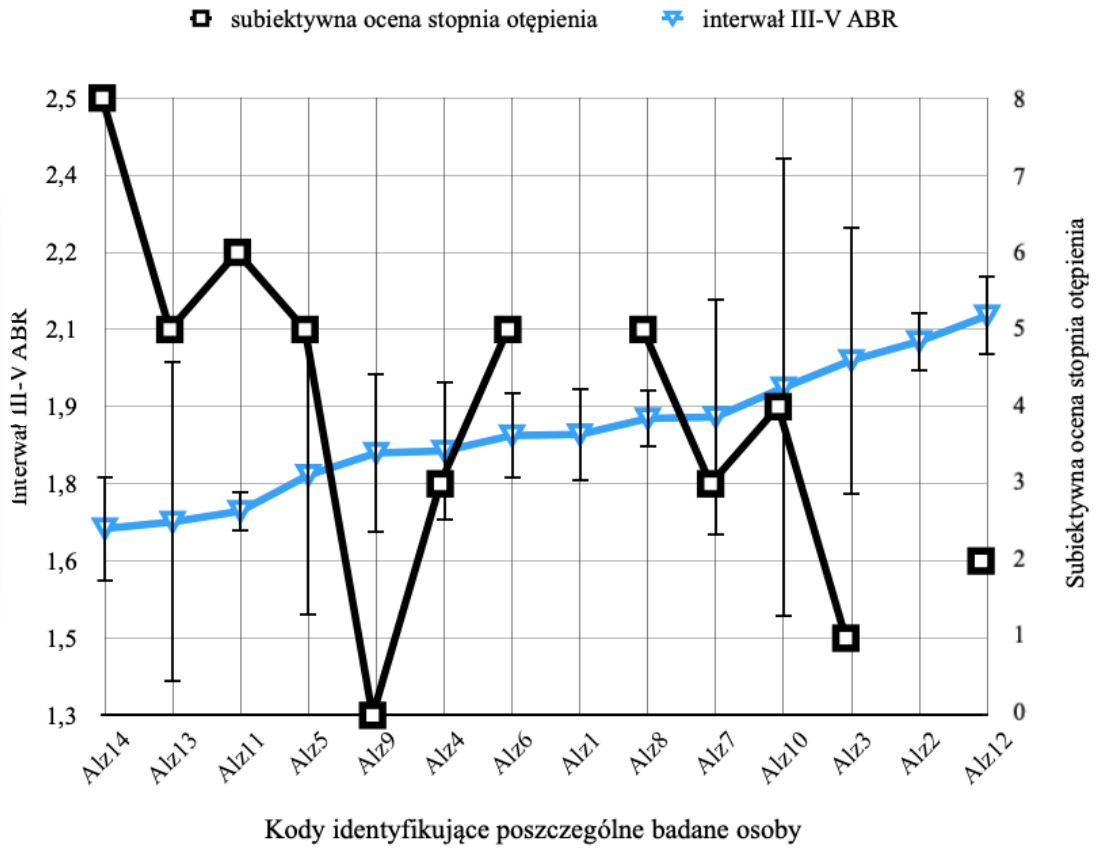
Rysunek 4.4. Porównanie wartości między wielkością interwału III-V ABR, a średnim poziomem słyszenia.



Rysunek 4.5. Porównanie wartości między wielkością interwału III-V ABR, a poziomem DPOAE.

Ilościowa analiza interwału III-V ABR oraz średniego poziomu słyszenia wykazała, że wzrost jednej z tych wartości jest skorelowany ze wzrostem drugiej (Rysunek 4.4.). Z kolei analiza interwału III-V oraz poziomu DPOAE (Rysunek 4.5.) wykazała, że wraz ze wzrostem interwału maleje poziom DPOAE. W celu ukazania zależności uszeregowano wartości interwału III-V w funkcji rosnącej.

Mimo, iż dane na temat subiektywnej oceny stopnia otępienia nie zostały zebrane od wszystkich osób, zauważono istotny statystycznie związek tego parametru z interwałem III-V. Tę zależność przedstawiono na wykresie na Rysunku 4.6.



Rysunek 4.6. Porównanie wartości między wielkością interwału III-V ABR, a subiektywną oceną stopnia otępienia osoby badanej.

Analizy dalszych czynników (pozostałych fal z badania późnych potencjałów, pozostałych fal z badania ABR) nie wykazały istotnych statystycznie związków z żadnym z pozostałych czynników.

4. Dyskusja i wnioski

Celem pracy było potwierdzenie dwóch hipotez:

Hipoteza 1: Wyniki badań fali P300 (która ma związek z umiejętnościami kognitywnymi) u osób w fazie prodromalnej choroby Alzheimera będą statystycznie istotnie różne od wyników badań osób zdrowych.

Hipoteza 2: Wyniki badania ABR (odpowiedzi z pnia mózgu) u osób w fazie prodromalnej choroby Alzheimera będą statystycznie istotnie różne od wyników badań u osób zdrowych.

Hipotezy próbowano potwierdzić, wykonując przesiewową słuchu (audiometrię tonalną oraz tympanometrię) oraz badania obiektywne, takie jak badanie otoemisji produktu zniekształceń nieliniowych (DPOAE), badanie odpowiedzi z pnia mózgu (ABR) oraz badanie potencjału P300. Wyniki tych badań poddano analizie wariancji ANOVA zestawiając je z subiektywną oceną stopnia otępienia osoby badanej. Parametr ten wyznaczono dzięki odpowiedziom z formularza dotyczącego oceny własnych umiejętności kognitywnych, na który osoby badane odpowiadały pod okiem lekarza psychiatry. Uzyskanie obiektywnych danych na temat stopnia neurodegeneracji z badań takich jak np. badanie płynu mózgowo-rdzeniowego niestety okazało się niemożliwe.

Ścieżka postępowania w ramach badania P300 nie doprowadziła do oczekiwanych rezultatów - nie zaobserwowano istotnych statystycznie związków między subiektywnym stopniem otępienia osoby badanej, a latencją żadnego z załamków fali P300.

Problem techniczny sprawiło samo wykonanie badania P300. W niektórych przypadkach pacjenci mieli problem ze skupieniem lub mimo wielokrotnego tłumaczenia nie mogli zrozumieć co jest ich zadaniem w tym badaniu (zliczanie bodźców za pomocą klikacza na palcu). Problemem była także sama metodologia. Badania profilaktyczne w kierunku choroby Alzheimera nie są powszechne, więc trudno jest znaleźć osoby, które zaliczają się do grupy zagrożonej wystąpieniem tej choroby, a tym bardziej mają potwierdzone wystąpienie wczesnej fazy choroby Alzheimera. Osoby te nierzadko przebywały długą drogę przez całą Polskę, a budżet nie pozwalał na zorganizowanie dla nich noclegu. Wszystkie badania były więc wykonywane jednego dnia. Czas trwania jednej sesji (z przerwami) wynosił ok. 3-4 godziny. Oprócz badań obiektywnych wykonywano także badania subiektywne słuchu oraz

badania wzroku. Niektórzy badani wykazywali oznaki zmęczenia, co mogło wpłynąć na wyniki. Badanie potencjału P300, które jest dość czasochłonne, wymagało także skupienia, a było wykonywane na końcu. Jednak inne badania subiektywne, które były przeprowadzane wcześniej, także wymagały skupienia.

Bardziej obiecujące wyniki uzyskano, analizując interlatencje z badania ABR - najważniejszym wynikiem był związek subiektywnej oceny stopnia otępienia z interwałem między falą III i V. Jest to sugestia, aby w dalszych badaniach skupić się na analizie interlatencji oraz kilkukrotnym wykonywaniu badania ABR dla jednego poziomu głośności. Dodatkowym atutem tego postępowania jest fakt, że to badanie jest szybsze w wykonaniu i mniej skomplikowane niż badanie potencjału P300.

Ubytki słuchu o podłożu neurodegeneracyjnym mogą także wystąpić w wyniku długotrwałego ubytku słuchu. Niekoniecznie muszą one mieć wtedy związek z chorobą Alzheimera. Zjawisko to może występować, jeśli dana osoba przez długi czas była pozbawiona bodźców słuchowych, a więc w przypadku np. osoby z długotrwałym ubytkiem słuchu nienoszącej aparatów słuchowych. Taka sytuacja mogła wystąpić np. u 83-letniej osoby badanej o kodzie Alz2. Większość pacjentów posiadała jednak brak, lub lekkie ubytki słuchu, co pozwala pominąć ten czynnik.

Głównymi perspektywami rozwoju badania są:

- uzyskanie wyników obiektywnych badań neurologicznych, dla potwierdzenia obecności wczesnej fazy choroby Alzheimera;
- wykonywanie większej liczby testów psychologicznych, takich jak test MMSE (*Mini Mental State Examination*), test zegara itp.;
- skrócenie czasu badań lub rozłożenie ich na kilka sesji;
- wykonanie badań longitudinalnych.

Literatura

1. Alzheimer's Association, *A blood test for Alzheimer's? Markers for tau take us a step closer,* Alzheimer's Association International Conference, 2020
2. Alzheimer's Association, *Stages of Alzheimer's*. Alzheimer's Disease and Dementia. https://www.alz.org/alzheimers-dementia/stages (dostęp 27.11.2022 r.), 2021
3. American Speech-Language-Hearing Association., *Short latency auditory evoked potentials [Relevant Paper],* 1987
4. Cecchi M.,, Moore D. K., Sadowsky C. H., Solomon P. R., Doraiswamy P. M., Smith C. D., Jicha G. A., Budson A. E., Arnold S. E., Fadem K. C., *A clinical trial to validate event-related potential markers of Alzheimer's disease in outpatient settings*. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 1 (2015) 387-394, 2015
5. Cichacz-Kwiatkowska, B., Sekita-Krzak, J., Kot-Bakiera, K., Jodłowska-Jędrych, B.& Wawryk-Gawda, E., *Choroba Alzheimera – rola badań immunohistochemicznych w diagnostyce choroby*. Journal of Education, Health and Sport. 6(2): 122-137, 2016
6. Humphries, C., Liebenthal, E., & Binder, J. R., *Tonotopic organization of human auditory cortex*. NeuroImage, 50(3), 1202–1211, 2010
7. Kenneth Kaitin, Christopher Milne. *Schizofrenia koncernów*. Świat Nauki nr. 9 (241), s. 18, 2011
8. Kochanek K., Śliwa L., *Metody obiektywne badania słuchu*, w: Hojan, E. (pod redakcją), Protetyka słuchu, Wydawnictwo Naukowe UAM, 2017
9. Kopka, M. & Truszczyński, O., *Przydatność potencjału P300 w diagnostyce zaburzeń poznawczych*, IV Ogólnopolskia Konferencja Naukowo-Szkoleniowa: Neurologia Medforum, 2013
10. Kraus, N. & McGee, T., *Mismatch negativity in the assessment of central auditory function*. Am J Audiol 3: 39-51, 1994
11. Lee, J., Lee, J., Shah, A., Ye, J., & ULAB PHHS., *Exploring the efficacy of P300 as a potential biomarker in detecting Alzheimer's disease: A replication study*. UC Berkeley: Public Health & Health Science Division, ULAB, 2021
12. Liu Y., Fang S., Liu L., Zhu Y., Chang R., Chen K., Zhao H., *Hearing loss is an early biomarker in APP/PS1 Alzheimer's disease mice*. Neurosci Lett. 2020 January 19; 717: 134705, 2020

13. Milner R., *Słuchowe potencjały korowe. Część II. Teoretyczne podstawy generacji oraz charakterystyka wybranych komponentów*, Nowa Audiofonologia 4(2), 2015
14. Nicholas, Marjorie & Obler, Loraine & Albert, Martin & Helm-Estabrooks, Nancy., Empty Speech in Alzheimer's Disease and Fluent Aphasia. Journal of speech and hearing research. 28. 405-10. 10.1044/jshr.2803.405
15. Ozimek, E., *Dźwięk i jego percepcja. Aspekty fizyczne i psychoakustyczne* (wydanie II rozszerzone), Wydawnictwo Naukowe PWN SA, 2018
16. Papadaniil, C. D., Kosmidou, V. E., Tsolaki, A., Tsolaki, M., Kompatsiaris, I. Y., & Hadjileontiadis, L. J., *Cognitive MMN and P300 in mild cognitive impairment and Alzheimer's disease: A high density EEG-3D vector field tomography approach.* Brain Research. 1648(A): 425-433, 2016
17. Pedroso, R. V., Fraga, F. J., Corazza, D. I., Andreatto, C. A., Coelho, F. G., Costa, J. L. & Santos-Galduróz, R. F., *P300 latency and amplitude in Alzheimer's disease: a systematic review*. Brazilian Journal of Otorhinolaryngology. 78(4): 126-32, 2012
18. Pistono, Aurélie & Senoussi, Mehdi & Guerrier, Lucile & Rafiq, M. & Gimeno, M. & Péran, Patrice & Jucla, Mélanie & Pariente, Jeremie., Language network connectivity increases in prodromal Alzheimer’s disease. 10.1101/2020.11.22.393199, 2010
19. Pruszewicz, A., Obrębowski, A., *Audiologia kliniczna - Zarys* (wydanie IV). Wydawnictwo Naukowe Uniwersytetu Medycznego im. Karola Marcinkowskiego w Poznaniu, 2010
20. Słotwiński, K. & Zagrajek, M., *Endogenne potencjały wywołane w zaburzeniach poznawczych*, Polski Przegląd Neurologiczny. 8 (3): 114-119, 2012
21. Strimbu, K., & Tavel, J. A., *What are biomarkers?*.Current opinion in HIV and AIDS, 5 (6): 463–466, 2010
22. Szymańska, A., Gryczyński, M. & Pajor, A., *Wywołane słuchowe potencjały stanu ustalonego (ASSR - auditory steady-state responses), dotychczasowy stan wiedzy.* Otolaryngol Pol 2010; 64 (5): 274-280), 2010
23. World Health Organization, *Dementia,*. https://www.who.int/news-room/fact-sheets/detail/dementia., 2020 (dostęp 27.11.2022 r.)

Aleksandra **SAWCZUK**[1], Bartłomiej **CHOJNACKI**[1]

METODY WIBROIZOLACJJI NISKOCZĘSTOTLIWOŚCIOWEJ DLA GRAMOFONÓW TYPU LEKKIEGO

LOW-FREQUENCY VIBROISOLATION METHODS FOR LIGHTWEIGHT TURNTABLES

[1] Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, al. Mickiewicza 30, 30-059 Kraków

asawczuk@student.agh.edu.pl

Streszczenie

Wibroizolacja to wszelkie działania, które mają na celu ograniczenie powstawania i przenoszenia się drgań pomiędzy ich źródłem a otoczeniem. Zachodzi ona jedynie powyżej częstotliwości rezonansowej, dlatego najczęściej redukcję drgań osiąga się przez zwiększanie masy układu. Jednakże takie rozwiązanie nie jest możliwe w przypadku delikatnych, lekkich urządzeń, do jakich zaliczają się gramofony o masie nieprzekraczającej 10 kg. W literaturze wykazano, że w urządzeniach tego typu szczególnie istotna jest kwestia przenoszenia drgań o niskich częstotliwościach z zakresu 5-15 Hz. Aby ograniczyć niepożądane dodatkowe wstrząsy na styku igły z płytą, należałoby odizolować gramofon od podłoża np. poprzez wykorzystanie platformy antywibracyjnej. W niniejszej pracy przedstawiona zostanie analiza wibroakustyczna gramofonu z wykorzystaniem różnorodnych metod pobudzenia wibroakustycznego. Korzystając z najlepszej wyznaczonej metody, przebadano różne rodzaje materiałów wibroizolacyjnych. Referat zaprezentuje porównanie badanych metod pomiarowych i metod wibroizolacji w celu wyznaczenia optymalnego rozwiązania do niwelacji przedstawionego problemu drgań niskoczęstotliwościowych.

1. Wprowadzenie

Gramofon, jako urządzenie służące do odtwarzania dźwięku, ponownie zyskuje na popularności w dzisiejszych czasach. Jego konstrukcja opiera się głównie na obracanym za pomocą napędu talerzu oraz nieruchomym ramieniu. Na końcu ramienia umieszczony jest przetwornik elektromechaniczny, czyli wkładka, do której przytwierdzona jest igła. Przesuwając się po rowkach płyty winylowej, zostaje ona wprowadzona w drgania, a jej ruch zamieniany jest przez wkładkę na sygnał elektryczny. Mechanizm ten zalicza się do bardzo

delikatnych układów i dlatego bardzo istotnym problemem jest odizolowanie drgającej igły od niepożądanych zakłóceń, które mogłyby zaburzyć pracę wkładki. Nie tylko spowoduje to utrudnienie w odtwarzaniu muzyki, a także doprowadzi do uszkodzenia płyt winylowych.

W celu ograniczenia występowania niepożądanych wstrząsów konieczne jest odizolowanie urządzenia od podłoża i wytłumienie wygenerowanych przez niego, i przenoszonych przez inne siły zewnętrzne drgań. Problem z doborem wibroizolacji dla gramofonów wynika z występowania rezonansu wkładki w bardzo niskich częstotliwościach. Wibroizolacja jest skuteczna jedynie powyżej częstotliwości rezonansowej [1], dlatego najczęściej redukcję wstrząsów dokonuje się poprzez zwiększanie masy układu. Rozwiązanie to nie jest zawsze możliwe do zastosowania, szczególnie w przypadku gramofonów lekkich. Z racji swojej niewielkiej masy, są bardziej podatne na niepożądane drgania niż cięższe układy. Z tego powodu dla lekkich gramofonów należałoby zastosować skuteczną wibroizolację w postaci np. platformy antywibracyjnej.

W poniższym referacie przedstawiona zostanie analiza źródeł i miejsc występowania drgań w gramofonie. Dodatkowo nastąpi omówienie metod pomiarów drgań dla tego typu układów oraz przedstawienie możliwych do zastosowania rodzajów wibroizolacji.

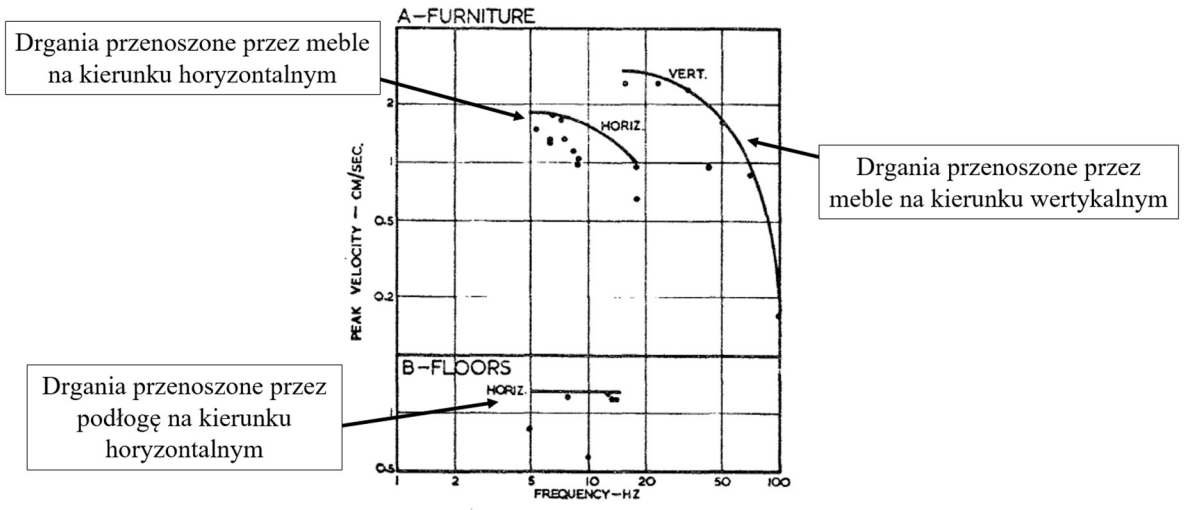
2. Źródła drgań w gramofonie i stosowane rozwiązania do ich redukcji

Źródła zakłóceń wibroakustycznych w gramofonie mogą być pochodzenia zarówno wewnętrznego, jak i zewnętrznego. W tej kwestii można wyróżnić dwie główne kategorie. Drgania występujące powyżej 20 Hz najczęściej wynikają z obecności sygnału w rowku płyty winylowej oraz z procesu jej odtwarzania [2]. Równocześnie, w okolicy 40 – 80 Hz pojawiają się wpływy działającego silnika gramofonu. Z tego powodu, bardziej znaczące są wibracje w niższych częstotliwościach, które pomimo leżenia poniżej zakresu częstotliwości słyszalnych silnie je modulują, wpływając na odbiór muzyki. Wynika to z faktu, iż w tych pasmach częstotliwości następuje problem ze stabilnością układu, spowodowany m.in. rezonansem wkładki oraz ramienia [2].

Największy wpływ na drgania gramofonu wywierają wstrząsy przenoszone przez podłoże i meble. Krytyczny zakres rozpoczyna się od 5 Hz i obejmuje wartości do 15,20 Hz, gdzie oba z wymienionych źródeł drgań nakładają się na siebie [3], co zostało przedstawione na rys. 1. Na wykresie zaznaczono drgania przenoszone przez podłogę w kierunku horyzontalnym oraz przez meble zarówno w kierunku horyzontalnym, jak i wertykalnym.

Dodatkowym źródłem ewentualnych zakłóceń w gramofonie są wypaczenia płyt winylowych, spowodowane zbyt wysoką temperaturą, ciśnieniem lub masą układu [4].

W literaturze przytoczono różne podejścia, które miały służyć rozwiązaniu omawianego problemu. Pierwszym z nich było zastosowanie dynamicznego stabilizatora [2]. W idealnych warunkach, rozwiązanie to miało na celu tłumienie rezonansu wkładki, co pozwalałoby na zachowanie stabilności centrowania igły w rowku, nawet w przypadku występowania w nim zmian jej wysokości. Zaproponowany system przypominał z wyglądu odwróconą osłonkę na igłę. Dodatkowo zawierał włókna umieszczone blisko igły, co ułatwiało prowadzenie ramienia nad płytą, nawet w przypadku jej odkształceń, poprzez zmniejszenie obciążenia igły. Ponadto, stabilizator amortyzował gwałtowne uderzenie spadającego ramienia na płytę [2]. Innym rozwiązaniem było zastosowanie silnika wyposażonego w urządzenie przeciwreakcyjne [5]. Drgania miały zostać wyeliminowane wprost w miejscu ich generowania. Pod głównym silnikiem talerza został zamontowany drugi silnik przeciwreakcyjny, działający w tej samej osi. Gdy momenty obrotowe generowane przez każdy z silników i oddziałujące na podstawę, były równe sobie, ale miały przeciwne zwroty, niwelowały się nawzajem, zmniejszając powstające drgania [5].



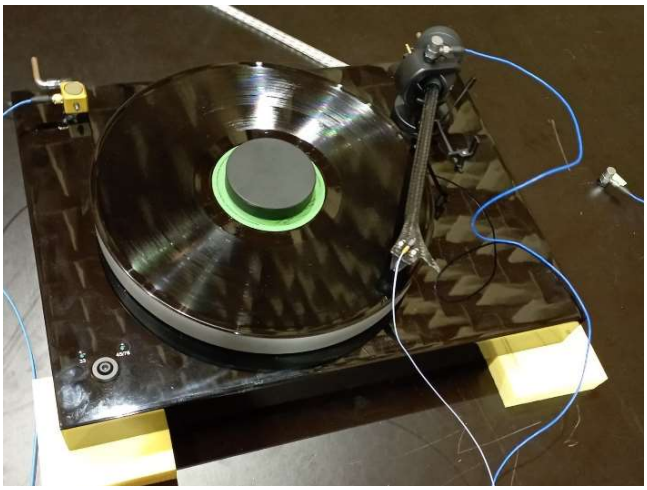
Rys. 1. Drgania przenoszone przez meble i podłogę w kierunkach wertykalnym i horyzontalnym [3].

3. Metodyka pomiarów wibroizolacji

Pomiary zostały przeprowadzone w komorze bezechowej. Na jednej płycie, umieszczonej na podłożu komory, ułożono gramofon, natomiast na drugiej stolik wraz z niezbędnym sprzętem pomiarowym obejmującym kartę RME Fireface UFX 1405, generator

szumu B&K 1405 oraz komputer. Przed przystąpieniem do pomiarów czujniki drgań zostały skalibrowane, a następnie przymocowane do gramofonu: na wkładce, ramieniu, podłożu i na obudowie (czujnik trójosiowy). W celu minimalizacji błędu pomiarowego, na wkładce umieszczono akcelerometr o niewielkiej masie, nieprzekraczającej 10% masy badanego elementu. Czujniki zostały podłączone do wejść karty pomiarowej, podobnie jak dwa kanały wyprowadzone z samego gramofonu. Ich umiejscowienie przedstawiono na rys. 2.

Do pobudzenia drgań zastosowano kilka metod dla wyodrębnienia najskuteczniejszej z nich. Wykorzystano stukacz młotkowy Norsonic i młotek impedancyjny (bez zamontowanego czujnika impedancji), aby zasymulować dźwięki uderzeniowe i drgania przenoszone przez podłoże i meble, które mogą występować w otoczeniu położenia gramofonu. Pobudzenie akustyczne szumem różowym, z wykorzystaniem głośnika Genelec 7050, zostało zastosowane w celu pobudzenia urządzenia większym zakresem częstotliwości w jednym momencie. Dodatkowo wykorzystano również pracę własną gramofonu zarówno bez dodatkowego szumu, jak i z nim.



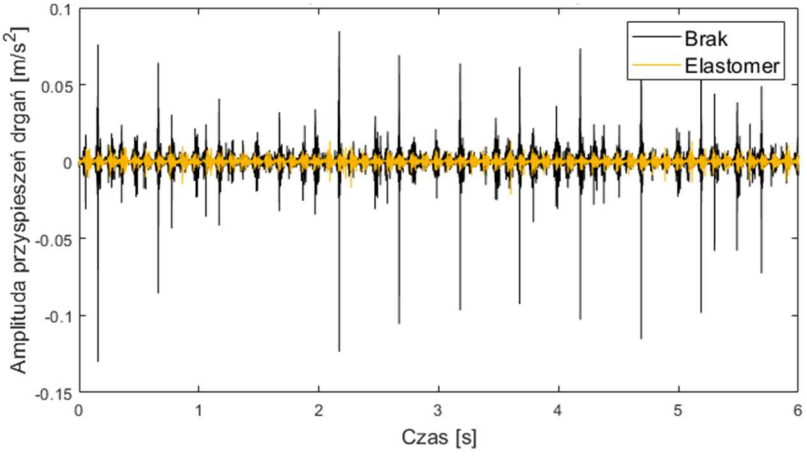
Rys. 2. Umiejscowienie czujników na badanym układzie.

Pierwszym etapem pomiarów było zbadanie drgań w gramofonie bez dodatkowej formy wibroizolacji. Następnie pod nóżki gramofonu podłożono elastomer, a kolejnymi zastosowanymi rozwiązaniami były 2 rodzaje wibroizolatorów – podłużny liniowy oraz okrągły trójosiowy, a także sprężyny. Wszystkie te elementy zostały umieszczone pod obudową gramofonu po odłączeniu fabrycznych nóżek. Wszystkie metody wibroizolacji zostały dobrane zgodnie z zasadami jej projektowania: dla częstotliwości rezonansowej 4-6 Hz przy obciążeniu 7-8 kg.

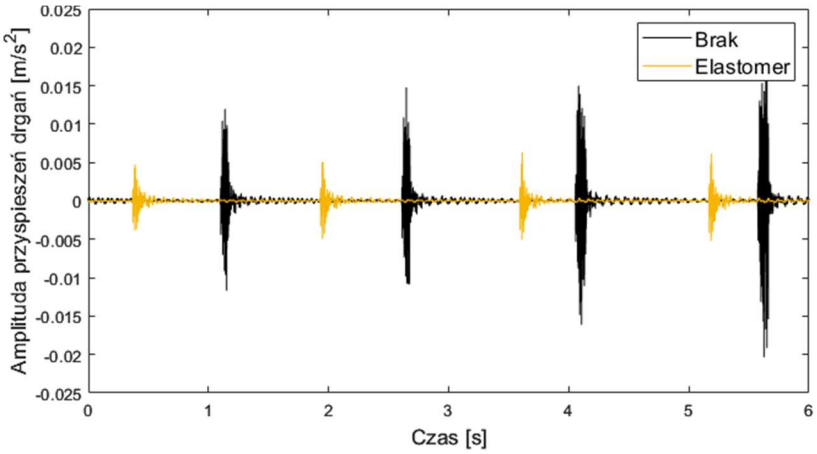
4. Porównanie wyników

Do przeanalizowania otrzymanych wyników wykorzystano analizy funkcji przejścia oraz analizę widmową. Analiza widmowa porównuje amplitudy w konkretnym punkcie badanego obiektu, w tym przypadku przyspieszenie drgań z i bez wykorzystania wibroizolacji. Przy zastosowaniu wibroizolacji amplitudy powinny być mniejsze. Wadą tej analizy jest potrzeba stałego, powtarzalnego pobudzenia, w celu wykonywania serii pomiarowych. Funkcja przejścia to iloraz widma policzonego np. na podłożu, a widmem na elemencie badanego układu. Przekazuje informację, ile drgań z podłoża przeniosło się na obiekt. Jeśli iloraz ten jest większy niż 1, to drgania zostały wzmocnione na obiekcie, a jeśli mniejszy od 1 – oznacza to zmniejszenie amplitudy drgań. Zaletą zastosowania takiej metody analizy jest mniejsza istotność bezwzględnej wartości widma w punkcie, więc sprawdzą się tu metody o zmiennym pobudzeniu, jednak wpływa to na większą problematyczność przy porównywaniu ze sobą konkretnych przypadków.

Na rys. 3 przedstawiono przebieg czasowy drgań zmierzonych na czujniku na wkładce, wykorzystując jako pobudzenie stukacz młotkowy. Otrzymano powtarzalny ciąg uderzeń o widocznym schemacie. Dzięki zastosowaniu wibroizolacji amplitudy drgań znacząco się zmniejszyły. Natomiast w przypadku użycia młotka powstały impulsy o różnych amplitudach, co przedstawiono na rys. 4. Nie jest to optymalne rozwiązanie dla wykorzystywania analizy widmowej oraz funkcji przejścia, ponieważ brak pomiaru siły uderzeń uniemożliwia poprawną analizę wpływu wibroizolacji.

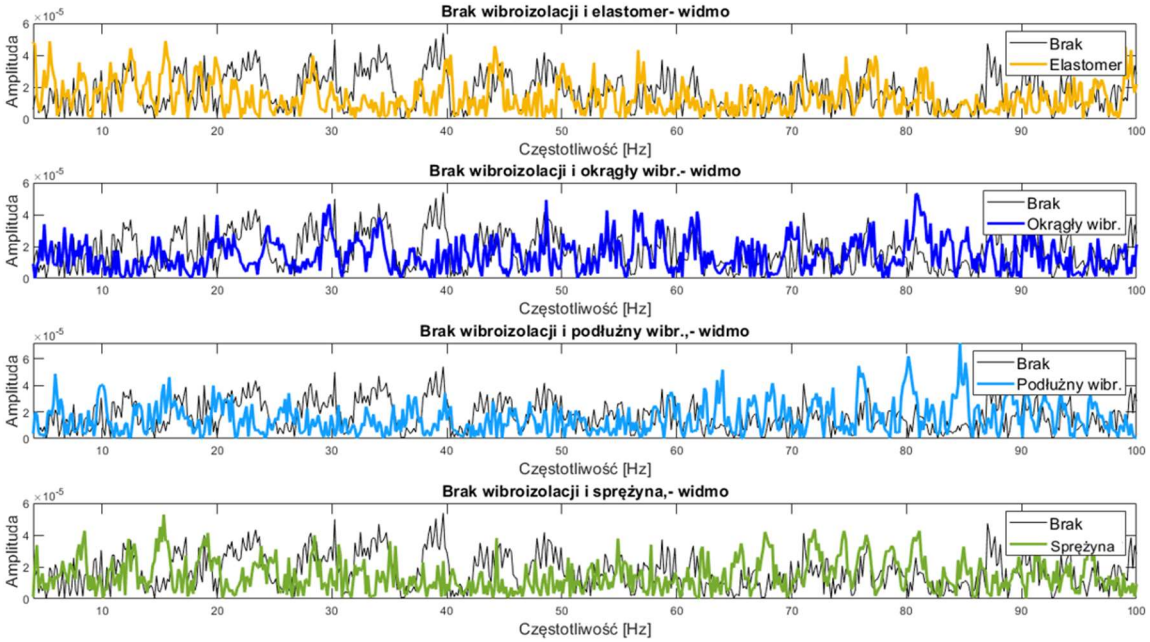


Rys. 3. Przebieg czasowy drgań zmierzonych na wkładce przy braku wibroizolacji i przy zastosowaniu elastomeru z wykorzystaniem stukacza młotkowego.



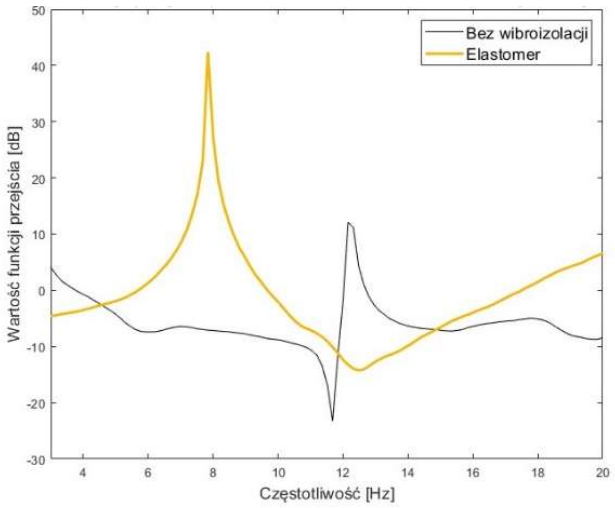
Rys. 4. Przebieg czasowy drgań zmierzonych na wkładce przy braku wibroizolacji i przy zastosowaniu elastomeru z wykorzystaniem młotka.

Podobny problem pojawił się przy pobudzeniu akustycznym, przy którym możliwe było jedynie względne ocenienie zmiany w wartościach amplitud drgań. Na rys. 5 przestawiono analizę widmową z czujnika na wkładce przy wykorzystaniu pobudzenia szumem różowym. Możliwe do zaobserwowania są jedynie zmniejszenia amplitud drgań w niektórych częstotliwościach, przy zastosowaniu wibroizolacji, jednakże niemożliwe jest jednoznaczne wyznaczenie częstotliwości rezonansowej. Natomiast przy pobudzeniu układu pracą własną gramofonu, zdecydowaną większością mierzonych drgań były te wynikające z odtwarzania płyty i obecności sygnału w rowku, stąd wykluczono te dane z dalszej analizy.



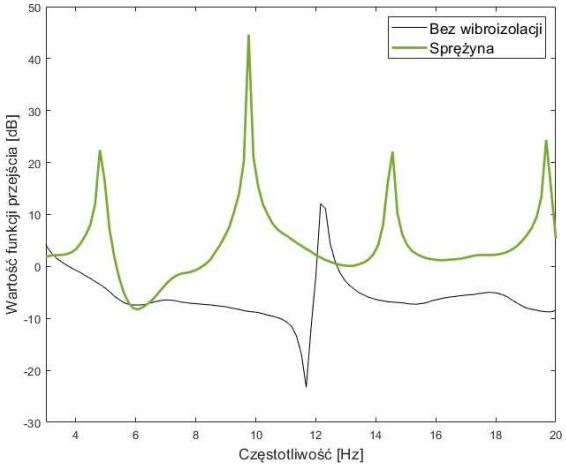
Rys. 5. Analiza widmowa na wkładce dla różnych rodzajów wibroizolacji, przy pobudzeniu akustycznym.

Z racji tego, iż niepożądane wstrząsy gramofonu najczęściej pochodzą z podłoża i drgań mebli, a także ze względu na otrzymywanie powtarzalnych i konkretnych pomiarów jedynie przy wykorzystaniu stukacza młotkowego, do dalszej analizy wykorzystano wyniki uzyskane dla tego pobudzenia. Na rys. 6 przestawiono funkcję przejścia między podłożem a wkładką przy zastosowaniu elastomeru i braku wibroizolacji. Częstotliwość rezonansowa układu zmniejszyła się o 4 Hz przy zastosowaniu wibroizolacji, co oznacza, iż zacznie ona skutecznie działać od niższych częstotliwości.

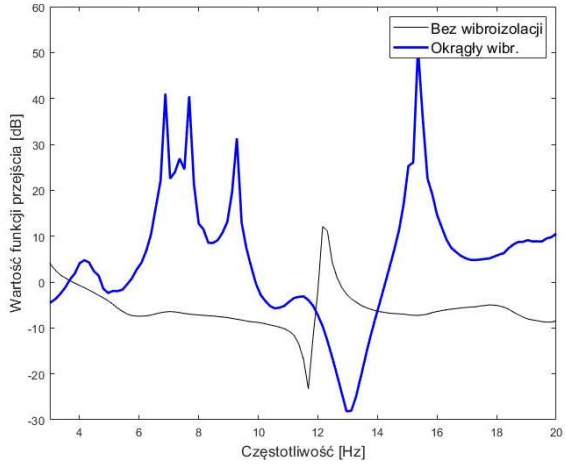


Rysunek 6. Funkcja przejścia podłoże – wkładka z wykorzystaniem stukacza młotkowego przy zastosowaniu elastomeru jako wibroizolatora.

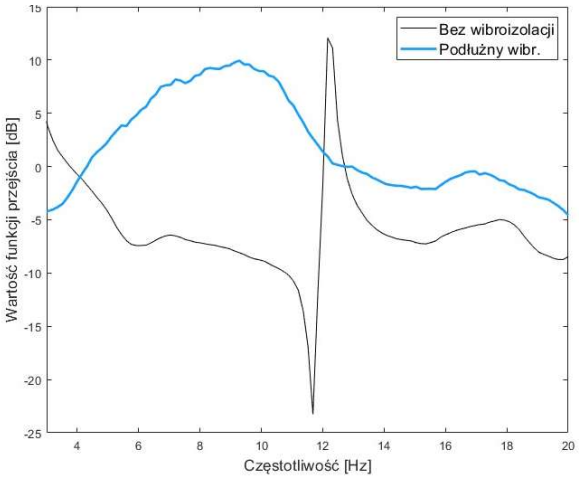
Funkcja przejścia między podłożem, a wkładką przy wykorzystaniu sprężyn oraz wibroizolatorów liniowych i trójosiowych została przedstawiona na rys. 7-9. Na wykresach zauważalne jest zjawisko wzmocnienia drgań w przypadku zastosowaniu wibroizolacji, a także brak wyraźnie zdefiniowanej częstotliwości rezonansowej. Taka charakterystyka sugeruje, że nie nastąpiła poprawa w zmniejszeniu przenoszenia drgań z podłoża na wkładkę.



Rys. 7. Funkcja przejścia podłoże – wkładka z wykorzystaniem stukacza młotkowego i sprężyny.

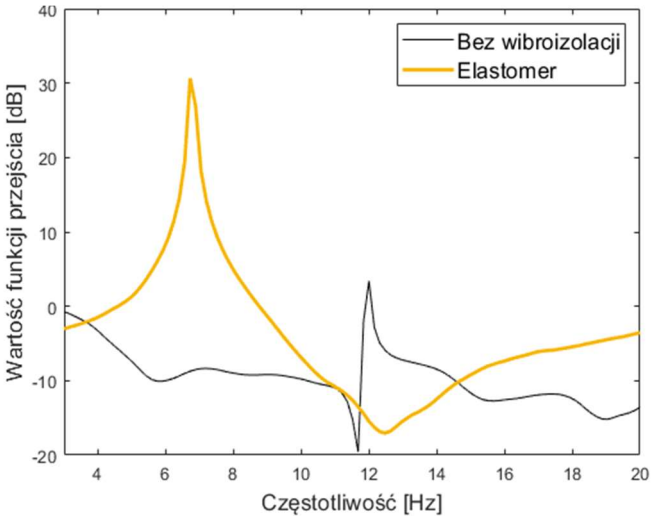


Rys. 8. Funkcja przejścia podłoże – wkładka z wykorzystaniem stukacza młotkowego i okrągłego wibroizolatora.

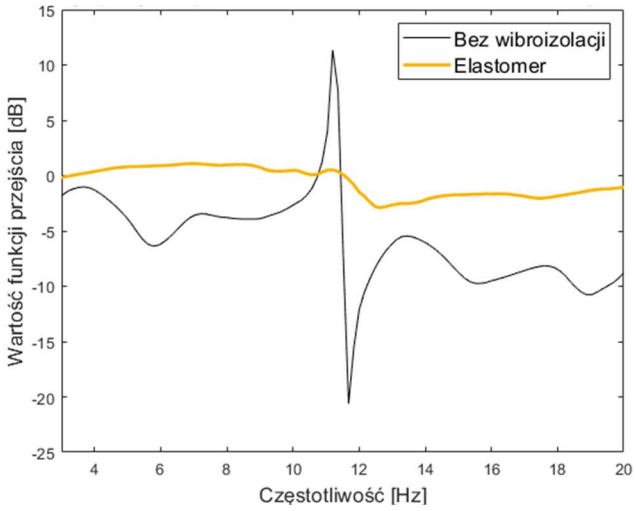


Rys. 9. Funkcja przejścia podłoże – wkładka z wykorzystaniem stukacza młotkowego i podłużnego wibroizolatora.

Funkcja przejścia między podłożem a ramieniem, przy wykorzystaniu stukacza młotkowego jako pobudzenia drgań, została przedstawiona na rys. 10. Ponownie zauważalne jest przesunięcie częstotliwości rezonansowej w kierunku niższych częstotliwości, co potwierdza skuteczność działania elastomeru w tłumieniu drgań. Na rys. 11 przedstawiono natomiast funkcję przejścia między podłożem, a obudową. W tym przypadku nie jest możliwe jednoznacznie wyznaczenie konkretnej częstotliwości rezonansowej układu. Funkcja przejścia ma w tym przypadku bardzo płaski przebieg, co świadczy o równomiernej reakcji układu na pobudzenie w niskich częstotliwościach.

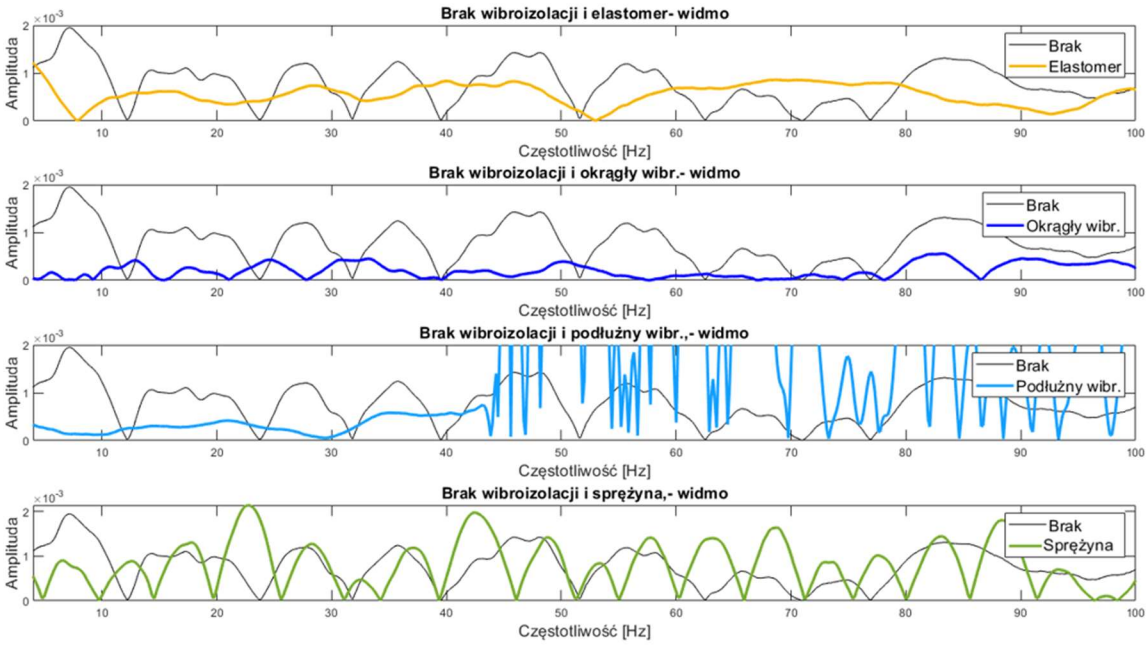


Rys. 10. Funkcja przejścia podłoże – ramię z wykorzystaniem stukacza młotkowego przy zastosowaniu elastomeru jako wibroizolatora.



Rys. 11. Funkcja przejścia podłoże – obudowa z wykorzystaniem stukacza młotkowego przy zastosowaniu elastomeru jako wibroizolatora.

Na rys. 12 przedstawiono analizę widmową z czujnika na wkładce dla różnych rodzajów wibroizolacji przy zastosowaniu stukacza młotkowego jako pobudzenia drgań. Analizując wykresy, można zauważyć redukcję drgań zarówno dla elastomeru, jak i wibroizolatorów w niskich częstotliwościach. Natomiast w przypadku sprężyny, w podanym zakresie, wystąpiło znaczące zwiększenie drgań. Na całym badanym paśmie częstotliwości najlepsze wyniki osiągnięto dla elastomeru oraz okrągłego wibroizolatora.



Rys. 12. Analiza widmowa na wkładce dla różnych rodzajów wibroizolacji, przy pobudzeniu stukaczem młotkowym.

5. Podsumowanie

Po przeanalizowaniu otrzymanych wyników wybrano stukacz młotkowy jako najskuteczniejszą metodę pobudzenia drgań i analizy skuteczności wibroizolacji gramofonu. Jego zastosowanie gwarantuje powtarzalność i precyzyjność pomiarów, co jest kluczowe dla wiarygodności wyników. Przy zastosowaniu młotka konieczne byłoby zadbanie o uwzględnienie pomiaru siły uderzenia, co umożliwiłoby dokładniejsze porównanie wyników z wibroizolacją i bez jej zastosowania. Bez wykorzystania czujnika impedancyjnego, pobudzenie drgań tą metodą jest równoznaczne z wykorzystywaniem stukacza młotkowego, ale bez zapewnionej powtarzalności uderzeń.

Najlepsze efekty w tłumieniu drgań w gramofonie uzyskano przy użyciu elastomeru, jednakże również przy zastosowaniu okrągłego wibroizolatora otrzymywano zadowalające wyniki w mniejszym zakresie częstotliwości. W przypadku wibroizolatora liniowego podłużnego i sprężyn w niektórych częstotliwościach następowało wzmocnienie drgań.

Wykonane pomiary były jedynie wstępną analizą postawionego problemu. W celu dalszych prac, należałoby wykonać pomiary, skupiając się na najlepszej z dotychczasowo wykorzystanych metod pobudzenia drgań, jaką jest zastosowanie stukacza młotkowego. Dodatkowo konieczne byłoby powtórzenie pomiarów z wykorzystaniem młotka impedancyjnego, równocześnie badając jego siłę uderzenia. Jako wibroizolację należałoby wykorzystać elastomery o różnych sztywnościach, jak i ich kombinację ze sprężynami i wibroizolatorami liniowymi.

Literatura

[1] Michalczyk J., Cieplok G., *Wysokoefektywne układy wibroizolacji i redukcji drgań*, 1999.

[2] Anderson C.R., *A Vibration-Stabilizer System for Phonograph Reproduction*, Journal of the Audio Engineering Society, vol. 27 (4), pp. 285-290, 1979.

[3] Barlow D.A., *The Performance of Pickups at Low Frequencies and Isolation from External Shock*, AES Convention, L-52, 1975.

[4] Happ L., Karlov F., *Record Warps and System Playback Performance*, Journal of the Audio Engineering Society, vol. 24 (8), pp. 630-638, 1976.

[5] Fujimoto Y., Suzuki M., Fujio K., Sasamoto K., Satoh Y., *A New Method of Reducing Direct-Drive Motor Vibration in Turntables*, Journal of the Audio Engineering Society, vol. 31 (4), pp. 246-252, 1983.

Emilia **STEFANOWSKA**[1], Janina **KRZYSIAK**[1], Adam **PILCH**[2]

GENEROWANIE TRÓJWYMIAROWYCH STRUKTUR AKUSTYCZNYCH Z WYKORZYSTANIEM SIECI NEURONOWYCH

METAMATERIALS AND ARTIFICIAL INTELLIGENCE – GENERATING 3D ACOUSTIC STRUCTURES USING NEURAL NETWORKS

[1] Hitachi Energy Research, Kraków

[2] Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

emilia.stefanowska@hitachienergy.com

Streszczenie

Metamateriały w dziedzinie akustyki mają ogromny potencjał w szerokim zakresie zastosowań, od maskowania akustycznego do specjalistycznej redukcji hałasu. Ich projektowanie często wymaga dużej znajomości fizyki dźwięku oraz technik symulacji komputerowej. Czy nowoczesne algorytmy i techniki sztucznej inteligencji mogą pozwolić na zautomatyzowanie i uproszczenie tego procesu? W tej pracy eksplorowano możliwości wykorzystania technik uczenia maszynowego do tworzenia trójwymiarowych struktur o zadanych właściwościach akustycznych. Przedstawiono innowacyjny sposób przygotowania geometrii, specjalistycznego zbioru treningowego oraz zaproponowano modele sieci neuronowych, które pozwalają na wygenerowanie odpowiednich struktur 3D. Opisano nowoczesne architektury takie jak warunkowe sieci autoenkodujące (cVAE) i warunkowe generujące sieci współzawodniczące (cGAN), wyłaniając najbardziej obiecujące modele do projektowania metamateriałów.

1. Wprowadzenie

W ostatnich latach można zaobserwować rosnące zainteresowanie metamateriałami akustycznymi [1, 2]. Ich właściwości manipulowania falami dźwiękowymi są wysoce zależne od struktury geometrycznej, która pozwala na osiągnięcie rezultatów nieobecnych w substancjach naturalnych [3].

Projektowanie metamateriałów wymaga jednak specjalistycznej wiedzy, a tym samym dużego nakładu czasu i kosztów. Jednym z rozwiązań tego problemu wydaje się być wykorzystanie algorytmów sztucznej inteligencji (SI) do generowania nowych struktur [3, 4]. Przegląd literatury dotyczącej uczenia maszynowego w metamateriałach ujawnia rosnące znaczenie SI w projektowaniu i optymalizacji struktur metamateriałowych. Obserwujemy

wykorzystanie zaawansowanych sieci neuronowowych zarówno do projektowania, jak i do dostosowywania właściwości metamateriałów pod zadane wymagania [3, 5].

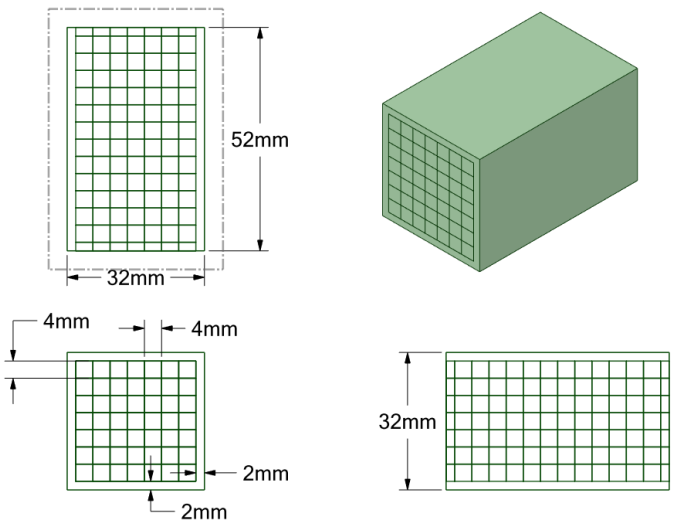
W tej pracy zaprezentowano rozwiązanie, które umożliwia projektowanie trójwymiarowych metamateriałów akustycznych korzystając z sieci neuronowych. Przedstawiono geometrię, która pozwala na innowacyjność tworzenia nowych struktur oraz zaproponowano architektury generatywnych sieci neuronowych do tworzenia nowych metamateriałów. Dodatkowo na podstawie wstępnych badań przedstawiono, który z algorytmów ma największy potencjał w tworzeniu poprawnych geometrii.

2. Przygotowanie treningowego zbioru danych

W ramach projektu przygotowano zbiór treningowy, zawierający geometrie i wyniki symulacji numeryczych struktur rezonansowych. W tej sekcji opisano sposób jego przygotowania i analizy.

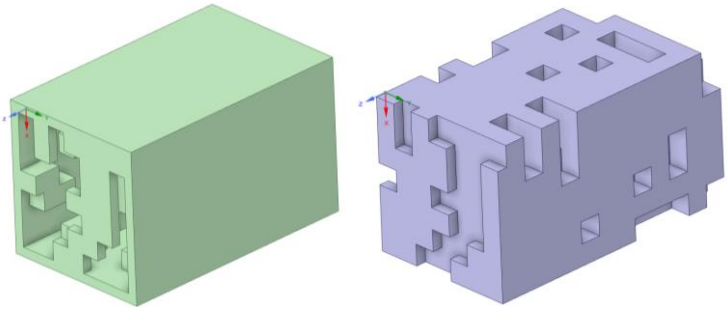
2.1 Geometria struktury rezonansowej

Pierwsza z proponowanych geometrii została przedstawiona na rysunku 2.1. Struktura ta została zainspirowana dwuwymiarowymi metamateriałami labiryntowymi opisanymi w [6]. Model składa się z 686 elementów jednostkowych umieszczonych w 14 warstwach. Każda warstwa jest ułożona w 7 kolumnach i 7 rzędach. Pierwszy i ostatni rząd są wykonane z elementów jednostkowych o wymiarach 2 mm x 4 mm x 4 mm. Pozostałych 12 rzędów składa się z elementów o wymiarach 4 mm x 4 mm x 4 mm. Utworzono obudowę o grubości 2 mm, aby zapewnić jednolite zakrycie czterech najdłuższych boków struktury.



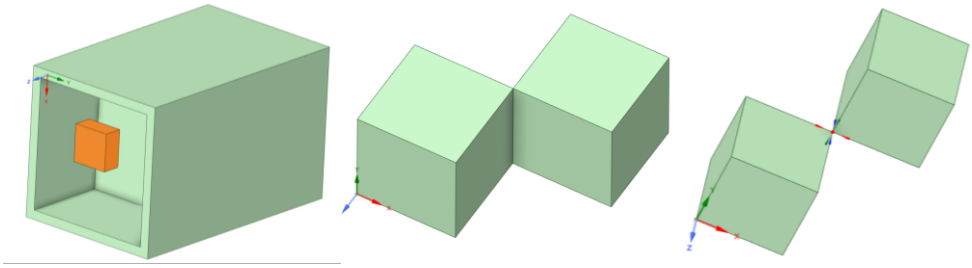
Rysunek 2.1. Podstawowa geometria struktury rezonansowej

Tworzenie każdej nowej struktury przbiegało w dwóch krokach. Pierwszym z nich było usunięcie wybranych elementów ze struktury, a drugim połączenie pozostałych sześcianów w jednolitą strukturę połączoną z obudową. W ramach tej części projektu przygotowano algorytm, który pozwalał na zautomatyzowanie procesu wybierania usuwanych elementów. Przykład stworzonej geometrii przedstawiono na rysunku 2.2.



Rysunek 2.2. Struktura wygenerowana przez pierwszy algorytm. Po lewej stronie przedstawiono geometrię rezonatora, a po prawej objętość przestrzeni powietrznej w jej wnętrzu.

Ze względu na ograniczenia oprogramowania Space Claim ANSYS, w którym wykonywano cyfrowe modele rezonatorów, algorytm przygotowujący kombinacje elementów jednostkowych został przygotowany w taki sposób, aby niedopuścić do utworzenia tzw. geometrii nieprawidłowych. Geometriami nieprawidłowymi nazywano modele, które zawierały elementy jednostkowe nieutwierdzone – czyli takie, które nie miały żadnego fizycznego połączenia z obudową, i/lub elementy jednostkowe styczne jedynie krawędziami lub wierzchołkami.



Rysunek 2.3. Fragmenty nieprawidłowych geometrii. Od lewej – nieutwierdzone elementy jednostkowe, elementy jednostkowe styczne krawędziami, elementy jednostkowe styczne wierzchołkami.

2.2 Symulacje numeryczne ANSYS

Stworzono skrypt w programie Space Claim, który pozwalał na automatyczne tworzenie nowych geometrii i zautomatyzowanie procesu przeprowadzania symulacji w programie ANSYS Workbench. Warunki brzegowe symulacji naśladowały przeprowadzanie eksperymentu w rurze impedancyjnej, wykorzystywanej w podobnych badaniach [7, 8]. Symulacje zostały przeprowadzone w zakresie częstotliwości 50 Hz – 3000 Hz

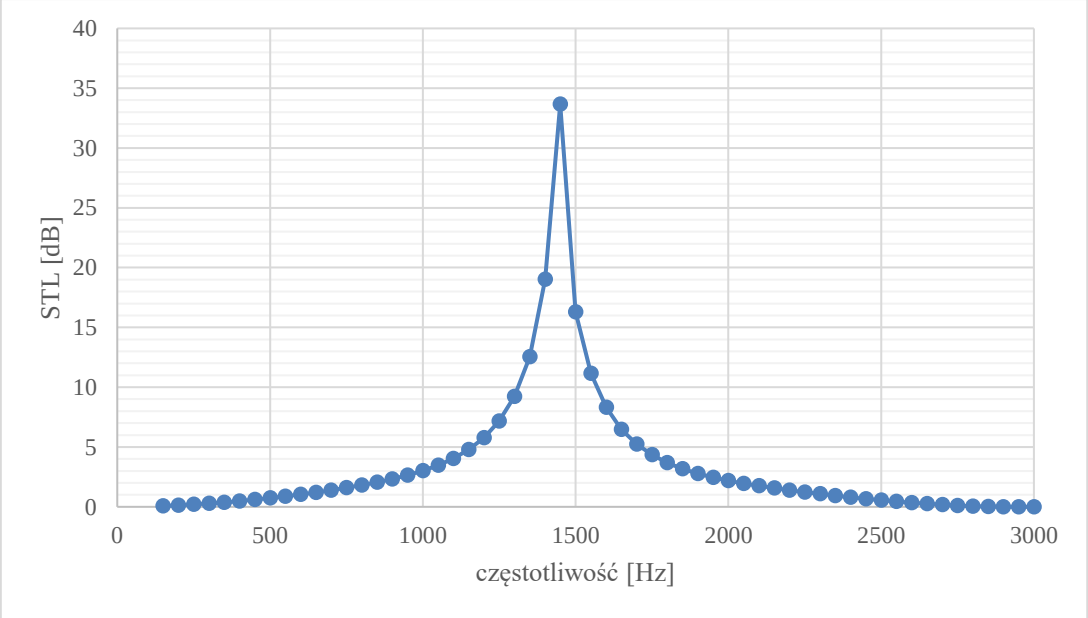
z rozdzielczością 50 Hz. Parametrem akustycznym analizowanym była strata transmisji dźwięku (ang. *Sound Transmission Loss*, STL), opisana zgodnie z dokumentacją programu ANSYS 2022R wzorem:

$$STL = 10 \log \left| \frac{W_i}{W_t} \right|, \qquad [\text{dB}] \tag{2.2.1}$$

gdzie

- W_i to moc fali padającej,
- W_t to moc fali transmitowanej.

Modele przygotowane w ramach tej części projektu wykazywały najczęściej pojedynczy szczyt charakterystyki STL. Wynikało to z niewielkiej liczby elementów bazowych i konieczności usuwania wielu z nich w celu uniknięcia nieprawidłowych geometrii. Przykładowy wykres charakterystyki STL przedstawiono na rysunku 2.4.



Rysunek 2.4. Przykładowa charakterystyka STL geometrii rezonansowej

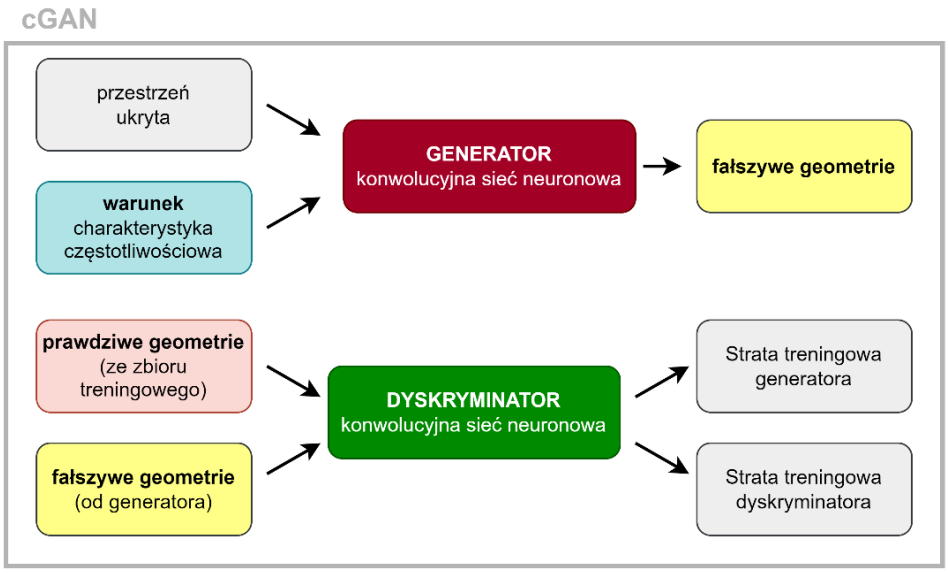
3. Modele sieci neuronowych

Stworzono i przetestowano wiele architektur sieci neuronowych, które mogłyby pozwolić na tworzenie nowych struktur rezonansowych w oparciu o pożądane właściwości akustyczne. Głównej analizie poddano architektury warunkowych generatywnych sieci współzawodniczących oraz warunkowych wariacyjnych sieci autokodujących. Proponowane architektury sieci neuronowych testowano na zbiorze treningowym złożym z 1500 modeli geometrycznych i korespondujących z nimi wyników STL uzyskiwanych w sposób opisany

w sekcji 2. Należy zaznaczyć, że opisane algorytmy należy traktować jako propozycje, które mogą zostać rozwinięte w kolejnych etapach projektu.

3.1 Generujące sieci współzawodniczące

Generujące sieci współzawodniczące (ang. *Generative Adversarial Networks*, GAN) to potężna architektura sieci neuronowych wykorzystywana w zadaniach generatywnych. Jej główna idea polega na stworzeniu dwóch sieci - generatora i dyskryminatora. Celem dyskryminatora jest zadecydowanie, czy geometrie proponowanych struktur akustycznych są prawdziwe (pochodzące ze zbioru danych treningowych) czy fałszywe (stworzone przez generator). Obie sieci są trenowane jednocześnie, tak aby dyskryminator jak najlepiej nauczył się rozróżniać geometrie prawdziwe od fałszywych, a generator tworzyć geometrie, które będą jak najbardziej imitować te prawdziwe [9]. W ramach tego projektu testowano warunkowe GAN, czyli cGAN. Zastosowanie warunku w tym projekcie polegało na wskazaniu, które częstotliwości powinny być tłumione przez model najlepiej. Docelowo sieć powinna generować geometrie metmateriału o zadanej przez użytkownika charakterystyce częstotliwościowej. W ramach tego projektu, cGAN miał za zadanie projektować nowe geometrie rezonansowe w oparciu o częstotliwość dla której zaobserwowano najwyższe wartości STL. Ogólna architektura modelu cGAN została przedstawiona na rysunku 3.1.1.



Rysunek 3.1.1. Schemat architektury sieci cGAN

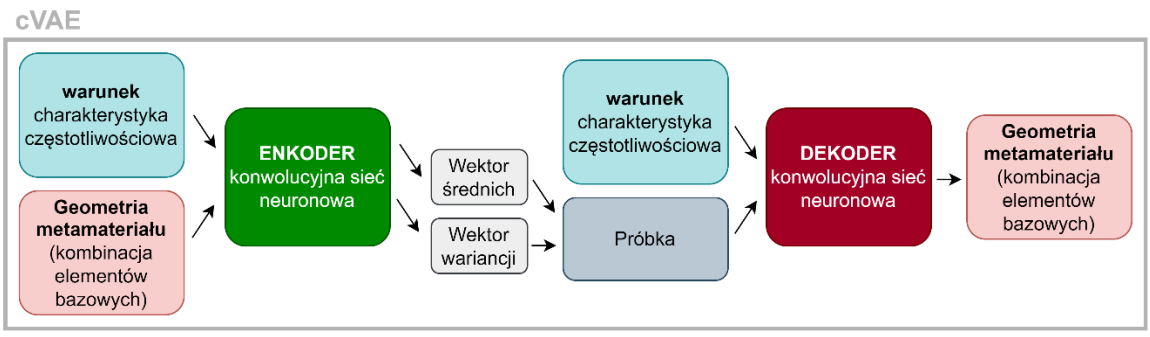
Podczas przygotowywania sieci napotkano na powszechny problem w sieciach GAN, nazywany *mode collapse*. Sieć generowała tylko jedną geometrię rezonansową, igorując zadawanie różnych warunków wejściowych. Podjęto próby zminimalizowania negatywnych działań tego mechanizmu, poprzez zastąpienie funkcji na stratę *Wasserstein*, zamianę

sigmoidalnych funkcji aktywacji na liniowe, dodanie ograniczeń wag sieci, zastąpienie optymalizatora *Adam* przez *RMSprop Stochastic Gradient Descent* oraz aktualizowanie sieci dyskryminatora częściej niż generatora [9, 10]. Po wprowadzeniu zmian uzyskano architekturę, która pozwalała na generowanie różnorodnych geometrii.

Testowane architektury cGAN wykorzystywały warstwy gęsto połączone, warstwy konwolucyjne dwuwymiarowe oraz trójwymiarowe, warstwy regularyzacyjne. Podjęto próbę dostrojenia hiperparametrów sieci, m.in. tempo treningu, liczbę epok, liczbę próbek podawanych jednocześnie, typ i liczbę warstw oraz ich komórek. Niestety mimo prób, nie uzyskano zadowalających wyników. Proponowane przez sieć rezonatory posiadały wiele nieprawidłowych geometrii (patrz sekcja 2.1). Dodatkowo, weryfikacja nowo wygenerowanych struktur metamateriałowych przeprowadzona za pomocą symulacji numerycznych, wskazywała na niewielką dokładność jeżeli chodzi o dopasowanie do docelowej częstotliwości rezonansowej. Brak dobrych efektów może wskazywać, że architektura cGAN jest zbyt zaawansowana na tak mały zbiór treningowy, który został wykorzystany do wstępnych testów.

3.2 Wariacyjne sieci autokodujące

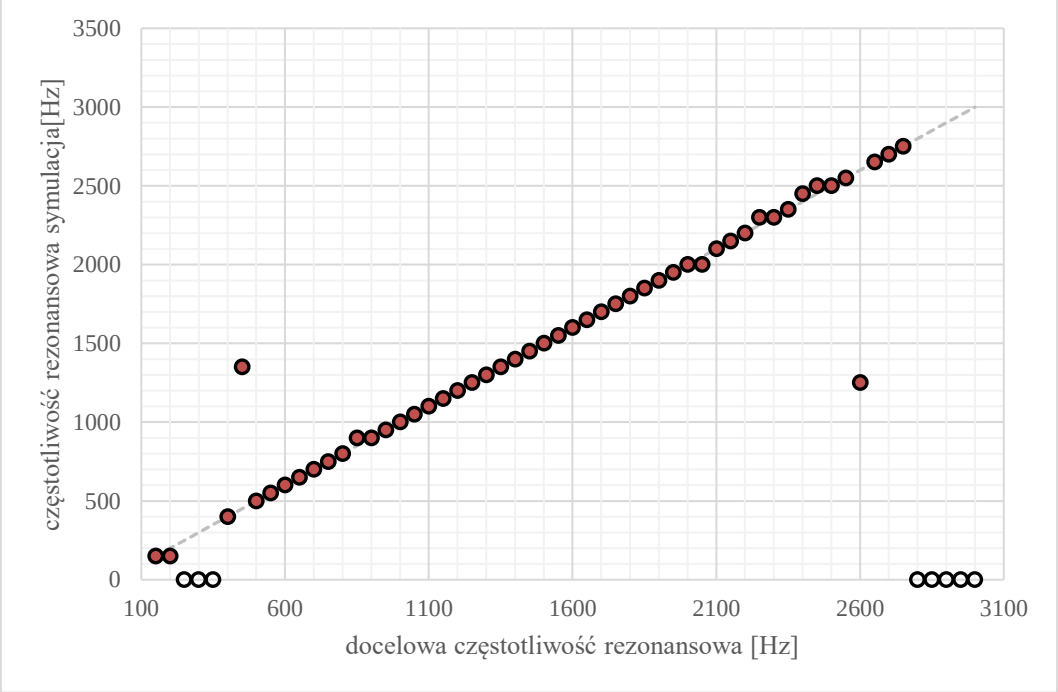
Ze względu na niepowodzenie z architekturą cGAN, zadecydowano na przetestowanie warunkowych wariacyjnych sieci autokodujących (ang. *conditional Variational AutoEncoder*, cVAE). Jest to kolejna popularna architektura sieci neuronowych wykorzystywana do zadań generatywnych. Podobnie jak w przypadku cGAN, cVAE składa się z dwóch sieci – kodera i dekodera. Koder ma za zadanie zmniejszyć wymiarowość danych, a enkoder jak najlepiej odzworować geometrię wejściową na podstawie tych zakodowanych informacji [9]. Ostatecznie do generowania nowych struktur wykorzystywany jest dekoder, któremu zadaje się losowy wektor inicjalizujący oraz pożądaną charakterystykę częstotliwościową metamateriału. Ogólna architektura tej sieci została przedstawiona schematycznie na rysunku 3.2.1.



Rysunek 3.2.1. Schemat architektury sieci cVAE

W ramach projektu testowano architektury cVAE, które generowały geometrie rezonatorów w oparciu o pojedynczą częstotliwość, dla której obserwowano największą wartość STL. Częstotliwości docelowe zostały przekazane jako warunkowe dane wejściowe w postaci zakodowanych wektorów. Po dostrojeniu hiperparametrów i architektury sieci, otrzymano model, który był w stanie wygenerować kombinacje elementów bazowych, które spełniały warunki geometryczne, a także posiadały największe wartości ΔL w pożądanych częstotliwościach docelowych. Enkoder składał się z ułożonych naprzemiennie warstw konwolucyjnych i gęsto połączonych, o całkowitej liczbie ok. 120 tysięcy parametrów. Dekoder wykorzystywał transponowane warstwy konwolucyjne oraz warstwy gęsto połączone i składał się z ponad 815 tysięcy parametrów. Model trenowano przez 3000 epok przy podawaniu 128 próbek jednocześnie. Analiza krzywych uczenia wskazywała na brak przetrenowania modelu.

W celu weryfikacji użyteczności przygotowanej sieci neuronowej, podjęto próbę wygenerowania 58 geometrii rezonatorów, po jednej dla każdej z docelowych częstotliwości rezonansowych między 150 Hz a 3000 Hz. Następnie w przeprowadzono symulację numeryczną każdej z wygenerowanych geometrii. Porównano wartości docelowej częstotliwości rezonansowej (warunek dla sieci) do częstotliwości rezonansowej zweryfikowanej poprzez symulację. Wyniki przedstawiono na rysunku 3.2.2.



Rysunek 3.2.2. Wyniki symulacji modeli stworzonych przez cVAE

Na 58 testowanych modeli, sieć wygenerowała 42 w pełni poprawne geometrie o odpowiedniej częstotliwości rezonansowej. Dla 6 geometri, częstotliwość rezonansowa struktury była przesunięta względem docelowej o 50 Hz. Model cVAE nie wygenerował poprawnych geometrii dla 8 modeli (oznaczono pustymi znacznikami na rysunku 3.2.2), a tylko 2 z 58 testowanych modeli nie było rezonatorami o zadanej częstotliwości tłumienia. Głębsza analiza wskazała jednak, że pierwszy z modeli, z częstotliwością docelową 450 Hz, wykazywał tłumienie dla czędstotliwości 1350 Hz, co odpowiada drugiej harmonicznej docelowej częstotliwości. Dla drugiego z błędnych modeli, który miał częstotliwością docelową równą 2600 Hz, wygenerowana geometria posiadała charakterystykę częstotliwościową z dwoma lokalnymi szczytami wartości STL – dla 1350 Hz oraz 2850 Hz. Analiza tych wyników wskazuje, że błędy mogły nie być w pełni przypadkowe, a wykorzystana sieć neuronowa po udoskonaleniu może mieć potencjał wychwytywania zjawisk fizycznych.

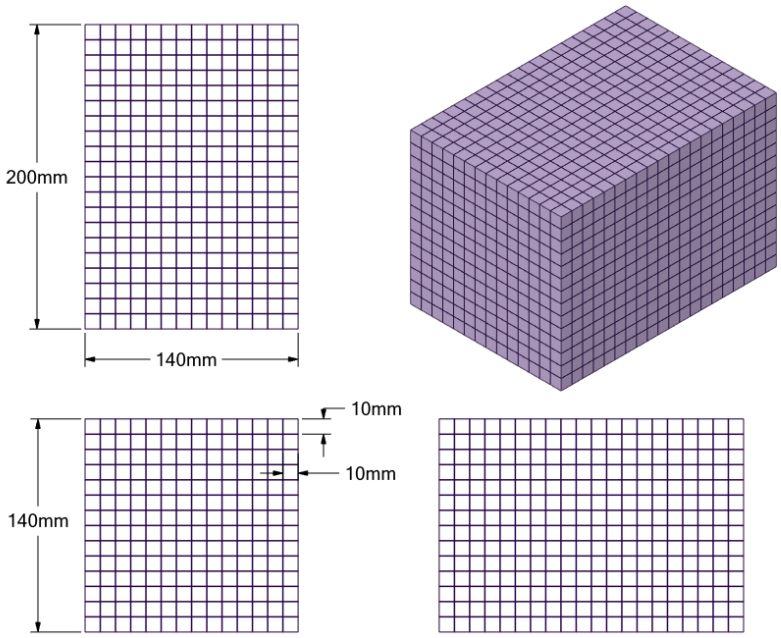
Należy zaznaczyć, że przeprowadzone testy należy traktować jedynie jako wstępne wskazówki do dalszych badań. Nie przeprowadzono usystematyzowanych badań na dużej próbie testowej, które pozwoliłyby na jednoznaczne wskazanie najlepszej architektury. Niemniej warto wskazać, że potencjał obu sieci został przetestowany, a obecne wyniki wskazują na obiecujący rozwój badań. Korzystając z większego zbioru treningowego i odpowiednio zaprojektowanych sieci neuronowych możliwe byłoby stworzenie modelu, który pozwoliłby na automatyczne projektowanie metamateriałów o zadanych właściwościach akustycznych.

4. Udoskonalenie badań i analiz

Przeprowadzone badania wykazywały potencjał przyjętego rozwiązania. Postanowiono udoskonalić geometrię i analizę akustyczną przygotowanego modelu, tak aby uzyskać bardziej zróżnicowane wyniki. W tej sekcji opisany jest nowo przyjęty model geometryczny, który będzie wykorzystywany w kolejnych etapach projektu.

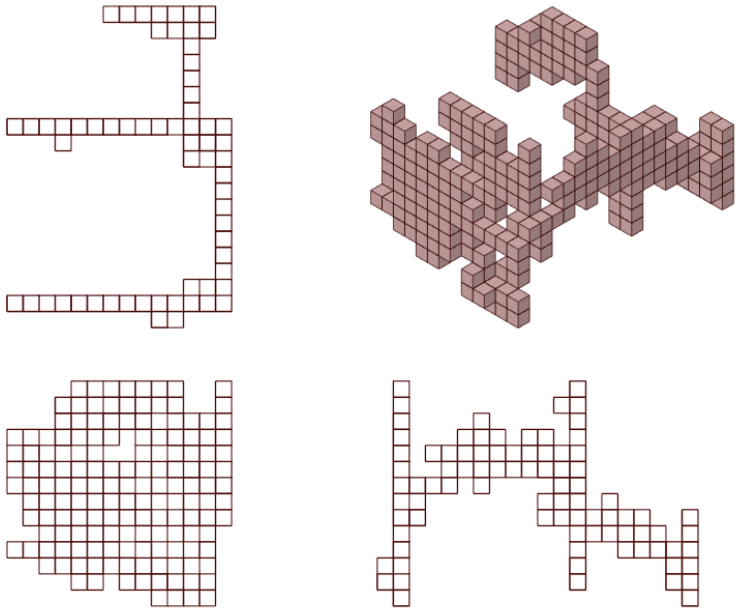
4.1 Nowy model geometryczny

Nowy model bazowej geometrii metamateriałów został przedstawiony na rysunku 4.1. Model ogólny zaproponowanej geometrii składa się z 3920 elementów bazowych, czyli sześcianów o wymiarach 10 mm x 10 mm x 10 mm. Elementy bazowe rozmieszczone są w 20 warstwach, składających się z 14 kolumn i 14 rzędów.



Rysunek 4.1.Geometria bazowa tworzonych metamateriałów

Tworzenie nowej struktury metamateriału rozpoczyna się od przeprowadzenia wieloetapowego procesu decyzyjnego, wybierającego które elementy bazowe mają być obecne, a które nie. Stworzone geometrie posiadają różnorodne, skomplikowane korytarze, które pozwalają na przepływ powietrza przez całą długość struktury. Duża liczba elementów bazowych pozwala na powstawanie bardzo zróżnicowanych przestrzeni rezonansowych, obecnych wewnątrz metamateriału. Przykład objętości przestrzeni powietrznej dla jednej ze stworzonych geometrii został przedstawiony na rysunku 4.2.



Rysunek 4.2. Przestrzeń powietrzna wewnątrz przykładowej struktury metamateriału

Nowy algorytm, który napisano w celu projektowania nowych metamateriałów, został oparty na algorytmie przeszukiwania wszerz (ang. *breadth-first search algorithm*, BFS). Proces tworzenia nowych geometrii można opisać jako rekurencyjny łańcuch decyzji, tworzący ścieżki przestrzeni powietrznych w strukturze. Oprócz dużej losowości w usuwaniu kolejnych elementów bazowych, zastosowano szereg warunków, które pozwalały na utworzenie różnorodnych, ale jednak posiadających wspólne cechy geometrii. W ramach opisanego w tej pracy modelu, przyjęto dwa warunki dla nowych struktur.

Pierwszy z warunków zakładał, że objętość przestrzeni powietrznej w ostatecznej geometrii nie powinna przekraczać 15% objętości całej struktury. Warunek ten wynikał z ograniczonych zasobów czasowych, które można było przeznaczyć dla symulacji numerycznej pojedynczego modelu. Warunki tej symulacji, szerzej opisane w sekcji 4.2, zakładały przeprowadzenie obliczeń dla przestrzeni powietrznych wewnątrz struktury. Większa objętość oznaczała zatem więcej punktów dla których wykonywano obliczenia, co prowadziło do długiego czasu symulacji.

Drugi warunek geometryczny określał, że powstająca przestrzeń powietrzna powinna pozwalać na swobodny przepływ powietrza przez całą długość struktury. Wynikało to z założeń projektowych, które miał spełniać proponowany metamateriał.

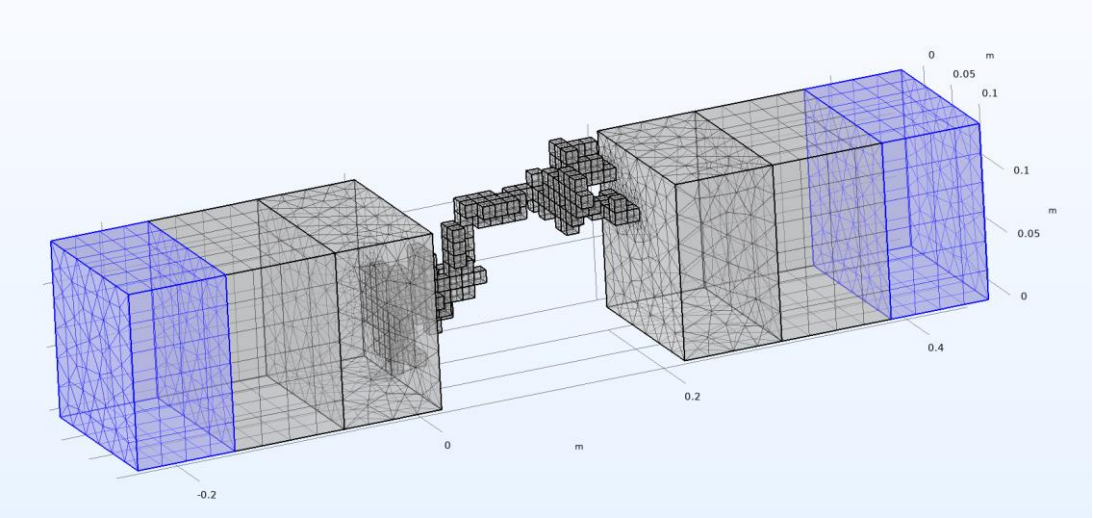
Należy zaznaczyć, że odpowiednie przygotowanie sieci neuronowej wymaga ogromnych zbiorów danych. Z tego względu, opisywany algorytm bazował w dużej mierze na losowości tworzenia nowych przestrzeni powietrznych wewnątrz metamateriałów. Pozwalało to na ogromną różnorodność powstających modeli, a tym samym bardzo zróżnicowane właściwości tłumienia hałasu poszczególnych struktur.

4.2 Symulacje numeryczne COMSOL

Dla każdej nowo stworzonej struktury przeprowadzano symulację numeryczną w programie COMSOL Multiphysics. Model obejmował analizę akustyczną dla przestrzeni powietrznych wewnątrz geometrii. Skorzystano z podejścia, w którym imitowano badanie metamateriału z wykorzystaniem rury impedancyjnej. Model zakładał wymuszenie falą płaską o zadanej częstotliwości i poziomie ciśnienia akustycznego równym $L_{in} = 91$ dB. Badano w jaki sposób geometria przestrzeni powietrznych struktury wpływa na poziom ciśnienia akustycznego po drugiej stronie rury L_{out}. Za parametr akustyczny ostatecznie opisujący model przyjęto $\Delta L = L_{in} - L_{out}$, czyli różnicę między poziomem ciśnienia akustycznego wymuszenia, a uśrednionionym na powierzchni poziomem ciśnienia akustycznego za strukturą

rezonansową. Badanie przeprowadzono w zakresie częstotliwości 250 Hz – 2500 Hz z krokiem 1/6 oktawy.

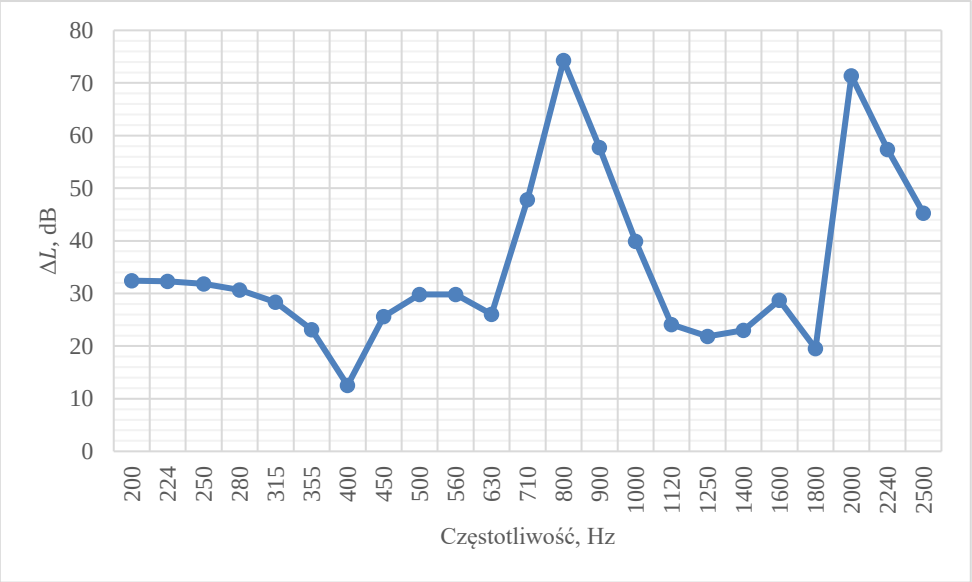
Warunki brzegowe modelu zakładały, że wszystkie ściany zewnętrzne w modelu są powierzchniami idealnie odbijającymi fale akustyczne. Wyjatkiem były skrajne fragmenty geometrii rury, które implementowały warunek idealnego pochłaniania fal akustycznych (ang. *perfectly matched layer*, PML). Takie rozwiązanie pozwoliło na zapewnienie, że wymuszona fala akustyczna przejdzie przez badaną strukturę tylko raz. W procesie tworzenia siatki modelu przyjęto zasadę, że największy wymiar siatki nie może przekraczać 1/6 najmniejszej długości fali. W symulacji wykorzystano również moduł termowiskotyczny, który zastosowano przy powierzchniach powietrznych wewnątrz struktury rezonansowej. Rysunek 4.2.1 przedstawia przykładową geometrię sumulacji numerycznej z przygotowaną siatką obliczeniową. Kolorem niebieskim zostały zaznaczone obszary PML.



Rysunek 4.2.1 Geometria przeprowadzanych symulacji z siatką

Przygotowano skrypt wykorzystujący moduł LiveLink for MATLAB, który pozwalał na zautomatyzowanie procesu symulacji. Kombinacje elementów bazowych były odczytywane przez model, a następnie na ich podstawie budowano geometrie w programie COMSOL po czym przeprowadzano symulację numeryczną. Zapisywane wyniki mogły być wykorzystywane i przetwarzane w dalszych etapach projektu.

Przykładowe wyniki przedstawiono na rysunku 4.2.2. Ogromna różnorodność generowanych przestrzeni pozwalała na tworzenie geometrii o wielu częstotliwościach rezonansowych. Tworzone modele mogłyby być zatem wykorzystywane do przeróżnych zadań tłumienia hałasu. Dodatkowo należy zaznaczyć, że zbiór treningowy który obejmuje tak duże zróżnicowanie, powinien być dobrym zbiorem treningowym dla przyszłych sieci neuronowych.



Rysunek 4.2.2 Przykładowy wykres różnicy poziomów ciśnienia akustycznego dla modelu treningowego

5. Podsumowanie i wnioski

W ramach pracy omówiono potencjał wykorzystania algorytmów sztucznej inteligencji, zwłaszcza sieci neruonowych, w projektowaniu metamateriałów akustycznych. Przedstawiono innowacyjny sposób przygotowania nowych struktur, umożliwiający automatyczne generowanie zróżnicowanych geometrii. Zaprezentowano modele symulacji numerycznych w programach ANSYS oraz COMSOL, które imitowały badania w rurze impedancyjnej. Korzystając z opisanych narzędzi stworzono zbiór danych złożony z 1500 geometrii metamateriałów oraz symulacyjnych wyników STL. Przygotowany zbiór pozwolił na ewaluację kilku architektur sieci neuronowych, generujących metamateriały o wybranych właściwościach akustycznych.

Przeprowadzone badania wskazują, że wykorzystanie sztucznej inteligencji może znacząco przyspieszyć i ułatwić proces projektowania metamateriałów akustycznych. Wykorzystanie odpowiednio dużej bazy danych pozwala na wytrenowanie popularnych architektur generatywnych sieci neuronowych. Mimo dużych ograniczeń w zaproponowanych testach, wykazano wstępnie że sieć cVAE pozwala na automatyzację projektowania metamateriałów dla zadanej częstotliwości tłumienia dźwięku. Uzyskane wyniki wskazują na potrzebę udoskonalenia bazy danych wykorzystywanej w treningu algorytmów. Analiza zaproponowanego procesu pozwoliła na identyfikację potencjału badań, ale również obszarów do poprawy w dalszym rozwoju projektu.

Literatura

[1] R. Wu, T. Liu i M. Jahanshahi, *Design of one-dimensional acoustic metamaterials using machine learning and cell concatenation*, Struct Multidisc Optim, tom 63, pp. 2399-2423, 2021.

[2] L. Xiang, N. Shaowu, L. Zhanli, Y. Ziming, L. Chengcheng i Z. Zhuo, *Designing phononic crystal with anticipated band gap through a deep learning based data-driven method*, Computer Methods in Applied Mechanics and Engineering, tom 361, 2020.

[3] Muhammad, J. Kennedy i C. Lim, *Machine learning and deep learning in phononic crystals and metamaterials - A review*, Materials Today Communications, tom 33, 2022.

[4] K. Mahesh, K. S. Ranjith i R. S. Mini, *Inverse design of a Helmholtz resonator based low-frequency acoustic absorber using deep neural network*, Journal of Applied Physics`, tom 129, p. 174901, 2021.

[5] N. Gao, M. Wang, B. Cheng i H. Hou, *Inverse design and experimental verification of an acoustic sink based on machine learning*, Applied Acoustics, tom 180, p. 108153, 2021.

[6] K. Donda, Y. Zhu, A. Merkel, F. Shi-Wang, L. Cao, S. Wan i B. Assouar, *Ultrathin acoustic absorbing metasurface based on deep learning approach*, Smart Materials and Structures, tom 30, nr 8, 2021.

[7] Herrero-Durá, A. Cebrecos, R. Picó, V. Romero-García, L. M. García-Raffi i V. J. Sánchez-Morcillo, *Sound Absorption and Diffusion by 2D Arrays of Helmholtz Resonators*, José , tom 10, nr 5, 2020.

[8] Y. Fei, W. Enshuai, X. Shen, Z. Xiaonan, Y. Qin, W. Xinqing, Y. Xiaocui, C. Shen i P. Wenqiang, *Optimal Design of Acoustic Metamaterial of Multiple Parallel Hexagonal Helmholtz Resonators by Combination of Finite Element Simulation and Cuckoo Search Algorithm*, Materials , tom 15, nr 18, 2022.

[9] R. Atienza, Advanced Deep Learning with TensorFlow 2 and Keras - Second Edition, Packt, 2020.

[10] „Google Developers - GANs Common Problems,” Google, 08 07 2022. [Online]. Available: https://developers.google.com/machine-learning/gan/problems. [Data uzyskania dostępu: 25 09 2023].

Julia **SZYMLA**[1], Karolina **PONDEL-SYCZ**[1]

BADANIA SYSTEMÓW ARM DLA POLSKIEJ MOWY O OBNIŻONEJ JAKOŚCI ORAZ WPŁYWU METOD NAPRAWCZYCH JAKOŚCI MOWY

RESEARCH ON ARM SYSTEMS FOR POLISH SPEECH OF REDUCED QUALITY AND THE IMPACT OF SPEECH QUALITY REPAIR METHODS

[1] Koło Naukowe Elektroakustyki Politechniki Warszawskiej

julia.szymla.stud@pw.edu.pl

Streszczenie

Zagadnienie automatycznego rozpoznawania mowy (ARM) dynamicznie rozwinęło się na przestrzeni ostatnich lat dzięki zastosowaniu podejścia End-to-end (E2E), które wykorzystuje techniki uczenia głębokiego. Specyfiką tej dziedziny jest mnogość języków, do których trzeba to rozwiązanie dostosować. W badaniu wybrano dwa modele E2E ARM dostosowane do pracy z językiem polskim – wielojęzykowy model Zoo z narzędzia ESPnet (dalej określany jako „ESPnet", o architekturze Speech-Transformer) oraz model z narzędzia NVIDIA NeMo zaadaptowany do rozpoznawania języka polskiego, FastConformer (o architekturze Conformer). Modele były trenowane na korpusach zawierających język polski, takich jak: Multilingual Librispeech, Mozilla Common Voice oraz VoxPopuli. Przeprowadzono weryfikację poprawności działania tych modeli za pomocą bazy Mobile Corpus (EMU) powstałej w ramach projektu Clarin PL. Próbki dźwiękowe z korpusu Mobile Corpus (EMU) nagrano podczas rozmowy telefonicznej, co wpływa na obniżenie jakości dźwięku poprzez różne zniekształcenia sygnału. W pracy podjęto próby poprawy jakości próbek, a następnie przeprowadzono ponowną weryfikację efektywności rozpoznawania mowy dla obu modeli po procesie naprawczym.

1 Wstęp

Podstawowym sposobem komunikacji ludzi jest mowa, którą można rozumieć jako informację niesioną w sygnale dźwiękowym oraz jako akt używania języka w procesie porozumiewania się. Ze względu na dualną naturę tego zagadnienia oraz mnogość języków używanych na świecie, technologie oparte o pracę z mową wymagają wielopłaszczyznowej analizy z perspektywy akustycznej oraz lingwistycznej.

Automatyczne Rozpoznawanie Mowy (ARM) to proces identyfikacji mowy ludzkiej przez komputer, a następnie przekonwertowanie jej treści do formatu tekstowego. Zastosowanie ARM umożliwia znaczne zmniejszenie dystansu na linii człowiek-komputer, co widoczne jest po wzroście popularności rozwiązań takich jak asystenci głosowi lub aplikacje sterowane głosem. Dostępne rozwiązania ARM obejmują szeroki zakres języków, jednak ze względu na swoją złożoną strukturę oraz ograniczone zasoby danych, język polski został zaimplementowany wybiórczo w niewielkiej liczbie produktów (np. Asystent Google), a sama technologia ARM nadal wymaga badań dla tego języka.

Celem niniejszego badania jest zbadanie dwóch modeli ARM przeznaczonych do rozpoznawania języka polskiego zrealizowanych w różnych architekturach oraz weryfikacja ich działania na próbkach mowy o różnej jakości dźwięku.

2 ARM

Na przestrzeni lat wyróżniono trzy podstawowe podejścia w projektowaniu systemów ARM: klasyczne, hybrydowe oraz E2E. Podejścia różnią się sposobem podawania danych i procesem ich przetwarzania, wytrzymałością na zniekształcenia sygnału mowy oraz elastycznością potencjalnych zastosowań.

Podejścia konwencjonalne (rysunek 1a), czyli klasyczne oraz hybrydowe opierają się na wyodrębnieniu trzech podstawowych modeli - akustycznego, językowego oraz leksykalnego. Każdy z modeli wymaga osobnego trenowania oraz odpowiedniego przygotowania danych. W podejściu E2E (rysunek 1b) wykorzystującym uczenie głębokie zrezygnowano z wyodrębniania modeli - proces przetwarzania odbywa się całkowicie w strukturze zintegrowanej, głębokiej sieci neuronowej. Proces treningowy wymaga obszernego zbioru zróżnicowanych danych, ale jest zintegrowany i wykonywany jednocześnie dla całego modelu, a nie jak w podejściu konwencjonalnym - oddzielnie dla modelu akustycznego, leksykalnego i językowego.

Rozpoznawanie mowy dzieli się na etap przyjęcia sygnału dźwiękowego na wejście systemu, przetwarzanie tego sygnału oraz wygenerowanie transkrypcji zawierającej informację zawartą w podanym sygnale mowy. Działanie systemu jest zależne od parametrów wykorzystywanego modelu oraz od parametrów sygnału mowy. Zniekształcenia sygnału akustycznego, obecność dźwięków tła zakłócających dźwięki mowy oraz cechy osobnicze takie jak silny akcent lub wady wymowy mogą znacząco wpłynąć na skuteczność rozpoznawania mowy przez system ARM [13].

(a) Podejście konwencjonalne

(b) Podejście E2E

Rysunek 1: Schematy blokowe podejść w realizacji systemów ARM

3 Metodologia badań

W badaniach wykorzystano dwa bezpłatne, dostępne na otwartej licencji modele przeznaczone do rozpoznawania języka polskiego - ESPnet [16] oraz STT-Pl-FastConformer Hybrid [1]. Modele różniły się podstawowymi założeniami takimi jak optymalizacja systemu do pracy z danym językiem, rozmiarem zestawu treningowego, zastosowaną architekturą oraz realizacją procesu tokenizacji, czyli sformułowania elementów reprezentujących mowę, z których tworzona jest transkrypcja [6]. Testy modeli przeprowadzono na korpusie, który nie wchodził w skład zbioru treningowego żadnego z nich, a nagrania mowy polskiej pobrane z tej bazy cechowały się zdegradowaną jakością sygnału mowy.

3.1 ESPnet

Pierwszym z testowanych modeli był wielojęzykowy model ESPnet [8] zrealizowany w architekturze Speech-Transformer. Wielojęzykowy aspekt tego modelu oznacza, że nie został on zoptymalizowany do pracy z konkretnym językiem, a proces działania systemu ARM dzieli się na etap identyfikacji języka, a następnie rozpoznania samej mowy.

Architektura Speech-Transformer opiera się na wykorzystaniu mechanizmu „uwagi" znanego z architektury Transformer [15]. Odpowiednie przypisanie wag i powiązań słowom w łańcuchu tekstowym pozwala na skuteczne modelowanie kontekstu globalnego analizowanego fragmentu. Mechanizm ten znajduje zastosowania w pracy z mową, pozwalając poprawić skuteczność rozpoznania wyrazów wypowiedzianych niewyraźnie lub zakłóconych szumem tła. Dodatkowo umożliwia to poprawny wybór sposobu zapisu homofonów - słów o różnym znaczeniu i zapisie reprezentowanych przez ten sam dźwięk.

Wykorzystano tokenizer BPE (Byte Pair Encoding)[11], który formułuje, a następnie rozbudowuje słownik tokenów na podstawie iteracyjnego łączenia najczęściej występujących par bajtów (lub znaków) w tekście w celu tworzenia nowych, dłuższych tokenów.

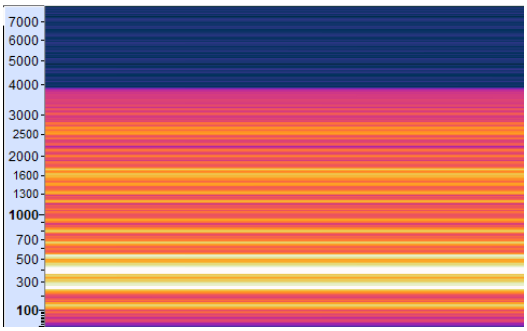
3.2 STT-Pl-FastConformer-Hybrid

Drugim z testowanych modeli był model STT-Pl-FastConformer Hybrid [1] opracowany przez firmę NVIDIA w ramach zestawu narzędzi NeMo [7]. Jest to model zoptymalizowany do pracy wyłącznie z językiem polskim. Model zrealizowano w architekturze Fast Conformer, który jest szybszą, zoptymalizowaną wersją architektury Conformer. Conformer bazuje na architekturze Speech-Transformer rozszerzonej o dodatkowe warstwy splotowe, realizując w ten sposób bloki Conformer [10].

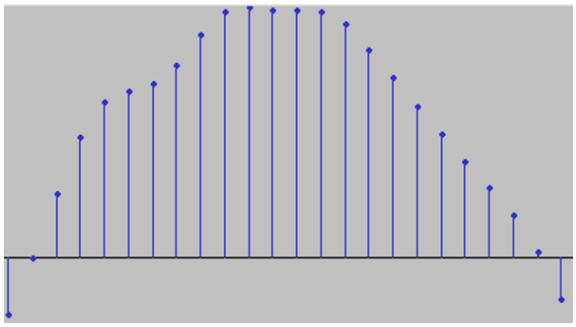
STT-Pl-FastConformer Hybrid wykorzystuje tokenizację Unigram, który w przeciwieństwie do tokenizera BPE, optymalizuje słownik tokenów poprzez iteracyjne redukowanie jego rozmiaru, wykluczając elementy z najmniejszym prawdopodobieństwem wystąpienia [3].

3.3 Mobile Corpus (EMU)

Skuteczność działania modeli w warunkach utrudnionej transmisji sygnału akustycznego testowano na próbkach nagrań mowy polskiej pobranych z bazy Mobile Corpus (EMU) opracowanej w ramach Konsorcjum CLARIN (Common Language Resources & Technology Infrastructure). Korpus zawiera 13,5 godzin nagrań rozmów telefonicznych w języku polskim [14]. Ze względu na ograniczony czas realizacji badań oraz okrojone zasoby sprzętowe, do badania wykorzystano podzbiór nagrań stanowiący 10% objętości korpusu. Po odrzuceniu uszkodzonych lub błędnie opisanych próbek, zbiór testowy składał się z 367 plików zapisanych z częstotliwością próbkowania 16 kHz na jednym kanale, w formacie *.wav*. Transmisja mowy w telefonii wymaga kompromisów ze względu na potrzebę ograniczenia ilości przesyłanych danych - w latach, kiedy baza była opracowywana (2015 r.) szerokopasmowe kodeki nie były powszechnie dostępne, a sygnał mowy charakteryzował się mocno ograniczonym pasmem i głęboką kompresją [9]. Sygnał mowy jest również zniekształcony przez zjawisko *hard clipping*, co skutkuje obcięciem szczytów sinusoid w miejscach, gdzie energia sygnału przekraczała próg maksymalnego napięcia mikrofonu. Opisane zjawiska przedstawiono na rysunkach 2a i 2b.



(a) Ograniczenie pasma do 4 kHz



(b) Zniekształcenia typu *hard clipping*

Rysunek 2: Analiza zniekształceń sygnału mowy próbek pobranych z korpusu testowego

3.4 WER

Skuteczność rozpoznawania mowy oceniano za pomocą współczynnika błędów transkrypcji WER (Word Error Rate), określonego wzorem 1. WER to iloraz sumy błędów - dodanie, usunięcie słowa lub zmiana słowa, względem całkowitej liczby wyrazów. Błędnie zapisane wyrazy określone są na podstawie porównania z transkrypcją referencyjną zawartą w metadanych korpusu. Pożądane są niskie wartości WER - dla bezbłędnej transkrypcji współczynnik będzie wynosił 0%, jednak nie określona jest wartość maksymalna. Uzyskanie WER większego niż 100% oznacza, że w wygenerowanej transkrypcji znajduje się więcej błędów niż poprawnych wyrazów. W celu wyliczenia wartości tego współczynnika skorzystano z biblioteki Python *jiwer* [2].

$$WER = \frac{I + D + S}{N} \cdot 100\% \tag{1}$$

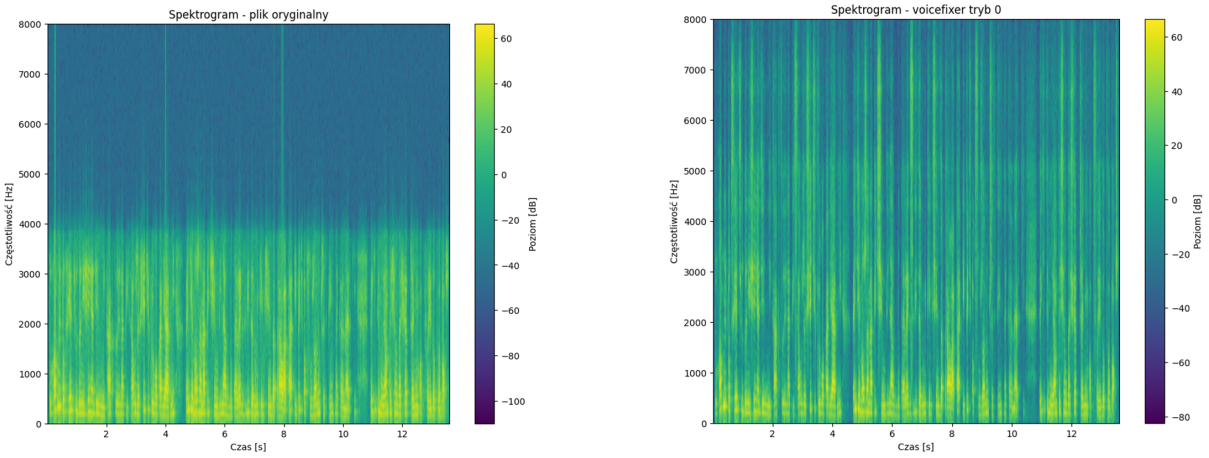
gdzie:

I dodanie (*insertion*),
D usunięcie (*deletion*),
S zamiana (*substitution*),
N łączna liczba wyrazów w próbce.

4 Badanie

Opisane wcześniej modele poddano dwóm testom w celu ewaluacji skuteczności ich działania na nagraniach o niskiej jakości dźwięku, które dodatkowo nie wchodziły w skład bazy treningowej. Następnie próbki zostały poddane naprawie z użyciem narzędzia open-source - modułu Voicefixer [12]. Otrzymane nagrania poddano testowi rozpoznania ponownie, a otrzymane wyniki porównano z rezultatami pierwszego testu.

Pasmo sygnału mowy po przetworzeniu przez moduł Voicefixer zostało poszerzone z 4 kHz do 8 kHz, porównanie spektrogramów obrazujące działanie modułu zawarto na rysunku 3.



(a) Oryginalna próbka

(b) Próbka z odtworzonym pasmem

Rysunek 3: Spektrogramy dla próbki pobranej z korpusu testowego przed i po działaniu modułu naprawczego

4.1 Test na oryginalnych próbkach

Pierwsza część badania dotyczyła testu działania każdego z modeli podczas pracy z próbkami pobranymi bezpośrednio z korpusu testowego. Wyniki uzyskane w badaniu przedstawiono w tabeli 1.

Tabela 1: Średnie wartości WER uzyskane dla próbek niskiej jakości

model	ESPnet		STT-Pl-FastConformer-Hybrid
zbiór	wszystkie próbki	próbki zidentyfikowane jako język polski	wszystkie próbki
średnia WER	104,26%	93%	14,47%

Analiza skuteczności działania wielojęzykowego modelu ESPnet skupiała się na identyfikacji języka, a następnie rozpoznawaniu treści mowy. 37 próbek z testowanego fragmentu bazy zostało zidentyfikowanych jako mowa polska, co stanowi jedynie około 10% tego zbioru. Średnia wartość uzyskanego WER dla ogółu próbek wynosiła 104,26%, w przypadku próbek poprawnie zidentyfikowanych jako mowa polska średnia WER wynosiła 93%.

Wysokie wartości współczynnika błędu uzyskane przez ten model mogą wynikać z faktu, iż błędnie zidentyfikowane próbki najczęściej były uznawane przez model za mowę rosyjską, której zapis realizowany jest w cyrylicy, a nie alfabecie łacińskim, jak w przypadku języka polskiego. Inny system znakowy wykorzystany do sformułowania transkrypcji próbek uniemożliwia poprawny zapis słowa nawet w przypadku identyfikacji w nagraniu wzorców fonetycznych, które są zgodne z brzmieniem języka polskiego. Potencjalnym rozwiązaniem tego problemu może być optymalizacja modelu pod kątem rozpoznawania tych dwóch języków poprzez zrównoważenie bazy, na której model jest trenowany.

Otrzymane wyniki znacząco odbiegają od wartości opisywanych w literaturze - dla tego modelu przewidywaną wartością przy pracy z językiem polskim jest WER na poziomie 15%-24%[4]. Tak duża rozbieżność może wynikać ze specyfiki badanego zbioru testowego - niekonsekwentnego zapisu transkrypcji oraz niskiej jakości próbek dźwiękowych.

Wyniki uzyskane dla próbek poprawnie zidentyfikowanych również nie są wartościami oczekiwanymi od sprawnie działającego systemu ARM - obecnie używane komercyjne rozwiązania ARM uzyskują WER na poziomie około 16% [5].

Model STT-Pl-FastConformer-Hybrid zoptymalizowany wyłącznie dla języka polskiego, osiągnął, w porównaniu z ESPnet, niższą średnią wartość WER, która wynosiłą 14,47%. Podkreśla to wyzwania związane z modelami wielojęzykowymi, ponieważ wyeliminowanie etapu identyfikacji języka uniemożliwiło wybór niewłaściwego słownika tokenów do sformułowania transkrypcji. Bazując na wynikach uzyskanych w przeprowadzonym teście, skuteczność działania tego modelu nie odbiega od skuteczności komercyjnie wykorzystywanych systemów ARM.

4.2 Test na próbkach naprawionych modułem Voicefixer

Proces naprawczy obejmował użycie oryginalnych plików z bazy Mobile Corpus (EMU), próbę poprawy ich jakości za pomocą modułu Voicefixer w domyślnym trybie 0. Pliki wynikowe, po przetworzeniu przez moduł naprawczy miały częstotliwość próbkowania równą 44,1 kHz, przed podaniem danych do modelu, należało ją obniżyć do 16 kHz, a następnie poddać nowe pliki procesowi rozpoznawania. Otrzymane wyniki są niejednoznaczne, wskazując rozbieżne trendy dla każdego z testowanych modeli. Wyniki badań zamieszczono w tabeli 2.

Tabela 2: Średnie wartości WER uzyskane dla próbek po jakości modułem Voicefixer

model	ESPnet		STT-Pl-FastConformer-Hybrid
zbiór	wszystkie próbki	próbki zidentyfikowane jako język polski	wszystkie próbki
średnia WER	97,9%	84,35%	29,64%

Poprawa jakości sygnału mowy miała pozytywny wpływ na model ESPnet, zwłaszcza na etap identyfikacji języka. Liczba próbek zidentyfikowanych jako mowa polska wzrosła ponad trzykrotnie, z 37 próbek do 126, co stanowi 34% zbioru testowego. Zagadnienie błędnej detekcji języka jako rosyjski jest nadal obecne wśród uzyskanych transkrypcji, jednak nasilenie tego problemu zmniejszyło się.

Uzyskane średnie wartości WER wynoszą 97,9% dla ogółu próbek oraz 84,35% dla próbek rozpoznanych jako mowa polska. Zaobserwowana poprawa skuteczności działania tego modelu na podstawie samego współczynnika WER jest nieznaczna, jednak z punktu widzenia modelu wielojęzykowego poprawiono istotny aspekt dotyczący detekcji języka.

Wyniki uzyskane dla modelu STT-Pl-FastConformer-Hybrid wykazują pogorszenie skuteczności rozpoznawania mowy. Średnia wartość WER wzrosła niemal dwukrotnie do wartości 29,64%. Pogorszenie skuteczności działania może być spowodowane zniekształceniami spektrogramu w paśmie wysokich częstotliwości. Analiza transkrypcji uzyskanych przez ten model wskazuje na to, że proces naprawczy spowodował niepoprawne rozpoznawanie spółgłosek przez ten model, co powodowało gwałtowny wzrost WER nawet w przypadku zmiany jednej litery w rozpoznanym wyrazie. Pomimo niekorzystnego wpływu modułu Voicefixer na działanie tego modelu, skuteczność rozpoznawania STT-Pl-FastConformer-Hybrid przeważa nad modelem ESPnet.

5 Podsumowanie

W badaniu porównano dwa modele ARM w podejściu E2E zaimplementowane w architekturach Fast Conformer oraz Speech-Transformer. Były to kolejno: zoptymalizowany do rozpoznawania języka polskiego STT-Pl-FastConformer-Hybrid oraz wielojęzykowy ESPnet. Modele przetestowano na próbkach pobranych z bazy nagrań rozmów telefonicznych Mobile Corpus (EMU). Żaden z modeli nie był trenowany na tej bazie, a próbki sygnału dźwiękowego charakteryzowały się niską jakością dźwięku oraz ograniczonym pasmem.

Pierwszy test przeprowadzono dla próbek w niskiej jakości pobranych bezpośrednio z bazy. Model STT-Pl-FastConformer-Hybrid uzyskał niższe średnie wartości WER (Word Error Rate) w porównaniu z modelem ESPnet. Ze względu na to, że ESPnet jest modelem wielojęzykowym, jego proces przetwarzania dzieli się na etap identyfikacji języka oraz rozpoznanie mowy. Nagrania często zostawały błędnie uznawane za mowę w języku rosyjskim, a transkrypcja była generowana w cyrylicy, co znacznie podnosiło wartość współczynnika błędu.

Odtworzenie widma sygnału mowy modułem Voicefixer wpłynęło na poprawę jakości dźwięku, jednak wyniki uzyskane dla modeli były dwuznaczne. Dla modelu ESPnet uzyskano niższe wartości WER w porównaniu z pierwszym badaniem, jednak skuteczność rozpoznania STT-Pl-FastConformer-Hybrid pogorszyła się dwukrotnie. Identyfikacja języka przez model ESPnet również się poprawiła, liczba próbek uznanych jako mowa polska wzrosła z 37 do 126. Rozbieżne wyniki, mogą wskazywać, że wpływ modułu Voicefixer może być zarówno pozytywny, jak i negatywny w zależności od parametrów systemu ARM.

Literatura

[1] Dokumentacja fast conformer. Dostęp zdalny (12.02.2024): https://docs.nvidia.com/deeplearning/nemo/user-guide/docs/en/main/asr/models.html#fast-conformer.

[2] Dokumentacja jiwer. Dostęp zdalny (01-02-2024): https://jitsi.github.io/jiwer/.

[3] Dokumentacja sentencepiece. Dostęp zdalny (12.02.2024): https://github.com/google/sentencepiece.

[4] Recent developments on espnet toolkit boosted by conformer. Dostęp zdalny (25-03-2024)): https://arxiv.org/pdf/2010.13956.pdf.

[5] Speech-to-text transcript accuracy rate among leading companies worldwide in 2021. Dostęp zdalny (12.02.2024): https://www.statista.com/statistics/1133833/speech-to-text-transcript-accuracy-rate-among-leading-companies/#statisticContainer.

[6] Tokenizers: How machines read. Dostęp zdalny (12.02.2024): https://blog.floydhub.com/tokenization-nlp/.

[7] Stt pl fastconformer hybrid transducer-ctc large pc - repozytorium, wersja 1.21.0, 2023. Dostęp zdalny (18.03.2024): https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_pl_fastconformer_hybrid_large_pc.

[8] Espnet - repozytorium, 2024. Dostęp zdalny (18.03.2024): https://github.com/espnet/espnet.

[9] 3GPP. Technical specifications and technical reports for a utran-based 3gpp system. Technical report, 3GPP, 2008.

[10] Chung-Cheng Chiu Niki Parmar Yu Zhang Jiahui Yu Wei Han Shibo Wang Zhengdong Zhang Yonghui Wu Ruoming Pang Anmol Gulati, James Qin. Conformer: Convolution-augmented transformer for speech recognition. 2020.

[11] Kaj Bostrom and Greg Durrett. Byte pair encoding is suboptimal for language model pretraining. *CoRR*, abs/2004.03720, 2020.

[12] Haohe Liu, Qiuqiang Kong, Qiao Tian, Yan Zhao, DeLiang Wang, Chuanzeng Huang, and Yuxuan Wang. Voicefixer: Toward general speech restoration with neural vocoder, 2021.

[13] Ryszard Makowski. *Automatyczne rozpoznawanie mowy - wybrane zagadnienia*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, Polska, 2011.

[14] Krzysztof Marasek, Danijel Koržinek, Łukasz Brocki, and Kamila Jankowska-Lorek. Clarin-PL mobile corpus (EMU), 2015.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[16] Shinji Watanabe, Takaaki Hori, Shigeki Karita, Tomoki Hayashi, Jiro Nishitoba, Yuya Unno, Nelson Enrique Yalta Soplin, Jahn Heymann, Matthew Wiesner, Nanxin Chen, Adithya Renduchintala, and Tsubasa Ochiai. Espnet: End-to-end speech processing toolkit, 2018.

Agata **ZATORSKA**[1], Michał **ŁUCZYŃSKI**[1]

POMIARY SŁUCHAWEK Z AKTYWNĄ REDUKCJĄ HAŁASU MEASUREMENTS OF HEADPHONES WITH ACTIVE NOISE REDUCTION

[1]Politechnika Wrocławska

agatazatorska01@gmail.com

Streszczenie

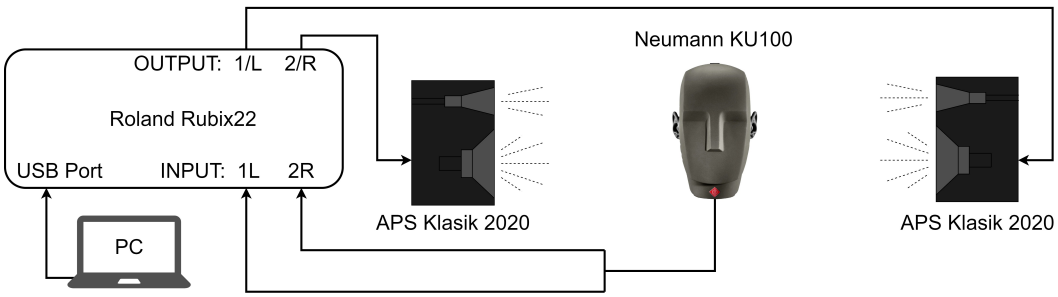
Praca opisuje obiektywną metodę pomiaru parametrów słuchawek z aktywną redukcją hałasu wykorzystując do pomiaru sztuczną głowę. W pracy przedstawiono metodę pomiaru tłumienia słuchawek (w tym tłumienia pasywnego, aktywnego oraz całkowitego) oraz czasu konwergencji. Pomiar czasu konwergencji bazuje na autorskiej metodzie, wprowadzając jednocześnie parametr skuteczności tłumienia sygnału przy załączeniu źródła dźwięku. Ten parametr umożliwia powiązanie dwóch istotnych wartości dla aktywnej redukcji hałasu, tj. wartości tłumienia aktywnego osiąganej przez słuchawki oraz czasu potrzebnego do osiągnięcia tej wartości. Zestawienie mierzonych parametrów pozwala na obiektywne porównanie słuchawek z ANC (aktywną redukcją hałasu, eng. Active Noise Cancelling).

1 Wstęp

Obecnie nie istnieje norma standaryzująca obiektywne pomiary słuchawek z ANC z wykorzystaniem testera akustycznego. Metoda pomiaru tłumienia przedstawiona w pracy bazuje na artykule Christophera J. Strucka "Objective Measurements of Headphone Active Noise Cancelation Performance"[1] oraz nocie technicznej Audio Precision "TN141 ANC Headphones: measuring insetrion loss"[2], które wykorzystują do pomiaru testery akustyczne. Obie prace opisują pomiar tłumienia aktywnego, pasywnego oraz całkowitego słuchawek. Jednak nie ma informacji o pomiarze czasu konwergencji na gotowych słuchawkach. W podpunkcie trzecim opisano autorską metodę pomiaru czasu konwergencji, wprowadzając jednocześnie parametr skuteczności tłumienia sygnału przy załączeniu źródła dźwięku. Ten parametr wyrażony jest jako stosunek osiągniętej wartości tłumienia do czasu konwergencji. Pomiary przeprowadzono na słuchawkach wokółusznych i dokanałowych.

2 Pomiar tłumienia

Norma PN-EN ISO 4869-3 [3] dotycząca pomiaru ochronników słuchu zaleca wykonanie pomiaru w polu dyfuzyjnym lub w polu bieżącej fali płaskiej. Z uwagi na ograniczone możliwości, pomiar wykonano w zaadaptowanym akustycznie pomieszczeniu o powierzchni $15m^2$. Sygnał generowano z pary urządzeń głośnikowych umieszczonych na osi testera (znajdującej się między jego lewym a prawym uchem) w odległości 1 metra od każdego ucha. Według noty technicznej Audio Precision [2] taka metoda pozwala uzyskać wiarygodne wyniki bez wykorzystania komory pogłosowej. Na rysunku 1 przedstawiono układ pomiarowy wykorzystany przy pomiarze tłumienia słuchawek. Rysunek 2 przedstawia rzeczywiste rozmieszczenie urządzeń podczas wykonywania pomiarów.
Pomiar wykonywany był przy użyciu szumu różowego. Zastosowano stosunkowo wysoki poziom sygnału z uwagi na fakt, że niektóre słuchawki z aktywną redukcją hałasu mogą posiadać progowe wartości aktywacji. Producenci urządzeń konsumenckich nie udostępniają takich informacji w dokumentacji. Niski poziom sygnału wejściowego mógłby nie spełnić warunków zadziałania układu aktywnej redukcji.



Rysunek 1: Układ pomiarowy wykorzystany do pomiaru tłumienia słuchawek

Poziom tła akustycznego dla pełnego pasma wynosił 22,6 dB(A), natomiast poziom sygnału testowego wynosił 86,3 dB(A). W badanym zakresie częstotliwości najmniejszy odstęp między poziomem tła a poziomem sygnału uzyskano w paśmie 63 Hz osiągając 43,7 dB. Oznacza to, że spełniony został warunek normy, która zaleca minimalny odstęp o wartości 10 dB.
Manekinem testowym wykorzystanym do pomiarów była sztuczna głowa Neumann KU100 [4]. Jest to rodzaj mikrofonu binauralnego imitujący ludzkie ucho i głowę, co pozwala na nagrywanie dźwięku w sposób zbliżony do tego, jak słyszy go słuchacz. Według producenta urządzenie może być wykorzystywane między innymi: do analizy wpływu hałasu (np. w zakładach przemysłowych), badań zrozumiałości mowy czy pomiarów słuchawek.

Rysunek 2: Rzeczywisty układ pomiarowy wykorzystany do pomiaru tłumienia słuchawek

W tabeli 1 zestawiono modele badanych słuchawek, w tym trzy modele wokółuszne firmy Sony oraz trzy modele dokanałowe różnych producentów. Wyboru słuchawek dokonano losowo, uwzględniając jedne, znacznie tańsze słuchawki, tj. model Sony WH-CH720N.

Tabela 1: Zestawienie badanych modeli słuchawek z podziałem na słuchawki wokółuszne i dokanałowe

Słuchawki wokółuszne	Słuchawki dokanałowe
Sony WH-1000XM3,	AirPods Pro2
Sony WH-CH720N,	Samsung Galaxy Buds 2
Sony WH-1000XM4	Sennheiser MOMENTUM TW3

Do wyznaczenia tłumienia konieczne jest dokonanie pomiarów w czterech różnych sytuacjach:

1. $L_{tła}(f)$ - pomiar bez założonych słuchawek przy wyłączonym źródle sygnału testowego,

2. $L_{sygnału}(f)$ - pomiar bez założonych słuchawek przy włączonym źródle sygnału testowego,

3. $L_{bezANC}(f)$ - pomiar z założonymi słuchawkami przy włączonym źródle sygnału testowego z wyłączoną aktywną redukcją hałasu,

4. $L_{zANC}(f)$ - pomiar z założonymi słuchawkami przy włączonym źródle sygnału testowego z włączoną aktywną redukcją hałasu.

Każdy pomiar wykonywano przez 30 sekund. Uzyskane sygnały poddano filtracji za pomocą banku filtrów 1/3 oktawowych w zakresie od 50 Hz do 10 kHz [1]. Pomiary z punktów 3 i 4 powtórzono pięciokrotnie, a uzyskane wyniki zostały uśrednione. Celem tego procesu było uśrednienie wartości tłumienia w zależności od ułożenia słuchawki w lub na uchu. Tłumienie (pasywne (wzór (1)), aktywne (wzór (2)) oraz całkowite (wzór (3))) w poszczególnych pasmach obliczono jako różnicę między poziomami sygnału zmierzonymi w każdym paśmie 1/3 oktawowym z poszczególnych pomiarów.

$$T_{pasywne}(f) = L_{sygnału}(f) - L_{bezANC}(f)\ [dB] \tag{1}$$

$$T_{aktywne}(f) = L_{bezANC}(f) - L_{zANC}(f)\ [dB] \tag{2}$$

$$T_{całkowite}(f) = L_{sygnału}(f) - L_{zANC}(f)\ [dB] \tag{3}$$

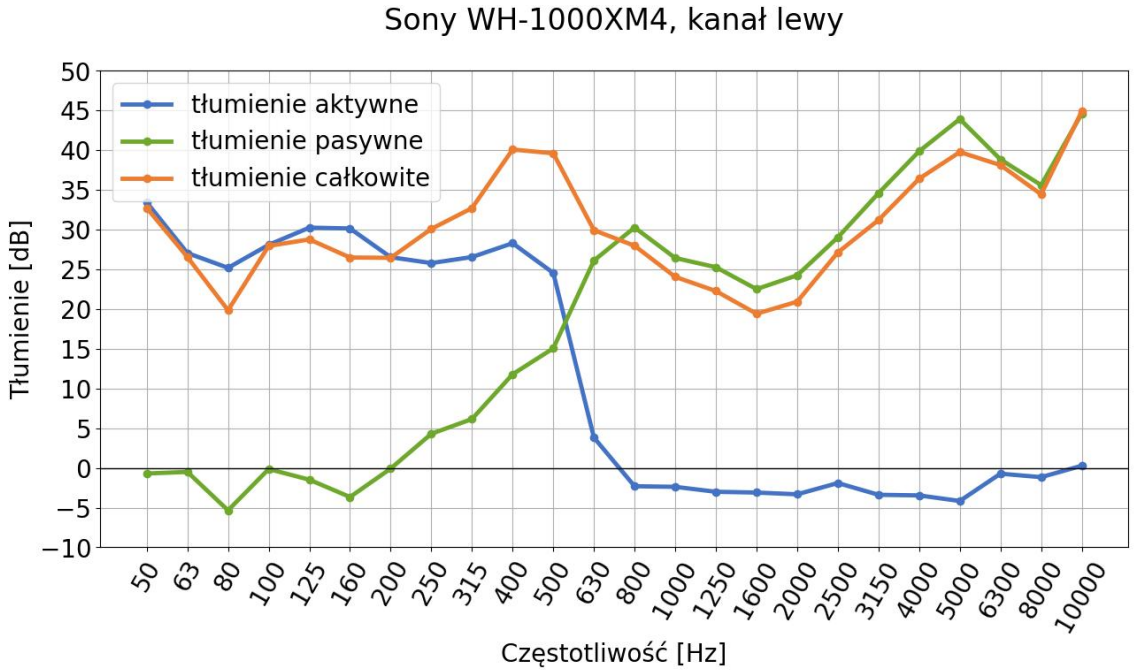
Uzyskane wyniki zestawiono w postaci graficznej dla poszczególnych modeli słuchawek. Zastosowana metoda zakłada przeprowadzenie pomiaru bez konieczności określania wartości ciśnienia akustycznego przy każdym pomiarze. Pomiary zostały zapisane jako pliki dźwiękowe w formacie wav, przy częstotliwości próbkowania 48 kHz i rozdzielczości bitowej 24 bity. Sygnał wejściowy został wysterowany w taki sposób, aby podczas pomiaru bez założonych słuchawek przy włączonym źródle sygnału testowego, wartość szczytowa sygnału nie przekraczała -5 dBFS.

[1] Zgodnie z wytycznymi normy PN-EN ISO 4869-3 zaleca się, aby zakres częstotliwości obejmował co najmniej częstotliwości środkowe z przedziału od 63 Hz do 8 kHz

Przetwarzanie danych oraz generowanie wykresów tłumienia poszczególnych słuchawek zostało zrealizowane przy pomocy stworzonego programu w języku Python.

2.1 Wyniki pomiaru tłumienia

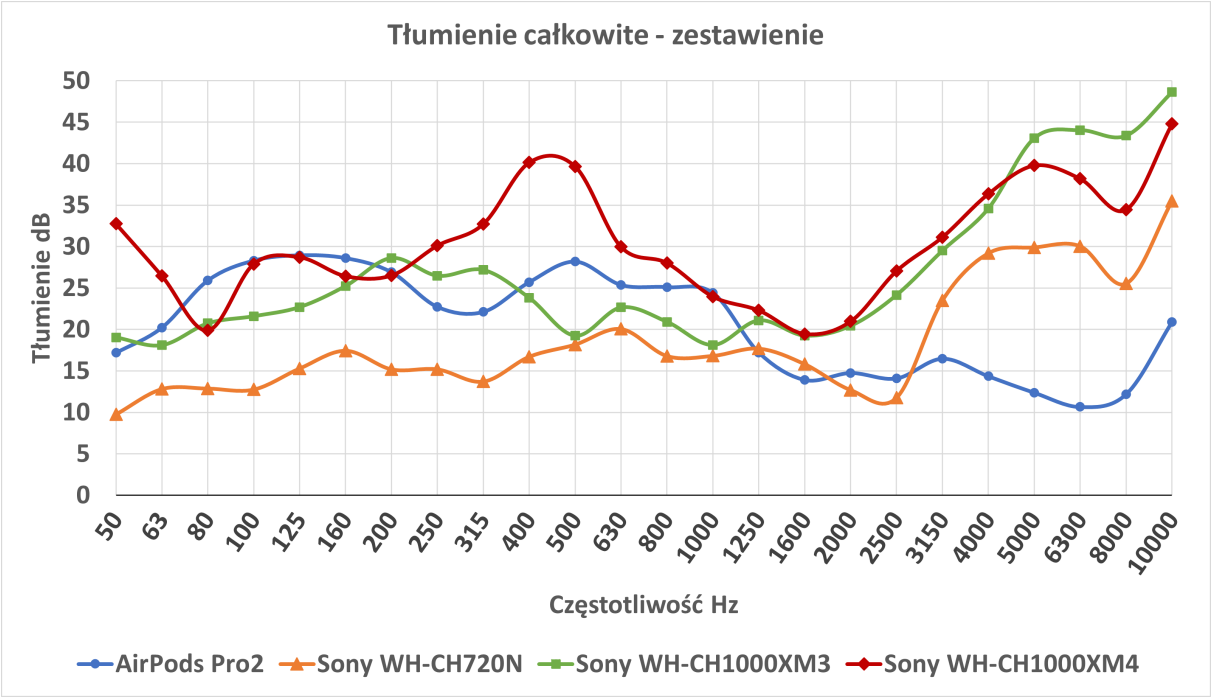
Na rysunku 3 przedstawiono przykładowy wykres tłumienia pasywnego, aktywnego oraz całkowitego uzyskany dla lewego kanału słuchawek Sony WH-1000XM4. Dane uzyskano poprzez przeprowadzenie obliczeń na poziomach sygnału w pasmach 1/3 oktawowych, korzystając z równań (1) - (3).



Rysunek 3: Zestawienie tłumienia aktywnego, pasywnego i całkowitego w funkcji częstotliwości dla kanału lewego słuchawek Sony WH-1000XM4

Uzyskane wyniki są zgodne z założeniami. Tłumienie aktywne (rozumiane jako tłumienie uzyskane w wyniku działania układu ANC) dominuje do 500 Hz, powyżej zaczyna dominować tłumienie pasywne (rozumiane jako tłumienie uzyskane poprzez konstrukcję słuchawki, która pełni rolę naturalnej bariery dźwiękowej). Tłumienie całkowite, które jest zestawieniem tłumienia aktywnego i pasywnego utrzymuje się na wysokich wartościach w całym badanym zakresie.

Zestawienie tłumienia całkowitego wszystkich modeli przedstawione na rysunku 4 umożliwia porównanie efektywności tłumienia oferowanego przez badane słuchawki. Można zauważyć, że tańsze słuchawki (w przypadku analizowanych modeli były to słuchawki Sony



Rysunek 4: Zestawienie tłumienia całkowitego dla badanych słuchawek

WH-CH720N) wykazują niższe wartości tłumienia całkowitego. W przypadku pozostałych (droższych) modeli wartości tłumienia całkowitego przekraczają 20 dB do częstotliwości 1 kHz, co gwarantuje efektywne tłumienie aktywne w tym zakresie częstotliwości. Wyniki AirPods Pro2 reprezentujących grupę słuchawek dokanałowych sugerują, że ten typ słuchawki uzyskuje gorsze wyniki tłumienia powyżej 3 kHz niż słuchawki wokółuszne. Aby potwierdzić tę hipotezę, należałoby zestawić więcej wyników dla słuchawek dokanałowych. Niestety AirPods Pro2 były jedynymi słuchawkami dokanałowymi, które udało się zmierzyć. Pozostałe słuchawki nie były dopasowane do kształtu ucha sztucznej głowy. Słuchawki wysuwały się z małżowiny usznej lub nie trafiały w kanał słuchowy. W kontekście słuchawek dokanałowych bardziej korzystne byłoby zastosowanie głowy pomiarowej z małżowinami usznymi o wymiennych rozmiarach. Takie podejście umożliwiłoby lepsze dopasowanie do kształtu badanych słuchawek, co może przyczynić się do uzyskania miarodajnych wyników pomiarów.

3 Pomiar czasu konwergencji

Czas konwergencji to czas, w jakim następuje dostrojenie współczynników filtrów układu ANC [5]. Czas dostrojenia filtrów przekłada się czas, w jakim dochodzi do stabilizacji wartości tłumienia. W przypadku tłumienia hałasu o szybkozmiennych parametrach istotne jest jak najszybsze dostrojenie układu do zmieniających się warunków. Chwi-

lowe niedostrojenia, zwłaszcza w przypadku hałasów impulsowych, mogą prowadzić do krótkotrwałego, niezamierzonego wzmocnienia sygnału. Dlatego też minimalizacja czasu konwergencji jest kluczowym kryterium w osiąganiu skutecznej redukcji hałasu w dynamicznie zmiennych środowiskach akustycznych.
Przedstawiona metoda stanowi autorską propozycję określenia czasu konwergencji przy użyciu szumu różowego. W literaturze można znaleźć publikacje, w których wyznacza się czas konwergencji poprzez przeprowadzenie symulacji tworzonego systemu [5][6]. Informacje związane z parametrami aktywnej redukcji hałasu nie są udostępniane przez producentów. Propozycja pomiarowa została stworzona w celu dokładnego ustalenia parametrów gotowego produktu, traktując go jako obiekt pomiarowy o ograniczonym dostępie do wewnętrznej struktury.

Do pomiaru czasu konwergencji użyto identycznego układu co w przypadku pomiaru tłumienia słuchawek (rysunek 1). W pierwszej części pomiarów wykorzystano tony proste o częstotliwościach 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz. W drugiej części pomiarów sygnałem testowym był szum różowy. Przy analizie danych wyniki przedstawiono dla poszczególnych pasm 1/3 oktawowych, których częstotliwości środkowe odpowiadały częstotliwościom sygnałów sinusoidalnych wykorzystanych w pierwszej części pomiaru.
Przy pomocy manekina testowego, wykonano pomiar w dwóch sytuacjach:

1. L_{bezANC} - pomiar z założonymi słuchawkami przy włączonym źródle sygnału testowego z wyłączoną aktywną redukcją hałasu,

2. L_{zANC} - pomiar z założonymi słuchawkami przy włączonym źródle sygnału testowego z włączoną aktywną redukcją hałasu.

Każdy pomiar był przeprowadzany przez 5 sekund. W celu zapewnienia synchronicznego załączenia sygnału testowego w trakcie obu pomiarów, jego odtworzenie następowało automatycznie w momencie rozpoczęcia rejestracji pomiaru.
Pomiar przy użyciu sygnału sinusoidalnego został przeprowadzony w celu monitorowania funkcjonowania układu reagującego na okresowy sygnał stacjonarny. Pomiar miał też służyć zaobserwowaniu wpływu długości ramki na dokładność charakterystyki tłumienia aktywnego w funkcji czasu w zależności od okresu sygnału testowego.
Dla każdej ramki obliczano początkowo wartość skuteczną wyrażoną w decybelach FS,

a następnie obliczano wartość tłumienia korzystając ze wzoru (4).

$$T_{aktywne} = L_{bezANC} - L_{zANC}\ [dB] \tag{4}$$

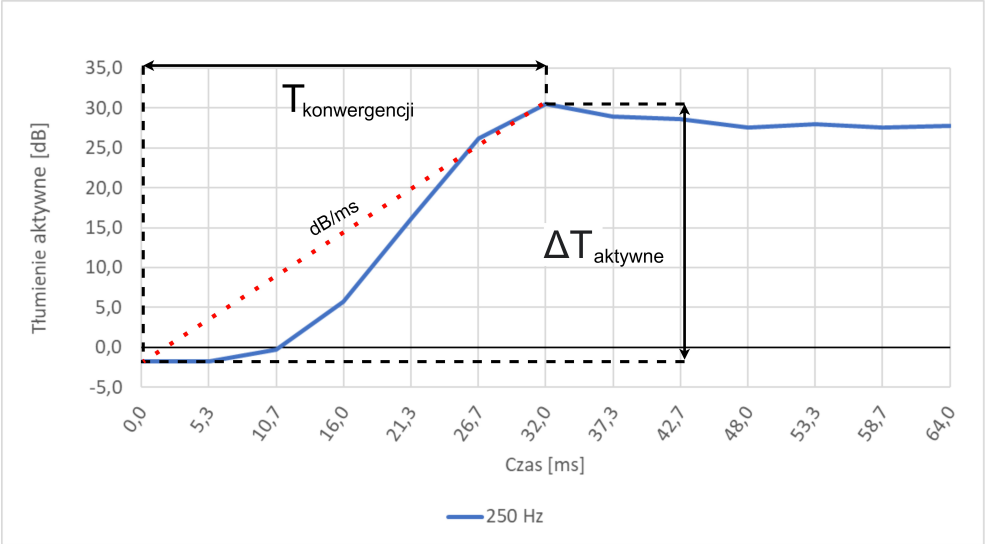
Obliczenia wartości skutecznej wykonywano bez nakładania ramek.
Pomiar sygnałem sinusoidalnym wykazał, że długość ramki musi wynosić co najmniej 0,7 okresu sinusa, aby otrzymać ustabilizowane wartości tłumienia.
Pomiar szumem różowym został wykonany w celu obserwacji zmian tłumienia aktywnego w czasie przy załączeniu sygnału. Po wstępnym przeanalizowaniu przebiegów wykresów zdecydowano się na wykorzystanie ramki o długości 256 próbek. W przypadku częstotliwości próbkowania równej 48 kHz ramka o długości 256 próbek uśrednia dane z zakresu 5,3 ms. Jest to wartość wystarczająca do obserwacji zmian tłumienia w czasie przy pomiarze szumem różowym. Na podstawie uzyskanych charakterystyk został wyznaczony czas konwergencji definiowany jako czas od załączenia sygnału do ustabilizowania charakterystyki tłumienia. Pomiar został wykonany analogicznie do pomiaru przy pomocy sygnałów sinusoidalnych. Podobnie jak w przypadku pomiaru tłumienia, aby wyeliminować niepewność związaną ze sposobem ułożenia słuchawek na manekinie, pomiar wykonano pięciokrotnie a wyniki uśredniono.

Sama wartość czasu zbieżności nie jest wystarczającym kryterium oceny efektywności tłumienia przy załączeniu sygnału testowego. Istotna jest również wartość tłumienia, jaką osiąga układ po dostrojeniu. W celu porównania obu parametrów wprowadzono parametr skuteczności tłumienia wyrażony jako stosunek osiągniętej wartości tłumienia do czasu konwergencji (wzór (5)).

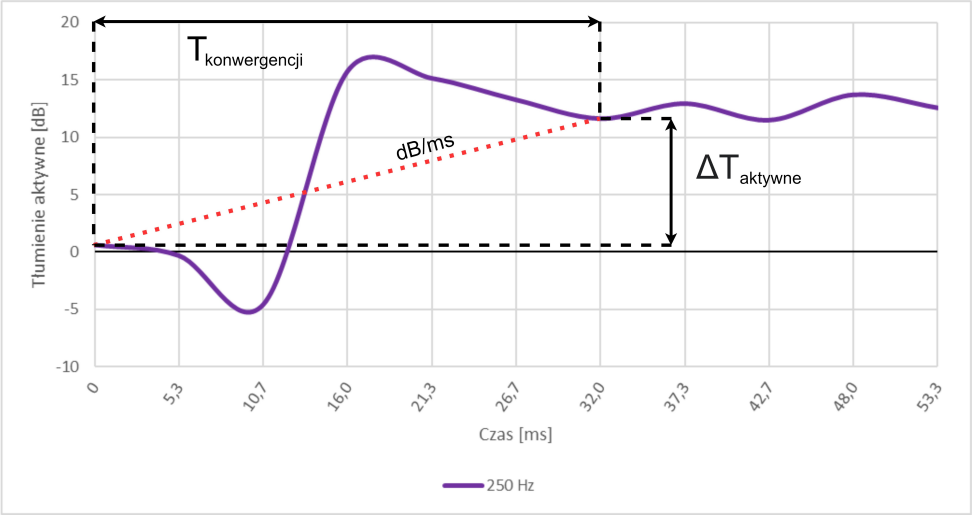
$$S_{tłumienia} = \frac{\Delta T_{aktywne}}{T_{konwergencji}} \tag{5}$$

Sposób wyznaczania obu wartości przedstawiono na rysunku 5. Jednostka parametru wyrażona jest w decybelach na milisekundę. Wartość dla poszczególnych modeli została przedstawiona jako średnia z 5 pomiarów.
Na rysunkach 5 i 6 czerwoną przerywaną linią zaznaczono liniową aproksymację charakterystyki tłumienia od momentu załączenia sygnału do stabilizacji tłumienia. Analizując wykres na rysunku 6 wydaje się, że bardziej adekwatne jest przeprowadzenie aproksymacji w obszarze charakteryzującym się większą liniowością. Jednakże nie uwzględnia ona czasu, jaki mija przed wejściem w obszar liniowy oraz przed dostrojeniem układu. Ujemne

wartości tłumienia, które układ uzyskuje w pierwszych 12 ms (rysunek 6) są spowodowane dobraniem zbyt długiego kroku adaptacji przez układ, co skutkuje brakiem stabilizacji. W celu powiązania wartości tłumienia w danym paśmie z czasem dostrojenia do tej wartości zdecydowano się na zastosowanie pewnego uśrednienia.



Rysunek 5: Wyznaczenie parametrów do obliczenia skuteczności tłumienia

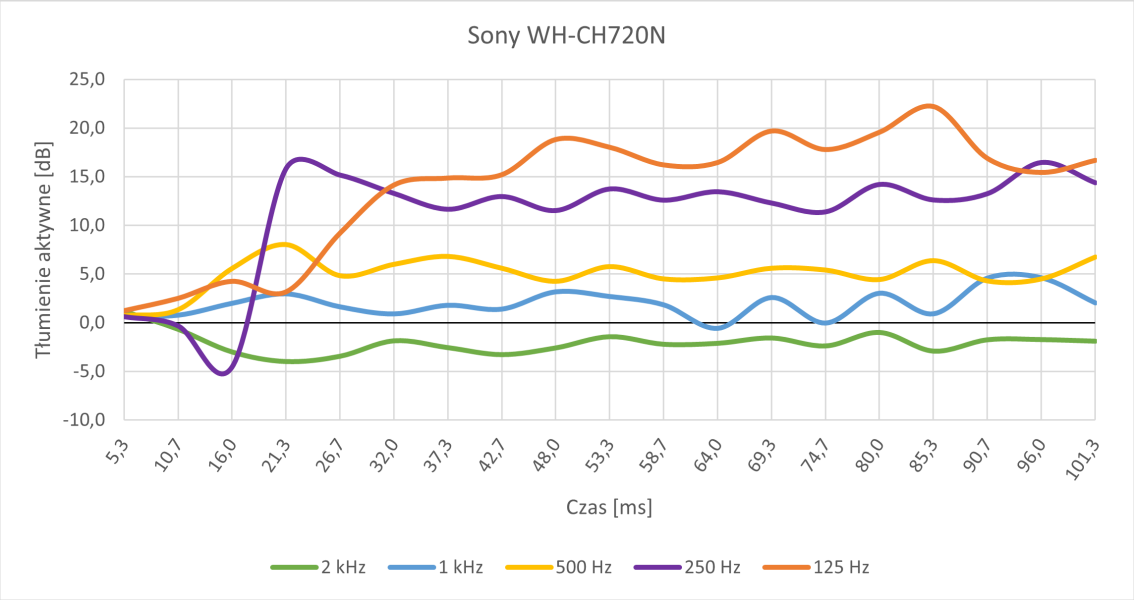


Rysunek 6: Liniowe uśrednienie tłumienia aktywnego przy wyznaczeniu skuteczności tłumienia

3.1 Wyniki pomiaru czasu konwergencji i skuteczności tłumienia

Na rysunku 7 przedstawiono przykładowy wykres tłumienia aktywnego w funkcji czasu dla słuchawek Sony WH-CH720N. Uzyskane wartości tłumienia aktywnego w badanych pasmach po stabilizacji pokrywają się z wynikami uzyskanymi przy pomiarze tłumienia.

Czas zbieżności wyznaczany przy pomiarze szumem różowym jest dłuższy niż w przypadku pomiaru przy użyciu sygnałów sinusoidalnych, co oznacza, że układ aktywnej redukcji hałasu dostraja się szybciej gdy sygnał pomiarowy ma właściwości deterministyczne.
Przy pomiarze szumem różowym układ jest o wiele mniej stabilny niż w przypadku pomiaru sygnałami sinusoidalnymi. Wynika to z faktu, że szum różowy ma charakter losowy, więc układ musi na bieżąco korygować swoje parametry.



Rysunek 7: Tłumienie aktywne w funkcji czasu dla przefiltrowanego szumu różowego. Pomiar wykonano na słuchawkach Sony WH-CH720N

W tabelach 2 i 3 przedstawiono zestawienie czasu konwergencji oraz skuteczności tłumienia dla badanych słuchawek. Parametry wyznaczono wyłącznie dla pasm, w których uzyskano wartości aktywnego tłumienia powyżej 5 dB - 125 Hz, 250 Hz, 500 Hz. Kolorem czerwonym zaznaczono wartości najmniej korzystne a kolorem zielonym wartości najbardziej korzystne w każdym badanym paśmie (125 Hz, 250 Hz, 500 Hz). W przypadku samego czasu konwergencji nie widać zależności między najszybszym czasem a konkretnym modelem. Gdy jednak spojrzymy na tabelę 3 zestawiającą wyniki skuteczności tłumienia wyraźnie widać, że słuchawki Sony WH-1000XM4 uzyskują najbardziej korzystne wyniki w każdym badanym paśmie.
We wszystkich badanych modelach zaobserwowano tendencję skrócenia czasu konwergencji wraz ze wzrostem częstotliwości tłumionego pasma. Niemniej jednak, skrócenie czasu konwergencji przy wyższych częstotliwościach nie przekłada się jednoznacznie na wartość skuteczności tłumienia.

Klasyfikacja słuchawek na podstawie czasu konwergencji i poziomu uzyskanego tłumienia aktywnego wydaje się być znacznie bardziej miarodajna.

Tabela 2: Zestawienie wartości czasu konwergencji dla badanych słuchawek w pasmach 125 Hz, 250 Hz i 500 Hz. Kolorem czerwonym zaznaczono wartości najmniej korzystne a kolorem zielonym wartości najbardziej korzystne w każdym badanym paśmie

Czas konwergencji [ms]				
f [Hz]	**Sony WH-CH720N**	**Sony WH-1000XM3**	**Sony WH-1000XM4**	**AirPods Pro2**
125	48,00	43,83	48,0	53,30
250	32,00	40,00	32,0	48,00
500	21,30	18,65	27,7	28,43

Tabela 3: Zestawienie wartości skuteczności tłumienia dla badanych słuchawek w pasmach 125 Hz, 250 Hz i 500 Hz. Kolorem czerwonym zaznaczono wartości najmniej korzystne a kolorem zielonym wartości najbardziej korzystne w każdym badanym paśmie

Skuteczność tłumienia [dB/ms]				
f [Hz]	**Sony WH-CH720N**	**Sony WH-1000XM3**	**Sony WH-1000XM4**	**AirPods Pro2**
125	0,38	0,55	0,9	0,26
250	0,47	0,59	1,0	0,25
500	0,45	0,84	1,2	0,20

4 Podsumowanie

Praca przedstawia metodę pomiaru umożliwiającą pomiar tłumienia słuchawek (w tym tłumienia pasywnego, aktywnego oraz całkowitego) oraz pomiar czasu konwergencji. Pomiary wykonywane są w sposób obiektywny przy wykorzystaniu sztucznej głowy.
W wyniku przeprowadzonej analizy badanych modeli stwierdzono, że słuchawki Sony WH-1000XM4 wykazują najlepsze tłumienie całkowite w badanym paśmie częstotliwości (50 Hz - 10 kHz). Słuchawki Sony WH-CH720N, które należą do niższej klasy cenowej uzyskały najgorsze wyniki. Ponadto, słuchawki dokanałowe AirPods Pro2 wykazały niższe wartości tłumienia całkowitego powyżej 3 kHz w porównaniu do badanych słuchawek

wokółusznych.

Wprowadzone pojęcie skuteczności tłumienia pozwala powiązać parametr czasu konwergencji i wartości tłumienia uzyskanej w stanie ustabilizowanym. Pod względem skuteczności tłumienia we wszystkich badanych pasmach 1/3 oktawowych, słuchawki Sony WH-1000XM4 osiągnęły najbardziej korzystne wartości.

Przedstawiona metoda pomiaru czasu konwergencji przy pomocy szumu różowego, uzyskane wykresy tłumienia w funkcji częstotliwości oraz wartości skuteczności tłumienia pozwalają na obiektywną ocenę badanych słuchawek pod względem szybkości dostrajania układu ANC do sygnału zewnętrznego.

Badane modele nie były porównywane subiektywnie przez słuchaczy, co może stanowić obszar dalszych badań. Taka analiza może pomóc w ustaleniu arbitralnej granicy tłumienia, powyżej której słuchacz uzna aktywne tłumienie za skuteczne.

Praca wskazuje, że można przeprowadzić pomiar czasu konwergencji na gotowym produkcie oraz porównać go między różnymi modelami słuchawek (np. za pomocą parametru skuteczności tłumienia). Obszar dalszych badań mógłby obejmować eksplorację zmiany rodzaju sygnału oraz analizę czasu konwergencji w szerszym zakresie pasma częstotliwościowego.

Literatura

[1] Christopher J. Struck, Objective Measurements of Headphone Active Noise Cancelation Performance, Audio Engineering Society conference paper, 2019

[2] Audio Precision, TN141 ANC Headphones: measuring insetrion loss, 2019

[3] Polski Komitet Normalizacyjny, PN-EN ISO 4869-3:2009 - Akustyka – Ochronniki słuchu – Część 3: Pomiary tłumienia wtrącenia nauszników przeciwhałasowych wykonywane z użyciem testera akustycznego, 2009

[4] Georg Neumann GmbH. KU100 - Product Information

[5] I. Tabatabaei Ardekani, W.H. Abdulla, On the convergence of real-time active noise control systems, Elsevier, 2011

[6] M. Pawełczyk, On convergence and stability of adaptive active noise control systems, Archives of Acoustics, 2006

Mateusz **ZYCH**[1], Agnieszka Paula **PIETRZAK**[1]

ANALIZA WPŁYWU BODŹCA KONTEKSTOWEGO NA DOKŁADNOŚĆ LOKALIZACJI PRZY BINAURALNYM ODSŁUCHU DŹWIĘKU AMBISONICZNEGO

ANALYSIS OF THE INFLUENCE OF THE CONTEXTUAL STIMULUS ON THE ACCURACY OF LOCALIZATION IN BINAURAL LISTENING OF AMBISONIC SOUND

[1] Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych, Instytut Radioelektroniki i Technik Multimedialnych, ul. Nowowiejska 15/19, 00-665 Warszawa

mateusz.zych7523@gmail.com

Streszczenie

Jedną z podstaw oceny jakości dźwięku przestrzennego w odsłuchu binauralnym jest wartość błędów lokalizacji wirtualnego źródła dźwięku. W niniejszej pracy skupiono się na analizie wpływu rodzaju bodźca wykorzystywanego w teście lokalizacji, a także na wpływie treningu na minimalizację błędów lokalizacji przy binauralnym odsłuchu dźwięku ambisonicznego. Wykorzystano dwa rodzaje bodźców – kontekstowy (głos ludzki) oraz neutralny (szum różowy). Opracowano test lokalizacji przy użyciu wirtualnie umiejscowionych próbek dźwiękowych oraz stworzono serię treningów przygotowawczych. Badanie zostało przeprowadzone dla dwóch grup osób, z których tylko jedna przystąpiła do serii treningów przed właściwym testem końcowym. Test lokalizacji był przeprowadzony osobno dla źródeł rozmieszczonych w płaszczyźnie poziomej i pionowej. Analiza wpływu bodźca kontekstowego na dokładność lokalizacji dla kątów elewacji wykazała wyższe średnie wartości błędów wśród odpowiedzi na próbkę głosu ludzkiego w porównaniu do szumu różowego, jednakże dla kątów azymutu nie można było jednoznacznie określić predyspozycji względem danego rodzaju próbki. Analizując wpływ treningu zaobserwowano nieznaczną poprawę wartości średnich błędów dla grupy, która przeszła serię treningów przed testem, w porównaniu do grupy która podjęła się testu bezpośrednio.

1 Wprowadzenie

W obszarach wykorzystujących wirtualną i rozszerzoną rzeczywistość dźwięk pełni fundamentalną rolę w percepcji prezentowanych wirtualnych scen. Stworzenie wiarygod-

nej przestrzennej sceny dźwiękowej jest kluczowe dla osiągnięcia zjawiska immersji, czyli wrażenia uczestniczenia w prezentowanych wirtualnych zdarzeniach. Do celów związanych z odtwarzaniem dźwięku w takich przypadkach stosuje się w większości odsłuch słuchawkowy. Aby osiągnąć wrażenie przestrzenności dźwięku z wykorzystaniem słuchawek stosuje się specjalne dekodery binauralne, których zadaniem jest konwersja reprezentacji dźwięku przestrzennego, np. ambisonicznego [1], zapisanego w wielu kanałach, na dwa kanały słuchawkowe (lewy i prawy). Metoda binauralnej reprodukcji dźwięku ambisonicznego [2] łączy ze sobą aspekty opisu pola akustycznego w oparciu o tak zwane harmoniki sferyczne oraz wpływ głowy i torsu na dźwięki docierające z różnych kątów w przestrzeni, określane jako HRTF (Head-Related Tranfer Function).

Jedną z podstawowych metod oceny jakości dźwięku przestrzennego w odsłuchu słuchawkowym jest zbadanie dokładności lokalizacji wirtualnych źródeł dźwięku [3]. Takie badanie pozwala zbadać, na ile dokładnie słuchacz jest w stanie wskazać, z jakiego kierunku dociera dźwięk, w danej realizacji rejestracji, dekodowania i odtwarzania dźwięku. W testach lokalizacji źródła dźwięku istotny może okazać się rodzaj bodźca wykorzystywany w badaniu [4]. Dźwięki o charakterze szerokopasmowym, takie jak głos ludzki, szumy i dźwięki otoczenia są lokalizowane lepiej niż dźwięki o charakterze tonalnym. Wpływ na percepcje lokalizacji dźwięku może mieć również to, czy wykorzystana próbka ma znaczenie emocjonalne lub semantyczne dla słuchacza.

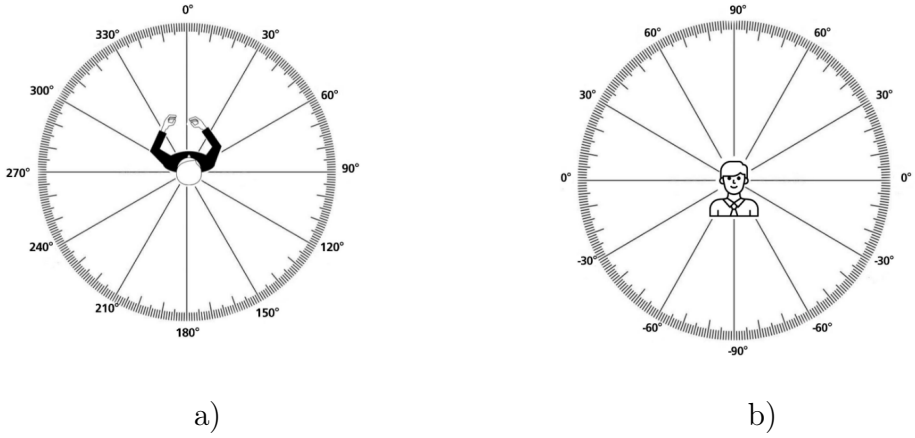
W niniejszej pracy przeprowadzono analizę wpływu rodzaju bodźca wykorzystywanego w teście lokalizacji na dokładność lokalizacji, a także zbadano wpływ treningu [5] lokalizacji na minimalizację błędów przy binauralnym odsłuchu dźwięku ambisonicznego.

2 Metodologia

Eksperyment został przeprowadzony z udziałem dwóch grup badawczych, w których skład weszli zarówno mężczyźni, jak i kobiety, znajdujący się w grupie wiekowej 18-28 lat. Proporcje płci kształtowały się na poziomie 3:1. Pierwsza grupa liczyła 14 osób i przystąpiła do testu bez wcześniejszego treningu. W skład drugiej grupy wchodziło 17 osób, które przed przystąpieniem do badania przeszły serię treningów przygotowawczy. W obu grupach znaleźli się głównie uczestnicy, którzy nie mieli doświadczenia w testach lokalizacji dźwięków przestrzennych. Badane osoby zadeklarowały brak problemów ze słuchem. Do treningu przystępowało od jednej do maksymalnie ośmiu osób w tym samym czasie.

Sesje przeprowadzane były w specjalnie przygotowanym stanowisku pomiarowym zlo-

kalizowanym w komorze bezechowej, przy użyciu bezprzewodowych słuchawek nausznych Audio Technica ATH-M50xBK. Warunki odsłuchowe zostały ustalone zgodnie z rekomendacją ITU-R BS.1116-3 [6]. W centrum komory bezechowej umieszczany był stolik, na którym znajdował się laptop i kartka A4 z testem. Każda sesja odbywała się indywidualnie, aby zapewnić maksymalną izolację i skupienie podczas testu. Treningi zostały przeprowadzone w specjalnej sali laboratoryjnej. Do określania lokalizacji źródła dźwięku skorzystano z metody wskazywania [5], polegającej na wskazaniu przez uczestnika kierunku, z którego pochodził dźwięk. Uczestnik po zidentyfikowaniu kierunku źródła dźwięku, za pomocą wzorcowej tarczy azymutalnej (rys. 1), dokonywał zapisu wyników na przygotowanym arkuszu.



a) b)

Rysunek 1: Poglądowe tarcze dla kątów a) azymutu b) elewacji

Za bodziec kontekstowy, posłużyło krótkie nagranie kobiecego głosu wymawiającego słowo "Hej", o częstotliwości podstawowej w przybliżeniu 230Hz. Wyrażenie to zostało wybrane, ze względu na jego krótki czas trwania, jego uniwersalne zrozumienie i nacechowanie emocjonalne [4]. Próbka trwała 1 sekundę i była prezentowana dwukrotnie z interwałem 1 sekundy między powtórzeniami. Bodźcem neutralnym, był wygenerowany wirtualnie szum różowy, z częstotliwością próbkowania 48kHz. Neutralny bodziec dźwiękowy składał się z czterech krótkich sygnałów szumu, trwających sumarycznie 1 sekundę, również powtarzanych dwukrotnie [5].

Po akwizycji próbek kontekstowych i neutralnych kolejną fazą było wyrenderowanie łącznie 74 bodźców wykorzystywanych do badania. Próbki były pozycjonowane względem punktu 0;0 a następnie rejestrowane w przestrzeni przez wirtualny mikrofon trzeciego

rzędu.

W ramach przeprowadzonego badania zrealizowano dwa typy testów: serię treningów oraz test końcowy. Obydwa segmenty były prowadzone przez syntezowany głos kobiecy, który w języku angielskim instruował uczestników, przedstawiając kolejne etapy badania i zapowiadając numery poszczególnych próbek dźwiękowych. W części testowej, uczestnicy mieli za zadanie określić kierunek źródła dźwięku, przy pomocy wzorcowych tarcz azymutalnych i elewacyjnych, a następnie spisać wyniki do tabeli w papierowym arkuszu. Kąty odpowiedzi w teście lokalizacji dźwięku należy interpretować zgodnie z przedstawionymi tarczami z kątami azymutu i elewacji (rys. 1)

Przed przystąpieniem do testu końcowego, jedna z grup badawczych przeszła serię pięciu sesji treningowych. Każdy trening trwał 7 minut i składał się z dwóch części: pierwszej, skoncentrowanej na lokalizacji źródła dźwięku w kątach azymutu, oraz drugiej, dotyczącej kątów elewacji. Na początku każdej części uczestnikom prezentowano dźwięki wprowadzające, zawierające zbiór próbek wykorzystywanych w teście i treningach. Dźwięki te krążyły dookoła głowy słuchacza, przygotowując go do badania i poszerzając jego percepcje przestrzenną. Następnie, rozpoczynała się właściwa część testowa, w której uczestnicy lokalizowali próbki dźwiękowe. Bodźce zawierające szum różowy i głos ludzki były prezentowane naprzemiennie. Po każdej próbce podawana była informacja zwrotna, która informowała uczestnika badania o poprawnej odpowiedzi. Pomagało to przygotować uczestnika do oficjalnego testu końcowego, oraz pozwalało zbadać własne zależności rozpoznawcze w percepcji lokalizacji źródła dźwięku.

Aby zminimalizować błędy Front-Back [7][8], czyli błędy lokalizacji przestrzennej, w których słuchacz mylnie odbiera źródło dźwięku znajdujące się przed nim jako pochodzące z tyłu, lub odwrotnie, bodźce dla kątów elewacji były prezentowane w kącie azymutu 90° lub 270° (rys. 1) w zależności od dominującego ucha osoby badanej, po czym zadaniem uczestników było określenie tylko i wyłącznie kąta elewacji danej próbki. Trening był przeprowadzany w oddzielnych salach laboratoryjnych, w której jednocześnie mogło trenować do 8 osób.

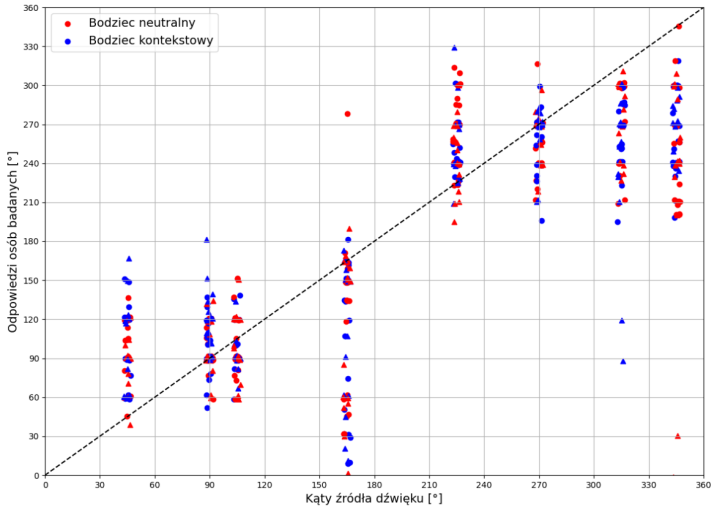
Struktura testu końcowego była bardzo podobna do treningów, jednak istotną różnicą był brak informacji zwrotnej po każdej prezentowanej próbce dźwiękowej. Pozbawienie tego elementu w teście końcowym miało na celu podniesienie poziomu wiarygodności oraz autentyczności rejestrowanych odpowiedzi.

W teście wykorzystano naprzemienną kolejność próbek, w celu maksymalizacji koncentracji uczestników, poprzez zapobieganie monotonii i krótkotrwałej adaptacji do tego samego typu bodźca. Dodatkowo starano się ograniczać następujące po sobie sąsiednie, lub blisko znajdujące się kąty z tej samej ćwiartki sfery.

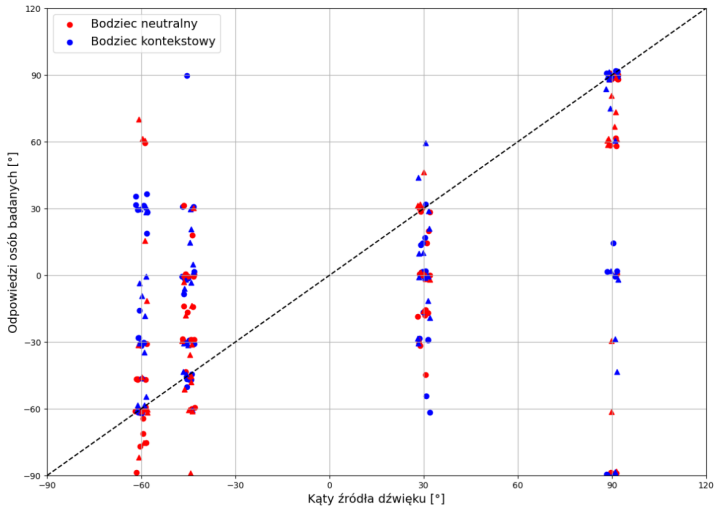
3 Analiza Wyników

W tym rozdziale zaprezentowano rezultaty przeprowadzonej analizy lokalizacji źródła dźwięku podczas binauralnego odsłuchu dźwięków ambisonicznych. Dane zostały przedstawione w zależności od rodzaju wykorzystanej próbki dźwiękowej: głosu ludzkiego oraz szumu różowego. Ponadto, rozdzielono odpowiedzi dwóch grup badanych: grupę, która uczestniczyła w testach po przeprowadzeniu serii treningów (oznaczoną jako T+T) oraz grupę, która przystąpiła do testu bez wcześniejszego treningu (oznaczoną jako T).

Po zakończeniu wszystkich sesji treningowych oraz testów przeprowadzonych w obu grupach, zebrano łącznie 788 odpowiedzi udzielonych w ramach testu końcowego. Z puli tej usunięto wyniki jednego z uczestników, który zgłosił znaczne zmęczenie poznawcze na etapie realizacji testu. Na poniższych wykresach (rys. 2, rys. 3) znajduje się rozmieszczenie wszystkich odpowiedzi uczestników w badaniu.



Rysunek 2: Dystrybucja wyników testu lokalizacji dźwięku dla kątów azymutu. Oznaczenie: kolor niebieski - bodziec kontekstowy, kolor czerwony - bodziec neutralny, koło - T+T, trójkąt - T.



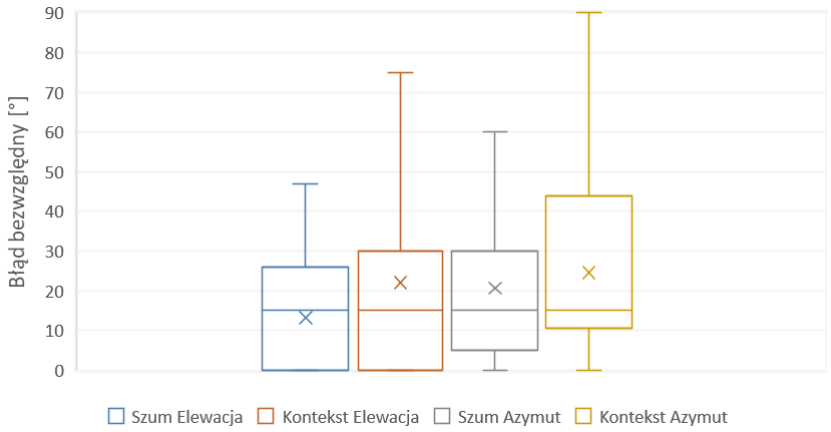
Rysunek 3: Dystrybucja wyników testu lokalizacji dźwięku dla kątów elewacji. Oznaczenie: Kolor niebieski - bodziec kontekstowy, Kolor czerwony - bodziec neutralny, Koło - T+T, Trójkąt - T.

Podczas analizy dystrybucji wskazywanych kątów podczas testu lokalizacji źródła dźwięku dla kątów azymutu (rys. 2) zaobserwowano zróżnicowane odpowiedzi w zależności od prezentowanego kąta. W przypadku kąta 45° zaobserwowano skłonność do lokalizacji dźwięku w obszarze znajdującym się wewnątrz stożka dezorientacji, czyli pomiędzy kątem źródła dźwięku a jego odbiciem lustrzanym (135°). Najwyższą precyzję lokalizacji odnotowano dla kąta 90°, co można tłumaczyć brakiem symetrycznego odpowiednika tego kąta w tej samej półkuli, co czyni go kątem łatwiejszym do lokalizowania. W przypadku kąta 165° zaobserwowano znaczącą ilość błędów Front-Back. Dla kątów 225°, 315° i 345° można zaobserwować podobną zależność jak dla kąta 45°, gdzie wszystkie odpowiedzi znajdują się wewnątrz stożka dezorientacji. Dla kąta 270°, znajdującego się bezpośrednio po lewej stronie słuchacza również zaobserwowano większą dokładność w określaniu źródła dźwięku, podobnie jak w przypadku kąta 90°.

Dystrybucja wyników testu lokalizacji źródła dźwięku dla kątów elewacji (rys. 3) również wykazała pewne schematy w odpowiedziach badanych uczestników. Najbardziej istotnym aspektem jest rozmieszczenie bodźców neutralnych bliżej linii diagonalnej, co wskazuje na wyższą dokładność właściwych kątów w porównaniu do bodźców kontekstowych. Największa liczba błędów Up-Down jest zauważalna dla kąta 30°. W przypadku kąta 90° istnieje też taka zależność wśród słuchaczy, że lokalizują dźwięk próbki kontekstowej w kącie 0° o wiele częściej, niż próbkę neutralną. Podobnie jak w przypadku analizy kątów

azymutu, dla elewacji również obserwujemy tedencję uczestników do postrzegania źródeł dźwięku wewnątrz stożka dezorientacji

W ramach zbadania wpływu próbki kontekstowej na percepcje lokalizacji źródła dźwięku, w badaniu dokonano porównania błędów bezwzględnych odpowiedzi dla wszystkich próbek, z podziałem na kąty azymutu i elewacji. Z analizy wykluczono próbki, dla których wystąpiły błędy Front-Back.

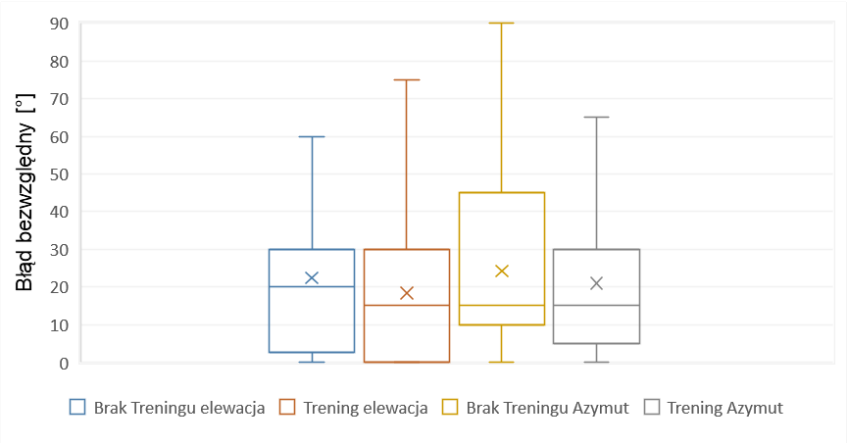


Rysunek 4: Zestawienie błędów bezwzględnych z wszystkich kątów w zależności od rodzaju bodźca

Podczas analizy błędów bezwzględnych lokalizacji dźwięku, uwzględniając różne rodzaje bodźców (rys. 4) oraz kąty azymutu i elewacji, ujawniła się interesująca zależność. Chociaż wartości median dla poszczególnych grup były podobne, wartości średnie wykazywały wyraźne różnice. Wskazuje to na to, że w kontekście analizy danych dotyczących lokalizacji dźwięku, średnia może lepiej odzwierciedlać ogólną tendencję obserwowaną w danych, podkreślając znaczenie skrajnych wartości w analizie. Mediana może nie odzwierciedlać pełnego rozkładu błędów, skupiając się tylko na środkowych wartościach zestawu danych. Zestawiając wykresy błędów w kątach azymutu, można zaobserwować wyższe wartości skrajne błędów dla próbki kontekstowej niż w przypadku próbki neutralnej, jednakże zarówno mediany i średnie są do siebie bardzo zbliżone, co może sugerować brak wpływu próbki kontekstowej na percepcje lokalizacji dźwięku w kątach azymutu.

Aby zbadać wpływ treningu na percepcje lokalizacji źródła dźwięku, w badaniu porównano błędy bezwzględne odpowiedzi osób, które podeszły bezpośrednio do testu, z błędami osób, które przeszły trening, dla różnych kątów azymutu.

W procesie analizy błędów kątowych bezwzględnych dla lokalizacji dźwięku, uwzględniając kąty azymutu i elewacji, dla grup z treningiem (T+T) oraz bez treningu (T) (rys.



Rysunek 5: Zestawienie błędów bezwzględnych z wszystkich kątów dla próbek treningu i jego braku

5), można zaobserwować pozytywny wpływ treningu na dokładność lokalizacji dźwięku. Parametry średnich i median prezentują nieznacznie niższe wartości dla grupy T+T, co świadczy o umiarkowanym zwiększeniu dokładności odpowiedzi uczestników. Niemniej jednak, obserwowana poprawa, choć istotna, nie jest tak znacząca jak mogłoby wynikać z początkowych oczekiwań.

Pod koniec badania przeprowadzono analizę wariancji w celu zbadania istotności statystycznej wyników. Rozdzielono odpowiedzi uczestników na dwie kategorie (badające wpływ treningu oraz rodzaju próbki na dokładność lokalizacji dźwięku), a następnie wyznaczone błędy bezwzględne odpowiedzi rozdzielono na płaszczyznę wertykalną i horyzontalną.

Przeprowadzona analiza wariancji potwierdziła obserwację z median i średnich błędów bezwzględnych odpowiedzi. W przypadku analizy wpływu bodźca kontekstowego w płaszczyznach azymutu, współczynnik F-Value, wyniósł 1.18, a współczynnik istotności statystycznej P-Value, wyniósł ponad 0.27. Wartości te oznaczają mało znaczące różnicę miedzy grupami a więc brak wpływu próbki kontekstowej na dokładność lokalizacji dźwięku przestrzennego w kątach azymutu. Natomiast dla kątów elewacji, wartość F-value wyniosła 5.17, a p-value w przybliżeniu 0.02. Niskie wartości współczynnika istotności statystycznej, przy wysokim współczynniku wariancji międzygrupowej do wewnątrzgrupowej potwierdzają statystyczną istotność rodzaju próbki na dokładność lokalizacji dźwięku przy odsłuchu binauralnego dźwięku ambisonicznego.

Dodatkowa analiza wariancji przeprowadzona dla grup T (test bez treningu) oraz T+T (test z treningiem) z podziałem na kąty azymutu i elewacji wykazała istotność wpływu tre-

ningu na dokładność lokalizacji dźwięku przestrzennego, ale tylko w odniesieniu do kątów azymutu. Wartość F-Value dla tej płaszczyzny, przekraczająca 4.28, oraz p-value poniżej 0.05 wskazują na znaczący wpływ treningu na poprawę percepcji lokalizacji dźwięku. Oznacza to, że uczestnicy, którzy przeszli serię treningową, wykazali lepszą zdolność do precyzyjnego określania położenia źródła dźwięku w płaszczyźnie azymutu.

Dla kątów elewacji, mimo obserwacji niższych wartości średnich i median wśród uczestników po treningu, wartości F-Value wynoszące zaledwie 1.21 i wysokie wartości współczynnika istotności statystycznej P-Value przekraczające 0.27, nie wykazały istotnych różnic między grupami. Takie wyniki sugerują, że trening nie miał statystycznie istotnego wpływu na dokładność lokalizacji dźwięku w płaszczyźnie elewacji. Może to wskazywać na to, że umiejętności lokalizacyjne w tej płaszczyźnie są trudniejsze do poprawy poprzez krótkotrwałe treningi

Niska wartość dla kątów elewacji mogła wynikać ze zbyt krótkich przerw miedzy treningami wśród grupy badanej, co powodowało wyższe zmęczenie poznawcze, co przełożyło się na wyniki testu lokalizacji źródła dźwięku.

4 Podsumowanie

Przeprowadzone badanie miało na celu zbadanie wpływu bodźca kontekstowego na dokładność lokalizacji przy odsłuchu słuchawkowym. Dodatkowo zbadano wpływ treningu na minimalizację błędów określania położenia źródła dźwięku.

W ramach realizacji tego zagadnienia zaprojektowano test lokalizacji źródła dźwięku, który został podzielony na dwie części: azymutu i elewacji. Opracowano również serię pięciu treningów, które, w przeciwieństwie do testu, zawierały informację zwrotną na temat prezentowanego kąta. Badanie zostało przeprowadzone na dwóch grupach liczących łącznie 31 osób z których pierwsza przystąpiła do testu z poprzedzającą go serią treningów, a druga grupa podeszła do testu bezpośrednio.

Analiza odpowiedzi uczestników wykazała skłonność do umieszczania źródła dźwięku w stożku dezorientacji, czyli obszarze między rzeczywistym źródłem a jego odbiciem lustrzanym. Zaobserwowano również wysoki odsetek błędów typu Front-Back oraz Up-Down. Po wykluczeniu tych błędów z analizy, zbadano wartości bezwzględne odpowiedzi w odniesieniu do rzeczywistego źródła dźwięku.

Porównując wpływ bodźca kontekstowego oraz neutralnego zaobserwowano, że szum różowy był lepiej lokalizowany niż próbka głosu ludzkiego. Widoczna była poprawa war-

tości średnich i median na korzyść szumu różowego dla kątów elewacji, jednakże dla kątów azymutu nie można było jednoznacznie określić predyspozycji względem danego rodzaju próbki, co zostało potwierdzone przeprowadzoną następnie analizą wariancji. Badając wpływ treningu na minimalizację błędów określania położenia dźwięku zaobserwowano nieznaczną poprawę wartości średnich dla grupy, która przeszła serię treningów przed testem, w porównaniu do grupy, która podjęła się testu bezpośrednio. Różnice między grupą która odbyła trening, a grupą która podeszła tylko do testu zostały określone jako znaczące tylko i wyłącznie w kątach azymutu, a dla kątów elewacji analiza nie wskazała na wyraźne różnice międzygrupowe.

Wysoka częstotliwość występowania błędów Front-Back i Up-Down mogła wynikać z wykorzystania uniwersalnych funkcji HRTF, stworzonych dla manekina akustycznego, którego kształt mógł wyraźnie różnić się od kształtu głowy uczestników, przez co dźwięki mogły być słyszane z innych kątów niż w rzeczywistości. Jedynie niewielka poprawa dokładności lokalizacji dla grupy, która przeprowadziła trening, mogła wynikać z zbyt krótkich przerw pomiędzy treningami, co mogło prowadzić do zawyżonego zmęczenia poznawczego. Zaniżona wartość dokładności lokalizacji dla próbki kontekstowej mógł wynikać też z samej jej struktury. Słowo "Hej" miało długo trwającą samogłoskę, co powodowało że próbka mogła mieć bardziej tonalny charakter, co mogło skutkować mniej dokładną lokalizacją próbek w przestrzeni.

Literatura

[1] D. Arteaga, "Introduction to ambisonics,"Escola Superior Politècnica, Universitat Pompeu Fabra, Barcelona, Spain, 2015, pp. 6-8.

[2] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares,"in Proceedings of the DAGA, vol. 44, 2018.

[3] J. Czajka and M. Niewiarowicz, "Localization of sound sources. Theoretical foundations and results of experimental investigations,"Postępy w chirurgii głowy i szyi/Advances in Head and Neck Surgery, vol. 4, no. 1, pp. 25-42.

[4] L. Fostick and N. Fink, "Situational awareness: The effect of stimulus type and hearing protection on sound localization,"Sensors, vol. 21, no. 21, 7044, 2021.

[5] S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners,"Hearing Research, vol. 114, no. 1-2, pp. 179-196, 1997.

[6] I. T. U. R. Recommendation, "1116-3. Methods for the Subjective Assessment of Small Impairments in Audio Systems,"International Telecommunication Union, Geneva, Switzerland, 2015.

[7] T. R. Letowski and S. T. Letowski, "Auditory spatial perception: Auditory localization,"Army Research Laboratory, Aberdeen Proving Ground, MD, Human Research and Engineering Directorate, 2012.

[8] T. Fischer, M. Caversaccio, and W. Wimmer, "A front-back confusion metric in horizontal sound localization: The fbc score,"in ACM Symposium on Applied Perception, 2020.

Patroni

Patroni medialni

Organizatorzy

Sponsorzy