https://doi.org/10.7494/978-83-67427-06-7_10

Wpływ modyfikacji fazy sygnału na percepcję pogłosowości w nagraniach audio

Teresa Makuch, Piotr Kleczkowski

AGH Akademia Górniczo-Hutnicza w Krakowie,
Katedra Mechaniki i Wibroakustyki

tmakuch@agh.edu.pl

Streszczenie

Zależności fazowe pomiędzy składowymi częstotliwościowymi fali akustycznej są w kontekście percepcji dźwięku zazwyczaj kojarzone przede wszystkim z barwą dźwięku oraz lokalizacją źródła fali o niskiej częstotliwości. Istnieją jednak badania sugerujące, że zmiany fazy w czasie mogą powodować odczucie pogłosowości. W artykule przedstawiono pierwszy etap badań nad wywoływaniem wrażenia pogłosowości przez losowe zmiany fazy składowych sygnału. W nagraniach bezechowych modyfikowano fazę w ramach przyjętych ograniczeń. Następnie przeprowadzono nieformalne i formalne testy słuchowe. Badano kilka parametrów modyfikacji, takich jak szerokość pasma, zakres zmian fazy oraz częstość ich występowania; niektóre z nich mają istotny wpływ na odczucie pogłosowości.

1. Wprowadzenie

Powszechnie stosowana w akustyce analiza widmowa dźwięku dostarcza informacji o amplitudach poszczególnych składowych oraz zależnościach fazowych pomiędzy nimi. Jednak zdecydowanie więcej uwagi poświęcane jest amplitudzie i energii sygnału (zazwyczaj mówimy o sygnale napięciowym, zarejestrowanym przez przetwornik elektroakustyczny) niż fazie. Zależności fazowe są uwzględniane m.in. przy projektowaniu filtrów – w tych przypadkach dąży się do osiągnięcia liniowej charakterystyki fazowej, aby możliwie zredukować zniekształcenia fazowe oraz wprowadzenie zakolorowania dźwięku.

Zależności fazowe są pomijane zazwyczaj w akustyce pomieszczeń – do jej oceny stosowane są głównie metody energetyczne, zarówno podczas pomiarów, jak i w symulacjach (metody geometryczne, analiza modalna). Należy jednak pamiętać, że jeden z kluczowych parametrów akustycznych powierzchni, fizyczny współczynnik pochłaniania dźwięku α, wywodzi się ze współczynnika odbicia fali R, który z kolei ma bezpośredni związek z impedancją akustyczną powierzchni [1]. Ta zaś jest wielkością zespoloną, z członem amplitudowym i fazowym. W chwili odbicia fali dźwiękowej od powierzchni następuje zmiana fazy sygnału; wartość wprowadzanego przesunięcia zależna jest od parametrów materiału, częstotliwości fali oraz kąta padania [2]. W polu pogłosowym przyjmuje się, że każdy kierunek padania fali jest jednakowo prawdopodobny, co oznaczałoby, że również zmianę fazy można potraktować jako zmienną losową. Efektem tych przybliżeń jest założenie o pomijalności zmiany fazy przy odbiciu; prowadzone są jednak badania związane z uwzględnieniem fazy np. w metodzie promieniowej [3]. W połączeniu

z rosnącymi mocami obliczeniowymi być może pozwoli to na korzystanie z dokładniejszych symulacji, uwzględniających także zjawiska fazowe.

Jednym z argumentów za pomijaniem zmian fazy jest ich znikoma zauważalność przez człowieka. Zgodnie z obecnym stanem wiedzy dla człowieka bardziej zauważalne są zmiany sygnału związane z amplitudą składowych, mniej z ich przesunięciami w czasie. Najlepiej zbadanym z mechanizmów percepcyjnych związanych z zależnościami fazowymi jest międzyuszna różnica czasów/faz (ITD, *interaural time difference*), którą możemy rozpatrywać tylko w kontekście pary sygnałów. Na podstawie różnicy w czasie dotarcia sygnału do lewego i prawego ucha układ słuchowy określa kierunek (w płaszczyźnie horyzontalnej), z którego dźwięk został wyemitowany. Mechanizm ITD jest najskuteczniejszy dla fal sinusoidalnych o częstotliwościach poniżej 1500 Hz; działa też w przypadku dźwięków złożonych, zawierających wyższe częstotliwości – wówczas jego skuteczność zależy od struktury czasowej (obwiedni czasowej) dźwięku. Jeśli obwiednia sygnału wykazuje periodyczność z powtarzalnością niższą niż 600 Hz, dźwięki mogą być lokalizowane także na podstawie ITD [4]. Poza tym aspektem percepcyjnie zjawiska fazowe przekładają się najbardziej na odbiór barwy dźwięku; w zależności od struktury dźwięku, przesunięcia między jego składowymi mogą być postrzegane jako szorstkość lub wywoływać wrażenie obecności dodatkowego sygnału tonalnego w widmie harmonicznym [5].

W kontekście odbioru przestrzeni dźwiękowej zależności fazowe nie są zbyt dobrze zbadane. Na podstawie subiektywnych eksperymentów D. Griesingera można wysnuć hipotezę, że relacje fazowe pomiędzy składowymi dźwięku o strukturze harmonicznej mają wpływ na to, czy jest on postrzegany jako pogłosowy, odległy od słuchacza. Wiąże się to z oceną pola akustycznego na podstawie stosunku dźwięku bezpośredniego i pogłosowego (DRR, *direct to reverberant ratio*) [6, 7]. Zgodnie z tą teorią w dźwięku pogłosowym pierwotne relacje między harmonicznymi są zaburzone na skutek odbić. Właściwości powierzchni odbijającej zmieniają się w funkcji częstotliwości, przesunięcia fazowe są zatem różne dla poszczególnych składowych. To z kolei wywołuje wrażenie „rozmytego" dźwięku, utratę klarowności.

Przedmiotem badań opisanych w niniejszym artykule jest zbadanie, czy randomizacja fazy sygnału faktycznie wywołuje wrażenie większego pogłosu. W tym celu w nagraniach dźwiękowych dokonano modyfikacji fazy bez ingerencji w amplitudę sygnału, a następnie przeprowadzono formalne testy odsłuchowe z udziałem 31 osób. W kolejnej sekcji opisano, w jaki sposób przygotowane zostały sygnały testowe; następnie przedstawiono przebieg i wyniki badań psychoakustycznych.

2. Metodyka

2.1. Przygotowanie próbek

W badaniach psychoakustycznych często wykorzystywane są dźwięki syntetyczne – pozwala to na precyzyjną kontrolę wszystkich elementów sygnału, jednak bywa nienaturalne dla słuchaczy. Z tego powodu zdecydowano się na wykorzystanie nagrań „żywych" sygnałów – instrumentów muzycznych oraz mowy.

Przy wyborze metody ich przetwarzania na potrzeby badań priorytetem była możliwość precyzyjnej zmiany fazowej przy minimalnej (najlepiej żadnej) ingerencji w amplitudę. Z tego powodu zdecydowano się na skorzystanie z transformaty Fouriera (STFT). W każdym okienku przekształcano otrzymaną oryginalną transformatę do postaci trygonometrycznej, po czym modyfikowano w niej jedynie widmo fazowe sygnału, pozostawiając widmo amplitudowe bez zmian. Tym samym zastosowano rodzaj filtracji wszechprzepustowej.

Widmo oryginalnego sygnału w m-tym okienku można zatem opisać wzorem:

$$\hat{S}_{0m(f)} = |\hat{S}_{0m(f)}| e^{i\varphi_{0m}(f)} \tag{1}$$

a po wprowadzeniu przesunięcia fazowego zależnością:

$$\hat{S}_{m(f)} = |\hat{S}_{0m(f)}| e^{i(\varphi_{0m}(f)+\Delta\varphi_m(f))} \tag{2}$$

gdzie:

φ_{m0} – oryginalna faza sygnału,
$\Delta\varphi_m$ – wprowadzane przesunięcie fazowe.

Wartość przesunięcia losowano z przedziału $\langle -\Delta\varphi_{max}; \Delta\varphi_{max}\rangle$ dla co drugiego pasma częstotliwości – następnie wartości przesunięć w pozostałych pasmach były interpolowane wielomianem 3. stopnia z uwzględnieniem wartości maksymalnych (zmodyfikowana metoda interpolacji Akimy) [8]. Taki zabieg miał na celu uniknięcie skokowych przesunięć między sąsiadującymi pasmami częstotliwości.

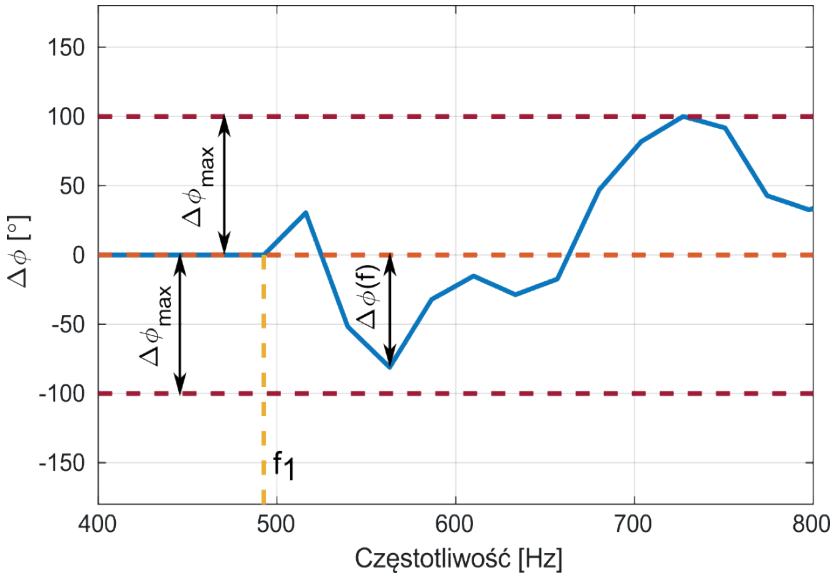
Drugim ograniczeniem zakresu losowania była różnica między przesunięciami w kolejnych oknach – jego wprowadzenie pozwoliło na uniknięcie bardzo gwałtownych zmian, które mogłyby być słyszane jako zniekształcenia sygnału. Te dwa warunki na losowaną wartość $\Delta\varphi_m(f)$ można zapisać jako:

$$\begin{cases} |\Delta\varphi_m(f)| \le \Delta\varphi_{max} \\ |\Delta\varphi_m(f) - \Delta\varphi_{m-1}(f)| \le \Delta\varphi_{win} \end{cases} \tag{3}$$

gdzie:

$\Delta\varphi_{max}$ – maksymalna dopuszczalna zmiana fazy,
$\Delta\varphi_{win}$ – maksymalna dopuszczalna różnica między kolejnymi oknami.

Na rysunku 1 przedstawiono przykładowy wynik losowania w jednym oknie.



Rys. 1. Przykład uzyskanych zmian fazy dla jednego okna sygnału; zaznaczono amplitudę zmian $\Delta\varphi_{max}$ (przedział losowania), dolną granicę zakresu częstotliwości podlegającego modyfikacjom f_1 oraz wartość wylosowaną w przypadku wybranego pasma częstotliwości $\Delta\varphi(f)$

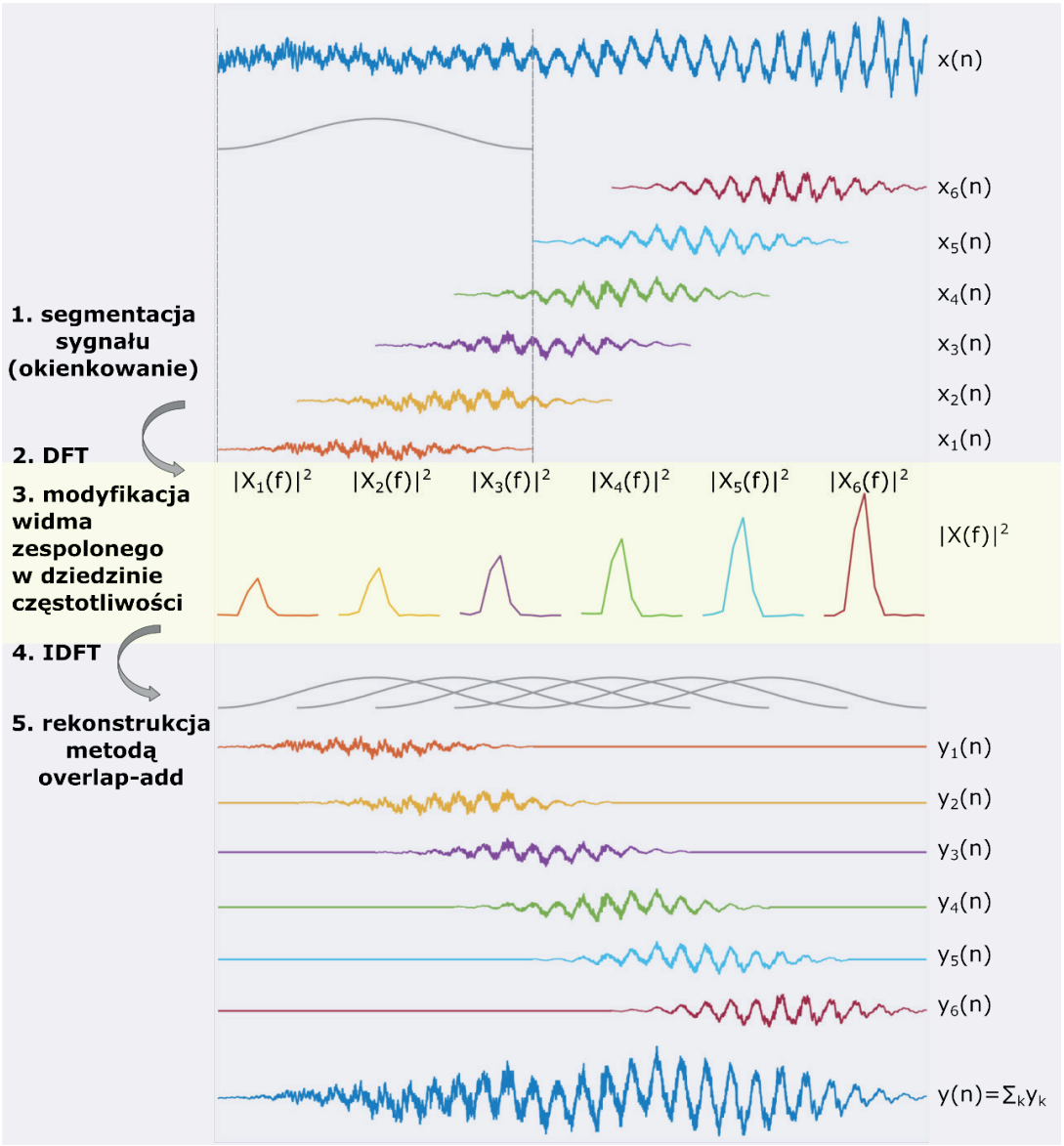
W kolejnym kroku dokonywano powrotnego przekształcenia zmodyfikowanego okienka transformaty do postaci algebraicznej:

$$\hat{S}_{m(f)} = Re_{m(f)} + \mathrm{i} \cdot Im_{m(f)} \tag{4}$$

gdzie:

$$\begin{aligned} Re_{m(f)} &= |\hat{S}_{0m(f)}|\cos(\varphi_{0m}(f) + \Delta\varphi_m(f)) \\ Im_{m(f)} &= |\hat{S}_{0m(f)}|\sin(\varphi_{0m}(f) + \Delta\varphi_m(f)). \end{aligned} \tag{5}$$

Następnie, korzystając z odwrotnej transformaty Fouriera i metody nakładkowania (*overlap-add*), rekonstruowano sygnał w dziedzinie czasu. Schemat analizy i rekonstrukcji sygnału z uwzględnieniem zastosowanych przekształceń przedstawiono na rysunku 2. Nakładkowanie okien wynosiło 75%, a długość okna w zależności od wariantu (opisane dalej) 1024 próbki lub 2048 próbek (tj. odpowiednio 21,3 ms i 42,7 ms – częstotliwość próbkowania w przypadku wszystkich sygnałów wynosiła 48 kHz).



Rys. 2. Schemat analizy i rekonstrukcji sygnału za pomocą STFT (na podstawie [9])

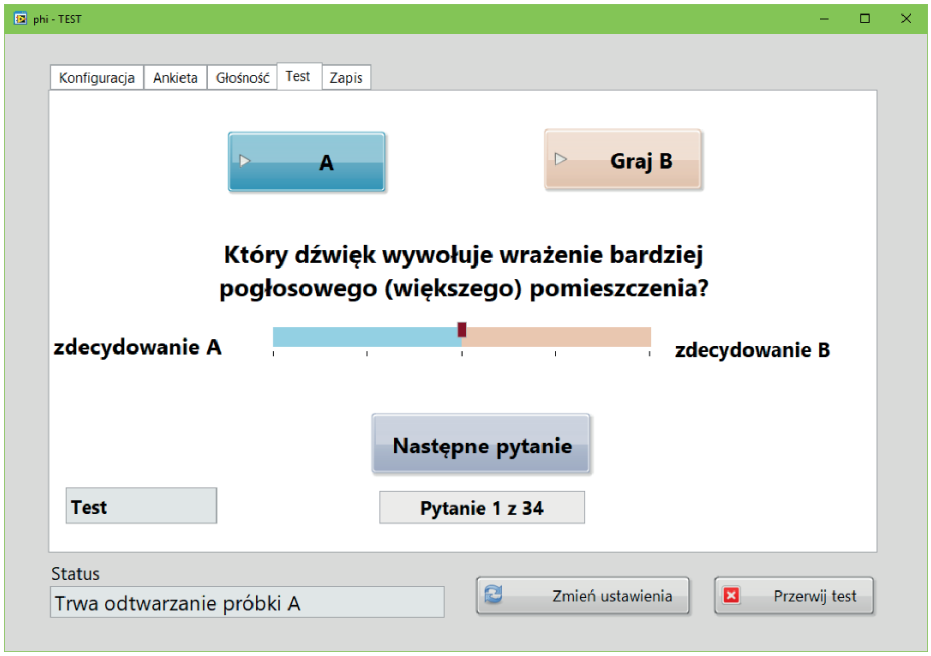
2.2. Badanie psychoakustyczne

Aby sprawdzić, czy wprowadzane modyfikacje wywołują wrażenie większego pogłosu, przeprowadzono test oparty na metodzie porównania dwóch bodźców (*paired scaling*), z zastosowaniem skali interwałowej [10]. Na potrzeby badania przygotowano aplikację komputerową w środowisku LabVIEW, służącą do przeprowadzenia testów, a także zbierania metryki statystycznej (wiek słuchacza, doświadczenie muzyczne, płeć). Podczas testu w każdej próbie słuchacze mieli za zadanie wysłuchać dwóch próbek dźwiękowych, a następnie odpowiedzieć na pytanie, **który dźwięk wywołuje wrażenie bardziej pogłosowego (większego) pomieszczenia.** Grupa badana składała się zarówno z osób zawodowo związanych z akustyką bądź muzyką, jak i mniej obeznanych w tej tematyce. Pytanie musiało być jasne dla wszystkich uczestników badania; określenie „większe" skierowane było do osób mniej obeznanych z terminologią akustyczną, gdyż zazwyczaj pogłos jest kojarzony z pomieszczeniem o dużej kubaturze; w przypadku zgłoszenia przez słuchacza wątpliwości zawsze precyzowano, że ocenianą cechą jest pogłosowość. Na rysunku 3 przedstawiono interfejs użytkownika, z którego korzystali słuchacze. Skalę odpowiedzi przedstawiono w tabeli 1 – górny wiersz to odpowiedzi dostępne dla słuchaczy, a dolny – ich interpretacja liczbowa na potrzeby analizy wyników.

Tabela 1

Reprezentacja liczbowa skali ocen użytej w testach odsłuchowych

zdecydowanie A	raczej A	brak różnicy	raczej B	zdecydowanie B
–2	–1	0	1	2



Rys. 3. Interfejs użytkownika aplikacji wykorzystywanej do testów odsłuchowych

Słuchacze mieli do czynienia z dwiema następującymi grupami bodźców: dźwiękami instrumentów oraz nagraniami mowy. W badaniach wykorzystano nagrania instrumentów pochodzące ze zbiorów autorów, które zostały wykonane w studiu nagrań Katedry Mechaniki i Wibroakustyki AGH, w warunkach śladowego pogłosu z bliskim mikrofonowaniem [11]. Są to kilkudźwiękowe sekwencje o niewielkiej rozpiętości wysokości, zagrane na flecie (Fl), oboju (Ob) i wiolonczeli (Vc) przez praktykujących muzyków. Użyte sygnały mowy pochodzą z dostępnej na licencji *Creative Commons* bazy nagrań *TSP Speech Database* przygotowanej na McGill University [12]; każde nagranie to jedno zdanie w języku angielskim, zarejestrowane w komorze bezechowej. Do badań wykorzystano jedno zdanie wypowiedziane przez kobietę (K) i jedno przez mężczyznę (M).

Przy wyborze sygnałów do badań kierowano się strukturą harmoniczną dźwięków oraz zakresem częstotliwości. W badaniach wykorzystano nagrania o najniższych częstotliwościach – odpowiednio flet – 297 Hz, obój – 293 Hz, wiolonczela – 98 Hz, głos kobiecy – 124 Hz, głos męski – 105 Hz. Flet i obój są instrumentami o wyraźnej strukturze harmonicznej dźwięku, natomiast widmo dźwięku wiolonczeli jest bogate w składowe przy jednoczesnej możliwości uzyskania dźwięków o niskiej częstotliwości. Sygnały mowy zostały włączone z dwóch powodów. Po pierwsze ich struktura czasowo-częstotliwościowa jest inna niż długich dźwięków instrumentów, gdyż w naturalnej wypowiedzi zmiany formantów następują często, z każdą wypowiadaną głoską. Po drugie w swoim eksperymencie D. Griesinger zastosował randomizację fazy właśnie na sygnale mowy, został on więc wybrany w celu weryfikacyjnym.

Pytania były podzielone na cztery kategorie. W pierwszej porównywano wpływ długości okna STFT (10 prób), w drugiej maksymalne odchylenie od pierwotnej wartości fazy (15 prób). W kategoriach trzeciej i czwartej badano odpowiednio maksymalną zmianę fazy pomiędzy kolejnymi oknami (cztery próby) i zakres częstotliwości, w którym faza była modyfikowana (pięć prób). Poza kategorią pierwszą wszędzie stosowano okno o długości 2048 próbek (42,7 ms). W teście znalazły się też trzy próby kontrolne, w których obie próbki były tym samym sygnałem, wplecione między pozostałe pytania. Próbki były ułożone losowo (w części przypadków bardziej zmodyfikowane było nagranie A, w części B), tak samo dla wszystkich słuchaczy; podobnie kolejność prób była jednakowa. Próby były pogrupowane według

wyżej wymienionych kategorii, a w obrębie każdej kategorii według źródła dźwięku; nie w każdej kategorii wykorzystano wszystkie dźwięki. Dokładne zestawienie użytych próbek przedstawiono w kolejnych podrozdziałach razem z wynikami.

Przed przystąpieniem do testu słuchacze wypełniali krótki kwestionariusz w celach statystycznych. Potem przystępowali do krótkiej sesji treningowej składającej się z pięciu prób, po jednej parze próbek każdego z użytych w badaniu sygnałów. Podczas treningu uczestnicy mieli możliwość zapoznania się z interfejsem aplikacji oraz specyfiką badanych próbek, a także zadawania pytań. Wyniki z tej sesji nie były zapisywane. Po niej następowało właściwe badanie, składające się z 34 prób. Słuchacze wiedzieli jedynie, że mają za zadanie porównać dźwięki i odpowiedzieć na pytanie (nie znali procedury przygotowania testu).

Czas trwania jednej próbki dźwiękowej wynosił 2–5 s (zależnie od sygnału – próbki mowy były krótsze). W takim czasie słuchacze mieli możliwość ocenić różnicę, a jednocześnie próbki były wystarczająco krótkie, aby je porównać. Każdą próbkę można było odtworzyć dowolną liczbę razy, jednak po udzieleniu odpowiedzi nie było możliwości powrotu do poprzednich pytań. Wypełnienie całego testu (kwestionariusz, pięć prób treningowych i 34 próby testowe) przeważnie zajmowało badanym około 20 minut.

Badanie odbywało się w sali komputerowej laboratorium wibroakustyki Katedry Mechaniki i Wibroakustyki AGH, w warunkach umożliwiających skupienie na badaniu. Wszyscy słuchacze korzystali z takiego samego wyposażenia, czyli słuchawek studyjnych Beyerdynamic DT770 Pro 250 Ω podłączonych do karty dźwiękowej zintegrowanej z płytą główną komputera Analog Devices.

W badaniu brało udział 31 osób w wieku od 16 do 58 lat, z czego 87% poniżej 28 roku życia; 58% badanych miało doświadczenie/wykształcenie muzyczne lub związane z inżynierią dźwięku. Trzy osoby zgłosiły, że zauważają u siebie problemy ze słuchem, pozostali ocenili swój słuch jako dobry. Słuchacze nie byli poddawani badaniu audiometrycznemu.

3. Wyniki

3.1. Analiza statystyczna

Na potrzeby analizy statystycznej przyjęto, że zawsze próbka A była sygnałem, w którym wprowadzono mniejsze modyfikacje fazy (lub nie wprowadzono żadnych) – odpowiedzi przeskalowano adekwatnie do tego założenia. W analizie wykorzystano test znakowanych rang Wilcoxona, który jest nieparametrycznym testem mediany i nie wymaga założenia o ciągłym rozkładzie badanej zmiennej [13]. Na poziomie ufności 95% testowano następujące hipotezy statystyczne:

- H_0: $Me = 0$ – udzielone odpowiedzi nie wskazują na występowanie różnic percepcyjnych między próbkami;
- H_1: $Me \neq 0$ – odpowiedzi wskazują na możliwość wystąpienia istotnych statystycznie różnic percepcyjnych między próbkami.

Wyznaczony w analizie współczynnik istotności p porównywano z przyjętym poziomem istotności $\alpha = 0{,}05$. Jeśli $p \geq \alpha$, to nie było podstaw do odrzucenia hipotezy zerowej. W przeciwnym razie uznawano, że zauważone przez słuchaczy różnice między próbkami są statystycznie istotne.

3.2. Próby kontrolne

Próby kontrolne potraktowano jako wyznacznik wiarygodności wyników danego słuchacza – odrzucono wyniki wszystkich badanych, którzy przynajmniej w dwóch próbach z trzech prób kontrolnych wskazali różnicę między dźwiękami, tworząc w ten sposób zawężony zbiór wyników. W każdej kategorii wykonano dwie jednakowe analizy – jedną na pełnej próbie słuchaczy, drugą na zawężonym zbiorze. W tabeli 2 przedstawiono wyniki testów istotności dotyczące prób kontrolnych z podziałem na zbiór pełny i zawężony. Jak widać, eliminacja części wyników poprawiła spójność odpowiedzi, nie wpłynęła jednak na wyniki testów istotności w pozostałych kategoriach.

Tabela 2

Wyniki analizy statystycznej w przypadku porównania jednakowych sygnałów.
Kolejno podano: m_1 – średnia arytmetyczna, *Me* – mediana, *W* – wartość statystyki Wilcoxona, *p* – współczynnik istotności (*p-value*)

Próba	Zbiór pełny, *N* = 31				Zbiór zawężony, *N* = 14			
	m_1	*Me*	*W*	*p*	m_1	*Me*	*W*	*p*
M 0(0)	–0,23	0	24	0,1431	–0,07	0	0	1,0000
Ob 0(0)	–0,03	0	64	0,8734	–0,07	0	0	1,0000
Vc 0(0)	0,42	0	201	**0,0525**	0,07	0	12	1,0000

3.3. Długość okna

Wstępne, nieformalne testy wykazały, że przy czasie trwania okna 21,3 ms (1024 próbki) nawet duże modyfikacje fazy są trudno zauważalne. Dlatego zbadano słyszalność zmian w przypadku tych samych parametrów zmian fazy, a różnych długości okna. Wyniki analizy statystycznej przedstawiono w tabeli 3. Przy opisie nagrań jako pierwszy parametr próbki podana jest amplituda zmian fazy w stopniach (zakres losowania), jako drugi (w nawiasie) – maksymalna dopuszczalna zmiana między kolejnymi oknami. W opisie każdej próbki dźwiękowej (A i B) podano długość okna (w milisekundach). Istotną różnicę między prezentowanymi próbkami słuchacze zauważyli w dwóch próbach z 10 (w przypadku zawężonego zbioru wyników – w trzech próbach). Kolorem szarym wyróżniono próby, w których wskazane przez słuchaczy różnice percepcyjne między nagraniami są statystycznie istotne.

Tabela 3

Wyniki analizy statystycznej w przypadku zestawienia długości okna analizy sygnału; symbole jak w tabeli 2

Próba			Zbiór pełny, *N* = 31				Zbiór zawężony, *N* = 14			
nagranie	próbka A	próbka B	m_1	*Me*	*W*	*p*	m_1	*Me*	*W*	*p*
K 90(10)	21,3 ms	42,7 ms	0,29	0	131	0,1511	0,50	1	37	0,1211
K 120(20)	21,3 ms	42,7 ms	0,19	0	119	0,3824	0,57	0	21	0,0313
Ob 90(10)	21,3 ms	42,7 ms	–0,13	0	47	0,5065	–0,07	0	2	1,0000
Ob 120(20)	21,3 ms	42,7 ms	–0,87	–1	27	0,0001	–0,79	–1	0	0,0039
Vc 120(20)	21,3 ms	42,7 ms	–0,55	–1	66	0,0131	–0,79	–1	5	0,0098
Vc 90(10)	21,3 ms	42,7 ms	0,10	0	109	0,5878	–0,07	0	6	1,0000
M 90(10)	21,3 ms	42,7 ms	0,23	0	130	0,3934	0,21	0	15	0,5313
M 120(20)	21,3 ms	42,7 ms	0,29	0	184	0,1575	0,36	0	33	0,2578
Fl 120(20)	21,3 ms	42,7 ms	–0,32	0	94	0,3166	–0,43	0	4	0,1094
Fl 90(10)	21,3 ms	42,7 ms	–0,16	0	63	0,3863	–0,07	0	2	1,0000

3.4. Maksymalna zmiana fazy

W tabeli 4 przedstawiono wyniki analizy statystycznej dotyczące kategorii drugiej, związanej z maksymalną wprowadzaną zmianą fazy (opis parametrów próbek A i B analogiczny jak w przypadku nagrań w podrozdziale 3.3). Statystycznie istotne różnice zaobserwowano odnośnie do tych nagrań mowy męskiej oraz wiolonczeli, w których zakres losowania był największy. Dla tych samych zakresów w przypadku nagrań oboju słuchacze nie wskazywali istotnych różnic pogłosowości.

Tabela 4

Wyniki analizy statystycznej w przypadku zestawienia maksymalnej zmiany fazy; symbole jak w tabeli 2

Próba			Zbiór pełny, $N = 31$				Zbiór zawężony, $N = 14$			
nagranie	próbka A	próbka B	m_1	*Me*	*W*	*p*	m_1	*Me*	*W*	*p*
M	0	60 (10)	0,45	1	199	**0,0507**	0,50	0,5	32	0,0625
M	0	90 (10)	0,68	1	192	0,0041	0,71	1	36	0,0078
M	0	120 (10)	1,35	2	436	0,0000	1,57	2	105	0,0001
M	60 (10)	120 (10)	1,35	2	397	0,0000	1,64	2	105	0,0001
Ob	0	60 (10)	–0,13	0	87	0,4951	0,07	0	9	1,0000
Ob	0	90 (10)	0,29	0	113	0,0791	0,07	0	4	1,0000
Ob	0	120 (10)	0,35	0	126	0,0795	0,00	0	5	1,0000
Ob	60 (10)	120 (10)	0,29	0	164	0,2238	0,21	0	20	0,4531
Vc	0	60 (10)	0,00	0	114	0,9999	0,21	0	29	0,6133
Vc	0	90 (10)	0,68	1	236	0,0091	0,93	1	45	0,0039
Vc	0	120 (10)	0,81	1	281	0,0006	0,86	1	55	0,0020
Vc	60 (10)	120 (10)	0,90	1	327	0,0004	0,71	1	48	0,0430

Ze względu na rzadsze zauważanie zmian przez słuchaczy w dźwiękach oboju wykonano dodatkową analizę mającą na celu zbadanie wpływu nagrania na wybór słuchaczy. Wykorzystując test znakowanych rang Wilcoxona dla par próbek, sprawdzono, czy jest istotna różnica w przypadku tej samej maksymalnej zmiany w zależności od użytych nagrań. Na poziomie ufności 95% testowano zatem następujące hipotezy:

– H0: odpowiedzi pochodzą z rozkładów statystycznych o porównywalnych parametrach,
– H1: odpowiedzi pochodzą z rozkładów statystycznych o istotnie różnych parametrach.

W tabeli przedstawiono uzyskane wyniki – jak widać, przypuszczenie o wpływie nagrania jest po części uzasadnione. W przypadku zestawień nagrań bez modyfikacji – maksymalne przesunięcie 60° (0, 60), sygnał użyty w badaniu nie ma znaczenia – test wskazuje, że rozkład wyników jest porównywalny. W tym zestawieniu również na podstawie analizy poszczególnych pytań wskazano, że słuchacze nie dostrzegają różnic między próbkami (por. tab. 4). Jednak przy wprowadzaniu większych zmian fazy w większości przypadków test wskazuje na różnice między rozkładem wyników dotyczących poszczególnych par instrumentów. Wykorzystany dźwięk może wpływać nie tylko na zauważalność różnicy (zestawienia głosu męskiego i oboju), ale także na jej wyrazistość – mediana wyników jest najwyższa w przypadku próbek z głosem męskim, co oznacza że w tym nagraniu różnice były najbardziej słyszalne dla uczestników badania.

Tabela 5

Wyniki analizy statystycznej w przypadku zestawienia par instrumentów przy jednakowej maksymalnej zmianie fazy; symbole jak w tabeli 2

Próbka	Porównywane nagrania						Zbiór pełny, $N = 31$		Zbiór zawężony, $N = 14$	
	nagranie	*Me*, $N = 31$	*Me*, $N = 14$	nagranie	*Me*, $N = 31$	*Me*, $N = 14$	*W*	*p*	*W*	*p*
0, 60(10)	M	1	0,5	Ob	0	0	108	0,3261	24,5	0,1094
	M	1	0,5	Vc	0	0	167	0,0673	24	0,5625
	Vc	0	0	Ob	0	0	121,5	0,4081	32	0,7930
0, 90(10)	M	1	1	Ob	0	0	168	0,0017	21	0,0313
	M	1	1	Vc	1	1	73,5	0,9072	4	0,4375
	Vc	1	1	Ob	0	0	203	0,0011	36	0,0078
0, 120(10)	M	2	2	Ob	0	0	290,5	0,0019	91	0,0002
	M	2	2	Vc	1	1	197,5	0,0198	60,5	0,0107
	Vc	1	1	Ob	0	0	182	0,0685	61,5	0,0088
60(10), 120(10)	M	2	2	Ob	0	0	269,5	0,0024	91	0,0002
	M	2	2	Vc	1	1	138,5	0,0733	51,5	0,0137
	Vc	1	1	Ob	0	0	182	0,0668	48	0,1855

3.5. Maksymalna zmiana fazy między oknami

Wstępne, nieformalne testy słuchowe sugerowały, że różnica zmian faz między kolejnymi oknami $\Delta\varphi_{win}$ ma wpływ na postrzeganie pogłosowości; jednak w badanym zakresie testy formalne nie potwierdziły tej zależności, jak można zauważyć na podstawie danych w tabeli 6. Badano dwie próbki głosu kobiecego i dwie fletu; w każdej maksymalna modyfikacja wynosiła 90°, w przypadku próbki A $\Delta\varphi_{win}$ wynosiła zawsze 10°, a w przypadku próbki B – 15° i 30°. Jedynie w przypadku pary nagrań fletu 90 (10) i 90 (30) postrzegane różnice były statystycznie istotne.

Tabela 6

Wyniki analizy statystycznej w przypadku zestawienia maksymalnej zmiany fazy między kolejnymi oknami; symbole jak w tabeli 2

Próba			Zbiór pełny, $N = 31$				Zbiór zawężony, $N = 14$			
nagranie	próbka A	próbka B	m_1	*Me*	*W*	*p*	m_1	*Me*	*W*	*p*
K	90(10)	90(15)	–0,06	0	79	0,8441	0,21	0	15	0,5313
K	90(10)	90(30)	–0,03	0	64	0,8909	0,07	0	4	1,0000
Fl	90(10)	90(15)	–0,13	0	79	0,5739	–0,07	0	6	1,0000
Fl	90(10)	90(30)	0,45	0	117	0,0092	0,57	0,5	28	0,0156

3.6. Zakres częstotliwości

Tę część badań przeprowadzono jedynie z użyciem nagrań wiolonczeli, których widmo obejmuje najszerszy zakres częstotliwości (najniższa częstotliwość w wykorzystanych nagraniach to 98 Hz). Parametry losowania były jednakowe w każdej próbie (maksymalna losowana zmiana 90°, maksymalna zmiana między oknami 15°), jednak zmiany wprowadzano w różnych zakresach częstotliwości. Wyniki przedstawiono w tabeli 7.

W pierwszych dwóch próbach porównywano efekty modyfikacji fazy w zakresach 0 – f_g oraz $(f_g - f_s)/2$, gdzie f_s to częstotliwość próbkowania (48 kHz), a f_g częstotliwość podziału wynosząca 400 Hz lub 1500 Hz. Za bardziej pogłosową słuchacze uznawali próbkę, w której faza była zmodyfikowana w zakresie wyższych częstotliwości czyli $(f_g - f_s)/2$. Wstępnie potwierdza to hipotezę D. Griesingera o największym znaczeniu tego zjawiska powyżej 1000 Hz [7], choć sformułowanie jednoznacznych wniosków wymaga przeprowadzenia badań na bardziej zróżnicowanych bodźcach i zbadania różnych kombinacji zakresów.

Przy porównaniu zmian w pełnym paśmie oraz w zakresie 1500 Hz – $f_s/2$ za bardziej pogłosową uznawano próbkę zmodyfikowaną w szerszym zakresie częstotliwości. Porównanie modyfikacji w zakresie 0–400 Hz i 0–1500 Hz oraz 400 Hz – $f_s/2$ i 1500 Hz – $f_s/2$ nie wykazało istotnych statystycznie różnic percepcyjnych.

Tabela 7

Wyniki analizy statystycznej w przypadku zestawienia zakresów częstotliwości poddawanych zmianie fazy; symbole jak w tabeli 2

Próba			Zbiór pełny, N = 31				Zbiór zawężony, N = 14			
nagranie	próbka A	próbka B	m_1	*Me*	*W*	*p*	m_1	*Me*	*W*	*p*
Vc 90(15)	0–1500 Hz	1500 Hz – $f_s/2$	1,45	2	412	0,0000	1,57	2	78	0,0005
Vc 90(15)	0–400 Hz	400 Hz – $f_s/2$	1,55	2	417	0,0000	1,57	2	78	0,0005
Vc 90(15)	0 – $f_s/2$	1500 Hz – $f_s/2$	–1,32	–2	31	0,0000	–1,36	–2	2	0,0020
Vc 90(15)	0–400 Hz	0–1500 Hz	–0,13	0	112	0,6527	–0,36	–0	10	0,1797
Vc 90(15)	1500 Hz – $f_s/2$	400 Hz – $f_s/2$	0,23	0	81	0,2664	0,21	0	5	0,5000

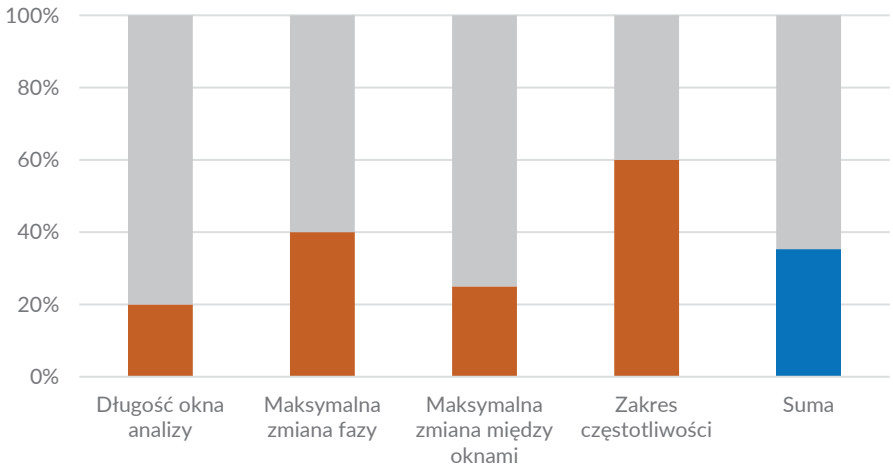
3.7. Dyskusja i podsumowanie

Przeprowadzono formalne testy słuchowe z wykorzystaniem nagrań instrumentów muzycznych i mowy, w których faza była losowo zmieniana w kolejnych oknach transformaty STFT. Sprawdzono wpływ różnych aspektów tej modyfikacji na wynik, czyli wartości wprowadzanego przesunięcia fazy, maksymalnej zmiany między kolejnymi oknami czasowymi, długości okna czasowego, w przypadku jednego nagrania także zakresu częstotliwości. Podsumowanie uzyskanych wyników przedstawiono w tabeli 8. Na rysunku 4 zebrano procentowy udział prób, w których zauważone różnice były istotne statystycznie, z podziałem na kategorie.

Tabela 8

Skrótowe omówienie wyników

Badany aspekt	Komentarz
Długość okna analizy sygnału	Nieformalne komentarze sugerują wpływ na naturalność brzmienia sygnału, badania formalne nie wykazały różnic w percepcji pogłosowości
Maksymalna zmiana fazy	Większa zmiana fazy zwiększa odczucie pogłosowości
Maksymalna zmiana fazy między kolejnymi oknami	W przypadku badanych wartości nie wykazano wpływu na percepcję pogłosowości
Zakres częstotliwości poddawany zmianie fazy	Wpływa na percepcję pogłosowości; wskazanie dokładniejszych zależności wymaga dalszych badań



Rys. 4. Procentowy udział prób, podczas których słuchacze zauważyli istotne różnice w pogłosowości między prezentowanymi próbkami dźwiękowymi

Przeprowadzono także analizę, której wyniki sugerują istotne znaczenie sygnału użytego do badań – zarówno w kwestii wykrywalności różnic, jak i ich wyrazistości. Fakt, że modyfikacje fazy nie były zauważalne w przypadku oboju, jest niezgodny z oczekiwaniami, gdyż dźwięk oboju charakteryzuje się harmoniczną strukturą widma, w której harmoniczne są łatwe do wyróżnienia, można się zatem było spodziewać, że zmiany fazy będą wyraźnie słyszalne. Na kształt widma wiolonczeli wpływają dodatkowo rezonanse korpusu i nieharmoniczność strun; jest ono więc bardziej „rozmyte". Z kolei widmo sygnału mowy opisywane jest przez układ formantów. Badania pokazują, że struktura widmowa ma wpływ na percepcję zależności fazowych w sygnale [5]; dobór nagrań do kolejnych etapów badań powinien być poprzedzony sprawdzeniem tej zależności.

Należy też pamiętać, że wprowadzane wartości losowano w przypadku każdej próbki z podanego zakresu, co nie daje jednak gwarancji, że maksymalna wartość przesunięcia została wylosowana. Dodatkowo każda próbka została zmodyfikowana w inny sposób (nowe losowanie) – w przyszłości należałoby rozważyć przeprowadzenie testów np. na kilku próbkach losowanych według tych samych kryteriów, aby zwiększyć wiarygodność wyników.

Po zakończeniu badania słuchacze byli poproszeni o komentarz i ewentualne uwagi odnośnie do jego przebiegu – kilka osób wskazało na istnienie różnic między próbkami, które jednak nie dawały wrażenia pogłosu, a także na nienaturalne brzmienie prezentowanych nagrań. Opinie te należy wziąć pod uwagę przy projektowaniu kolejnych eksperymentów; być może należy rozważyć inne metody wprowadzania modyfikacji fazy (np. filtr wszechprzepustowy o określonej charakterystyce fazowej). Niektórzy zwrócili uwagę na ogólną trudność wskazania różnic między próbkami – te obserwacje pokrywają się z wynikami testów statystycznych, które w 65% przypadków nie wykazały istotnych różnic.

W przeprowadzonych badaniach wykazano, że losowe zmiany fazy sygnału mogą wywoływać wrażenie pogłosu, muszą jednak być wystarczająco zauważalne. Przedstawione wyniki skłaniają do postawienia hipotezy, że istnieje pewna wartość zmiany fazy, która jest zauważalna. To zagadnienie wymaga kontynuacji badań; z pewnością warta zbadania jest także kwestia zakresów częstotliwości, w których wprowadzane są zmiany fazy.

Bibliografia

[1] A. Snakowska, *Teoria pola akustycznego zastosowana do badania układów o symetrii cylindrycznej*, Wydawnictwa AGH, Kraków, 2018

[2] C.-H. Jeong, D. Lee, S. Santurette, J.-G. Ih, *Influence of Impedance Phase Angle on Sound Pressures and Reverberation Times in a Rectangular Room*, The Journal of the Acoustical Society of America 2014, vol. 135, s. 712–723

[3] C.-H. Jeong, J.-G. Ih, J. Rindel, *A study on the characteristics of phased beam tracing method for the acoustic simulation of an enclosure at mid frequencies*, w: *Proceedings of the WESPAC IX 2006, The 9th Western Pacific Acoustics Conference: Better life through acoustics*, Seoul, 2006
[4] E. Ozimek, *Dźwięk i jego percepcja. Aspekty fizyczne i psychofizyczne*, Wydawnictwo Naukowe PWN, Warszawa, 2018
[5] M.-V. Laitinen, S. Disch, V. Pulkki, *Sensitivity of Human Hearing to Changes in Phase Spectrum*, The Journal of the Audio Engineering Society 2013, vol. 61, s. 860–877
[6] D. Griesinger, *Pitch Coherence as a Measure of Apparent Distance in Performance Spaces and Muddiness in Sound Recordings*, w: *Audio Engineering Society Convention Paper 6917*, San Francisco, CA, 2006
[7] D. Griesinger, *Pitch, Timbre, Source Separation, and the Myths of Loudspeaker Imaging*, w: *Audio Engineering Society Convention Paper 8610*, Budapest, 2012
[8] MathWorks, Matlab Documentation: interp1, 2019, https://uk.mathworks.com/help/matlab/ref/interp1.html, dostęp 10.05.2019
[9] MathWorks, Matlab Signal Processing Toolbox Documentation: istft, 2019, https://uk.mathworks.com/help/signal/ref/istft.html, dostęp 8.05.2019
[10] F.A. Kingdom, N. Prins, *Psychophysics. A Practical Introduction*, Elsevier Academic Press, 2010
[11] T. Makuch, Z. Kusal, K. Mikulec, *Śpiewające instrumenty*, w: *Studium badawcze młodych akustyków 2017*, Akademia Górniczo-Hutnicza im. Stanisława Staszica, Kraków, 2018
[12] P. Kabal, TSP Speech Database, 2018, http://www-mmsp.ece.mcgill.ca/Documents/Data/, dostęp 6.05.2019
[13] V.K. Rohatgi, A.K.M. Ehsanses Saleh, *An Introduction to Probability and Statistics*, John Wiley & Sons, New Jersey, 2015