



KRAKOWSKA AKADEMIA
im. Andrzeja Frycza Modrzewskiego

Wydział Ekonomii i Zarządzani

Kierunek: Informatyka i ekonometria

Specjalność: Informatyka zarządcza

Konrad Wroński

**ZASTOSOWANIE ROZKŁADU BENFORDA DO ANALIZY
PLIKÓW DŹWIĘKOWYCH**

Praca dyplomowa

licencjacka

napisana pod kierunkiem

prof. zw. dr hab. Tadeusza Grabińskiego

Kraków 2012

Spis treści

Wstęp.....	3
1. Prawo Benforda i jego zastosowania do analizy plików dźwiękowych.....	4
1.1. Historia i specyfika rozkładu Benforda	4
1.2. Zastosowania rozkładu Benforda do analizy danych	5
1.3. Wybory prezydenckie w Iranie w 2009 roku	6
1.4. Badania Marka Nigeriniego przy użyciu Prawa Benforda	7
1.5. Analiza poprawności zebranych danych na temat wydatków konsumentów w USA.....	7
1.6. Opis wybranych funkcji programu „Audit Commander”	8
1.7. Opis działania programu „Benford Sound Tester”	10
1.8. Lista funkcji programu „Benford Sound Tester”	13
2. Zastosowanie rozkładu Benforda do analizy dźwięku	16
2.1. Charakterystyki dźwięku	16
2.2. Cyfrowa metoda zapisu dźwięku.....	16
2.3. Zastosowanie rozkładu Benforda do analizy dźwięku zapisanego cyfrowo	17
2.4. Metodologia.....	17
2.5. Porównanie ogólnego odchylenia od rozkładu Benforda.....	17
2.6. Porównanie odchyłeń poszczególnych cyfr w badanych próbkach dźwięku.....	19
2.7. Porównanie odchyłeń po odjęciu składowej stałej oraz odsumieniu.....	24
2.8. Porównanie odchyłeń poszczególnych cyfr dla odsumionych próbek	26
3. Badanie odchyłeń od rozkładu Benforda wskutek wycinania lub wklejania dźwięków.....	33
3.1. Porównanie odchyłeń przed i po usunięciu fragmentu dźwięku	33
3.2. Porównanie odchyłeń przed i po doklejeniu fragmentu dźwięku.....	37
4. Badanie odchyłeń od rozkładu Benforda wskutek zmiksowania dwóch dźwięków	41
4.1. Porównanie odchyłeń przed i po zmiksowaniu na równym poziomie głośności	41
4.2. Porównanie odchyłeń przed i po zmiksowaniu na różnych poziomach głośności.....	43
Podsumowanie	47
Spis tabel	48
Spis rysunków	50
Bibliografia.....	52

Wstęp

Praca niniejsza ma na celu sprawdzenie, czy cyfrowe pliki dźwiękowe podlegają rozkładowi Benforda. Rozkład Benforda służy do weryfikacji autentyczności zbiorów danych statystycznych. Jest stosowany dla bardzo wielu rodzajów danych, lecz nie był do tej pory stosowany do weryfikacji oryginalności dźwięku zapisanego cyfrowo. W pracy tej podejmuję próbę stwierdzenia, czy cyfrowe pliki dźwiękowe podlegają wyżej wymienionemu rozkładowi oraz jak wpływa na ten rozkład manipulacja zawartością plików dźwiękowych.

W rozdziale pierwszym omówione zostały następujące zagadnienia: 1) Istota jak również zastosowanie rozkładu Benforda do weryfikacji autentyczności różnych danych statystycznych, 2) Ogólne zastosowania rozkładu Benforda, 3) Zastosowanie rozkładu Benforda do sprawdzenia autentyczności wyborów w Iranie w 2009 roku, 4) Badania Marka Nigeriniego oraz Zastosowanie Prawa Benforda do analizy wydatków konsumentów w USA. W tym rozdziale zostały również opisane zostały dwa programy przeznaczone do weryfikacji zgodności danych z rozkładem Benforda: dostępny w internecie Audit Commander oraz stworzony na potrzeby tej pracy Benford Sound Tester.

W rozdziale drugim została omówiona ogólna charakterystyka dźwięku oraz wielkości które są z nim związane, opis techniki cyfrowego zapisu dźwięku, której charakter sprawia iż można poddawać go analizie statystycznej przy pomocy rozkładu Benforda, oraz metoda analizy cyfrowych plików dźwiękowych, która umożliwia sprawdzenie przy pomocy przygotowanego do tego programu komputerowego stopnia zgodności dźwięku z rozkładem Benforda i dzięki któremu możliwe jest zweryfikowanie kilku pytań badawczych opisanych w trzech ostatnich rozdziałach tej pracy. W rozdziale tym również zbadany został sam stopień zgodności czterech rodzajów dźwięku z rozkładem Benforda oraz rozkład poszczególnych jego cyfr od 1 do 9. To samo porównanie zostało powtórzone po wyczyszczeniu dźwięku z szumów.

W rozdziale trzecim została podjęta próba weryfikacji zmian w rozkładzie Benforda nagrania dźwiękowego w skutek wycięcia oraz doklejenia fragmentu dźwięku.

W rozdziale czwartym weryfikacji rozkładu pierwszych cyfr został poddany dźwięk, który powstał w skutek zmiksowania dwóch dźwięków. Analiza została przeprowadzona dla dwóch różnych proporcji głośności obu dźwięków.

W podsumowaniu starałem się zebrać wszystkie rezultaty przedstawić warte zauważenia wnioski.

1. Prawo Benforda i jego zastosowania do analizy plików dźwiękowych

1.1. Historia i specyfika rozkładu Benforda

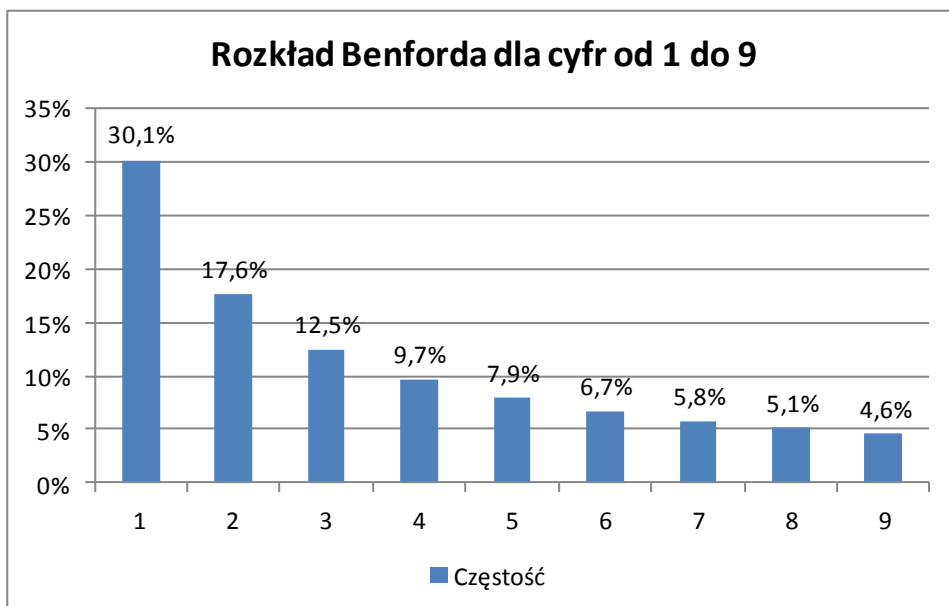
„Frank Benford był fizykiem pracujący w zakładach General Electric. W 1930 roku zauważył coś niezwykłego przeglądając tablice logarytmiczne. Pierwsze strony tych tablic były bardziej zużyte i na tych pierwszych stronach najczęstszymi cyframi były jedyńki. Były częstsze niż cyfry od 2 do 9. Benford spędził lata na kolekcjonowaniu różnych danych statystycznych dotyczących różnorodnych zjawisk. W 1938 roku opublikował około 20 000 różnych wartości takich jak masy atomowe, liczby w magazynach i artykułach, statystyki baseballa, długości rzek. Okazało się, że ta zasada potwierdza się dla większości z nich.”¹

Rozkład Benforda uzyskuje się poprzez zliczenie cyfr najbardziej znaczących w liczbach wielocyfrowych, które są obserwacjami dowolnych zjawisk np. dotyczących powierzchni jezior w Polsce. Specyficzny rozkład tych cyfr jest nazywany prawem Benforda. Prawdopodobieństwo P wystąpienia poszczególnych cyfr jest wyrażone wzorem:

$$P_k = \log_{10} \frac{k+1}{k}$$

Gdzie k jest daną cyfrą np. od 1 do 9.

Rysunek 1. Rozkład Benforda.



źródło: opracowanie własne

Rozkład ten został po raz pierwszy odkryty przez astronoma Simona Newcomba, w 1881 r. Prawo to zostało ponownie dokładnie przeanalizowane przez Franka Benforda w 1938 r. Dowód twierdzenia przedstawił w 1995 Theodor Hill. Rozkład ten jak się później okazało pojawia się w bardzo wielu danych statystycznych pod warunkiem, że nie były przeprowadzane

¹ Steven W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, 1998

na nich modyfikacje. W związku z tym rozkład ten zaczął być stosowany do weryfikacji wielu nadużyć związanych z materiałem statystycznym. Z czego najślynniejsze to wykrycie fałszerstwa księgowego Jamesa Nelsona, który wystawiał fałszywe czeki na sumę ok. 1,8 mln dolarów oraz sfałszowane wybory w Iranie w 2009 roku. Rozkład Benforda znalazł również zastosowanie do weryfikacji oryginalności fotografii zapisanych cyfrowo, rozmiaru plików zapisanych na dysku. Weryfikowane były również dokumenty tekstowe. Jednak rozkład ten nie doczekał się jeszcze zastosowania w weryfikacji nagrań filmowych oraz dźwięków zapisanych cyfrowo.

1.2. Zastosowania rozkładu Benforda do analizy danych

„Rozkład Benforda jest wykorzystywany w naukach ekonomicznych do:

- wykrywania fałszywych lub błędnych w księgowości, rozpoznawania oszustw podatkowych,
- analizy danych z giełdy papierów wartościowych,
- analizy cen wylicytowanych na aukcjach internetowych,
- szacowania wysokości odszkodowań w firmach ubezpieczeniowych,
- oceny najbardziej dostosowanych wysokości kar finansowych orzekanych w procesach sądowych.

W naukach ilościowych rozkład Benforda jest używany do:

- testowania poprawności modeli ekonometrycznych,
- optymalizacji obliczeń rozwiązywaniu matematycznego problemu transportowego komiwojażera oraz weryfikacji autentyczności danych statystycznych.

W naukach informatyczno-technicznych rozkład ten jest wykorzystywany do:

- projektowania architektury pamięci masowych,
- rozpoznawania autentyczności fotografii zapisanych cyfrowo,
- testowania rozmiarów plików komputerowych,
- analizy wydajności algorytmów numerycznych.

W zakresie nauk przyrodniczych do:

- oceny skuteczności działania leków,
- do badania autentyczności komunikatów o poziomach zanieczyszczenia.

W pozostałych naukach rozkład jest wykorzystywany do:

- badania wielości dotacji dla rządzących partii,
- liczby odnośników do przeczytanych publikacji,
- liczby rannych i zabitych w wypadkach,
- liczby powrotów klientów do sklepów,

- sprawdzania liczby zakupionych produktów,
- do statystyk sportowych².

1.3. Wybory prezydenckie w Iranie w 2009 roku

W wyborach zapowiedzieli start następujący kandydaci Mohammad Chatami, Mir-Hossein Mousavi, Mehdi Karroubi oraz urzędujący prezydent Iranu Mahmud Ahmadineżad. Mohammad Chatami wycofał się z wyborów popierając kandydaturę Mira Hosseina Moussaviego. Do głosowania uprawnionych było 46 mln obywateli z czego 39 mln wzięło udział w głosowaniu (82%). Mir-Hossein Mousavi do swojej kampanii używał Internetu, jednak władze zablokowały dostęp do Internetu na kilka dni. W trakcie kampanii wyborczej w dniach 5 i 6 czerwca doszło do starć na ulicach Teheranu pomiędzy zwolennikami Mahmud Ahmadineżada a Mehdiego Karroubiego. Z powodu wysokiej frekwencji zbieranie głosów nie zostało zakończone o 18:00 tylko o 21:30. Gdy ogłoszono wyniki doszło do zamieszek z powodu podejrzeń o możliwe fałszerstwa w czasie przeprowadzania wyborów. Na skutek zamieszek władze postanowiły ponownie przeliczyć 10% losowo wybranych głosów. Po przeliczeniu 10% głosów władze odrzuciły wszelkie wnioski o nieprawidłowościach w przebiegu wyborów i ogłosiła zwycięstwo Ahmadineżada³.

Dane na temat wyników z poszczególnych okręgów wyborczych przy użyciu Prawa Benforda przeanalizował Boudewijn F. Roukema. „Zauważył on, że w 15 okręgach wyborczych cyfra 7 była dwa razy liczniejsza niż zakłada rozkład Benforda. Nadmiarowość ta dotyczyła tylko jednego z kandydatów Karroubiego i jest prawdopodobne, iż dotyczyła ok 1 mln głosów⁴. Wynik ten został potraktowany sceptycznie przez niektórych naukowców a także Radę Strażników.

Tabela 1. Liczba głosów oddanych na poszczególnych kandydatów.

Kandydat (13.VI.2009)	Głosy (tys.)	% głosów
Mahmud Ahmadineżad	24 527	63,3%
Mir-Hossein Mousavi	13 216	34,1%
Mohsen Rezaee	678	1,8%
Mehdi Karroubi	333	0,9%
Frekw. = 82%	38 755	100,0%

źródło: Marzena Farbaniec, Tadeusz Grabiński, Bartłomiej Zabłocki, Waław Zajac, *Wykorzystanie prawa Benforda do wykrywania nieprawidłowości w zbiorach danych*. 2011.

² Marzena Farbaniec, Tadeusz Grabiński, Bartłomiej Zabłocki, Waław Zajac, *Wykorzystanie prawa Benforda do wykrywania nieprawidłowości w zbiorach danych*. 2011.

³ Marzena Farbaniec, Tadeusz Grabiński, Bartłomiej Zabłocki, Waław Zajac, *Wykorzystanie prawa Benforda do wykrywania nieprawidłowości w zbiorach danych*. 2011.

⁴ Roukema, BF. *Benford's Law Anomalies in the 2009 Iranian presidential election.*, 2009.

1.4. Badania Marka Nigieriniego przy użyciu Prawa Benforda⁵

Nowe zastosowania prawa Benforda rozpoczęły się 1992, kiedy Mark Nigirini zbadał rozkład pierwszych cyfr w danych o płacach i kosztach w swoim doktoracie. Dane które przeanalizował zgadzały się dość dokładnie z rozkładem Benforda. Po tym sukcesie Nigierini zaczął używać prawa Benforda do badania innych biznesowych i finansowych danych. Na przykład zbadał koszty wniosków o odszkodowanie z międzynarodową siecią moteli, gdzie wykrył fałszywe kwoty na około milion dolarów.

W 1996 Nigirini przeanalizował dane zwrotach podatku w Urzędzie podatkowym USA (IRS) i zbadał pierwsze cyfry w kategoriach „Zapłacone odsetki” oraz „Odebrane odsetki”. Badania te opublikował w *Journal of the American Taxation Association*. Następnie Nigierini zbadał wiodące cyfry zawarte w zeznaniach podatkowym z lat 1977-1992 Prezydenta Bila Clintona. Nigierini wykazał na podstawie wiodących cyfr, że Prezydent Clinton był uczciwy w swych zeznaniach podatkowych.

1.5. Analiza poprawności zebranych danych na temat wydatków konsumentów w USA⁶

U.S. Bureau of Labor Statistics (BLS) wykonało to badanie aby sprawdzić na co konsumenci amerykańscy wydają pieniądze. Jak w każdym badaniu rzetelność przedstawionych danych zależy od dokładności zebranych danych. CE (Consumer Expenditure Survey) „Badanie wydatków konsumentów” posiada kilka procedur używanych do sprawdzenia rzetelności publikowanych przybliżonych wydatków. Przeprowadzają ponowne wywiady z niektórymi respondentami, sprawdzanie logicznej spójności danych. Z uwagi na rosnącą popularność metody weryfikacji przy użyciu Prawa Benforda BLS zdecydowało użyć go w swoich badaniach.

Poniższa tabela przedstawia rozkład pierwszych cyfr w kwartalnych wywiadach CEQ z badania przeprowadzonego w roku 2000. Dane zawierają 734 684 różnych wydatków. W tabeli można zauważyć, że rozkład pierwszych cyfr zgadza się dość dokładnie z rozkładem Benforda. Cyfr 1 jest 30,5% a rozkład Benforda zakłada 30,1%. Liczebność dwójek w CEQ odpowiada 19,3% natomiast w rozkładzie Benforda jest ich 17,6%.

Tabela 2. Porównanie kwartalnych kwestionariuszy wydatków z Prawem Benforda.

⁵ Swanson, D, Cho, MJ and Eltinge, J. *Detecting possibly fraudulent or error-prone survey data using Benford's law*. pp 4172-4177 in: 2003 Joint Statistical Meetings - Section on Survey Research Methods, Proceedings of the American Statistical Association, 2003.

⁶ Swanson, D, Cho, MJ and Eltinge, J. *Detecting possibly fraudulent or error-prone survey data using Benford's law*. pp 4172-4177 in: 2003 Joint Statistical Meetings - Section on Survey Research Methods, Proceedings of the American Statistical Association, 2003.

Comparison of CEQ Data with Benford's Law

Leading Digit (<i>d</i>)	Reported Expenditures		Benford's Law
	Number	Percent (SE)	$\log_{10}\left(\frac{d+1}{d}\right) \times 100\%$
1	223,776	30.5 (.063)	30.1
2	141,992	19.3 (.053)	17.6
3	90,589	12.3 (.045)	12.5
4	66,266	9.0 (.040)	9.7
5	76,473	10.4 (.044)	7.9
6	50,024	6.8 (.034)	6.7
7	35,019	4.8 (.029)	5.8
8	32,294	4.4 (.028)	5.1
9	18,251	2.5 (.021)	4.6
Total	734,684	100.0	100.0

Źródło: Swanson, D, Cho, MJ and Eltinge, J. *Detecting possibly fraudulent or error-prone survey data using Benford's law*. pp 4172-4177 in: 2003 Joint Statistical Meetings - Section on Survey Research Methods, Proceedings of the American Statistical Association, 2003.

Całościowo dane CEQ nie odbiegają od rozkładu Benforda, lecz cyfr 2 (19,3%) i 5 (10,4%) jest w nadmiarze a cyfr 9 (2,5%) jest zdecydowanie mniej niż przewiduje Prawo Benforda. Prawdopodobnie często respondenci zaokrąglali kwoty do \$25 oraz \$50 i stąd nadmiar cyfr 2 i 5.⁷

1.6. Opis wybranych funkcji programu „Audit Commander”

Program „Audit Commander” służy do przeprowadzania wielu analiz statystycznych. Jednak opiszę tylko wybrane jego funkcje: Benford by Variable Analysis oraz Benford's Law. Program można pobrać ze strony internetowej <http://www.ezrstats.com/>.

Plik zawiera ok. 200 megabajtowe archiwum z bardzo dużą ilością plików. Należy je rozpakować do jakiegoś folderu. Polecam wyeksportowanie danych, które będą analizowane do pliku tekstowego z nazwami zmiennych w pierwszym wierszu i z kolumnami rozdzielonymi tabulatorami. Program Microsoft Excel posiada funkcję eksportu danych do pliku tekstowego. Tak przygotowany plik należy skopiować do folderu htdocs/upload/audit1 w rozpakowanym folderze. Następnie w głównym folderze należy uruchomić plik Server2go.exe. Gdy otworzy się strona w przeglądarce należy funkcję Audits/Sign In (jeżeli User Id nie zawiera nazwy audit1, to analizowany plik należy skopiować do folderu htdocs/upload/(User Id)) i wcisnąć górny przycisk „Sign-in” na nowo otwartej stronie. Następnie bez zmian jakichkolwiek opcji wciskamy przycisk „Process”.

⁷ Swanson, D, Cho, MJ and Eltinge, J. *Detecting possibly fraudulent or error-prone survey data using Benford's law*. pp 4172-4177 in: 2003 Joint Statistical Meetings - Section on Survey Research Methods, Proceedings of the American Statistical Association, 2003.

Od tego momentu możemy importować pliki zawarte w folderze htdocs/Data. W tym celu wybieramy funkcję Processes/Data Import/Load data from text file. W okienku wybieramy naszą tabelę i wciskamy przycisk „Process”. W następnym oknie w dolnej części „Create new table” możemy zmieniać formaty poszczególnych zmiennych. Zdarza się, że któraś ze zmiennych nie ma odpowiednio rozpoznanego formatu. Jeżeli zmienna jest typu tekstowego to należy wybrać format „Variable Character”, jeżeli zawiera liczby całkowite to „Integer” a dla rzeczywistych „Float” lub „Double Precision”. Dodatkowo jest możliwość wybrania rodzaju zmiennych typu „Date” - data i czasu „Time”. Po wykonaniu tej operacji wciskamy przycisk „New Table”. Jeżeli wszystko przebiegło bez zakłóceń to w tym momencie mamy możliwość dokonywania analiz na wybranym pliku.

Funkcja „Benford’s Law” - wybieramy ją poprzez kliknięcie na Functions/Numeric Functions/Benford’s Law. Służy ona do sprawdzania zgodności rozkładu pierwszych cyfr w wielocyfrowych danych z rozkładem Benforda. W polu „Numeric Column” wybieramy zmienną, której zgodność chcemy zbadać. W polu „Select type of Benford test to be run” wybieramy rodzaj testu: „First digit” - test pierwszej najbardziej znaczącej cyfry, „First two digits” - test 99 kombinacji dwóch pierwszych cyfr, „First three digits” - test pierwszych trzech cyfr, „Last digit” - test ostatniej najmniej znaczącej cyfry, „Last two digits” - test dwóch najmniej znaczących cyfr, „Second” - test tylko drugiej cyfry oraz „Third digit” - test trzeciej z najbardziej znaczących cyfr. Po wybraniu funkcji „Process” dla każdego testu są wypisywane teoretyczne i zaobserwowane liczebności poszczególnych kombinacji cyfr. Jest również podawany wynik testu χ^2 , liczba stopni swobody oraz wartość graniczna po przekroczeniu której wynik testu χ^2 oznacza niezgodność danych z rozkładem Benforda przy wartości krytycznej 95%.

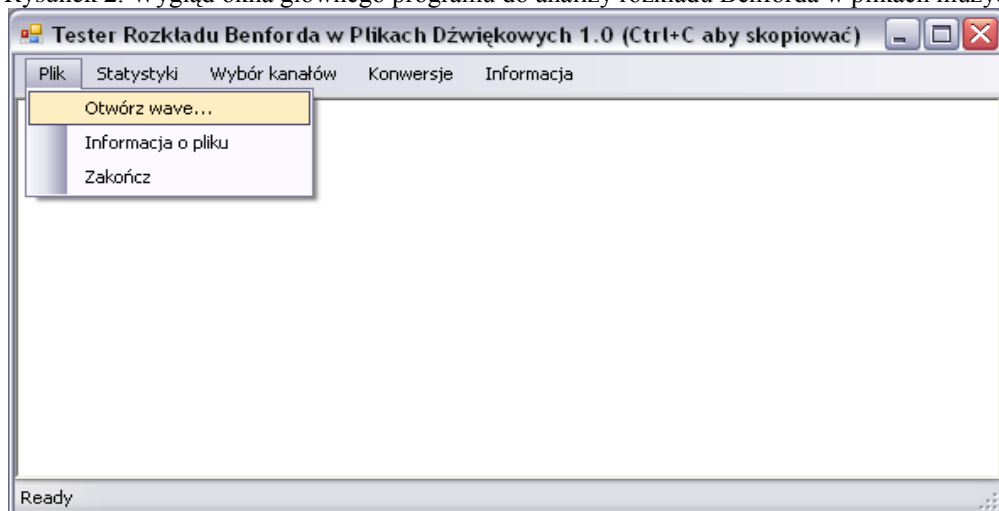
Funkcja „Benford by Variable Analysis” - wybieramy z menu Functions/Commonly Used. Funkcja służy do wypisania samych wyników testu χ^2 zgodności z rozkładem Benforda dla różnych grup danych wybranych poprzez dodatkową zmienną grupującą dane. W polu „Numeric Column” wybieramy zmienną, którą chcemy analizować. W polu „Analysis (by) column” wybieramy zmienną grupującą. Po wciśnięciu przycisku Process program przedstawi wyniki testu χ^2 zgodności z rozkładem Benforda dla poszczególnych grup. Wybór testowanych cyfr jest ten sam co w funkcji „Benford’s Law”.

Dodatkowo w obydwu funkcja jest możliwość dodania szczególnych kryteriów w celu wyboru części danych.

1.7. Opis działania programu „Benford Sound Tester”

Program został stworzony na potrzeby niniejszej pracy i jest przeznaczony do analizy plików dźwiękowych typu Wave pod kątem zbieżności całego rozkładu z rozkładem Benforda lub stopnia zbieżności poszczególnych jego cyfr. W celu analizy pliku muzycznego należy wybrać menu Plik/Otwórz wave... i wybrać lokalizację pliku. Następnie należy wcisnąć standardowy przycisk Otwórz. Jeżeli wszystko przebiegło pomyślnie to na dolnym pasku statusu programu powinien ukazać się napis: "Plik <<lokalizacja\nazwa>> został pomyślnie otwarty."

Rysunek 2. Wygląd okna głównego programu do analizy rozkładu Benforda w plikach muzycznych.

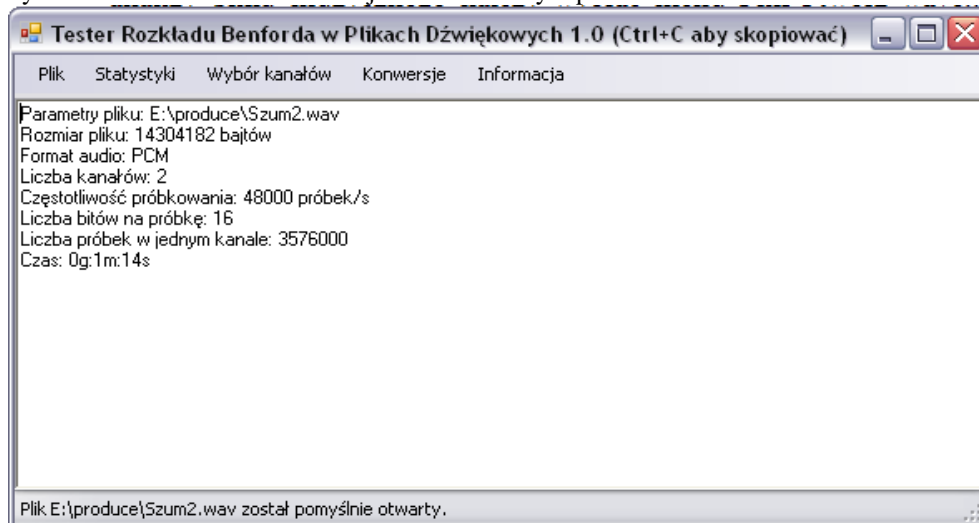


źródło: opracowanie własne

Po wykonaniu powyższej operacji można zająć się zbieraniem danych nt. otwartego pliku.

Funkcja "Informacja o pliku" przedstawia bardzo ogólne informacje o badanym pliku: nazwa, długość, liczba kanałów, częstotliwość próbkowania, liczbę bitów na próbkę, liczbę próbek na kanał oraz czas odtwarzania dźwięku.

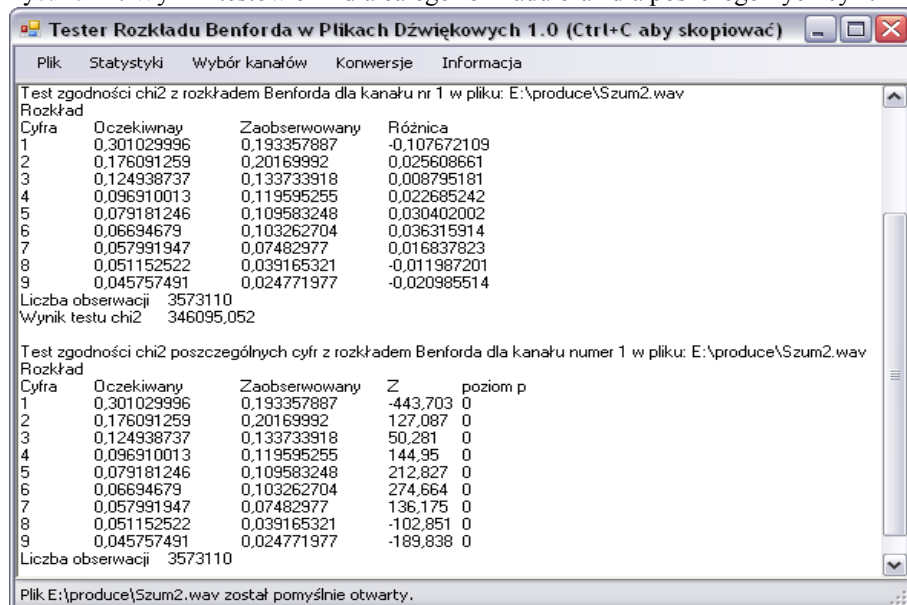
Rysunek 3. Podstawowe informacje o analizowanym pliku.



źródło: opracowanie własne

Menu "Statystyka" zawiera cztery opcje przeznaczone do analizy zbieżności pliku z rozkładem Benforda: 1) test zgodności χ^2 całego rozkładu z rozkładem Benforda, 2) test zgodności χ^2 dla poszczególnych cyfr, 3) wartość jednego z mierników służących do opisu rozkładu Benforda oraz 4) standardowe parametry rozkładu próbek. Na poniższym rysunku przedstawiam wyniki dwóch pierwszych statystyk z powyższego menu.

Rysunek 4. Wyniki testów χ^2 dla całego rozkładu oraz dla poszczególnych cyfr.



źródło: opracowanie własne

Pierwszy ogólny test zgodności z rozkładem Benforda wypisuje procentowe częstotliwości ich odchylenie od rozkładu, ogólną liczebność oraz wynik testu χ^2 wg wzoru⁸:

$$\chi^2 = \sum_{i=1}^9 \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^9 \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$$

Gdzie X^2 - to wynik testu χ^2 , n_i - liczebność zaobserwowana i-tej cyfry, n - liczebność wszystkich cyfr, oraz p_i - zakładane prawdopodobieństwo wystąpienia i-tej cyfry wg rozkładu Benforda.

Test zgodności poszczególnych cyfr podaje rozkład oczekiwany, zaobserwowany, wyniki testu χ^2 dla poszczególnych cyfr oraz poziomy istotności ich różnic od zakładanych liczebności. Test ten opiera się na następującym wzorze⁹:

⁸ Tadeusz Grabiński, *Prawo Benforda i Jego Wykorzystanie w Analizie Wyników Wyborów Prezydenckich W Polsce*, 2009 r.

⁹ Tadeusz Grabiński, *Prawo Benforda i Jego Wykorzystanie w Analizie Wyników Wyborów Prezydenckich W Polsce*, 2009 r.

$$z_i = \frac{n_i / n - p_i}{\sqrt{\frac{p_i(1-p_i)}{n}}}$$

Gdzie z_i to wynik testu dla i -tej cyfry rozkładu, n_i to liczebność empiryczna i -tej cyfry, n – ogólna liczebność wszystkich cyfr, p_i – prawdopodobieństwo wystąpienia i -tej cyfry w zakładanym rozkładzie w tym przypadku jest to rozkład Benforda

Wynik testu dla Miernika 1 jest podawany wraz z przedstawieniem oczekiwanego i zaobserwowanego rozkładu oraz różnicą pomiędzy tymi rozkładami. Jest on obliczony wg wzoru¹⁰:

$$M_1 = \frac{1}{9} \sum_{i=1}^9 \left| \frac{n_i - \hat{n}_i}{\hat{n}_i} \right|$$

Gdzie M_1 – to wyrażone w procentach odchylenie od zakładanego rozkładu, n_i – liczebność empiryczna i -tej cyfry, \hat{n}_i - liczebność i -tej cyfry wynikająca z rozkładu teoretycznego Benforda

Ostatni test pokazuje takie parametry ogólnego rozkładu jak średnia arytmetyczna, wariancja oraz odchylenie standardowe.

Rysunek 5. Wynik wartości miernika 1 oraz podstawowych parametrów podstawowego rozkładu próbek (nie pierwszych cyfr).

Cyfra	Oczekiwany	Zaobserwowany	Różnica
1	0,301029996	0,193357887	-0,107672109
2	0,176091259	0,20169992	0,025608661
3	0,124938737	0,133733918	0,008795181
4	0,096910013	0,119595255	0,022685242
5	0,079181246	0,109583248	0,030402002
6	0,06694679	0,103262704	0,036315914
7	0,057991947	0,07482977	0,016837823
8	0,051152522	0,039165321	-0,011987201
9	0,045757491	0,024771977	-0,020985514

Liczba obserwacji: 3573110
Wartość Miernika 1 0,30192

Parametry rozkładu dla kanału numer 1 w pliku: E:\produce\Szum2.wav
Liczba obserwacji: 3573110
Średnia: -0,00737639821029083
Wariancja: 595420513609,39
Odchylenie standardowe: 771634,961370589

Plik E:\produce\Szum2.wav został pomyślnie otwarty.

źródło: opracowanie własne

¹⁰ Tadeusz Grabiński, *Prawo Benforda i Jego Wykorzystanie w Analizie Wyników Wyborów Prezydenckich W Polsce*, 2009 r.

Kolejne menu "Wybór kanałów" służy do wybrania analizowanego kanału. Menu "Konwersje" w ogólnym sensie służy do zmniejszania liczebności badanego rozkładu na 3 różne sposoby.

Pierwszy sposób to podzielenie całego zakresu próbek na zadaną liczbę części, z których obliczana jest średnia arytmetyczna.

Kolejny podział przez częstotliwość umożliwia obcięcie wszystkich częstotliwości wyższych od wybranej i podjęcie analizie wyłącznie częstotliwości niższych od zakładanej. Poszczególne próbki są uzyskiwane również przez wyliczenie średnich i dodanie stałej jak w poprzednim przypadku jednak ich ilość powinna być równa iloczynowi wybranej częstotliwości oraz długości dźwięku w sekundach.

Ostatni rodzaj podziału w zasadzie nie jest podziałem, ponieważ rozkład nie jest zmieniany. Jedyną rzeczą, która się zmienia to liczebność do której stosowany jest rozkład. Uzyskane wyniki nie mają wartości statystycznej, lecz umożliwiają porównanie dwóch rozkładów, których długości różnią się znacznie poprzez założenie, że dotyczyłyby konkretnej jednej liczebności.

1.8. Lista funkcji programu „Benford Sound Tester”

Program "Benford Sound Tester" jest złożony z czterech aktywnych menu:

1. Menu Plik (File) zawiera w sobie następujące podmenu:
 - 1.1. Otwórz wave... (Open wave...) - służy do otwierania plików dźwiękowych w formacie Wave z rozszerzeniem .wav. W obecnej wersji program otwiera pliki o rozmiarze ograniczonym do ok. 200 MB,
 - 1.2. Informacja o pliku (File info) - wypisuje podstawowe informacje o badanym pliku:
 - 1.2.1. Rozmiar pliku (File size) - nazwa pliku wraz z jego lokalizacją i rozmiar pliku wyrażonym w bajtach,
 - 1.2.2. Format audio (Audio format) - format zapisu dźwięku (PCM),
 - 1.2.3. Liczba kanałów (Number of channels) - 1 - mono, 2 - stereo, 4 - kwadro,
 - 1.2.4. Częstotliwość próbkowania (Sample rate) - ilość próbek używanych w ciągu jednej sekundy do odwzorowania kształtu fali dźwiękowej w hercach (Hz),
 - 1.2.5. Ilość bitów na próbkę (Bits per sample) - ilość bitów używanych do zapisu próbek określających kształt fali dźwiękowej. Najczęściej wartości to: 8 bitów - 256 możliwych poziomów w zakresie od -128 do +127 oraz 16 bitów - 65536 poziomów zakres od -32768 do +32767. Od ilości poziomów zależy dokładność odwzorowania dźwięku,

- 1.2.6. Ilość próbek na jeden kanał (Number of samples for one channel) - ilość próbek w jednym ze współbieżnych kanałów. Częstotliwość próbkowania * czas odtwarzania dźwięku,
 - 1.2.7. Czas (Time) - czas całego utworu,
 - 1.2.8. Zamknij (Exit) - opuszczenie programu.
2. Menu Statystyki (Statistics) zawiera:
- 2.1. Test zgodności χ^2 z rozkładem Benforda (Chi square test for Benford distribution) - przedstawia następujące dane:
 - 2.1.1. Nazwę pliku wraz z jego lokalizacją oraz numer badanego kanału,
 - 2.1.2. Oczekiwany (Expected) oraz zaobserwowany (Observed) rozkład cyfr od 1 do 9 - Cyfra (First digit) oraz odchylenie rozkładu zaobserwowanego od oczekiwanego - Różnica (Difference),
 - 2.1.3. Liczba obserwacji (Number of observations) - jest to liczba sampli w danym kanale
 - 2.1.4. Wynik testu χ^2 (Chi square test result) - przedstawia wynik testu porównania rozkładu empirycznego z hipotetycznym rozkładem Benforda.
 - 2.2. Test zgodności χ^2 poszczególnych cyfr z rozkładem Benforda (Benford distribution Chi square test for specific digits) - przedstawia następujące dane:
 - 2.2.1. Nazwę i lokalizację pliku oraz numer badanego kanału,
 - 2.2.2. Oczekiwany (Expected), zaobserwowany (Observed) rozkład cyfr od 1 do 9, wynik testu Z dla każdej z cyfr w rozkładzie oraz poziom istotności (poziom p, p-level) odchylenia każdej z cyfr od rozkładu Benforda,
 - 2.2.3. Liczba obserwacji (Number of observations) - jest to liczba sampli w danym kanale.
 - 2.3. Miernik 1 dla rozkładu Benforda (Benford distribution Meter 1) - przedstawia następujące dane:
 - 2.3.1. Nazwę pliku wraz z lokalizacją i numer badanego kanału,
 - 2.3.2. Oczekiwany i zaobserwowany rozkład cyfr od 1 do 9,
 - 2.3.3. Wartość miernika 1, który przedstawia sumę odchyleń wszystkich cyfr od rozkładu Benforda,
 - 2.4. Parametry rozkładu (Distribution parameters) - przedstawia podstawowe parametry rozkładu wartości próbek (samples) takie jak: Liczba obserwacji (Number of observations), Średnia (Average), Wariancja (Variance), Odchylenie standardowe (Standard Deviation).

3. Menu Wybór kanałów (Channels selection) umożliwia wybór analizowanego kanału (ścieżki).
4. Menu Konwersje (Conversions) posiada następujące funkcje:
 - 4.1. Podziel na określoną liczbę części (Divide by part number) - dzieli cały zakres próbek (sampli) na określoną ilość uśrednionych bloków, które od tej pory są traktowane jakby były jednostkowymi próbkami,
 - 4.2. Podziel przez częstotliwość Hz (Divide by frequency Hz) - dzieli plik muzyczny na taką ilość części aby w każdej sekundzie utworu znajdowała się ich określona ilość (funkcja obcina częstotliwości wyższe niż częstotliwość podziału wyrażona w hercach (Hz),
 - 4.3. Funkcja - Minimalizuj liczbę obserwacji (Minimize number of observations) - traktuje zaobserwowany pełny rozkład tak jakby dotyczył mniejszej zadanej liczebności niż zaobserwowana,
 - 4.4. Wyczyść wszystkie podziały (Clear all divisions) - likwiduje wszelkie zastosowane podziały pliku dźwiękowego.

2. Zastosowanie rozkładu Benforda do analizy dźwięku

2.1. Charakterystyki dźwięku

Dźwięk - „zaburzenie fal. W ośrodku spręż. Zdolne do wywołania wrażenia słuchowego a także wrażenie słuchowe wywołane tym zjawiskiem...”¹¹. Dźwięk jest falą rozchodzącą za pośrednictwem określonego ośrodka np. powietrza. Jak każda fala posiada dwie podstawowe własności częstotliwość, amplitudę. Częstotliwość jest liczbą drgań ośrodka, które odbywają się w ciągu jednej sekundy. Jest przyjęte, że ludzkie ucho potrafi rejestrować dźwięki w zakresie od 16 do 20000 Hz czyli drgań na sekundę. Częstotliwość dźwięku jest ściśle związana z jego wysokością. Dwukrotny wzrost częstotliwości oznacza nutę wyższą o całą oktawę. Amplituda dźwięku jest związana z jego natężeniem i jest mierzona w decybelach dB. Dźwięk który słyszymy jest złożony z bardzo wielu częstotliwości o różnych amplitudach, które dodatkowo zmieniają się w czasie. Dźwięk można przetworzyć na postać cyfrową.

2.2. Cyfrowa metoda zapisu dźwięku

Cyfrowa metoda zapisu dźwięku polega na podzieleniu całego jego czasu odtwarzania na równe bardzo krótkie odcinki, w których amplituda fali dźwiękowej jest rejestrowana w postaci konkretnej liczby. Odcinki, które są rejestrowane nazywane są próbkami lub z angielska *samplami*. Do zapisu cyfrowego dźwięku stosuje się różne częstotliwości próbkowania. Najczęściej obecnie stosowane częstotliwości próbkowania to 44100 Hz oraz 48000 Hz. Jak łatwo zauważyć są one wyższe od maksymalnych częstotliwości słyszalnych przez człowieka jednak oznacza to, że częstotliwości najwyższe są zapisywane przy pomocy zaledwie trzech próbek, które odwzorowują kształt fali dźwiękowej.

Kolejną charakterystyką dźwięku zapisanego cyfrowo jest ilość bitów stosowanych do zapisania wartości sampla lub próbki. Liczba 2 podniesiona do potęgi ilości bitów wyraża ilość możliwych poziomów dźwięku, które można zapisać przy pomocy sampla. Najczęściej stosowana w tej chwili ilość bitów to 16 choć jeszcze dość często można się spotkać się z 8 bitowymi samplami. Na 8 bitach można zapisać 256 poziomów głośności od -128 do +127 natomiast przy użyciu 16 bitów można uzyskać 65536 poziomów od -32768 do 32767 z zerem włącznie. Rzadziej stosowane choć dużo wierniejsze formy zapisu sampli to 64 bity na sampel oraz stosowanie liczb zmiennoprzecinkowych, które dają dużo dokładniejsze odwzorowanie kształtu fali dźwiękowej.

¹¹ PAN *Encyklopedia fizyki* str. 448

2.3. Zastosowanie rozkładu Benforda do analizy dźwięku zapisanego cyfrowo

Zastosowanie rozkładu Benforda do analizy dźwięku zapisanego cyfrowo zasadza się na charakterystyce próbek dźwięku, które są zapisane w postaci liczb wielocyfrowych. Każda z tych próbek jeżeli tylko jej wartość nie wynosi zero co zdarza się stosunkowo rzadko ma cyfrę najbardziej znaczącą która jest zliczana do kategorii od 1 do 9. W tej procedurze nie ma różnic pomiędzy innymi rodzajami danych analizowanych przy pomocy Prawa Benforda. Różnica jednak dość istotna pojawia się w liczebności tych próbek. W kilkuminutowym nagraniu w zależności od częstotliwości próbkowania liczebność ta dochodzi do rzędu milionów. Rzadko kiedy jakiegokolwiek inne dane statystyczne są tak liczne. Dlatego test χ^2 dla tych danych ma bardzo małą szansę na wykazanie zgodności z rozkładem Benforda. Nie przeszkadza to jednak w analizie danych, ponieważ samo zastosowanie rozkładu do analizy opiera się na sprawdzeniu czy jakieś konkretne cyfry są nadmiernie liczne kosztem innych lub na ile cały rozkład odbiega od założonego w stosunku do innych danych.

2.4. Metodologia

Do analizy plików muzycznych (*.wav) była potrzeba napisania oddzielnej aplikacji. Uzyskane przez program wyniki zostały zebrane dla 35 analizowanych próbek dźwięku (telewizyjnego programu informacyjnego), z których później zostały wycięte trzy różne fragmenty: mowa prezentera, fragment muzyki oraz prognoza pogody (mowa + podkład muzyczny). Próbki te zostały przeanalizowane na różne sposoby: odchylenie od rozkładu Benforda całego rozkładu próbki, różnica mierzona miernikiem 1, który pokazuje procentowe odchylenie od całego rozkładu oraz procentowe odchylenie poszczególnych cyfr od tego rozkładu. Tak uzyskane różnice procentowe zostały poddane porównaniom testem dla prób zależnych Wilcoxon'a możliwym dzięki odpowiedniej liczebności 35 próbek oraz dla poszczególnych cyfr testem t porównującym rozkład z zadaną stałą (procentowa liczebność cyfry w rozkładzie Benforda). Ta metoda została zastosowana w celu większego uwiarygodnienia uzyskanych wyników.

2.5. Porównanie ogólnego odchylenia od rozkładu Benforda

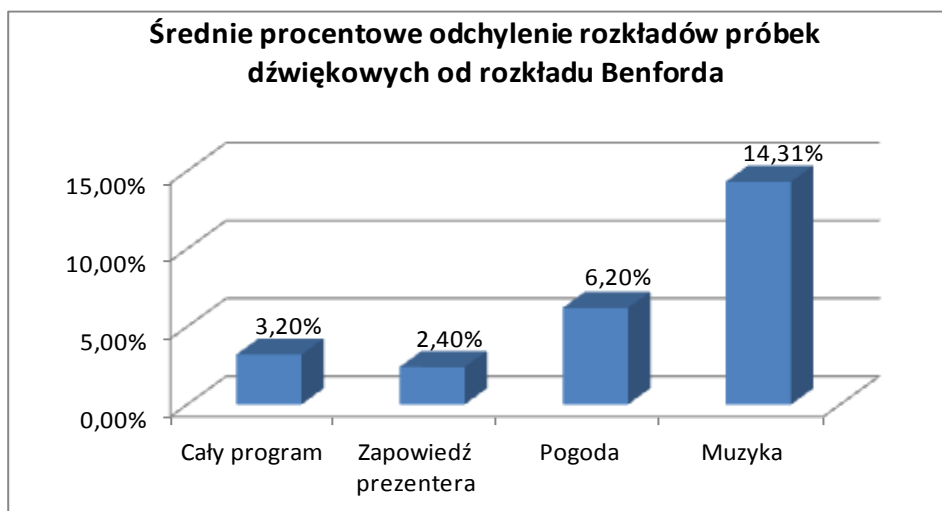
Wyodrębniono cztery kategorie próbek dźwięku: Cały program, Zapowiedź prezentera, Pogoda (z podkładem muzycznym) oraz Muzyka (fragment muzyki). Wyodrębnione grupy poddano porównaniu z zakładanym rozkładem Benforda.

2.5.1. Porównanie procentowych odchyleń poszczególnych próbek

Jak widać w tabeli 3. procentowe odchylenie od wszystkich cyfr rozkładu Benforda pomiędzy tymi kategoriami jest zróżnicowane. Zauważono istotne różnice w całkowitym

procentowym odchyleniu od Rozkładu Benforda pomiędzy wszystkimi tymi próbkami. Jak wynika z tabeli 3. najbardziej odbiegającym od rozkładu Benforda jest nagranie Muzyka (14,31%), zaraz potem ilustrowana podkładem muzycznym Pogoda (6,20%), następnie różnego rodzaju posklejane dźwięki przedstawione w Całym programie (3,20%), natomiast najbliższej rozkładu Benforda znajduje się nagranie głosu ludzkiego zawarte w Zapowiedzi prezentera 2,40%.

Rysunek 6. Średnie procentowe odchylenie rozkładów próbek dźwiękowych od rozkładu Benforda.



źródło: opracowanie własne

Tabela 3. Rozkład procentowego odchylenia próbek dźwięku od rozkładu Benforda.

Procentowe odchylenia próbek od rozkładu Benforda	N	Średnia	Minimum	Maksimum	Odchylenie Standardowe
Cały program (próbki)	35	3,20%	2,06%	5,73%	0,95%
Zapowiedź prezentera	35	2,40%	1,02%	6,47%	1,33%
Pogoda	32	6,20%	2,75%	9,52%	1,50%
Muzyka	35	14,31%	7,70%	20,27%	3,14%

źródło: opracowanie własne

2.5.2. Badanie istotności różnic w odchyleniach procentowych przy pomocy testu prób zależnych Wilcoxon'a

Za pomocą testu Wilcoxona zmierzono czy różnice pomiędzy średnimi procentowymi odchyleniami od rozkładu Benforda są istotne statystycznie. Różnice zostały obliczone na podstawie zestawienia procentowych odchyleń 35 lub 32 próbek dźwięku w każdej kategorii. Różnice te przedstawione w tabeli numer 4. Okazało się, iż wszystkie kategorie różnią się od siebie w sposób bardzo istotny. Najmniejsza różnica $p = 0,002$ została zaobserwowana pomiędzy nagraniem 'Całego programu' a 'Zapowiedzią prezentera'. Prawdopodobnie większość materiałów zawartych w programie nie jest w znaczny sposób przetworzona choć niewątpliwie stopień tego hipotetycznego przetworzenia jest wyższy niż w przypadku naturalnego głosu

ludzkiego. Dwie największe różnice, które można wywnioskować po wysokim wyniku testu $Z = 5,159$ zaobserwowane zostały przy porównaniu 'Całego programu' oraz 'Zapowiedzi prezentera' z nagraniem 'Muzyki'. Być może w nagraniach muzycznych stosowane były zabiegi mające na celu poprawę jakości dźwięku, które zmieniły oryginalną jego formę.

Tabela 4. Porównanie procentowych odchyłeń od rozkładu Benforda pomiędzy poszczególnymi rodzajami próbek dźwięku za pomocą testu Wilcoxon.

Porównania poszczególnych próbek	N	T	Z	poziom p
Cały program i Zapowiedź prezentera	35	126	3,096	0,002
Cały program i Pogoda	32	0	4,937	<0,001
Cały program i Muzyka	35	0	5,159	<0,001
Zapowiedź prezentera i Pogoda	32	8	4,787	<0,001
Zapowiedź prezentera i Muzyka	35	0	5,159	<0,001
Pogoda i Muzyka	32	0	4,937	<0,001

źródło: opracowanie własne

2.6. Porównanie odchyłeń poszczególnych cyfr w badanych próbkach dźwięku

To badanie ma na celu obserwację, czy któreś z poszczególnych cyfr rozkładu Benforda w zaobserwowanych próbkach różnią się istotnie, odchylają się w jakimś kierunku oraz czy układają się w jakimś szczególnym porządku.

2.6.1. Porównanie procentowych odchyłeń czterech w czterech grupach próbek dźwięku

Analizując odchylenia poszczególnych cyfr w 'Całym programie' (Tab. 5) od rozkładu Benforda można dojść do ogólnego wniosku iż niedobór wystąpień dotyczy cyfr 1 i 2, cyfra 3 biorąc pod uwagę wartości minimalną i maksymalną jest zgodna z rozkładem, cyfry 4 do 7 występują w nadmiarze aby cyfry 8 i 9 mogły pozostać mniej więcej zgodne z tym rozkładem.

Tabela 5. Zestawienie odchyłeń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach Całego programu.

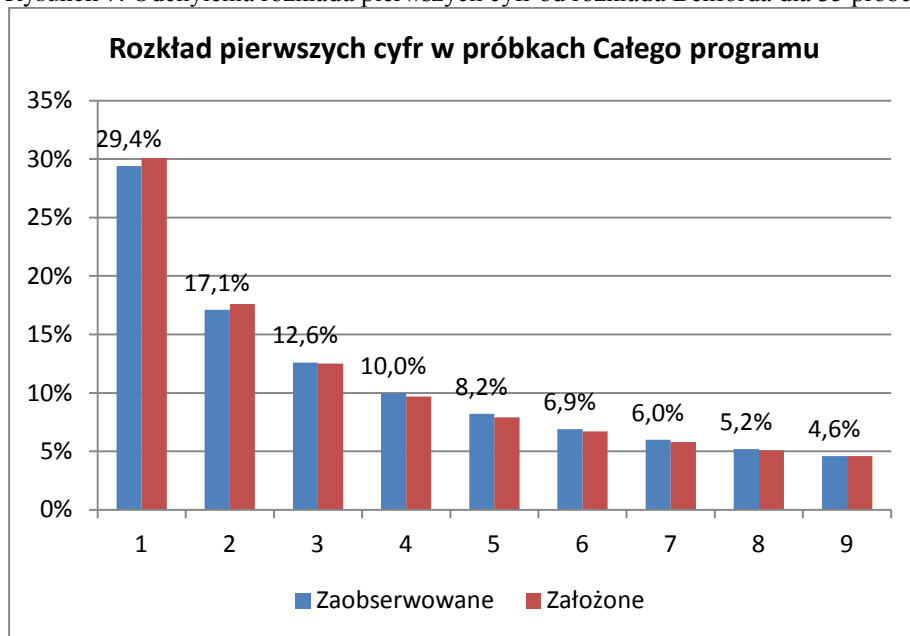
	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	35	-0,008	-0,021	0,008	0,006
Odchylenie 2 cyfry	35	-0,005	-0,015	0,006	0,005
Odchylenie 3 cyfry	35	0,001	-0,009	0,008	0,004
Odchylenie 4 cyfry	35	0,003	-0,002	0,007	0,002
Odchylenie 5 cyfry	35	0,003	0,001	0,007	0,001
Odchylenie 6 cyfry	35	0,002	-0,001	0,005	0,001
Odchylenie 7 cyfry	35	0,002	-0,002	0,005	0,002
Odchylenie 8 cyfry	35	0,001	-0,003	0,004	0,002
Odchylenie 9 cyfry	35	<0,001	-0,003	0,003	0,002

źródło: opracowanie własne

Jak widać w zestawieniu w zawartym tabeli 7 cyframi które najmniej różnią się od rozkładu Benforda w próbkach Całego programu są 3 i 9. Rozkład zaobserwowany jest

przesunięty w kierunku cyfr wyższych. Cyfry poniżej 3 mają wyraźny niedobór liczebności. Cyfra 3 jest w sposób istotny statystycznie zgodna ze średnią z rozkładu Benforda. Cyfry powyżej 3 aż do 8 mają nadmiar przy czym cyfra 5 posiada największy zważywszy na wynik testu $t = 15,408$. Cyfra 9 z dużym prawdopodobieństwem jest zgodna z rozkładem Benforda (patrz Tab. 9).

Rysunek 7. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek całego programu.



źródło: opracowanie własne

Średnie wartości odchyłeń od rozkładu Benforda dla Zapowiedzi prezentera są praktycznie znikomo różne. Na podstawie poniższego zestawienia można wywnioskować iż jedyną cyfrą która ma istotny niedobór jest 1. W pewnym stopniu różne są cyfry 3 do 5 a pozostałe nie różnią się wcale (Tab. 6).

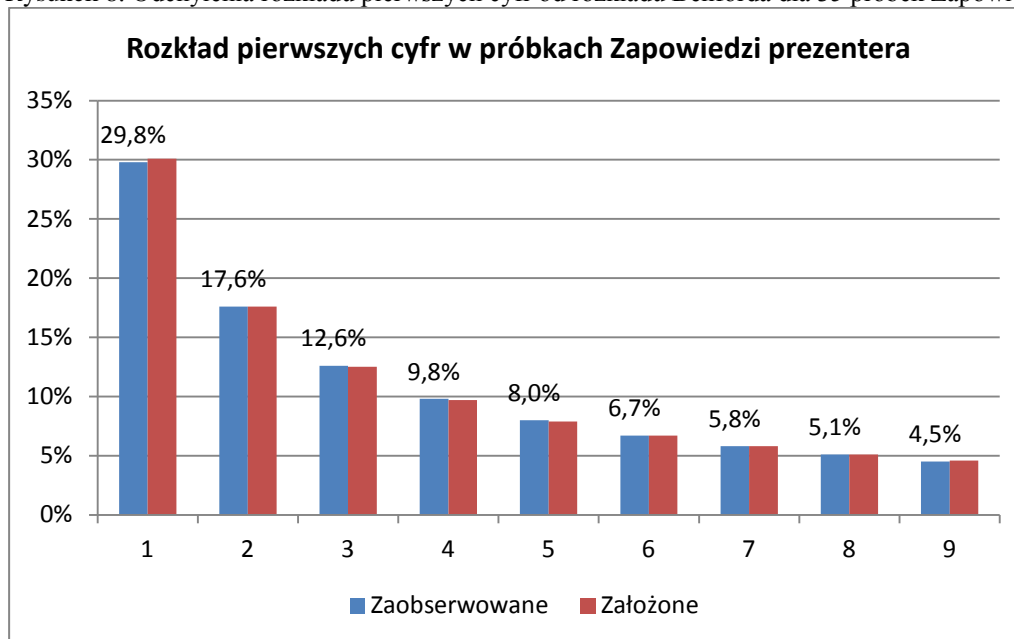
Tabela 6. Zestawienie odchyłeń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach Zapowiedzi prezentera.

	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	35	-0,003	-0,024	0,010	0,007
Odchylenie 2 cyfry	35	<0,001	-0,015	0,014	0,006
Odchylenie 3 cyfry	35	0,001	-0,007	0,008	0,003
Odchylenie 4 cyfry	35	0,001	-0,003	0,011	0,003
Odchylenie 5 cyfry	35	0,001	-0,003	0,009	0,002
Odchylenie 6 cyfry	35	<0,001	-0,003	0,005	0,002
Odchylenie 7 cyfry	35	<0,001	-0,004	0,005	0,002
Odchylenie 8 cyfry	35	<0,001	-0,005	0,004	0,002
Odchylenie 9 cyfry	35	<0,001	-0,005	0,004	0,002

źródło: opracowanie własne

Jak widać w tabeli nr 9 dla próbek 'Zapowiedzi prezentera' niedomiar liczebności występuje dla cyfry 1, cyfra 2 jest praktycznie zgodna z rozkładem Benforda $p = 0,974$, nieznaczny choć istotny nadmiar dotyczy cyfr 3, 4, 5. Pozostałe liczebności cyfr od 6 do 9 nie różnią istotnie od rozkładu Benforda. Przy czym prawie żadnych różnic nie odnotowano w przypadku cyfr 6 i 7. Cyfry 8 i 9 są w mniejszym stopniu zgodne z rozkładem Benforda.

Rysunek 8. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Zapowiedzi prezentera.



źródło: opracowanie własne

Oceniając poniższe wyniki (Tab. 7) dotyczące próbek 'Pogody' można dojść do wniosku iż tak jak w przypadku 'Całego programu' występuje niedobór cyfr 1 i 2 oraz, że cyfry 8 oraz 9 wyglądają na bardzo zbliżone do rozkładu Benforda.

Tabela 7. Zestawienie odchyłeń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 32 próbkach Pogody.

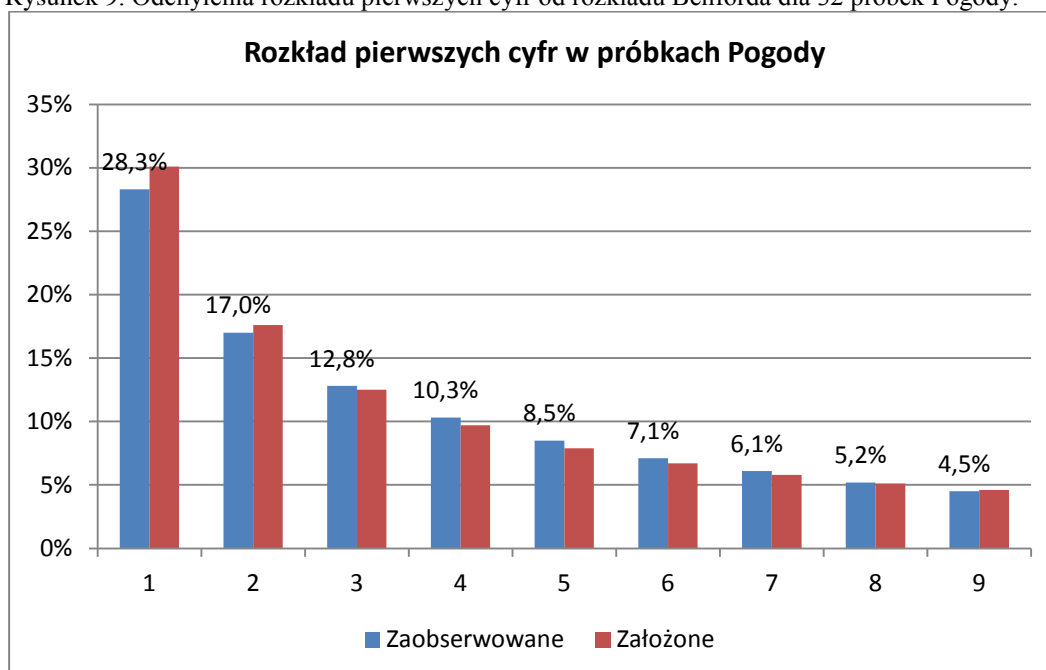
	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	32	-0,018	-0,036	0,016	0,013
Odchylenie 2 cyfry	32	-0,006	-0,023	0,012	0,008
Odchylenie 3 cyfry	32	0,003	-0,016	0,014	0,007
Odchylenie 4 cyfry	32	0,006	-0,004	0,013	0,004
Odchylenie 5 cyfry	32	0,006	-0,002	0,012	0,003
Odchylenie 6 cyfry	32	0,005	-0,004	0,009	0,003
Odchylenie 7 cyfry	32	0,003	-0,004	0,007	0,003
Odchylenie 8 cyfry	32	0,001	-0,005	0,007	0,003
Odchylenie 9 cyfry	32	<0,001	-0,006	0,006	0,003

źródło: opracowanie własne

Tabela 9. Pokazuje różnice pomiędzy średnimi częstościami poszczególnych cyfr w próbkach 'Pogody'. Jedynymi cyframi, których średnie nie różnią się od rozkładu Benforda są

3, 8 oraz 9. Tak jak w przypadku całego programu niedobór dotyczy cyfr 1 i 2 a najbardziej odbiegającą od rozkładu cyfrą jest 5 (Wykres 9).

Rysunek 9. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 32 próbek Pogody.



źródło: opracowanie własne

Dla próbek 'Muzyki' (Tab. 8) rozkład poszczególnych cyfr wydaje się być zbliżony do 'Całego programu'. Cyfry 1 oraz 2 występują w niedoborze, cyfra 3 różni się mało natomiast cyfry 8 oraz 9 różnią się bardzo mało. Cyfry od 4 do 7 reprezentują również tak jak 'Całym programie' największy nadmiar.

Tabela 8. Zestawienie odchyłeń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach Muzyki.

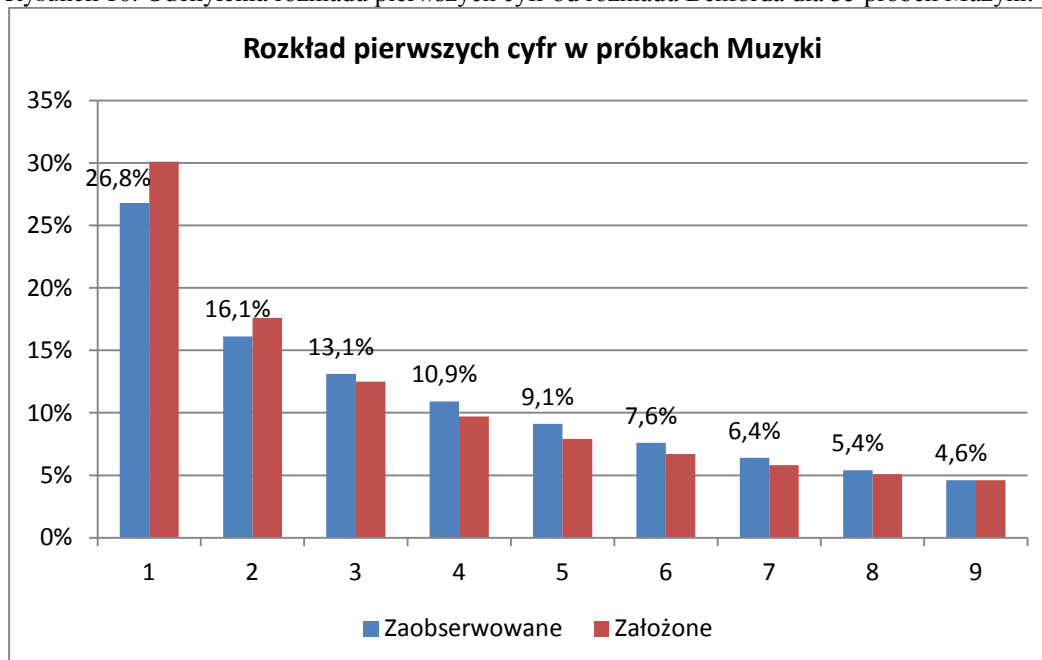
	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	35	-0,033	-0,075	0,027	0,025
Odchylenie 2 cyfry	35	-0,015	-0,054	0,029	0,026
Odchylenie 3 cyfry	35	0,006	-0,028	0,030	0,016
Odchylenie 4 cyfry	35	0,012	-0,008	0,026	0,008
Odchylenie 5 cyfry	35	0,011	-0,009	0,020	0,006
Odchylenie 6 cyfry	35	0,009	-0,011	0,020	0,008
Odchylenie 7 cyfry	35	0,006	-0,011	0,017	0,009
Odchylenie 8 cyfry	35	0,003	-0,010	0,014	0,009
Odchylenie 9 cyfry	35	0,001	-0,012	0,014	0,008

źródło: opracowanie własne

Tabela 9 przedstawia wyniki testów odchylenia poszczególnych cyfr w próbkach 'Muzyki' od idealnego rozkładu Benforda. Jedyne cyfry, które są zgodne z rozkładem Benforda to 8 i 9. Wszystkie pozostałe cyfry odbiegają od rozkładu Benforda w sposób istotny

statystycznie. Można zauważyć tak jak to miało miejsce we wcześniejszych zestawieniach iż cyfra 3 zdaje się być najbliższej zgodności w porównaniu z resztą niezgodnych cyfr. Dodatkowo cyfra 5 również jest najdalej od zakładanego rozkładu.

Rysunek 10. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Muzyki.



źródło: opracowanie własne

Podsumowując odchylenia cyfr w poszczególnych kategoriach od rozkładu Benforda najmniej podatną na odchylenie cyfrą w rozkładzie jest 9. Najczęściej dwie pierwsze cyfry występują w niedomiarze na korzyść cyfr od 4 do 7. Najbardziej odbiegającą cyfrą od rozkładu najczęściej jest 5. W przypadku głosu ludzkiego rozkład jest bardziej przesunięty ku mniejszym cyfrom i tylko cztery cyfry 1, 3, 4, 5 są różne od założeń rozkładu Benforda.

Tabela 9. Porównanie procentowych liczebności poszczególnych cyfr z wartością średnią rozkładu Benforda za pomocą testu t-Studenta dla 4 rodzajów próbek.

	Cały program		Zapowiedź prezentera		Pogoda		Muzyka	
	t	p	t	p	t	p	t	p
Odchylenie 1 cyfry	-7,715	<0,001	-2,394	0,022	-7,728	<0,001	-7,812	<0,001
Odchylenie 2 cyfry	-5,189	<0,001	-0,033	0,974	-4,358	<0,001	-3,33	0,002
Odchylenie 3 cyfry	1,736	0,092	2,524	0,016	2,591	0,014	2,303	0,028
Odchylenie 4 cyfry	9,988	<0,001	3,079	0,004	8,119	<0,001	8,712	<0,001
Odchylenie 5 cyfry	15,408	<0,001	2,492	0,018	10,649	<0,001	10,943	<0,001
Odchylenie 6 cyfry	9,74	<0,001	0,261	0,796	8,14	<0,001	6,865	<0,001
Odchylenie 7 cyfry	5,738	<0,001	-0,065	0,948	5,274	<0,001	3,858	<0,001
Odchylenie 8 cyfry	3,046	0,004	-1,141	0,262	2,023	0,052	1,955	0,059
Odchylenie 9 cyfry	0,978	0,335	-1,755	0,088	-0,575	0,569	0,54	0,592

(źródło: opracowanie własne)

2.7. Porównanie odchylen po odjęciu składowej stałej oraz odszumieniu

Cztery kategorie dźwięku z poprzedniego rozdziału zostały poddane operacji przepuszczenia przez filtr górnoprzepustowy w celu odszumienia oraz odjęcia składowej stałej. Tak zmienione pliki zostały poddane powtórnej analizie.

2.7.1. Porównanie procentowych odchylen poszczególnych próbek

Po operacji odszumienia zależności pomiędzy rodzajami dźwięku nie zmieniły się. Nadal najbardziej odbiegającym od rozkładu Benforda rodzajem dźwięku jest próbka Muzyki, po czym kolejno coraz bliżej tego rozkładu są Pogoda, Cały program i Zapowiedź prezentera (Tab. 10).

Tabela 10. Rozkład procentowego odchylenia odszumionych próbek dźwięku od rozkładu Benforda.

Procentowe odchylenia odszumionych próbek od rozkładu Benforda	N	Średnia	Minimum	Maksimum	Odchylenie Standardowe
Cały program (próbki)	35	1,61%	0,28%	4,35%	0,75%
Zapowiedź prezentera	35	1,47%	0,25%	3,81%	0,84%
Pogoda	32	4,22%	2,03%	6,63%	1,35%
Muzyka	35	11,60%	7,09%	25,21%	3,64%

(źródło: opracowanie własne)

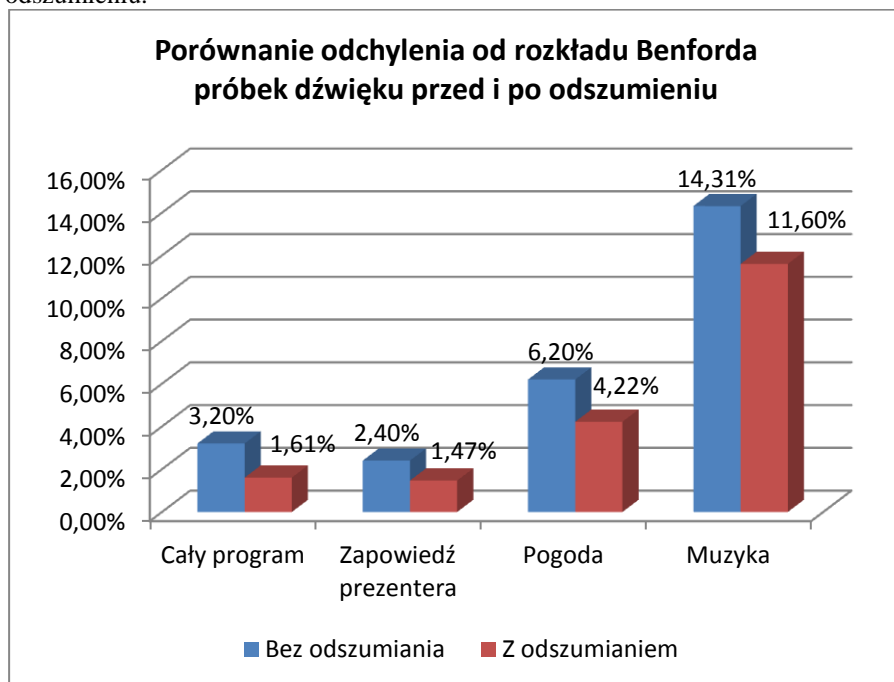
Porównując odchylenia od rozkładu Benforda plików podstawowych i odszumionych można dojść do wniosku iż operacja odszumienia spowodowała zbliżenie się próbek do rozkładu Benforda. Przy czym największe procentowe zbliżenie, bo o 50% nastąpiło w przypadku Całego programu, następnie Zapowiedzi prezentera o 39%, następnie 32% dla Pogody a najmniej dla Muzyki 19%.

Tabela 11. Rozkład procentowego odchylenia nieodszumionych i odszumionych próbek dźwięku od rozkładu Benforda.

Porównanie podstawowych oraz odszumionych próbek	N	Średnia		Minimum		Maksimum		Odchylenie Standardowe	
		Norm	Odszum	Norm	Odszum	Norm	Odszum	Norm	Odszum
Cały program (próbki)	35	3,20%	1,61%	2,06%	0,28%	5,73%	4,35%	0,95%	0,75%
Zapowiedź prezentera	35	2,40%	1,47%	1,02%	0,25%	6,47%	3,81%	1,33%	0,84%
Pogoda	32	6,20%	4,22%	2,75%	2,03%	9,52%	6,63%	1,50%	1,35%
Muzyka	35	14,31%	11,60%	7,70%	7,09%	20,27%	25,21%	3,14%	3,64%

(źródło: opracowanie własne)

Rysunek 11. Średnie procentowe odchylenie rozkładów próbek dźwiękowych od rozkładu Benforda przed i po odszumieniu.



źródło: opracowanie własne

2.7.2. Badanie istotności różnic procentowych odchyłeń przy pomocy testu prób zależnych Wilcoxon'a

Wykonane zostały dwa rodzaje porównań. Jedno dotyczące czterech różnych rodzajów dźwięku i drugie dotyczące wyniku przeprowadzonej operacji odszumienia. W pierwszym przypadku większość próbek różni się od siebie w sposób istotny statystycznie podczas gdy średnia z Całego programu jest równa średniej z Zapowiedzi prezentera $p = 0,422$.

Wszystkie operacje odszumienia dały istotną zmianę zgodności z rozkładem Benforda. Wszystkie te rozkłady uległy istotnemu przybliżeniu do rozkładu Benforda po operacji odszumienia (Tab. 13). Trudno jest powiedzieć coś o przyczynie tego zjawiska.

Tabela 12. Porównanie procentowych odchyłeń od rozkładu Benforda pomiędzy poszczególnymi rodzajami próbek dźwięku za pomocą testu Wilcoxon'a.

Porównania poszczególnych próbek	N	T	Z	poziom p
Cały program i Zapowiedź prezentera	35	266	0,803	0,422
Cały program i Pogoda	32	0	4,937	<0,001
Cały program i Muzyka	35	0	5,159	<0,001
Zapowiedź prezentera i Pogoda	32	4	4,862	<0,001
Zapowiedź prezentera i Muzyka	35	0	5,159	<0,001
Pogoda i Muzyka	32	0	4,937	<0,001

źródło: opracowanie własne

Z pewnością w trakcie odszumiania z dźwięku usuwane są pewne częstotliwości. Nie wiadomo również czy kolejne przeprowadzone tego typu operacje na dźwięku dalej w istotny sposób zbliżają do rozkładu Benforda. Być może należałoby w celu ustalenia przyczyny tego

zjawiska wykonać dodatkowe testy. To co zostało udowodnione w tej pracy to wpływ jednokrotnej operacji odsumiania poprzez zastosowanie filtra górnoprzepustowego w programie do edycji dźwięku „Audicity 1.2.6” w sposób istotny zbliża rozkład dźwięku do rozkładu Benforda.

Tabela 13. Porównanie procentowych odchyłeń od rozkładu Benforda próbek przed i po odsumieniu za pomocą testu Wilcoxon.

Porównania poszczególnych próbek	N	T	Z	poziom p
Cały program i Cały program odsumiony	35	0	5,159	<0,001
Zapowiedź prezentera i Zapowiedź odsumiona	35	104	3,456	0,001
Pogoda i Pogoda odsumiona	32	33	4,319	<0,001
Muzyka i Muzyka odsumiona	35	62	4,144	<0,001

źródło: opracowanie własne

2.8. Porównanie odchyłeń poszczególnych cyfr dla odsumionych próbek

2.8.1. Porównanie procentowych odchyłeń cyfr w czterech grupach odsumionych próbek dźwięku

Odchylenia Całego programu dla odsumionych próbek programu jak widać w tabeli nr 14 są bardzo małe ich średnie nie przekraczają 0,3%. Największe dodatnie odchylenie dotyczy cyfry 2 (0,24%) następnie cyfry 1 (0,22%). Dodatnie odchylenie ma jeszcze tylko cyfra 3 i jest ono bardzo nieznaczne 0,05%. Pozostałe cyfry występują w niedomiarze odwrotnie w stosunku do nieodsumionego materiału. Tak jak przed odsumieniem cyfry 5 oraz 6 są w dużym relatywnie stopniu odchyłone a cyfra 9 tak jak przedtem ma odchylenie najmniejsze.

Tabela 14. Zestawienie odchyłeń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach odsumionego Całego programu.

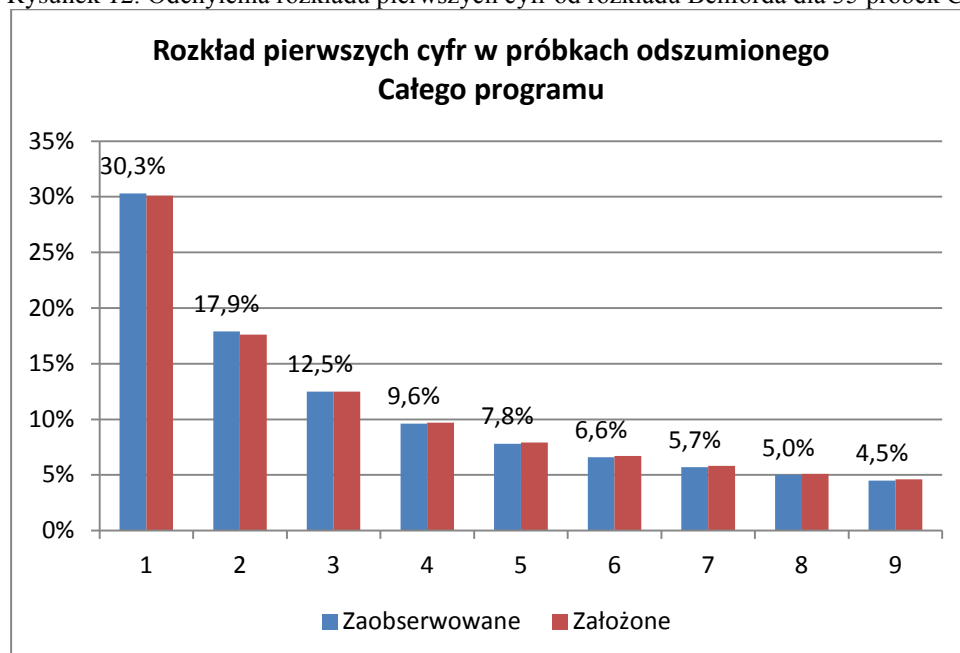
	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	35	0,0022	-0,0070	0,0111	0,0039
Odchylenie 2 cyfry	35	0,0024	-0,0030	0,0099	0,0025
Odchylenie 3 cyfry	35	0,0005	-0,0040	0,0058	0,0023
Odchylenie 4 cyfry	35	-0,0006	-0,0039	0,0029	0,0015
Odchylenie 5 cyfry	35	-0,0011	-0,0040	0,0003	0,0010
Odchylenie 6 cyfry	35	-0,0011	-0,0042	0,0007	0,0009
Odchylenie 7 cyfry	35	-0,0010	-0,0036	0,0011	0,0009
Odchylenie 8 cyfry	35	-0,0008	-0,0029	0,0014	0,0009
Odchylenie 9 cyfry	35	-0,0005	-0,0022	0,0012	0,0009

źródło: opracowanie własne

W porównaniu procentowych odchyłeń poszczególnych cyfr od 1 do 9 od średnich założonych przez prawo Benforda dla Całego programu można zauważyć, że jedyną cyfrą zgodną z tym rozkładem jest cyfra 3 (Tab. 18). Dwie pierwsze cyfry występują w nadmiarze.

Cyfrы od 4 do 9 występują w niedomiarze. Wygląda na to, że operacja odszumiania metodą filtra górnoprzepustowego powoduje przesunięcie rozkładu w kierunku mniejszych cyfr. Dodatkowo jego środek pozostaje niezmienny przy cyfrze 3.

Rysunek 12. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Całego programu.



źródło: opracowanie własne

Po analizie odchylenia od rozkładu Benforda odszumionych Zapowiedzi prezentera zauważono iż dwie pierwsze cyfry występują w nadmiarze, pozostałe w niedomiarze. Dodatkowo największe odchylenie dotyczy cyfry 1, natomiast najmniejsze dotyczą cyfr 8 i 9. Należy również nadmienić iż to zróżnicowanie odchyłeń może wynikać z samych wielkości częstości, które dla 1 wynosi 0,301 a dla cyfry 9 (0,046). Gdyby wziąć pod uwagę i odfiltrować tę zależność to najbardziej odchyłonymi oprócz cyfry 1 cyframi byłyby 4 i 5.

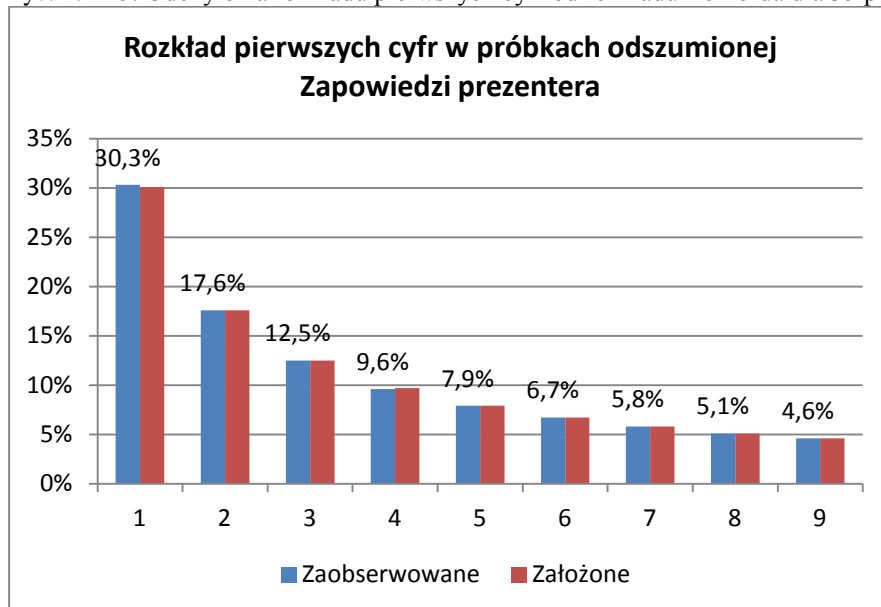
Tabela 15. Zestawienie odchyłeń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach odszumionej Zapowiedzi prezentera.

	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	35	0,0018	-0,0077	0,0169	0,0055
Odchylenie 2 cyfry	35	0,0004	-0,0067	0,0073	0,0029
Odchylenie 3 cyfry	35	-0,0004	-0,0060	0,0036	0,0025
Odchylenie 4 cyfry	35	-0,0006	-0,0050	0,0021	0,0019
Odchylenie 5 cyfry	35	-0,0005	-0,0042	0,0030	0,0016
Odchylenie 6 cyfry	35	-0,0003	-0,0040	0,0024	0,0012
Odchylenie 7 cyfry	35	-0,0002	-0,0028	0,0020	0,0010
Odchylenie 8 cyfry	35	-0,0001	-0,0024	0,0023	0,0009
Odchylenie 9 cyfry	35	-0,0001	-0,0017	0,0025	0,0009

(źródło: opracowanie własne)

Gdy porównamy średnie częstości cyfr rozkładu odsumionej Zapowiedzi prezentera z rozkładem Beforda, to okaże się, iż wszystkie cyfry są zgodne z tym rozkładem. Najbardziej skłonną do odejścia od rozkładu nadal pozostaje cyfra 1 a najmniej cyfra 9. Cyfra 4 wykazuje najwyższą po jedynce skłonność do odejścia od rozkładu Benforda, choć to w tych rozkładach nie następuje.

Rysunek 13. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Zapowiedzi prezentera.



źródło: opracowanie własne

Średnie wartości odchyień od rozkładu Benforda dla 32 próbek z odsumionym nagraniem Pogody wyróżnia relatywnie duża nadmiarowość cyfry 2 (0,77%) oraz 3 (0,45%), niedomiar cyfry 1 oraz cyfr od 5 do 9. Najmniej odchylną od zakładanego rozkładu jest cyfra 4 (0,08%). Mimo tego, że odchylenia od zakładanego rozkładu są małe to i tak są znacznie większe od również odsumionych próbek z zapisem Zapowiedzi prezentera oraz Całego programu.

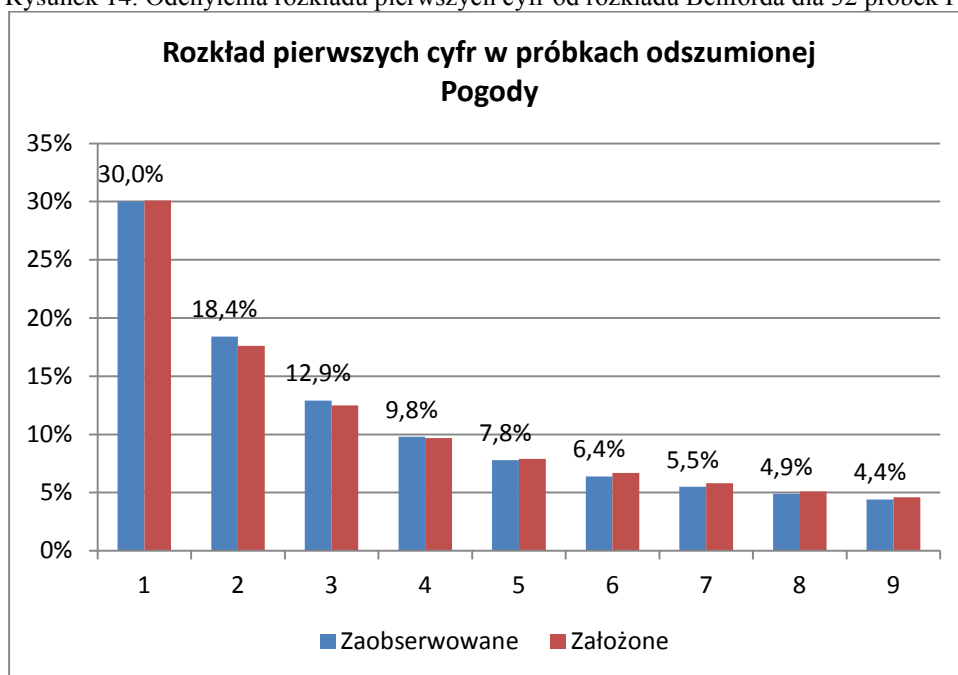
Tabela 16. Zestawienie odchyień poszczególnych cyfr od rozkładu Benforda zaobserwowane w 32 próbkach odsumionej Pogody.

	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	32	-0,0014	-0,0286	0,0120	0,0096
Odchylenie 2 cyfry	32	0,0077	-0,0034	0,0168	0,0048
Odchylenie 3 cyfry	32	0,0045	-0,0021	0,0112	0,0034
Odchylenie 4 cyfry	32	0,0008	-0,0046	0,0101	0,0037
Odchylenie 5 cyfry	32	-0,0015	-0,0056	0,0083	0,0033
Odchylenie 6 cyfry	32	-0,0025	-0,0062	0,0053	0,0025
Odchylenie 7 cyfry	32	-0,0027	-0,0055	0,0025	0,0017
Odchylenie 8 cyfry	32	-0,0026	-0,0049	-0,0002	0,0013
Odchylenie 9 cyfry	32	-0,0022	-0,0043	0,0002	0,0011

źródło: opracowanie własne

W plikach z odszumioną Pogodą w wyniku porównania cyfr z rozkładem Benforda cyfry 1 oraz 4 okazują się być z nim zgodne. Cyfry 2 i 3 reprezentują nadmiar natomiast cyfry od 5 do 9 niedomiar.

Rysunek 14. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 32 próbek Pogody.



źródło: opracowanie własne

Analizując tabelę z odchyleniami cyfr od 1 do 9 dla odszumionych nagrań Muzyki rzucają się w oczy dwa relatywnie duże odchylenia dla cyfr 2 (2,01%) i 3 (1,01%). Tak samo jak w przypadku próbek odszumionej pogody nieznacznie odchylna ujemnie jest cyfra 1 (0,17%) oraz również nieznacznie dodatnio cyfra 4 (0,07%). Pozostałe cyfry od 5 do 9 są tak samo jak w przypadku Pogody w mniej więcej równym stopniu odchylnone od rozkładu Benforda.

Tabela 17. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach odszumionej Muzyki.

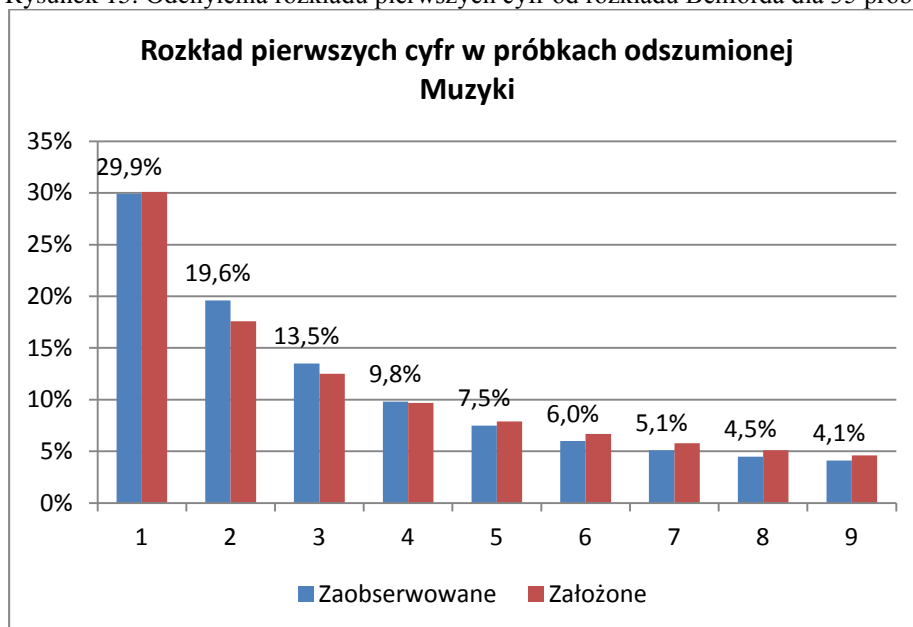
	N	Średnia	Minimum	Maksimum	Odch. std.
Odchylenie 1 cyfry	35	-0,0017	-0,0521	0,0464	0,0333
Odchylenie 2 cyfry	35	0,0201	0,0041	0,0379	0,0094
Odchylenie 3 cyfry	35	0,0101	-0,0159	0,0414	0,0139
Odchylenie 4 cyfry	35	0,0007	-0,0213	0,0299	0,0137
Odchylenie 5 cyfry	35	-0,0043	-0,0210	0,0115	0,0095
Odchylenie 6 cyfry	35	-0,0065	-0,0159	0,0059	0,0056
Odchylenie 7 cyfry	35	-0,0069	-0,0182	0,0011	0,0041
Odchylenie 8 cyfry	35	-0,0062	-0,0196	0,0019	0,0041
Odchylenie 9 cyfry	35	-0,0052	-0,0167	0,0036	0,0044

źródło: opracowanie własne

Analizując wyniki testów t-Studenta odchyleń poszczególnych cyfr dla 35 próbek odszumionej Muzyki zauważalna jest zgodność cyfr 1 ($p = 0,764$) i 4 ($p = 0,767$) z rozkładem

Benforda, duże dodatnie odchylenie cyfr 2 i 3 oraz ujemne odchylenie cyfr od 5 do 9. Najbardziej odchylną cyfrą spośród ujemnie odchylonych jest cyfra 7 ($t = -10,007$) natomiast najmniej 5 ($t = -2,692$). Wyniki przedstawiono w tabeli 18.

Rysunek 15. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Muzyki.



źródło: opracowanie własne

Tabela 18. Porównanie procentowych liczebności poszczególnych cyfr z wartością średnią rozkładu Benforda za pomocą testu t-Studenta dla 4 rodzajów próbek.

	Cały program		Zapowiedź prezentera		Pogoda		Muzyka	
	t	p	t	p	t	p	t	p
Odchylenie 1 cyfry	3,281	0,002	1,894	0,067	-0,852	0,401	-0,303	0,764
Odchylenie 2 cyfry	5,727	<0,001	0,809	0,424	9,115	<0,001	12,64	<0,001
Odchylenie 3 cyfry	1,263	0,215	-0,939	0,354	7,562	<0,001	4,318	<0,001
Odchylenie 4 cyfry	-2,489	0,018	-1,787	0,083	1,168	0,252	0,298	0,767
Odchylenie 5 cyfry	-6,433	<0,001	-1,687	0,101	-2,633	0,013	-2,692	0,011
Odchylenie 6 cyfry	-7,617	<0,001	-1,573	0,125	-5,813	<0,001	-6,903	<0,001
Odchylenie 7 cyfry	-6,438	<0,001	-1,457	0,154	-8,888	<0,001	-10,01	<0,001
Odchylenie 8 cyfry	-5,024	<0,001	-0,718	0,477	-11,440	<0,001	-9,034	<0,001
Odchylenie 9 cyfry	-3,781	0,001	-0,401	0,691	-11,400	<0,001	-7,045	<0,001

źródło: opracowanie własne

Podsumowując wyniki odchylenia poszczególnych cyfr od rozkładu Benforda po odsumieniu dla wszystkich czterech kategorii dźwięku należy zauważyć iż wszystkie próbki oprócz Całego programu mają nieodbiegające od rozkładu Benforda cyfry 1 i 4. Zapowiedź prezentera ma wszystkie cyfry nieodbiegające, natomiast Cały program ma tylko jedną zgodną cyfrę tj. 3 ($p = 0,215$).

2.8.2. Test wpływu operacji odsumiania na odchylenia od rozkładu poszczególnych cyfr

Tabela 19. Porównanie wpływu odsumiania na zmianę rozkładu poszczególnych cyfr dla Całego programu za pomocą testu Wilcoxon'a.

Porównania próbek Całego programu	N	T	Z	poziom p
1 cyfra nieodsumiona i odsumiona	35	45	4,422	<0,001
2 cyfra nieodsumiona i odsumiona	35	63	4,128	<0,001
3 cyfra nieodsumiona i odsumiona	35	267	0,786	0,432
4 cyfra nieodsumiona i odsumiona	35	31	4,652	<0,001
5 cyfra nieodsumiona i odsumiona	35	0	5,159	<0,001
6 cyfra nieodsumiona i odsumiona	35	11	4,979	<0,001
7 cyfra nieodsumiona i odsumiona	35	44	4,439	<0,001
8 cyfra nieodsumiona i odsumiona	35	117	3,243	0,001
9 cyfra nieodsumiona i odsumiona	35	198	1,916	0,055

źródło: opracowanie własne

Dla Całego programu wszystkie cyfry z wyjątkiem 3 i 9 poddały się procesowi zbliżania do rozkładu Benforda pod wpływem odsumiania (Tab. 19).

Tabela 20. Porównanie wpływu odsumiania na zmianę rozkładu poszczególnych cyfr dla Zapowiedzi prezentera za pomocą testu Wilcoxon'a.

Porównania próbek Zapowiedzi prezentera	N	T	Z	poziom p
1 cyfra nieodsumiona i odsumiona	35	169	2,391	0,017
2 cyfra nieodsumiona i odsumiona	35	302	0,213	0,831
3 cyfra nieodsumiona i odsumiona	35	193	1,998	0,046
4 cyfra nieodsumiona i odsumiona	35	139	2,883	0,004
5 cyfra nieodsumiona i odsumiona	35	159	2,555	0,011
6 cyfra nieodsumiona i odsumiona	35	265	0,819	0,413
7 cyfra nieodsumiona i odsumiona	35	280	0,573	0,566
8 cyfra nieodsumiona i odsumiona	35	293	0,360	0,719
9 cyfra nieodsumiona i odsumiona	35	267	0,786	0,432

źródło: opracowanie własne

Dla Zapowiedzi prezentera procesowi zbliżenia poddały się cyfry 1, 3, 4, 5. Pozostałe cyfry nie zbliżyły się do rozkładu Benforda patrz (Tab. 20).

Tabela 21. Porównanie wpływu odsumiania na zmianę rozkładu poszczególnych cyfr dla Pogody za pomocą testu Wilcoxon'a.

Porównania próbek Pogody	N	T	Z	poziom p
1 cyfra nieodsumiona i odsumiona	32	71	3,609	<0,001
2 cyfra nieodsumiona i odsumiona	32	15	4,656	<0,001
3 cyfra nieodsumiona i odsumiona	32	237	0,505	0,614
4 cyfra nieodsumiona i odsumiona	32	85	3,347	0,001
5 cyfra nieodsumiona i odsumiona	32	23	4,506	<0,001
6 cyfra nieodsumiona i odsumiona	32	11	4,731	<0,001
7 cyfra nieodsumiona i odsumiona	32	13	4,693	<0,001
8 cyfra nieodsumiona i odsumiona	32	42	4,151	<0,001
9 cyfra nieodsumiona i odsumiona	32	115	2,786	0,005

źródło: opracowanie własne

Dla próbek Pogody wszystkie cyfry oprócz 3 poddały się zbliżeniu do rozkładu Benforda pod wpływem odsumiania (Tab. 21). Taki sam efekt został osiągnięty w przypadku próbek Muzyki (Tab. 22).

We wszystkich czterech rodzajach próbek cyfra 3 najczęściej nie poddaje się procesowi odsumiania. Dodatkowo można wyciągnąć wniosek iż im dalej od rozkładu Benforda jest plik muzyczny tym bardziej poddaje się zbliżaniu. Dla próbek Zapowiedzi prezentera, która jest najbliższej zakładanego rozkładu tylko mniej więcej połowa cyfr poddała się procesowi zbliżania do rozkładu. Dla próbek dalekich od tego rozkładu zbliżaniu poddały się wszystkie cyfry oprócz trójki.

Tabela 22. Porównanie wpływu odsumienia na zmianę rozkładu poszczególnych cyfr dla Muzyki za pomocą testu Wilcoxon'a.

Porównania próbek Pogody	N	T	Z	poziom p
1 cyfra nieodsumiona i odsumiona	35	124	3,128	0,002
2 cyfra nieodsumiona i odsumiona	35	29	4,684	<0,001
3 cyfra nieodsumiona i odsumiona	35	263	0,852	0,394
4 cyfra nieodsumiona i odsumiona	35	141	2,850	0,004
5 cyfra nieodsumiona i odsumiona	35	4	5,094	<0,001
6 cyfra nieodsumiona i odsumiona	35	3	5,110	<0,001
7 cyfra nieodsumiona i odsumiona	35	23	4,783	<0,001
8 cyfra nieodsumiona i odsumiona	35	86	3,751	<0,001
9 cyfra nieodsumiona i odsumiona	35	151	2,686	0,007

źródło: opracowanie własne

Podsumowując cały rozdział można wyciągnąć następujące wnioski: 1) Oddalenie od rozkładu Benforda jest różne dla różnych rodzajów dźwięków, 2) Głos ludzki jest bliżej tego rozkładu niż inne analizowane w tej pracy rodzaje dźwięku, 3) Muzyka jest dźwiękiem, który najbardziej odbiega od rozkładu, 4) Poszczególne cyfry w różnym stopniu odbiegają od zakładanego rozkładu, 5) Najmniej od rozkładu odbiegają cyfry 3 oraz 9, 6) Jednokrotne odsumianie przy pomocy filtra górnoprzepustowego powoduje zbliżenie zarówno całego rozkładu jak i poszczególnych cyfr do rozkładu Benforda, 7) Zbliżenie to jest tym większe im dalej dana cyfra jest od zakładanego rozkładu.

3. Badanie odchylenia od rozkładu Benforda wskutek wycinania lub wklejania dźwięków

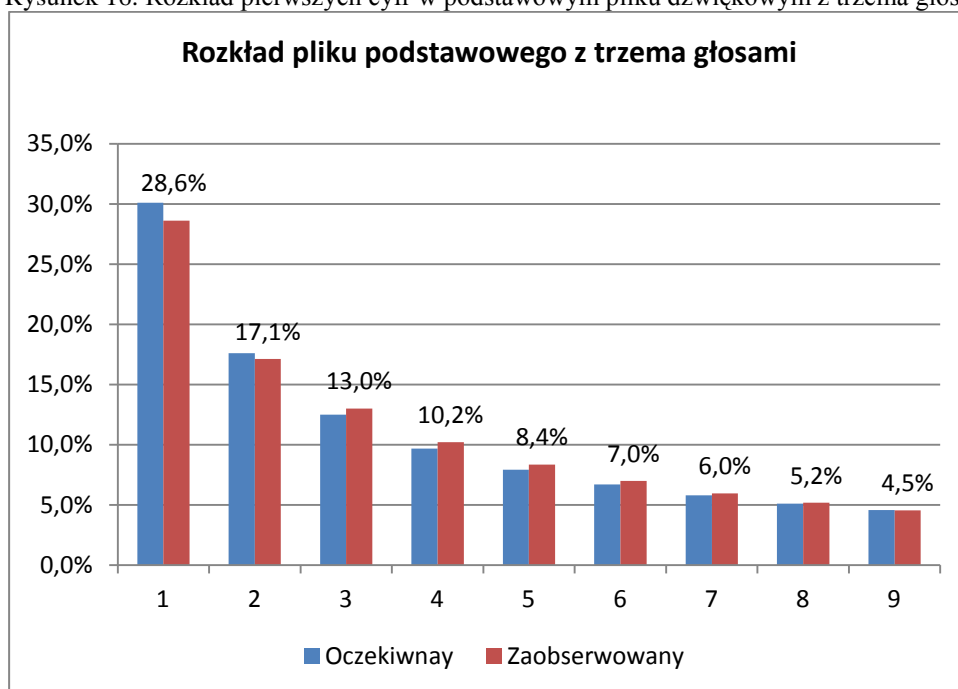
Do wykonania testów wycinania i wklejania dźwięku użyto pliku dźwiękowego z trzema różnymi głosami ludzkimi. W podrozdziale pierwszym jeden z tych głosów został wycięty, po czym zostało porównane odchylenie próbki przed i po operacji wycięcia. W drugim podrozdziale do trzech podstawowych głosów został dodany jeszcze jeden. Wzięte zostało również pod uwagę odchylenie od rozkładu wycinanych i wklejanych próbek.

3.1. Porównanie odchylenia przed i po usunięciu fragmentu dźwięku

3.1.1. Porównanie odchylenia ogólnych rozkładów

Wykres 12 prezentuje odchylenia podstawowego pliku z trzema głosami z którego został usunięty jeden z głosów. Wynik testu χ^2 który mierzy jego odchylenie od rozkładu Benforda wyniósł 4980,774 i jest to dość duże odchylenie od zakładanego rozkładu. Natomiast wartość miernika M1 dla tego rozkładu wyniosła 3,58% więc wynik ten jest porównywalny z odchyleniem, które dotyczy Całego programu bez usunięcia szumów.

Rysunek 16. Rozkład pierwszych cyfr w podstawowym pliku dźwiękowym z trzema głosami.

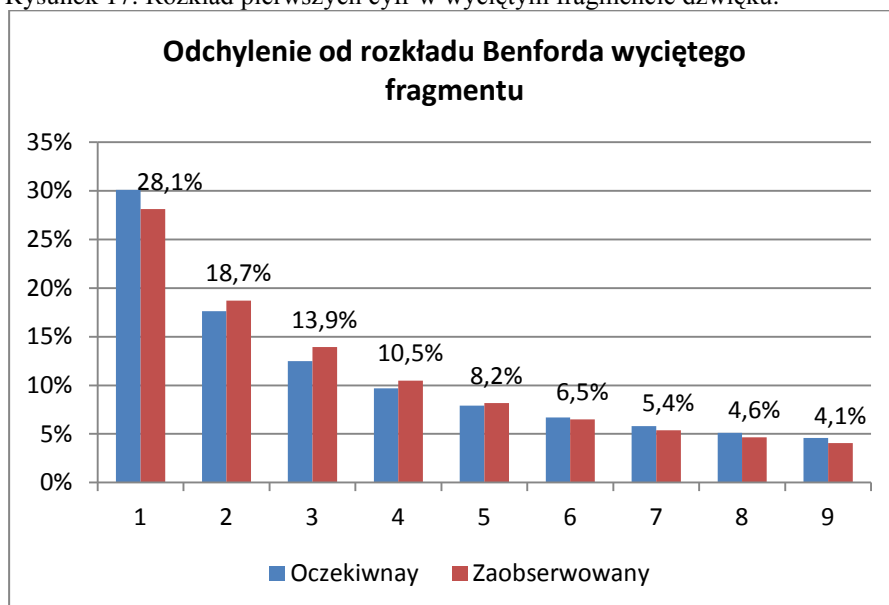


źródło: opracowanie własne

Wycięty fragment jest odchylony od rozkładu o 7,43% a mimo to z powodu jego mniejszej długości wartość testu χ^2 wynosi 3292,630 czyli mniej niż w pliku podstawowym. Biorąc pod uwagę bardziej istotne w tym porównaniu odchylenie mierzone miernikiem M1

można stwierdzić, iż wycięty fragment bardziej odbiegał od rozkładu Benforda niż podstawowa próbka z trzema głosami.

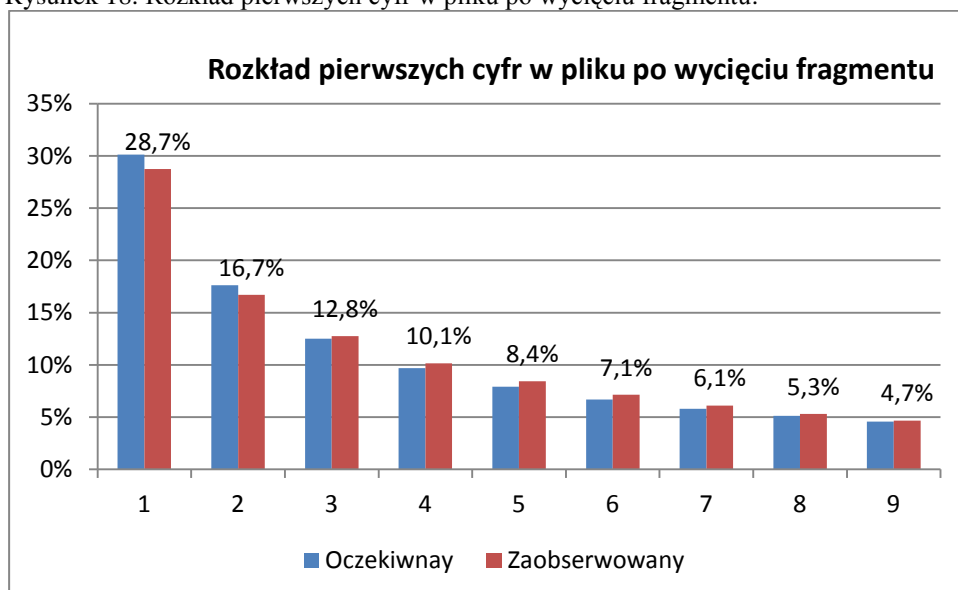
Rysunek 17. Rozkład pierwszych cyfr w wyciętym fragmencie dźwięku.



źródło: opracowanie własne

Podstawowa 56 sekundowa próbka dźwięku po wycięciu 11 sekundowego fragmentu charakteryzowała się odchyleniem od rozkładu Benforda mierzonym testem χ^2 na poziomie $t = 4823,743$ czyli bliskim podstawowego. Natomiast odchylenie zmierzone miernikiem M1 wyniosło 4,54%. Mimo iż wynik testu χ^2 obniżył się nieznacznie, to nominalne procentowe odchylenie od rozkładu Benforda zwiększyło się po operacji usunięcia fragmentu. Gdyby rozkład ten rządził się prawami rozkładu normalnego, to można by po usunięciu fragmentu bardziej odchylonego oczekiwać zbliżenia do zakładanego rozkładu. Jednak w przypadku rozkładu Benforda nastąpiło oddalenie próbki po wycięciu fragmentu od zakładanego rozkładu.

Rysunek 18. Rozkład pierwszych cyfr w pliku po wycięciu fragmentu.

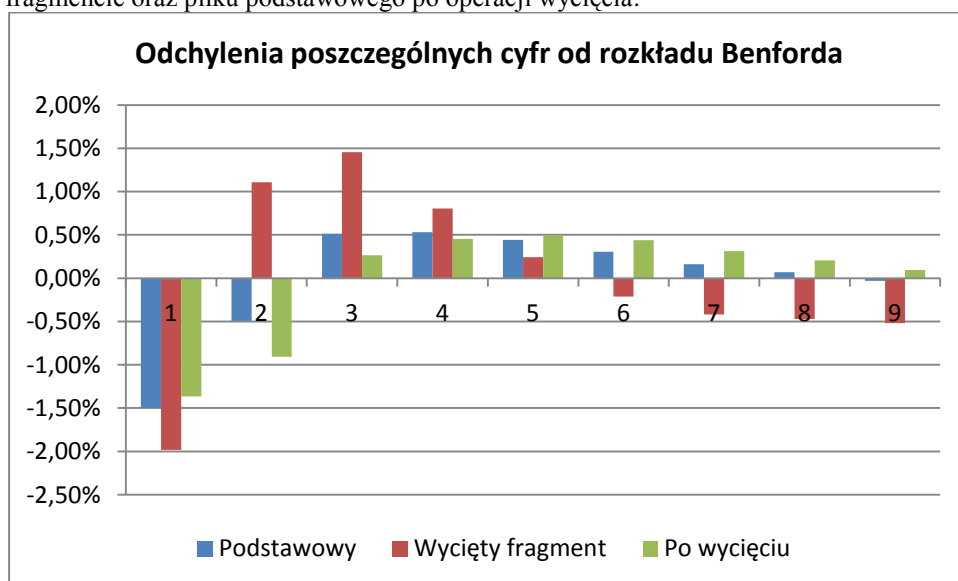


źródło: opracowanie własne

3.1.2. Porównanie odchyłeń poszczególnych cyfr

Porównanie poszczególnych cyfr rozkładu przed i po usunięciu fragmentu dźwięku zostało zestawione z porównaniem rozkładu poszczególnych cyfr pliku podstawowego z wycinanym fragmentem. W tabeli 23 przedstawione są odchylenia wszystkich trzech próbek od rozkładu Benforda oraz różnice pomiędzy plikiem podstawowym a wyciętym fragmentem i plikiem po usunięciu fragmentu. Analizując rysunek 19 można dojść do wniosku, iż jeżeli cyfra w wyciętym fragmencie była bardziej odchylna do rozkładu Benforda od cyfry w pliku podstawowym w tym samym kierunku to w pliku z usuniętym fragmentem cyfra ta ulega zbliżeniu. Jeżeli cyfra w wyciętym fragmencie była bardziej oddalona od rozkładu Benforda lecz w przeciwnym kierunku, to wynikowe odchylenie ulega zwiększeniu. Natomiast jeżeli odchylenie od rozkładu Benforda w wyciętym fragmencie jest mniejsze od odchylenia w pliku podstawowym to jeżeli nie jest ono odwrotne to następuje oddalenie od rozkładu.

Rysunek 19. Porównanie odchyłeń od rozkładu Benforda poszczególnych cyfr w pliku podstawowym, wyciętym fragmencie oraz pliku podstawowego po operacji wycięcia.



źródło: opracowanie własne

W tabeli 23 przedstawiona jest wartość oddalenia wyciętego fragmentu od dźwięku podstawowego. Jest ona również przedstawiona po przedzieleniu przez arbitralną stałą 3,8. Jak widać po tym zabiegu odchylenia od rozkładu Benforda są prawie identyczne jak w podstawowym pliku po wycięciu fragmentu lecz ze znakiem przeciwnym. Zależność ta została przedstawiona na wykresie nr 16. Podane wartości są prawie idealnie ujemnie skorelowane. Współczynnik korelacji wynosi $-0,99987$. Można więc wywnioskować iż liczebności cyfr odjętych przez wycięcie fragmentu dźwięku zmniejszają ogólną liczbę zliczeń cyfr w całym podstawowym dźwięku.

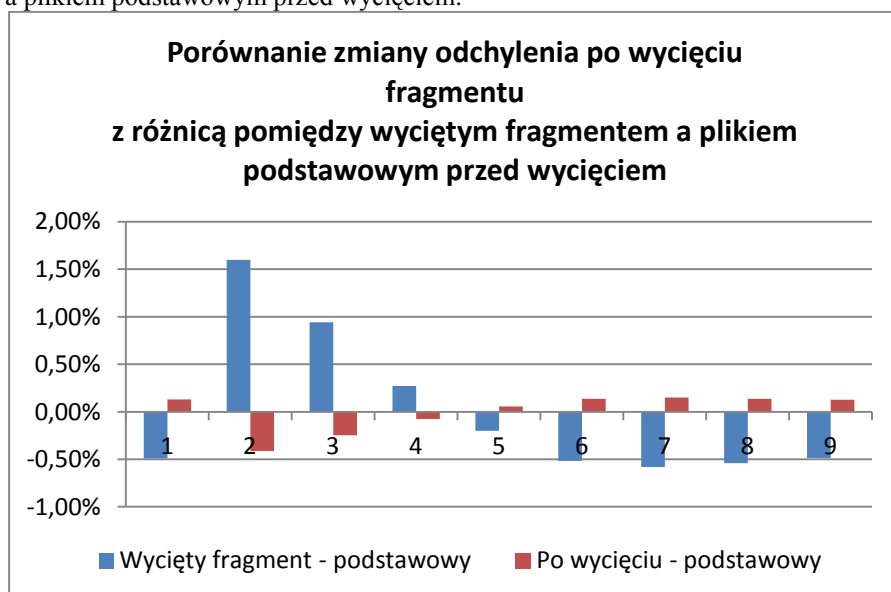
Tabela 23. Odchylenia pierwszych cyfr trzech analizowanych próbek dźwięku od rozkładu Benforda oraz dwie różnice pomiędzy plikiem podstawowym i pozostałymi.

	Podstawowy	Wycięty fragment	Po wycięciu	Wycięty fragment - podstawowy	Po wycięciu - podstawowy	Wycięty fragment - podstawowy /3,8
1	-1,50%	-1,98%	-1,36%	-0,49%	0,13%	-0,13%
2	-0,49%	1,11%	-0,91%	1,60%	-0,41%	0,42%
3	0,51%	1,45%	0,26%	0,94%	-0,25%	0,25%
4	0,53%	0,80%	0,45%	0,27%	-0,08%	0,07%
5	0,44%	0,24%	0,50%	-0,20%	0,06%	-0,05%
6	0,30%	-0,21%	0,44%	-0,52%	0,14%	-0,14%
7	0,16%	-0,42%	0,31%	-0,58%	0,15%	-0,15%
8	0,07%	-0,47%	0,20%	-0,54%	0,14%	-0,14%
9	-0,03%	-0,52%	0,10%	-0,49%	0,13%	-0,13%
M1	3,59%	7,43%	4,54%			

źródło: opracowanie własne

Dlatego liczebności poszczególnych cyfr zachowują się zgodnie ze zwykłą arytmetyką. Jeżeli w odejmowanym rozkładzie liczebność jakiejś cyfry była większa to ogólna liczebność po odjęciu zmniejszy się. Jeżeli liczebności w odejmowanym fragmencie będą równe z tymi, które należą do całego analizowanego dźwięku to wynik tej operacji będzie z prawie stu procentowym prawdopodobieństwem taki sam. Ta zasada jednak nie dotyczy całościowego odchylenia od rozkładu Benforda ponieważ mimo odjęcia fragmentu bardziej odchylonego w ogólny sposób odchylenie od rozkładu zwiększyło się. Nie ma zastosowania w tym przypadku zwykła arytmetyka ponieważ nigdy nie wiadomo nie tylko jakie będą różnice pomiędzy liczebnościami cyfr ale też jakie będą ich znaki w wycinanym fragmencie w stosunku do pliku podstawowego.

Rysunek 20. Porównanie zmiany odchylenia po wycięciu fragmentu z różnicą pomiędzy wyciętym fragmentem a plikiem podstawowym przed wycięciem.



źródło: opracowanie własne

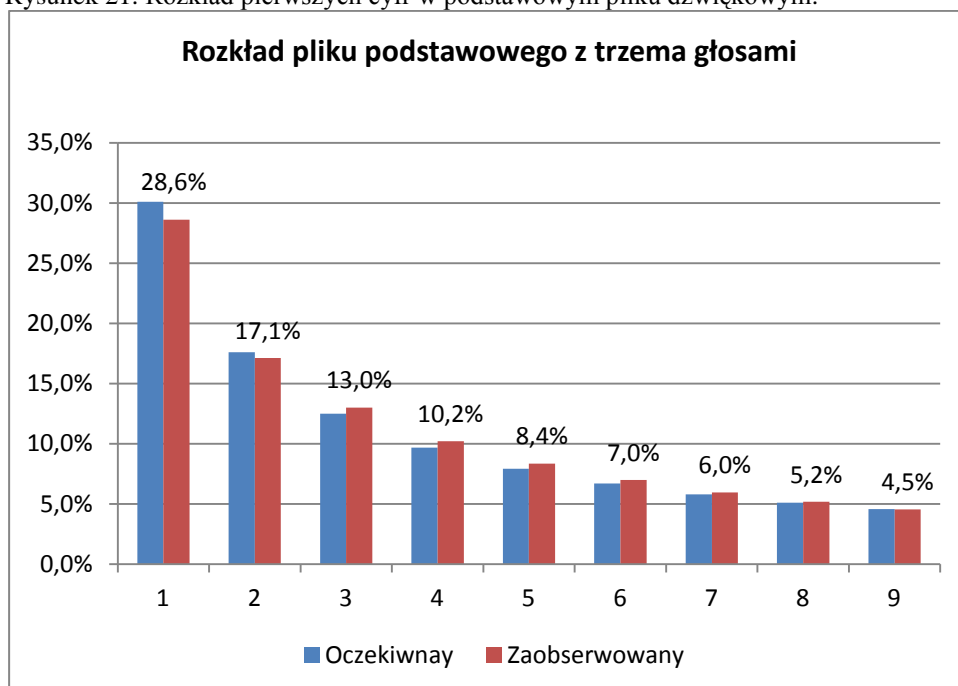
3.2. Porównanie odchyłeń przed i po doklejeniu fragmentu dźwięku

Do sprawdzenia wpływu doklejenia fragmentu dźwięku na rozkład pliku użyto tego samego pliku podstawowego co w poprzednim podrozdziale. Doklejony fragment zawiera w sobie głos jednej z osób których głosy są w pliku podstawowym.

3.2.1. Porównanie odchyłeń ogólnych rozkładów

Plik podstawowy jest tym samym plikiem, który został użyty w rozdziale 3.1. Charakteryzuje się on wynikiem testu $\chi^2 = 4980,774$ oraz wartością $M1 = 3,58\%$.

Rysunek 21. Rozkład pierwszych cyfr w podstawowym pliku dźwiękowym.

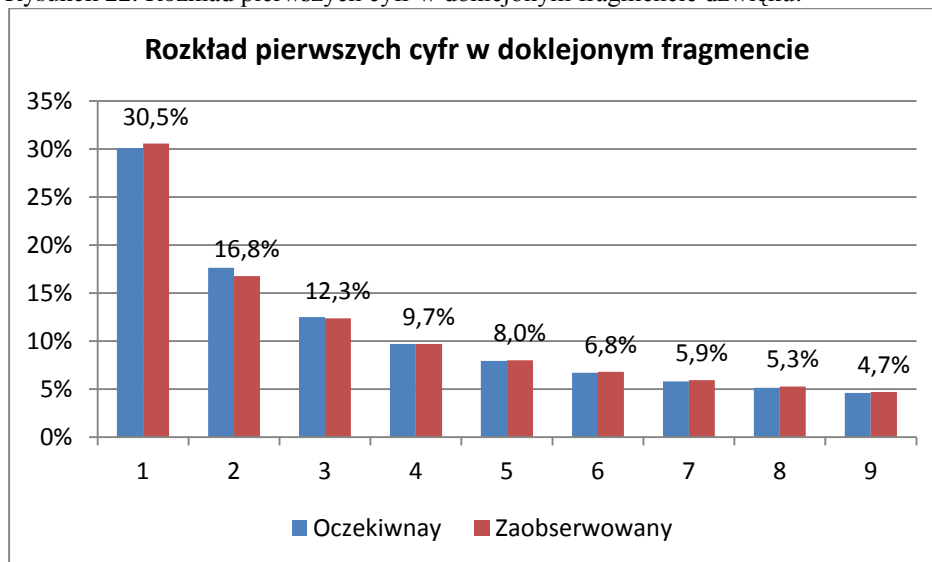


źródło: opracowanie własne

Doklejany fragment charakteryzuje się wynikiem χ^2 na poziomie 517,906 czyli dość dobrym w porównaniu z podstawowym plikiem. Wartość $M1$ to 1,95% czyli około półtora raza mniej niż w pliku podstawowym. Po doklejeniu plik wynikowy charakteryzował się wynikiem testu $\chi^2 = 4043,037$ oraz $M1 = 2,94\%$. W tym wypadku zarówno wartość miernika $M1$ jak i testu χ^2 zachowały się zgodnie z zasadami zwykłej arytmetyki. W tym przypadku różnica w odchyleniach mierzona testem χ^2 jest znacznie większa niż w poprzednim i końcowy wynik testu zbliżył się znacznie do rozkładu. Miernik $M1$ również zachował się przewidywalnie zbliżając się do rozkładu. Doklejony fragment ma długość 17 s, więc jest prawie dwa razy dłuższy od wycinanego. Być może ten test pokazuje, że w przypadku dużych różnic w odchyleniu od rozkładu Benforda pomiędzy składanymi plikami daje przewidywalne zmiany w rozkładzie wynikowym. Trzeba nadmienić iż zarówno doklejanie jak i wycinanie dźwięku nie

jest ingerencją w samą strukturę dźwięku. Być może gdyby połączyć dwie fotografie ze sobą w większy obraz dawałoby taki sam efekt jak powyższe działania zastosowane na plikach dźwiękowych. Trzeba zwrócić uwagę iż zmiany w rozkładzie Benforda w plikach graficznych są powodowane zmianami w istniejących pikselach a nie łączeniem kilku mniejszych obrazów w jeden lub wycinaniem fragmentu większego obrazu.

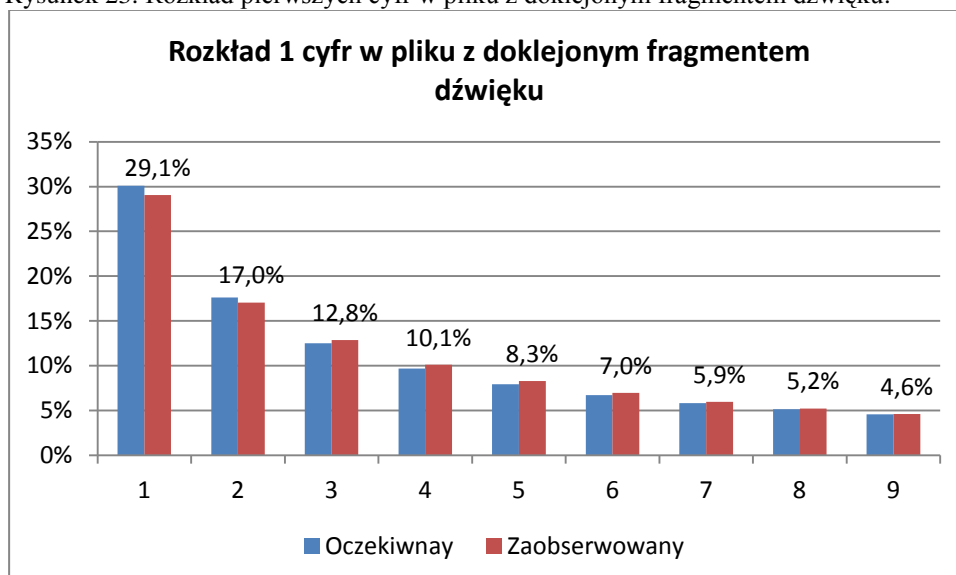
Rysunek 22. Rozkład pierwszych cyfr w doklejonym fragmencie dźwięku.



źródło: opracowanie własne

Poniżej dołączam jeszcze wykres rozkładu pierwszych cyfr w pliku będącym wynikiem złączenia podstawowego z fragmentem.

Rysunek 23. Rozkład pierwszych cyfr w pliku z doklejonym fragmentem dźwięku.

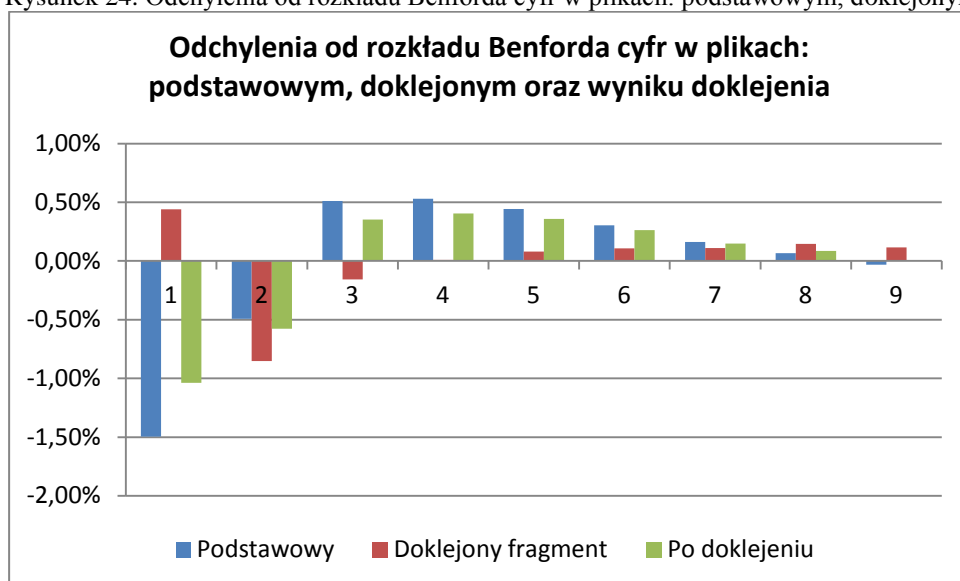


źródło: opracowanie własne

3.2.2. Porównanie odchyłeń poszczególnych cyfr

Analizując wykres odchyłeń w pliku podstawowym, doklejanym fragmencie oraz wyniku ich połączenia można zauważyć iż odchylenie w pliku wynikowym jest sumą odchylenia w pliku podstawowym oraz odchylenia w doklejanym fragmencie lecz podzielonego przez pewną stałą. Stała ta jest wynikiem różnicy w długości pliku podstawowego oraz doklejanego fragmentu. Zależność tą można przedstawić poprzez korelację różnic w odchyleniach pomiędzy plikiem podstawowym a doklejanym fragmentem oraz odchyleniach pomiędzy plikiem podstawowym a wynikowym. Korelacja ta jest bardzo bliska jedności (0,99993381) więc można założyć, że ilości poszczególnych cyfr w prosty sposób się sumują.

Rysunek 24. Odchylenia od rozkładu Benforda cyfr w plikach: podstawowym, doklejanym oraz wyniku doklejenia.



źródło: opracowanie własne

W tym przykładzie nie doszło do oddalenia się rozkładu od rozkładu Benforda tak jak to było w poprzednim tylko przeciwnie. Wynik miernika 1 zmniejszył się z 3,59% do 2,94%.

Tabela 24. Odchylenia pierwszych cyfr trzech analizowanych próbek dźwięku od rozkładu Benforda oraz dwie różnice pomiędzy plikiem podstawowym i pozostałymi.

	Podstawowy	Doklejony fragment	Po doklejeniu	Doklejony fragment - podstawowy	Po doklejeniu - podstawowy	Doklejony-podstawowy/4,1
1	-1,50%	0,44%	-1,04%	1,94%	0,46%	0,46%
2	-0,49%	-0,85%	-0,58%	-0,36%	-0,08%	-0,09%
3	0,51%	-0,16%	0,35%	-0,67%	-0,16%	-0,16%
4	0,53%	0,01%	0,41%	-0,52%	-0,12%	-0,12%
5	0,44%	0,08%	0,36%	-0,36%	-0,09%	-0,09%
6	0,30%	0,11%	0,26%	-0,20%	-0,04%	-0,05%
7	0,16%	0,11%	0,15%	-0,05%	-0,01%	-0,01%
8	0,07%	0,15%	0,08%	0,08%	0,02%	0,02%
9	-0,03%	0,12%	0,00%	0,15%	0,03%	0,03%
M1	3,59%	1,9%	2,94%			

źródło: opracowanie własne

Wniosek jaki można wyciągnąć na podstawie tych dwóch prób wklejenia i wycięcia fragmentu to brak reakcji rozkładu na te operacje. Jedyną rzeczą jaką można zaobserwować to proste sumowanie się bądź odejmowanie liczebności poszczególnych cyfr. Nie ma możliwości ustalenia na podstawie wyraźnego zwiększenia się różnicy pomiędzy rozkładem empirycznym a zakładanym czy do danego pliku zostało coś wklejone czy zostało coś z niego wycięte. Być może jednak jest możliwość ustalenia, czy dwa dźwięki zostały zmiksowane czy też nie.

4. Badanie odchylen od rozkładu Benforda wskutek zmiksowania dwóch dźwięków

W celu sprawdzenia wpływu miksowania dźwięku na odchylenie od rozkładu Benforda zmiksowano każdy z 34 plików z głosem człowieka (Zapowiedź prezentera) z jednym najdłuższym plikiem Zapowiedzi prezentera. Ten najdłuższy plik był za każdym razem skracany do długości konkretnego miksowanego pliku. Następnie w celu uzyskania równych głośności pliki te były poddane normalizacji (funkcja programu Audicity) dla miksowania przy tych samych głośnościach. Dla miksowania na różnych głośnościach najdłuższy plik za każdym razem był następnie ścisany o 3dB przy pomocy operacji (Wzmacniaj) z parametrem -3,0. Miksowanie zostało jeszcze powtórzone, przy czym operacji ścisania został poddany każdy z 34 plików użytych do miksowania.

4.1. Porównanie odchylen przed i po zmiksowaniu na równym poziomie głośności

4.1.1. Porównanie odchylen ogólnych rozkładów

Najdłuższy plik zapowiedzi prezentera został zmiksowany z każdym z pozostałych na równym poziomie głośności. Analizując średnie odchylenia w plikach przed zmiksowaniem oraz po zmiksowaniu można dojść do wniosku, że po zmiksowaniu oddalenie od rozkładu Benforda prawie się podwoiło. Zarówno 34 pliki użyte do miksowania jak i jeden najdłuższy są bliższe rozkładu Benforda niż wyniki operacji miksowania.

Tabela 25. Poziomy odchylenia od rozkładu Benforda w 34 plikach zapowiedzi prezentera oraz tych samych plikach zmiksowanych z 35 najdłuższym z nich na tym samym poziomie głośności.

	N	Średnia	Minimum	Maksimum	Odch. Std.
Zapowiedź prezentera	34	2,440%	1,016%	6,470%	1,329%
Zmiksowana zapowiedź prezentera	34	4,149%	1,634%	7,919%	1,446%
Domiksowany fragment	1	1,448%			

źródło: opracowanie własne

Jak można się było domyślić operacja miksowania w sposób istotny statystycznie oddaliła odchylenie od rozkładu Benforda w stosunku do miksowanych plików. Poziom istotności jest bardzo wysoki i z szansą na błąd mniejszą niż 1/1000 potwierdza oddalenie od zakładanego rozkładu.

Tabela 26. Porównanie odchylen od rozkładu Benforda w plikach Zapowiedź prezentera oraz Zmiksowana zapowiedź prezentera przy pomocy testu Wilcoxon.

	N	T	Z	poziom p
Zapowiedź prezentera oraz Zapowiedź prezentera zmiksowana	34	17,000	4,796	<0,001

źródło: opracowanie własne

Test powyższy wykazał iż zmiksowanie dwóch plików na tym samym poziomie głośności w sposób istotny oddala całościowy wynikowy rozkład od rozkładu Benforda. Oddalenie to jest większe niż w każdym z dwóch plików użytych do miksowania.

4.1.2. Porównanie odchyłeń poszczególnych cyfr

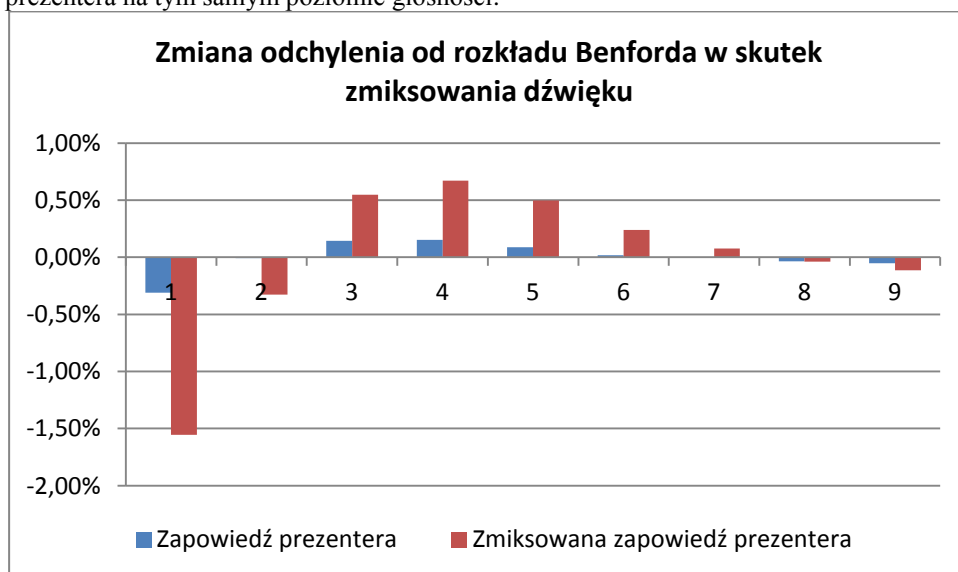
Odchylenia poszczególnych cyfr od rozkładu Benforda zostały porównane pomiędzy Zapowiedziami prezentera przed i po zmiksowaniu. Wyniki porównań zawiera tabela 27. Jak widać w niej wszystkie cyfry oprócz 8 w sposób istotny zmieniły odchylenie od rozkładu Benforda. Przy czym cyfry 1, 2, 8 oraz 9 były ujemnie odchylone od rozkładu więc ich bezwzględne oddalenie od rozkładu Benforda wzrosło.

Tabela 27. Wyniki porównania odchyłeń poszczególnych cyfr od rozkładu Benforda pomiędzy Zmiksowanymi zapowiedziami prezentera a Zapowiedziami prezentera niezmiksowanymi testem t-Studenta.

	1	2	3	4	5	6	7	8	9
t	-11,626	-3,914	8,766	14,024	10,443	6,051	2,753	-0,051	-3,212
P	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	0,010	0,959	0,003

źródło: opracowanie własne

Rysunek 25. Porównanie odchyłeń od rozkładu Benforda w Zapowiedzi prezentera oraz Zmiksowanej zapowiedzi prezentera na tym samym poziomie głośności.

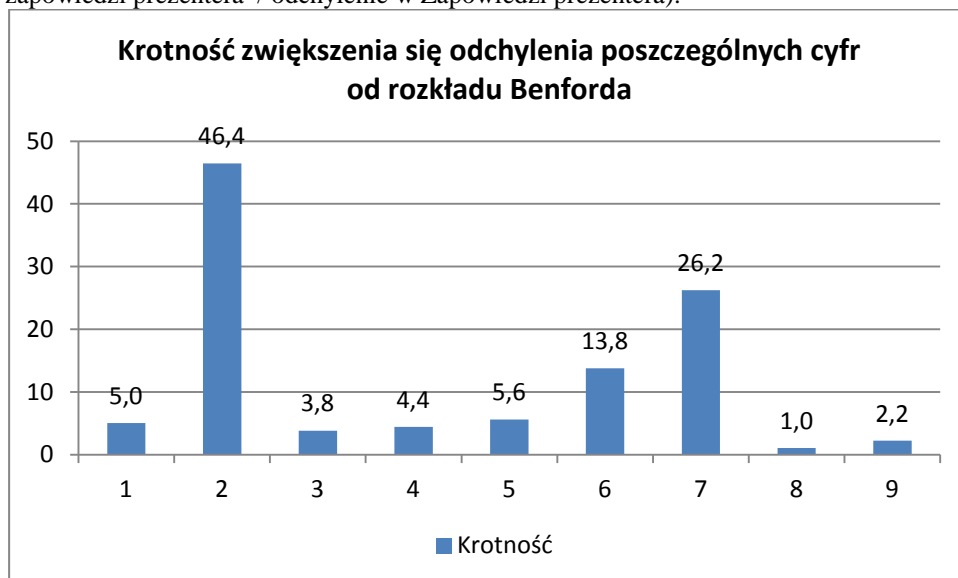


źródło: opracowanie własne

Biorąc pod uwagę nominalne odchylenia od rozkładu Benforda cyfrą która najbardziej oddaliła się od rozkładu jest 1, cyfrą, która w ogóle się nie oddaliła jest 8. Na rysunku 25. można też zauważyć, że wszystkie cyfry oprócz 8 w jakimś stopniu oddaliły się od zakładanego rozkładu.

Jeżeli weźmiemy pod uwagę stosunek wielkości odchylenia cyfry po zmiksowaniu oraz przed, to okaże się, że najbardziej oddaliła się od rozkładu cyfra 2 ok. 46 razy bardziej oddalona, następnie 7 oraz 6. Pozostałe cyfry oddaliły się około 4 krotnie i tylko jedna 8 nie oddaliła się wcale.

Rysunek 26. Krotność zwiększenia się odchylenia cyfr od rozkładu Benforda (odchylenie w Zmiksowanej zapowiedzi prezentera / odchylenie w Zapowiedzi prezentera).



źródło: opracowanie własne

Podsumowując ten podrozdział należy zaznaczyć, iż zmiksowanie dźwięku w sposób istotny wpłynęło na odchylenie od rozkładu Benforda. Zarówno całość rozkładu jak i poszczególne cyfry oddaliły się od niego przy czym jedna z cyfr (8) pozostała niezmienną.

4.2. Porównanie odchyłeń przed i po zmiksowaniu na różnych poziomach głośności

To porównanie przeprowadzono w dwóch wariantach: ze ściszym o 3dB plikiem, który był miksowany ze wszystkimi innymi oraz ze ściszymi o tą samą wartość wszystkimi, które zostały użyte do miksowania oprócz tego jednego najdłuższego użytego do wszystkich miksowań.

4.2.1. Porównanie odchyłeń ogólnych rozkładów

Zmiksowanie dwóch plików na różnych poziomach głośności tak samo jak przy tym samym poziomie spowodowało prawie dwukrotne oddalenie się ogólnego rozkładu od rozkładu Benforda. Widoczna jest niewielka różnica pomiędzy ściszym tym samym plikiem w każdym miksowaniu a ściszymi wszystkimi oprócz domiksowanego pliku. Jednak ta różnica nie wygląda na istotną statystycznie.

Tabela 28. Poziomy odchylenia od rozkładu Benforda w 34 plikach zapowiedzi prezentera oraz tych samych plikach zmiksowanych z 35 najdłuższym z nich na różnych poziomach głośności.

	N	Średnia	Minimum	Maksimum	Odch. Std.
Pliki użyte do miksowania	34	2,440%	1,016%	6,470%	1,329%
Po zmiksowaniu ze ściszym tym samym	34	3,859%	0,843%	7,347%	1,565%
Po zmiksowaniu ze ściszymi różnymi	34	4,158%	2,105%	6,996%	1,172%
Domiksowany fragment	1	1,448%			

źródło: opracowanie własne

W obu miksowaniach z różnymi głośnościami efektem miksowań było oddalenie się w sposób istotny rozkładu pierwszych cyfr od rozkładu Benforda (Tab. 29). Porównując efekty miksowania na różnych poziomach głośności z miksowaniem na tym samym poziomie albo nie widać żadnych albo różnice są na granicy akceptowanego poziomu istotności lub są to różnice małe. Przy ściszym pliku użytym do każdego miksowania a miksowaniem nieściszym wystąpiła różnica na poziomie $p=0,037$ a więc choć istotna to dość mało znacząca. Przy ściszych różnych plikach nie ma praktycznie żadnej różnicy ($p=0,817$). Nie ma również istotnej różnicy pomiędzy różnymi rodzajami ściszenia. Należy wziąć pod uwagę iż poziom ściszenia plików był niewielki, więc dla większych zmian w głośności plików być może zostaną uzyskane inne może bardziej różniące wyniki.

Tabela 29. Porównanie odchylenia od rozkładu Benforda w plikach Zapowiedź prezentera oraz Zmiksowana zapowiedź prezentera przy pomocy testu Wilcoxon na różnych poziomach głośności.

Porównania	N	T	Z	poziom p
Zapowiedź prezentera i efekt miksowania z tym samym ściszym plikiem	34	88,235	4,287	<0,001
Zapowiedź prezentera i efekt miksowania z różnymi 34 ściszymi plikami	34	88,235	4,287	<0,001
Zmiksowane zapowiedzi prezentera na równych głośnościach i efekt miksowania z tym samym ściszym plikiem	34	164,000	2,082	0,037
Zmiksowane zapowiedzi prezentera na równych głośnościach i efekt miksowania z różnymi 34 ściszymi plikami	34	284,000	0,231	0,817
Efekt miksowania z tym samym ściszym plikiem i efekt miksowania z różnymi 34 ściszymi plikami	34	203,000	1,616	0,106

źródło: opracowanie własne

4.2.2. Porównanie odchylenia poszczególnych cyfr

Porównując odchylenia poszczególnych cyfr od rozkładu Benforda przed i po zmiksowaniu na różnych poziomach głośności można zauważyć, iż nie wszystkie cyfry w sposób istotny oddaliły się od tego rozkładu. Cyfry 2 i 6 nie oddaliły się w sposób istotny od rozkładu Benforda. Pozostałe cyfry uległy oddaleniu od rozkładu Benforda w sposób bardzo istotny. Cyfra 7 oddaliła się w sposób umiarkowany choć istotny. W tabeli 30 zawarte są wyniki tego porównania.

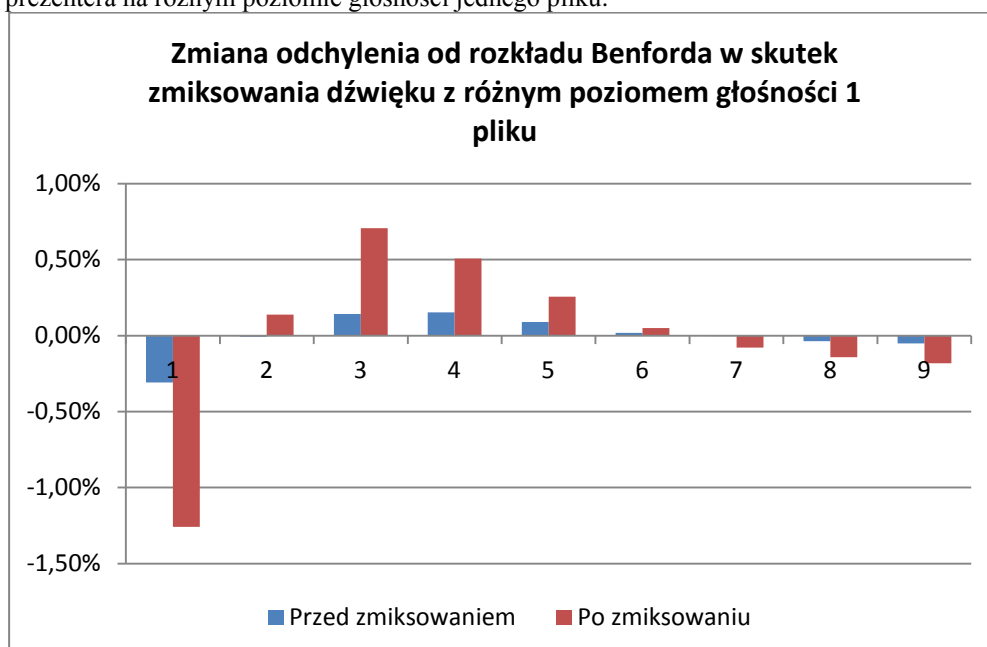
Tabela 30. Wyniki porównania odchyłeń poszczególnych cyfr od rozkładu Benforda pomiędzy Zmiksowanymi zapowiedziami prezentera z przyciszonym jednym plikiem a Zapowiedziami prezentera niezmiśkowanymi przy pomocy testu t-Studenta.

	1	2	3	4	5	6	7	8	9
t	8,779	-1,628	-13,045	-8,447	-4,118	-0,888	2,562	4,542	7,595
P	<0,001	0,113	<0,001	<0,001	<0,001	0,381	0,015	<0,001	<0,001

źródło: opracowanie własne

Nominalnie najbardziej oddaloną cyfrą od rozkładu okazała się cyfra 1, następnie cyfry 3, 4, 5 oraz 8 i 9. Nie było żadnej cyfry, która zbliżyłaby się do zakładanego rozkładu. Najmniej oddaliła się cyfra 6.

Rysunek 27. Porównanie odchyłeń od rozkładu Benforda w Zapowiedzi prezentera oraz Zmiksowanej zapowiedzi prezentera na różnym poziomie głośności jednego pliku.

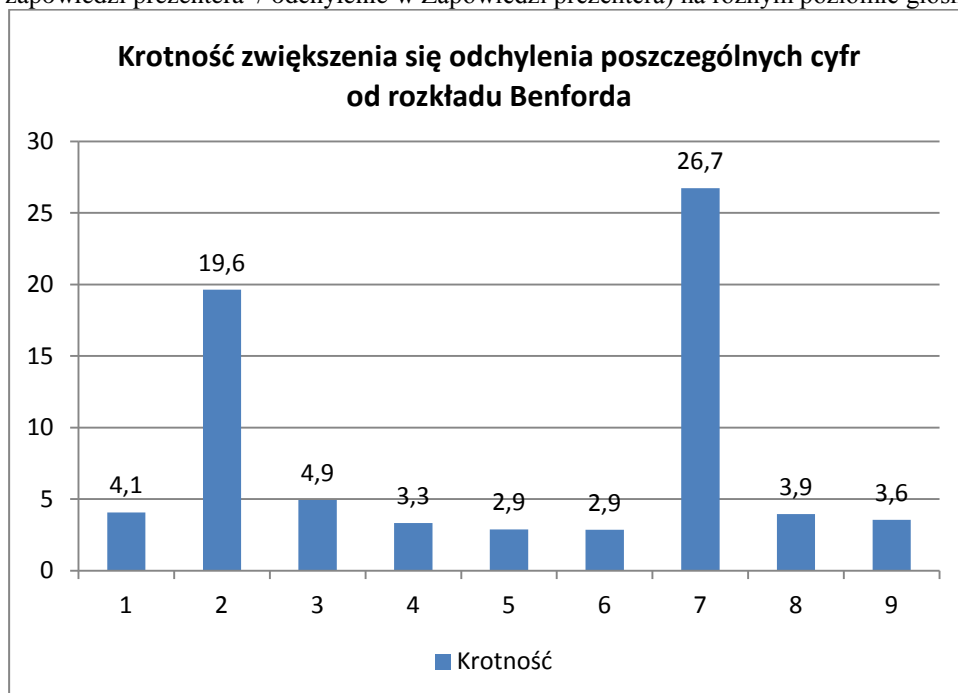


źródło: opracowanie własne

Jeżeli chodzi o krotność oddalenia się od rozkładu Benforda (Wykres 24) to cyfra 7 jest najbardziej oddalona. 27 razy zwiększyła swoje oddalenie. Kolejne znacznie oddalona cyfra to 2 (ok. 20 razy). Pozostałe cyfry oddaliły się około trzykrotnie.

Wszystkie cyfry oprócz 2 i 6 w sposób istotny oddaliły się od zakładanego rozkładu. Niewątpliwie jeżeli chodzi o krotność oddalenia się poszczególnych cyfr od rozkładu to największa zarówno przy tym samym poziomie głośności jak i różnym była w przypadku cyfr 2 i 7. Pozostałe cyfry oddaliły w stosunku do nich o niewielką ilość razy.

Rysunek 28. Krotność zwiększenia się odchylenia cyfr od rozkładu Benforda (odchylenie w Zmiksowanej zapowiedzi prezentera / odchylenie w Zapowiedzi prezentera) na różnym poziomie głośności jednego pliku.



źródło: opracowanie własne

Podsumowanie

Rozkład Benforda był już wielokrotnie stosowany w różnych dziedzinach. Jego charakter pozwalała na weryfikację autentyczności przeróżnych zjawisk: ekonomicznych, politycznych jak i dokumentów tekstowych, obrazków zapisanych cyfrowo.

Analiza plików dźwiękowych typu Wave wykazała ogólną zgodność ich z rozkładem Benforda. Największą zgodność z tym rozkładem wykazało nagranie nieprzetworzonego głosu ludzkiego. Najbardziej oddalonym od tego rozkładu dźwiękiem było nagranie muzyczne. Operacja odszumiania poprzez obcięcie niskich tonów przy pomocy filtra górnoprzepustowego spowodowało zbliżenie rozkładu pierwszych cyfr do rozkładu Benforda. Poszczególne cyfry w różnym stopniu zbliżały się do tego rozkładu.

Operacja zarówno doklejenia jak i wycięcia części dźwięku nie skutkowało oddaleniem się dźwięku od rozkładu Benforda. Wszelkie różnice przy okazji tej operacji wynikały z prostej arytmetyki dodawania i odejmowania liczebności poszczególnych cyfr w rozkładach.

Operacja miksowania spowodowała oddalenie się rozkładu pierwszych cyfr od rozkładu Benforda. W pliku wynikowym poszczególne cyfry były bardziej oddalone od tego rozkładu niż w obydwu plikach użytych do miksowania. Analizy powyższe miały na celu bardzo ogólne sprawdzenie, czy Prawo Benforda może być stosowane w przypadku plików dźwiękowych. Z pewnością nie są one wyczerpujące i zagadnienie to wymaga dużo głębszych studiów w celu weryfikacji przydatności poszczególnych stworzonych narzędzi do analizy konkretnych przypadków. Powinny zostać powtórzone na dużo większej ilości różnorodnych dźwięków. Również ilość analiz, które można przeprowadzić w celu weryfikacji tego narzędzia daleko wykracza poza ramy niniejszej pracy.

Niewątpliwie istotne znaczenie ma fakt powstania programu służącego do analizy plików typu Wave. Może on być wykorzystany przez różne osoby w celu poszerzenia wiedzy na temat analizy zgodności dźwięków zapisanych cyfrowo z rozkładem Benforda. Mam nadzieję, że przedstawione tutaj analizy będą stanowić inspirację do powstawania nowych bardziej szczegółowych prac. Mam nadzieję, że udało mi się otworzyć drogę dla stworzenia nowego narzędzia do analizy dźwięku.

Spis tabel

Tabela 1. Liczba głosów oddanych na poszczególnych kandydatów.	6
Tabela 2. Porównanie kwartalnych kwestionariuszy wydatków z Prawem Benforda.	7
Tabela 3. Rozkład procentowego odchylenia próbek dźwięku od rozkładu Benforda.	18
Tabela 4. Porównanie procentowych odchyleń od rozkładu Benforda pomiędzy poszczególnymi rodzajami próbek dźwięku za pomocą testu Wilcoxon.	19
Tabela 5. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach Całego programu.	19
Tabela 6. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach Zapowiedzi prezentera.	20
Tabela 7. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 32 próbkach Pogody.	21
Tabela 8. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach Muzyki.	22
Tabela 9. Porównanie procentowych liczebności poszczególnych cyfr z wartością średnią rozkładu Benforda za pomocą testu t-Studenta dla 4 rodzajów próbek.	23
Tabela 10. Rozkład procentowego odchylenia odsumionych próbek dźwięku od rozkładu Benforda.	24
Tabela 11. Rozkład procentowego odchylenia odsumionych próbek dźwięku od rozkładu Benforda.	24
Tabela 12. Porównanie procentowych odchyleń od rozkładu Benforda pomiędzy poszczególnymi rodzajami próbek dźwięku za pomocą testu Wilcoxon.	25
Tabela 13. Porównanie procentowych odchyleń od rozkładu Benforda próbek przed i po odsumieniu za pomocą testu Wilcoxon.	26
Tabela 14. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach odsumionego Całego programu.	26
Tabela 15. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach odsumionej Zapowiedzi prezentera.	27
Tabela 16. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 32 próbkach odsumionej Pogody.	28
Tabela 17. Zestawienie odchyleń poszczególnych cyfr od rozkładu Benforda zaobserwowane w 35 próbkach odsumionej Muzyki.	29
Tabela 18. Porównanie procentowych liczebności poszczególnych cyfr z wartością średnią rozkładu Benforda za pomocą testu t-Studenta dla 4 rodzajów próbek.	30

Tabela 19. Porównanie wpływu odszumienia na zmianę rozkładu poszczególnych cyfr dla Całego programu za pomocą testu Wilcoxon'a.	31
Tabela 20. Porównanie wpływu odszumienia na zmianę rozkładu poszczególnych cyfr dla Zapowiedzi prezentera za pomocą testu Wilcoxon'a.	31
Tabela 21. Porównanie wpływu odszumienia na zmianę rozkładu poszczególnych cyfr dla Pogody za pomocą testu Wilcoxon'a.	31
Tabela 22. Porównanie wpływu odszumienia na zmianę rozkładu poszczególnych cyfr dla Muzyki za pomocą testu Wilcoxon'a.	32
Tabela 23. Odchylenia pierwszych cyfr trzech analizowanych próbek dźwięku od rozkładu Benforda oraz dwie różnice pomiędzy plikiem podstawowym i pozostałymi.	36
Tabela 24. Odchylenia pierwszych cyfr trzech analizowanych próbek dźwięku od rozkładu Benforda oraz dwie różnice pomiędzy plikiem podstawowym i pozostałymi.	39
Tabela 25. Poziomy odchylenia od rozkładu Benforda w 34 plikach zapowiedzi prezentera oraz tych samych plikach zmiksowanych z 35 najdłuższym z nich na tym samym poziomie głośności.	41
Tabela 26. Porównanie odchyłeń od rozkładu Benforda w plikach Zapowiedź prezentera oraz Zmiksowana zapowiedź prezentera przy pomocy testu Wilcoxona.	41
Tabela 27. Wyniki porównania odchyłeń poszczególnych cyfr od rozkładu Benforda pomiędzy Zmiksowanymi zapowiedziami prezentera a Zapowiedziami prezentera niezmiśowanymi testem t-Studenta.	42
Tabela 28. Poziomy odchylenia od rozkładu Benforda w 34 plikach zapowiedzi prezentera oraz tych samych plikach zmiksowanych z 35 najdłuższym z nich na różnych poziomach głośności.	44
Tabela 29. Porównanie odchyłeń od rozkładu Benforda w plikach Zapowiedź prezentera oraz Zmiksowana zapowiedź prezentera przy pomocy testu Wilcoxona na różnych poziomach głośności.	44
Tabela 30. Wyniki porównania odchyłeń poszczególnych cyfr od rozkładu Benforda pomiędzy Zmiksowanymi zapowiedziami prezentera z przyciszonym jednym plikiem a Zapowiedziami prezentera niezmiśowanymi przy pomocy testu t-Studenta.	45

Spis rysunków

Rysunek 1. Rozkład Benforda.....	4
Rysunek 2. Wygląd okna głównego programu do analizy rozkładu Benforda w plikach muzycznych.....	10
Rysunek 3. Podstawowe informacje o analizowanym pliku.....	10
Rysunek 4. Wyniki testów χ^2 dla całego rozkładu oraz dla poszczególnych cyfr.	11
Rysunek 5. Wynik wartości miernika I oraz podstawowych parametrów podstawowego rozkładu próbek (nie pierwszych cyfr).....	12
Rysunek 6. Średnie procentowe odchylenie rozkładów próbek dźwiękowych od rozkładu Benforda.	18
Rysunek 7. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek całego programu.	20
Rysunek 8. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Zapowiedzi prezentera.	21
Rysunek 9. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 32 próbek Pogody.	22
Rysunek 10. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Muzyki.	23
Rysunek 11. Średnie procentowe odchylenie rozkładów próbek dźwiękowych od rozkładu Benforda przed i po odszumieniu.....	25
Rysunek 12. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Całego programu.	27
Rysunek 13. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Zapowiedzi prezentera.	28
Rysunek 14. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 32 próbek Pogody.	29
Rysunek 15. Odchylenia rozkładu pierwszych cyfr od rozkładu Benforda dla 35 próbek Muzyki.	30
Rysunek 16. Rozkład pierwszych cyfr w podstawowym pliku dźwiękowym z trzema głosami..	33
Rysunek 17. Rozkład pierwszych cyfr w wyciętym fragmencie dźwięku.....	34
Rysunek 18. Rozkład pierwszych cyfr w pliku po wycięciu fragmentu.....	34
Rysunek 19. Porównanie odchyleń od rozkładu Benforda poszczególnych cyfr w pliku podstawowym, wyciętym fragmencie oraz pliku podstawowego po operacji wycięcia.....	35

Rysunek 20. Porównanie zmiany odchylenia po wycięciu fragmentu z różnicą pomiędzy wyciętym fragmentem a plikiem podstawowym przed wycięciem.	36
Rysunek 21. Rozkład pierwszych cyfr w podstawowym pliku dźwiękowym.	37
Rysunek 22. Rozkład pierwszych cyfr w doklejonym fragmencie dźwięku.	38
Rysunek 23. Rozkład pierwszych cyfr w pliku z doklejonym fragmentem dźwięku.	38
Rysunek 24. Odchylenia od rozkładu Benforda cyfr w plikach: podstawowym, doklejonym oraz wyniku doklejenia.	39
Rysunek 25. Porównanie odchyłeń od rozkładu Benforda w Zapowiedzi prezentera oraz Zmiksowanej zapowiedzi prezentera na tym samym poziomie głośności.	42
Rysunek 26. Krotność zwiększenia się odchylenia cyfr od rozkładu Benforda (odchylenie w Zmiksowanej zapowiedzi prezentera / odchylenie w Zapowiedzi prezentera).	43
Rysunek 27. Porównanie odchyłeń od rozkładu Benforda w Zapowiedzi prezentera oraz Zmiksowanej zapowiedzi prezentera na różnym poziomie głośności jednego pliku.	45
Rysunek 28. Krotność zwiększenia się odchylenia cyfr od rozkładu Benforda (odchylenie w Zmiksowanej zapowiedzi prezentera / odchylenie w Zapowiedzi prezentera) na różnym poziomie głośności jednego pliku.	46

Bibliografia

1. Albrecht, WS and Albrecht, CC. *Root out financial deception*. Journal of Accountancy 193(4), 30-34, 2002,
2. Bhattacharya, P, Chatterjee, A and Chakrabarti, BK. *A common mode of origin of the power law distributions in models of market and earthquake*. <http://arxiv.org/abs/physics/0510038>, 2005,
3. Bhattacharya, S, Kumar, K and Smarandache, F. *Conditional probability of actually detecting a financial fraud – a neutrosophic extension to Benford’s law*. International Journal of Applied Mathematics 17(1), 7-14, 2004,
4. Bourke, N and van Peurse, K. *Detecting fraudulent financial reporting: teaching the 'watchdog' new tricks*. Working paper 79, University of Waikato. ISSN:1173-7182, 2004,
5. Durtschi, C, Hillison, W and Pacini, *The effective use of Benford’s law to assist in detecting fraud in accounting data*. Journal of Forensic Accounting 1524-5586/Vol. V, 17-34, 2004,
6. Farbaniec M., Grabiński T., Zabłocki B., Zając W., *Analiza wpływu przekształceń matematycznych na zbiory o zadanym rozkładzie cyfr*,
7. Farbaniec M., Grabiński T., Zabłocki B., Zając W., *Wykorzystanie prawa Benforda do wykrywania nieprawidłowości w zbiorach danych*. 2011,
8. Grabiński T., *Prawo Benforda i Jego Wykorzystanie w Analizie Wyników Wyborów Prezydenckich W Polsce*, 2009,
9. Harrington, JE. *Detecting Cartels*. Economics Working Paper 526, Johns Hopkins University, to appear in: Advances in the Economics of Competition Law, 2005,
10. Li, Z, Cong, L and Wang, H. *Discussion on Benford’s law and its application*. <http://arxiv.org/abs/math.ST/0408057>, 2004,
11. Lindsay, DH, Foote, PS, Campbell, A and Reilly, DP. *Detecting fraud in the data using automatic intervention detection*. Fraud Magazine. A Publication of the Association of Certified Fraud Examiners, January/February 2004,
12. Murphy, J, Baxter, R, Eyerman, J, Cunningham, D and Kennet, J. *A system for detecting interviewer falsification*. RTI International at 59th Annual AAPOR Conference, Phoenix, Arizona, 2004,
13. Newcomb, S. *Note on the frequency of use of the different digits in natural numbers*. American Journal of Mathematics 4(1), 1881,
14. PAN *Encyklopedia Fizyki*, 1974,
15. Petucci, SD. *Benford’s Law: Can It Be Used to Detect Irregularities in First Party Automobile Insurance Claims?*. Journal of Economic Crime Management 3(1), 2005,

16. Posch, PN and Kreiner, WA. *A general approach to digital analysis exemplified by stock market indices. unpublished manuscript.*, 2005,
17. Roukema, BF. *Benford's Law Anomalies in the 2009 Iranian presidential election.*, 2009,
18. Schäfer, C, Schräpler, JP and Müller, KR. *Identification, Characteristics and Impact of Faked and Fraudulent Interviews in Surveys.*, 2004,
19. Schäfer, C, Schräpler, JP, Müller, KR and Wagner GG. *Automatic Identification of Faked and Fraudulent Interviews in Surveys by Two Different Methods.* Discussion paper 441, DIW Berlin (German Institute for Economic Research), 2004,
20. Smith, CA. *Detecting Anomalies in Your Data Using Benford's Law.* Paper 249 (Statistics and Data Analysis) in: Proceedings of SUGI 27, Orlando, USA, April 14-17, 2002,
21. Steven W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, 1998,
22. Swanson, D, Cho, MJ and Eltinge, J. *Detecting possibly fraudulent or error-prone survey data using Benford's law.* pp 4172-4177 in: 2003 Joint Statistical Meetings - Section on Survey Research Methods, Proceedings of the American Statistical Association, 2003,
23. Taylor, RN, McEntegart, DJ and Stillman, EC. *Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data.* Drug Information Journal 36, 115-125, 2002,
24. Theodor Hill, *A Statistical Derivation of the Significant-Digit Law*, 1995.