KONSTRUKCJA CHARAKTERYSTYK DŹWIĘKU „MAŁO WRAŻLIWYCH" NA TRANSFORMACJE NIELINIOWE.

Tadeusz ZIĘBAKOWSKI
Instytut Technologii Mechanicznej, Politechnika Szczecińska
Al. Piastów 19, 70-310 Szczecin
Tel.: (+4891)4494880, fax: (+4891)4343507, e-mail: harmat@safona.ps.pl

1. Wstęp

Jedną z trudniejszych do wyjaśnienia własności mechanizmu percepcji dźwięku jest stosunkowo niewielki wpływ zniekształceń nieliniowych dźwięku na uświadamiane przez człowieka wrażenia słuchowe. Wiadomo przecież, że wrażenia słuchowe są silnie powiązane z widmem dźwięku, które z kolei jest bardzo podatne na transformacje nieliniowe. W dodatku zakres natężeń dźwięku odbierany przez człowieka jest znaczny, ponad 100 dB, tak więc duży wpływ będą miały również nieliniowości rzędów większych niż 2.

W swojej pracy zaprezentuję konstrukcję charakterystyk „mało wrażliwych" na transformacje nieliniowe, utworzonych na bazie dyskretnego widma dźwięku. Konstrukcja ta ma charakter algebraiczny i wykorzystuję w niej w zakresie podstawowym elementy teorii reprezentacji grup skończonych.

Posługując się tymi charakterystykami pokażę, jak można wyjaśnić niektóre zjawiska psychoakustyczne jak np. zjawisko residuum, które polega na utrzymywaniu się słyszenia tonu podstawowego po jego usunięciu, czy zagadnienie niesłyszenia większości tonów nieliniowych (tzw. kombinacyjnych).

2. Transformacje nieliniowe

Dźwięk jest zjawiskiem mechanicznym. Jego rozchodzenie i oddziaływanie na inne obiekty materialne można opisać matematycznie wychodząc z zasad dynamiki. Zagadnienie to sprowadza się w końcu do rozwiązywania równań różniczkowych maksymalnie 2-go rzędu.

Dźwięk dla człowieka jest ważnym nośnikiem informacji. W wielu przypadkach w rozchodzeniu się sygnału dźwiękowego można wyodrębnić etapy pośrednie związane z formą przenoszenia informacji, np. rozchodzenie dźwięku w różnych ośrodkach czy drgania elementów sztywnych. Przejście z jednego etapu do drugiego będziemy określać właśnie mianem transformacji.

Precyzyjne matematyczne określenie klasy takich transformacji nie jest proste – są to w ogólności klasy operatorów różniczkowo-całkowych. Nie będziemy tu się tym zajmować, wyróżnimy jednak pewne podklasy, a mianowicie transformacje liniowe, które spełniają zasadę superpozycji i nieliniowe, które jej nie spełniają.

Rozważmy dla przykładu transformację nieliniową postaci:

❄$(x) := x^k$

Będziemy mówić w tym przypadku o nieliniowości rzędu k. Rozważmy sygnał akustyczny zawierający n składowych sinusoidalnych o częstościach $\omega_1, \omega_2, \ldots, \omega_n$:

$$p(t) = c_1 e^{i\omega_1 t} + c_{-1} e^{-i\omega_1 t} + c_2 e^{i\omega_2 t} + c_{-2} e^{-i\omega_2 t} + \\ + \ldots + c_n e^{i\omega_n t} + c_{-n} e^{-i\omega_n t} \qquad \text{gdzie} \quad c_{-i} = \overline{c_i}$$

Składową sinusoidalną będziemy nazywać dalej tonem.

Po podstawieniu drugiego wyrażenia do pierwszego otrzymamy:

$$❄(p(t)) := \sum_{\substack{\text{po wszystkich ciągach} \\ k\text{-elementowych liczb od } 1 \text{ do } n \\ \text{z kombinacjami znaków } \pm}} c_{\pm i_1} c_{\pm i_2} \cdot \ldots \cdot c_{\pm i_k} e^{i(\pm\omega_{i1} \pm \omega_{i2} \ldots \pm \omega_{ik})t}$$

Widzimy więc, że tego typu transformacje nieliniowe wprowadzają do widma wiele nowych tonów. Dla przykładu gdy mamy początkowo 2 tony o częstościach ω_1, ω_2, gdy $k = 2$ dodatkowo pojawią się tony:

$$2\omega_1, 2\omega_2, \text{oraz } \omega_1 + \omega_2 \text{ i } \omega_1 - \omega_2,$$

zaś dla $k = 3$:

$$3\omega_1, 3\omega_2, \text{oraz } 2\omega_1 + \omega_2, 2\omega_1 - \omega_2, 2\omega_2 + \omega_1, 2\omega_2 - \omega_1$$

Zjawisko to powoduje szereg trudności w zrozumieniu słyszenia wysokości dźwięku, o czym mowa jest w dalszej części pracy.

3. Opis konstrukcji

Powszechnie wiadomo, że wrażeniu wysokości dźwięków periodycznych jednoznacznie odpowiada fizyczny parametr jakim jest częstotliwość. Można by zatem przypuszczać, że w przypadku dźwięku złożonego z kilku dźwięków periodycznych będziemy słyszeć swego rodzaju superpozycję wrażeń wysokości poszczególnych składowych periodycznych. I na pierwszy rzut oka tak rzeczywiście jest. Jednak począwszy od 19. stulecia psychoakustycy znaleźli sposoby, aby wrażenie

wysokości mniej lub bardziej obiektywnie mierzyć (określać ilościowo), i okazało się, że (przy założeniu słyszenia poszczególnych składowych harmonicznych poprzez wysokość) w wielu konfiguracjach dźwięków nie wszystkie składowe są słyszalne, a zdarza się również tak, że właśnie składowe nieobecne są słyszalne (por.[1],[2]).
Do dziś nie ma w pełni zadowalającego wyjaśnienia tych zjawisk. Dotychczasowe wysiłki idą w kierunku szukania przyczyn tych osobliwości w zjawiskach mechanicznych zachodzących w uchu człowieka, bądź w funkcjonowaniu receptorów słuchowych i nerwu słuchowego.
W niniejszej pracy prezentuję pogląd, że **wrażenia słuchowe nie odpowiadają bezpośrednio składowym widma dźwięku** lecz są **odzwierciedleniem istnienia pewnej dodatkowej transformacji,** która następuje po analizie widmowej zachodzącej w uchu człowieka. Główne zadanie tej transformacji to wyodrębnienie takiej informacji z dźwięku, która w możliwie najmniejszym stopniu ulega zniekształceniom pod wpływem transformacji nieliniowych. W wyniku tej transformacji otrzymujemy pewną nową charakterystykę dźwięku „mało wrażliwą" na transformacje nieliniowe. Pokażę, że dzięki niej można bardziej precyzyjnie opisać wrażenia słuchowe, wyjaśniając osobliwości o których była mowa powyżej.

W celu łatwiejszego zrozumienia szczegółów proponowanej przeze mnie konstrukcji zacznę od przykładu.

PRZYKŁAD 1

Niech W będzie 2-wymiarową przestrzenią liniową nad ciałem $K=Z_7$ (ciało reszt modulo 7) i wybierzmy bazę $\{e_1,e_2\}$, co pozwali utożsamiać odwzorowania liniowe z macierzami.

Niech H będzie grupą generowaną przez macierze:

$$\lambda=\begin{bmatrix}1 & -1\\ 1 & 0\end{bmatrix} \qquad \sigma=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$$

Można sprawdzić, że $\lambda^6 = 1$, $\sigma^2 = 1$ i że jest to grupa izomorficzna z $Z_2{\times}S_3$.

Jeśli potraktujemy współrzędne w bazie $\{e_1,e_2\}$jako wartości częstości to grupa H „naśladuje" transformacje nieliniowe 2-rzędu, np.:

$$\lambda\cdot\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix}=\begin{bmatrix}\omega_1-\omega_2\\ \omega_1\end{bmatrix}$$

$$\lambda\cdot\begin{bmatrix}\omega_1\\ -\omega_2\end{bmatrix}=\begin{bmatrix}\omega_1+\omega_2\\ \omega_1\end{bmatrix}$$

$$\lambda\cdot\begin{bmatrix}\omega_1\\ -\omega_1\end{bmatrix}=\begin{bmatrix}2\omega_1\\ \omega_1\end{bmatrix}$$

Grupa H wyznacza w W zbiór 8 orbit, który oznaczymy standardowo przez W/H :

$\mathcal{O}$ = $\{[0, 0]\}$
$\mathcal{A}_\omega$= $\{[-\omega, -\omega], [-\omega, 0], [0, -\omega], [0, \omega], [\omega, 0], [\omega, \omega]\}$ gdzie $\omega = 1,2,3$
$\mathcal{B}_\omega$= $\{[-2\omega, -\omega], [-\omega, -2\omega], [-\omega, \omega], [\omega, -\omega], [\omega, 2\omega], [2\omega, \omega]\}$ gdzie $\omega = 1,2,3$
$\mathcal{C}$ = $\{[-3, -2], [-3, -1], [-2, -3], [-2, 1], [-1, -3], [-1, 2], [1, -2], [1, 3], [2, -1], [2, 3], [3, 1], [3, 2]\}$

Orbity $\mathcal{A}_\omega$ są współzmiennicze z częstością:

$$k\,\mathcal{A}_\omega = \mathcal{A}_{k\omega}$$

podobnie $\mathcal{B}_\omega$, orbita $\mathcal{C}$ jest niezmiennicza względem mnożenia przez stałą. Orbita $\mathcal{O}$ jest trywialna i nie będziemy się nią zajmować.
Orbity posłużą nam do scharakteryzowania wrażeń słuchowych prostych. Podamy teraz przykładową „abstrakcyjną" interpretację psychoakustyczną:

$\mathcal{A}_\omega \leftrightarrow$ wysokość
$\mathcal{B}_\omega \leftrightarrow$ współbrzmienia oktawy
$\mathcal{C}$ $\leftrightarrow$ fonem samogłoski „a" (lub np. barwa piszczałki)

Na koniec podamy konstrukcję miary, która pozwoli nam powiedzieć „ile" każdego z wrażeń słyszymy w dźwięku o widmie $F(\omega)$.

Niech Z będzie podzbiorem W. Dla danego F zdefiniujemy skończoną rzeczywistą miarę I_F na W następująco:

$$I_F(Z) := \sum_{[\omega_1,\omega_2]\in Z} \log^+|F(\omega_1)F(\omega_2)|$$

gdzie:

$$\log^+(x) := \begin{cases}\log(x) & dla \quad x \geq 1\\ 0 \quad dla & 0 \leq x < 1\end{cases}$$

W definicji tej miary zawiera się m.in. prawo Webera-Fechnera. Zawartość poszczególnych wrażeń prostych określają liczby $I_F(\mathcal{A}_\omega)$, $I_F(\mathcal{B}_\omega)$ dla $\omega = 1,2,3$ oraz $I_F(\mathcal{C})$. □

Uogólnimy teraz ten przykład.

Zacznijmy od tego, że człowiek rozróżnia co do częstotliwości około 1400 składowych widma (tonów). Dlatego, będziemy rozważać widmo dyskretne. Podobnie jak w przykładzie, do numerowania składowych dyskretnego widma dźwięku, będziemy używać odpowiednio dużego ciała skończonego prostego ***K***. Uporządkujemy to ciało według ciągu:

$$-l, -l+1, \ldots, -1, 0, 1, \ldots, l-1, l$$

gdzie l takie, że $2l + 1 = \mathrm{char}\, \boldsymbol{K} > 2$.

W dalszej części pracy określimy charakterystykę tego ciała na podstawie przesłanek psychoakustycznych. Użycie ciała skończonego można uzasadnić tym, że w wyniku transformacji nieliniowych nowe częstotliwości powstają na zasadzie dodawania i odejmowania częstotliwości i mnożenia ich przez liczby naturalne (co prawda ma tę własność już pierścień, jednak ze względu na dalszą konstrukcję rozważamy od razu ciało).

Krok 1. Utworzenie n-wymiarowej przestrzeni częstotliwościowej $W = \boldsymbol{K}^n$.
W tej przestrzeni będziemy rozważać grupę macierzową H, której zadaniem jest pogrupowanie tonów pierwotnych i pochodzących od nich tonów nieliniowych. Taka grupa w pewien sposób naśladuje dowolną transformację nieliniową.
Przez problem nieliniowy rozumieć będziemy wyznaczenie częstotliwości tonów nieliniowych na podstawie tonów pierwotnych w postaci n współrzędnych wektora z przestrzeni W. Wymiar przestrzeni W decyduje o tym w jakim stopniu rozwiązywany jest problem nieliniowy. Jeśli wymiar jest n, to możemy w pełni rozwiązać problemy nieliniowe rzędów do n włącznie i częściowo problemy nieliniowe rzędów wyższych.
Przestrzeń ilorazowa W/H to matematyczny model wrażeń słuchowych

Krok 2. Przyporządkowanie dowolnej charakterystyce widmowej F pewnej rzeczywistej miary I_F na W, a mianowicie dla $Z \subset W$:

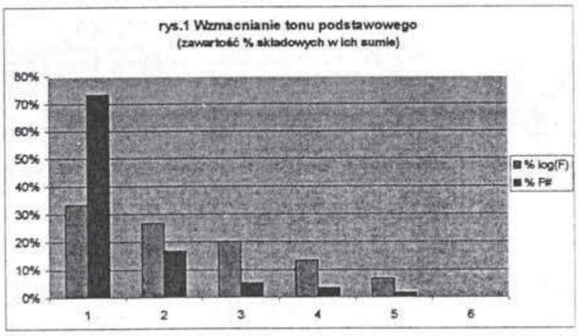
$$I_F(Z) := \sum_{\{\omega_1, \omega_2, \ldots, \omega_n\} \in Z} \log^+ \left| F(\omega_1) F(\omega_2) \cdot \ldots \cdot F(\omega_n) \right|$$

$$\text{gdzie: } \log^+(x) := \begin{cases} \log(x) & dla \quad x \geq 1 \\ 0 & dla \quad 0 \leq x < 1 \end{cases}$$

Miara I_F wyznacza funkcję:

$$F^{\#} : W/G \ni \Omega \rightarrow I_F(\Omega) \in \boldsymbol{R}.$$

Funkcję $F^{\#}$ można również zdefiniować następująco:



$$F^{\#}(Hx) := \frac{1}{|St_H(x)|} \sum_{h \in H} \log^+ \left| \underbrace{F \otimes F \otimes \ldots \otimes F}_{n}(hx) \right|$$

gdzie $|St_H(x)|$ rząd stabilizatora elementu x.

Przestrzeń W/H wraz z funkcją $F^{\#}$ stanowią podstawę opisu zjawisk psychoakustycznych w powyższym modelu.

4. Opis niektórych zjawisk psychoakustycznych przy użyciu charakterystyki $F^{\#}$

Pokażemy teraz na przykładach, że posługując się charakterystykami F# można wyjaśnić niektóre trudne z teoretycznego punktu widzenia zjawiska psychoakustyczne.
Niech $\boldsymbol{K}$ będzie odpowiednio dużym ciałem skończonym, a $W := K \oplus K$. Tym razem, dla uproszczenia rozważań, nie będziemy konstruować grupy H a złożymy jedynie, że istnieje taka grupa, która wyznacza następującą rodzinę orbit:

$$\begin{aligned} A_\omega = \{ & (\pm\omega, \pm\omega), (\pm\omega, \pm 2\omega), (\pm\omega, \pm 3\omega), \\ & (\pm\omega, \pm 5\omega), (\pm 2\omega, \pm 3\omega), (\pm 2\omega, \pm 5\omega), \\ & (\pm 3\omega, \pm 5\omega), (\pm 2\omega, \pm\omega), (\pm 3\omega, \pm\omega), \\ & (\pm 5\omega, \pm\omega), (\pm 3\omega, \pm 2\omega), \\ & (\pm 5\omega, \pm 2\omega), (\pm 5\omega, \pm 3\omega) \} \end{aligned}$$

gdzie $\omega = 1,2,3,\ldots$
Orbity te zinterpretujemy jako wrażenia wysokości tonów o częstotliwości ω.
Charakterystykę widmową F zapiszemy w postaci ciągu:

$$F = (q_{-l}, q_{-l+1}, \ldots, q_{-1}, q_0, q_1, \ldots, q_{l-1}, q_l)$$

$$\text{gdzie} \quad q_{-i} := \overline{q}_i$$

PRZYKŁAD 2 ***Wzmacnianie tonu podstawowego***

Niech:

$|q_1| = 10^5$ (50dB) $\quad |q_2| = 10^4$ (40 dB)

$|q_3| = 10^3$ (30 dB) $\quad |q_4| = 10^2$ (20 dB)

$|q_5| = 10$ (10 dB) $\quad$ pozostałe q_i są równe 0

Wykonany na podstawie tych danych wykres 1 pokazuje, że charakterystyki $F^{\#}$ maja własność wzmacniania tonu podstawowego i zagłuszania kolejnych harmonicznych.

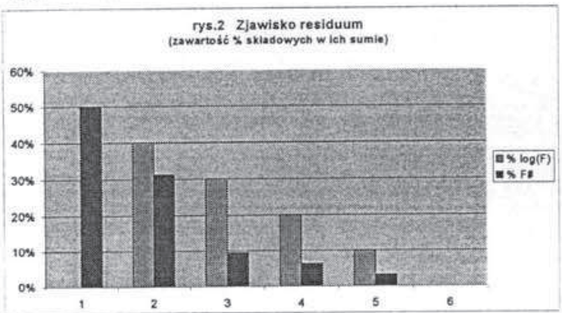
PRZYKŁAD 3 ***Zjawisko residuum.***

Polega ono na słyszeniu wysokości tonu podstawowego w dźwięku złożonym z tonów harmonicznych ($2\omega, 3\omega, 4\omega, \ldots$), z którego usunięto właśnie ton podstawowy ω. Można by sądzić, że jest to efekt nieliniowy. Stwierdzono jednak, że występuje on nawet wtedy, gdy w uchu nie ma takiego tonu nieliniowego. Charakterystyki $F^{\#}$ dobrze opisują to zjawisko co pokazuje wykres 2, który został otrzymany dla następujących danych:

$|q_1| = 0$ $\qquad$ $|q_2| = 10^4$ (40 dB)

$|q_3| = 10^3$ (30 dB) $\qquad$ $|q_4| = 10^2$ (20 dB)

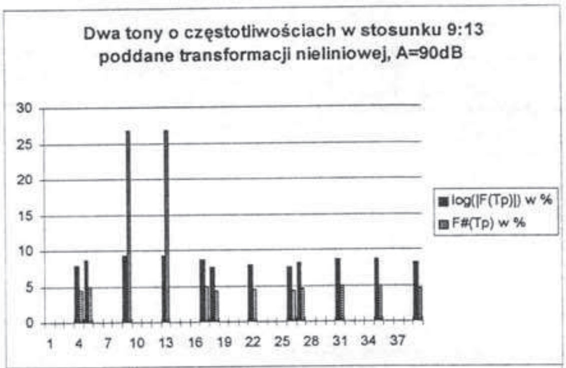
$|q_5| = 10$ (10 dB) $\qquad$ pozostałe q_i są równe 0

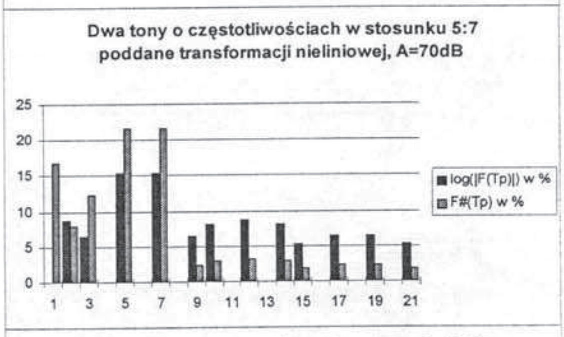


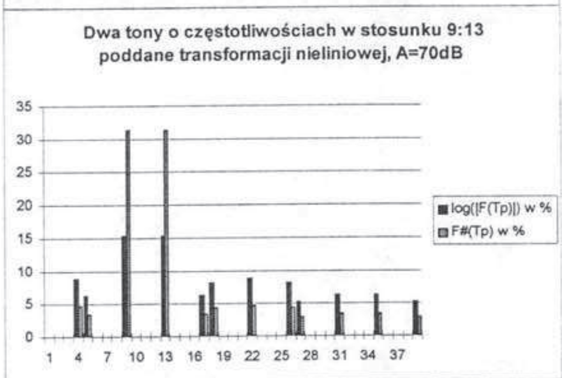
PRZYKŁAD 4 *Nieliniowe własności ucha - tony kombinacyjne [1][2][3].*

Tak jak już wspomniano, przy założeniu, że słyszymy poprzez wrażenie wysokości składniki widma trudno jest wyjaśnić osobliwości słyszenia tonów nieliniowych tzw. kombinacyjnych. Ogromnej większości z nich po prostu nie słychać. Najbardziej słyszalnym jest ton kombinacyjny pochodzący od nieliniowości 3 stopnia: gdy weźmiemy dwa tony o częstotliwościach f_1 i f_2 to przy pewnych stosunkach częstotliwości $f_2 : f_1$ z zakresu od 1,08 do 1,5 słychać ton o częstotliwości $2f_1 - f_2$. Co więcej nigdy nie było doniesień co do słyszenia tonów sumacyjnych $2f_1 + f_2, 2f_2 + f_1$ i towarzyszącego mu tonu różnicowego $2f_2 - f_1$.

Wykresy 3,4,5 otrzymano dla dźwięku postaci: $p(t) = A\cos(\omega_1 t) + A\cos(\omega_2 t)$ i transformacji postaci: $Tp(t) = p(t) + \alpha p(t)^2 + \beta p(t)^3$ i przedstawiono porównanie charakterystyk widmowych F i $F^{\#}$ dla sygnału $Tp(t)$. W wyniku zastosowania charakterystyki $F^{\#}$ uzyskujemy znaczące wytłumienie składowych „nieliniowych". Dla niektórych stosunków częstotliwości np.: 5:7 (por. rys. 2.6.4) pojawiają się "składowe osobliwe", które w widmie wyjściowym F są znacznie mniejsze. Wykresy poniżej wykonano przyjmując $\alpha = 10^{-10} i \beta = 10^{-18}$.







Widzimy więc, że charakterystyki $F^{\#}$ istotnie zmniejszają wpływ zniekształceń nieliniowych. W przykładach 2,3 i 4 przyjęto orbity które rozwiązywały niewiele problemów nieliniowych. Obecnie autor finalizuje prace nad konstrukcją grupy H rozwiązującej znacznie więcej problemów nieliniowych i opisującej precyzyjnie wspomniane zjawiska psychoakustyczne.

LITERATURA

[1] de Boer E.: *Auditory Physics. Physical Principles in Hearing Theory II.* Physics Reports 62 No.2 (1980), 105 No.3 (1984), 203 No.3 (1991)

[2] Smoorenburg G.F.: *Audibility region of combination tones*, Journal of the Acoustical Society of America 52 (1972) 603-614

[3] Smoorenburg G.F.: *Combination tones and their origin*, Journal of the Acoustical Society of America 52 (1972) 615-632