

PERCEPCJA DŹWIĘKU¹

ALEKSANDER P. SĘK*

Instytut Akustyki

UAM Poznań

AUDITORY PERCEPTION

Summary. This paper is concerned with basis of physiology and psychology of hearing. It presents the most important mechanisms in hearing including middle and inner ear mechanics and coding of acoustic stimuli in the auditory nerve. It also shows some consideration connected with cochlear amplifier that probably exists in the inner ear. The amplifier enhances sensitivity of the auditory system to very weak sounds and brings about high frequency selectivity of the system and its nonlinear properties. The amplifier is based on properties of the outer hair cells and may be a source of energy delivered to the basilar membrane. This mechanism is probably responsible for otoacoustic emission, i.e. very weak sounds generated by the auditory system. This paper presents the most important relations between physical parameters of sound and its loudness and pitch not only for pure tones but also for complex signals. It also shows some details of the auditory filter concept and some aspects of the frequency selectivity of the auditory system. The most important findings concerned with binaural perception of sounds are also included.

Wstęp

Słuch umożliwia przetworzenie bodźców akustycznych na użyteczną informację. Potrafi odebrać bodźce dźwiękowe o częstotliwościach od 20 do 20 000 Hz, i natężeniach pozostających w stosunku jak 1 : 1000 000 000 000 : najgłośniejszy dźwięk jaki odbieramy (bez ryzyka uszkodzenia słuchu) ma natężenie 1 bilion razy większe od natężenia najcichszego dźwięku, jaki potrafimy usłyszeć. Jednak jedną z najważniejszych właściwości układu słuchowego jest to, że funkcjonuje on w sposób podobny do analizatora częstotliwości, bowiem umożliwia rozłożenie złożonego bodźca akustycznego na składowe tonalne. Ta cecha układu słuchowego, nazywana rozdzielczością częstotliwościową, umożliwia nam usłyszenie dwóch prezentowanych jednocześnie tonów, odległych w dziedzinie częstotliwości o np. pół oktawy, jako dwóch oddzielnych dźwięków.

Oprócz percepcyjnego rozseparowania dźwięków układ słuchowy potrafi także rozdzielić bodźce występujące po sobie w czasie, co nazywane jest czasową zdolnością rozdzielczą, a z drugiej strony potrafi też kumulować informację/energię niesioną przez bo-

¹ Pracę wykonano w ramach grantu Komitetu Badań Naukowych, projekt nr 7 TO7B 052 16.

* Korespondencję kierować pod adresem: Aleksander P. Sęk, Instytut Akustyki, UAM Poznań, ul. Umultowska 85, 61-614 Poznań.

dzień akustyczny, co nazywane jest sumowaniem w czasie. Inną cechą układu słuchowego jest zdolność uporządkowania dźwięków od najniższego do najwyższego, czyli przypisanie im wysokości oraz możliwość rozróżnienia barwy dźwięków. Potrafimy również zlokalizować położenie źródła dźwięku.

Wszystkie właściwości układu słuchowego wykorzystywane jednocześnie w pewien spójny sposób stanowią o jakości naszego słuchu i o jego znakomitym przystosowaniu do najróżniejszych bodźców akustycznych. W szczególności nasz słuch potrafi przetworzyć złożony bodziec akustyczny jakim jest mowa, tak że w konsekwencji jest ona interpretowana jako konkretna informacja lingwistyczna.

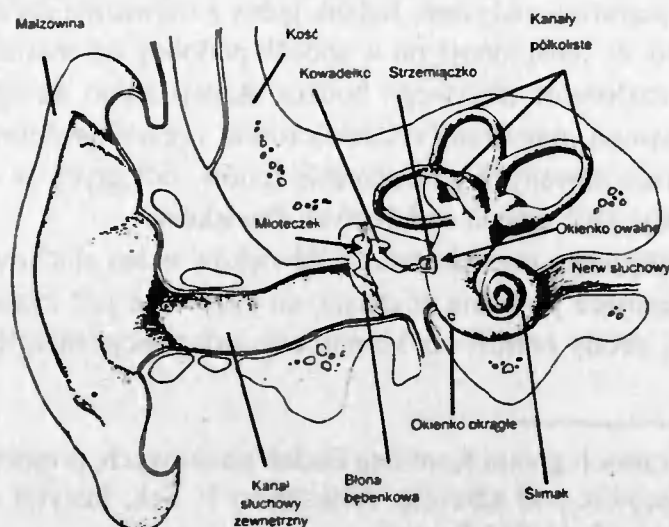
Praca niniejsza stanowi próbę syntetycznego opisu podstawowych zasad funkcjonowania peryferyjnego układu słuchowego oraz wskazania związków pomiędzy fizycznymi parametrami dźwięku a wrażeniem przez ten dźwięk wywołanym.

1. BUDOWA UKŁADU SŁUCHOWEGO

Rozważania dotyczące zasad funkcjonowania układu słuchowego nie są możliwe bez opisu anatomicznej struktury tego narządu. Toteż zmierzając do odpowiedzi na pytanie: Jak ucho zamienia falę akustyczną na wrażenie słuchowe? Przedstawić należy przede wszystkim anatomię tzw. peryferyjnego układu słuchowego.

1.1. Droga fali akustycznej w układzie słuchowym

W peryferyjnym układzie słuchowym wyróżnia się trzy części: ucho zewnętrzne, środkowe i wewnętrzne, które zilustrowano na rys. 1 (Pickles, 1988). Ucho zewnętrzne składa się z małżowiny usznej, przewodu słuchowego zewnętrznego i zakończone jest błoną bębenkową. Do błony bębenkowej przymocowana jest pierwsza z kosteczek słuchowych – młoteczek. Do młoteczka przymocowane jest kowadełko, a do kowadełka – strzemiączko. Kosteczki słuchowe stanowią ucho środkowe. Strzemiączko swoją podstawą dotyka błony w tzw. okienku owalnym, stanowiącym początek ślimaka (ucho wewnętrzne).



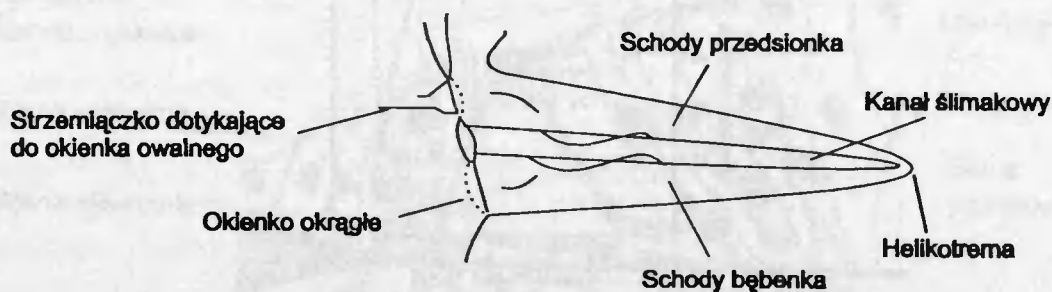
Rys. 1. Peryferyjny układ słuchowy człowieka: ucho zewnętrzne, środkowe i wewnętrzne.

Dźwięk, czyli zaburzenie ciśnienia atmosferycznego, pada na małżowinę, przechodzi przez przewód słuchowy zewnętrzny i dociera do błony bębenkowej, wprawiając ją w drgania. Drgania błony bębenkowej przekazywane są przez łańcuch kosteczek słuchowych do ślimaka.

Zasadniczym zadaniem ucha zewnętrznego i środkowego jest przeniesienia energii dźwięku z powietrza do ślimaka. Ucho zewnętrzne i środkowe wypełnione są powietrzem, zaś ucho wewnętrzne – cieczami. Opór stawiany falom akustycznym przez ciecze jest większy od oporu stawianego przez powietrze, lub ściślej – impedancja cieczy jest większa od impedancji powietrza. Dlatego też zadaniem ucha środkowego jest takie przetworzenie dźwięku z powietrza, aby wywołać efektywne ruchy cieczy w ślimaku. Jest to możliwe dzięki temu, że powierzchnia okienka owalnego jest kilkadziesiąt razy mniejsza od powierzchni błony bębenkowej: małe ciśnienie padające na „dużą” błonę bębenkową jest zamieniane na duże ciśnienie na „małym” okienku owalnym. Dzięki temu, oraz z uwagi na różną długość ramion kowadłka i młoteczka, ciśnienie wywierane przez strzemiączko na okienko owalne jest około 200 razy większe od ciśnienia padającego na błonę bębenkową (Pickles, 1988).

Ucho środkowe wyposażone jest w mechanizm chroniący ślimak przed dźwiękami o zbyt dużym natężeniu. Jest to tzw. odruch strzemiączkowy (Durrant i Lovrinic, 1995). Kosteczki słuchowe podtrzymywane są przez mięśnie, które kurczą się, gdy do ucha dochodzą bardzo głośne dźwięki. Kurczenie się tych mięśni prowadzi do ograniczenia wychylenia kosteczek słuchowych, a tym samym do tłumienia drgań przekazywanych do ucha wewnętrznego. Odruch ten występuje z pewnym opóźnieniem i dlatego jest nieefektywny w przypadku dźwięków impulsowych, takich jak np. strzał z pistoletu.

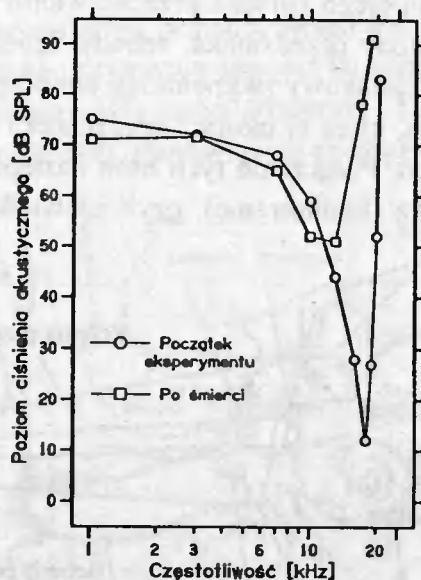
Ślimak uformowany jest z kości czaszki. Pomimo, że jest on obiektem o strukturze spiralnej, to najczęściej analizuje się go w wersji „rozwiniętej”, co znacznie ułatwia jego opis. Przekrój podłużny rozwiniętego ślimaka przedstawiono na rys. 2. Jest on podzielony na trzy komory (partycje): schody przedsionka, schody bębenka (wypełnione perylimfą o potencjale ok. 0 mV) i kanał ślimakowy (wypełniony endolimfą o potencjale ok. +80 mV) błonami Reissnera i podstawną, które w okolicy wierzchołka ślimaka łączą się, zamykając przestrzeń kanału ślimakowego. Połączenie tych błon następuje jednak przed końcem ślimaka, tworząc szparę osklepka (helikotrema), czyli niewielki otwór pomiędzy schodami przedsionka i bębenka.



Rys. 2. Podłużny przekrój rozwiniętego ślimaka.

Gdy okienko owalne porusza się do wewnątrz ślimaka, to chwilowa gęstość cieczy ślimaka w bezpośrednim sąsiedztwie tego okienka wzrasta. To „zaburzenie” ciśnienia rozchodzi się schodami bębienka w kierunku szpary osklepka i dalej schodami przedsionka do tzw. okienka okrągłego, usytuowanego poniżej okienka owalnego. Ruch okienka owalnego (strzemiączka) „do wewnątrz” powoduje ruch błony okienka okrągłego „na zewnątrz” ślimaka i na odwrót.

Zaburzenie gęstości cieczy, które propaguje się w schodach bębienka i w schodach przedsionka jest przyczyną lokalnych, chwilowych różnic ciśnienia perylimfy po obu stronach błony podstawnej. Różnice te powodują powstawanie fali biegnącej na błonie podstawnej. Błona ta podlega niewielkim odkształceniom, które przemieszczają się wraz z zaburzeniami ciśnienia perylimfy od okienka owalnego do okienka okrągłego (por. rys. 2). W przypadku, gdy dźwiękiem jest ton, to cykliczne zagęszczenia i rozrzedzenia perylimfy powodują cykliczne wychylenia znacznych obszarów błony podstawnej. To, która część błony podstawnej drga, zależy od częstotliwości dźwięku; amplituda tych drgań zależy natomiast od natężenia dźwięku. Dźwięki o dużych częstotliwościach najsilniej wzbudzają obszary w pobliżu okienka owalnego, a dźwięki o małych częstotliwościach – w pobliżu szpary osklepka (Pickles, 1988; Palmer, 1995). Dzięki temu ślimak zachowuje się tak jak analizator dźwięku: sygnał akustyczny złożony z tonów o różnych częstotliwościach i amplitudach pobudza w różnym stopniu różne obszary błony podstawnej. Ta właściwość ślimaka nazywa się selektywnością (rozdzielczością) częstotliwościową (Moore, 1997). Analiza dźwięku zachodząca na błonie podstawnej nie jest doskonała: jeśli częstotliwości dwóch tonów niewiele różnią się od siebie, to mechanizm analizy nie pozwala na ich rozseparowanie. Jeśli drgania dostarczane są do ślimaka nie za pomocą strzemiączka, ale poprzez kości czaszki (tzw. przewodnictwo kostne), to obwiednie drgań błony podstawnej nie zmieniają się znacząco.



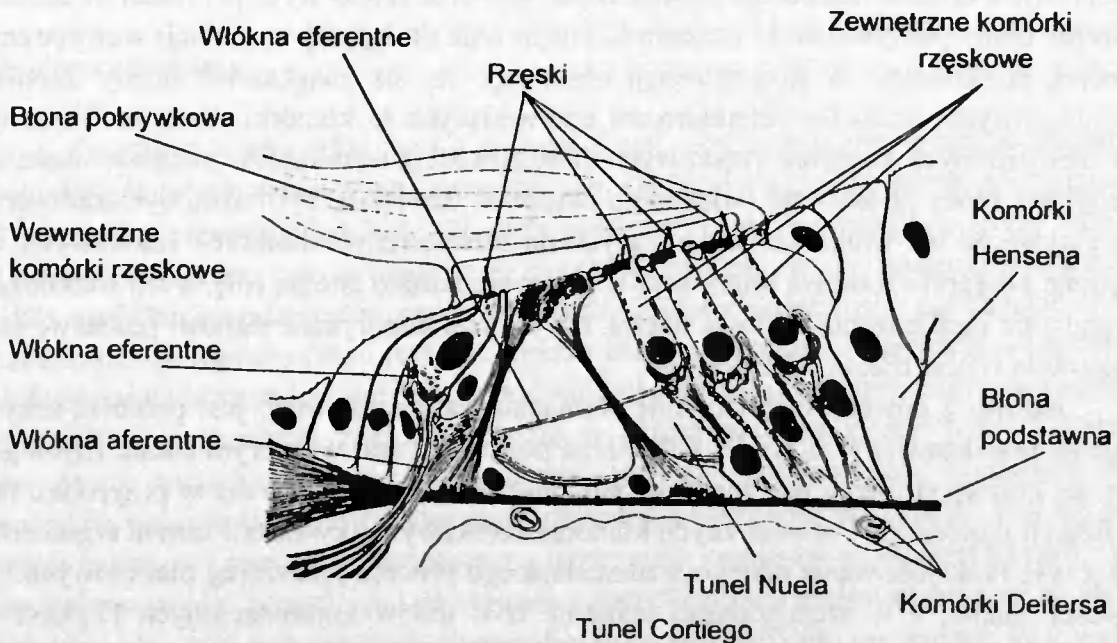
Rys. 3. Krzywe strojenia błony podstawnej mierzone w punkcie najbardziej wrażliwym na 18 kHz. Krzywa wyznaczona przez kółka ilustruje ostre strojenia błony żywej świnki morskiej. Kwadraty wyznaczają krzywą strojenia uzyskaną po śmierci zwierzęcia.

Selektywność częstotliwościową błony podstawnej przedstawia się często za pomocą tzw. krzywych strojenia poszczególnych jej punktów (Pickles, 1988). Krzywą strojenia wyznacza się obserwując wychylenia jednego, ustalonego punktu błony podstawnej, pobudzanej tonem o zmienianej częstotliwości. Natężenie tonu dobiera się każdorazowo w ten sposób, aby amplituda wychylenia badanego punktu błony podstawnej była zawsze taka sama. W poprawnie funkcjonującym układzie słuchowym krzywa strojenia charakteryzuje się bardzo ostro zarysowanym minimum i stromo nachylonymi zboczami (tzw. ostre strojenie). Przykład takiej krzywej, uzyskanej dla świnki morskiej, przedstawiono na rys. 3 za pomocą kółek.

Krzywa strojenia zmienia się znacząco, gdy wyznaczona jest po śmierci zwierzęcia, a więc wówczas, gdy ustały wszystkie czynności biologiczne organizmu, co przedstawiono na rys. 3 za pomocą kwadratów. Krzywa strojenia charakteryzuje się w tym przypadku znacznie mniej ostro zarysowanym minimum (tzw. szerokie strojenie), które jest przesunięte w kierunku mniejszych częstotliwości. Ta krzywa strojenia odwzorowuje wyłącznie mechaniczne właściwości błony podstawnej. Wysoka czułość błony podstawnej oraz ostre strojenie poprawnie funkcjonującego układu słuchowego nie wynika zatem wyłącznie z mechanicznych właściwości ślimaka i przypisuje się je biologicznie aktywnemu procesowi, nazywanemu często wzmacniaczem ślimakowym.

1.2. Proces przetwarzania. Wzmacniacz ślimakowy

Drgania błony podstawnej są zmieniane na potencjały czynnościowe włókien nerwu słuchowego (impulsy neuronowe) w narządzie Cortiego, usytuowanym wzdłuż błony podstawnej w środowisku dodatnio spolaryzowanej endolimfy (+80 mV). Przekrój poprzeczny narządu Cortiego przedstawiono na rys. 4. Zasadniczymi jego elementami są komórki rzęskowe wewnętrzne i zewnętrzne, umieszczone po obu stronach tunelu Cortiego.



Rys. 4. Przekrój narządu Cortiego.

Wewnętrzne komórki rzęskowe u ludzi umieszczone są w jednym rzędzie, po wewnętrznej stronie tunelu Cortiego. Jest ich około 3500, a każda z nich ma około 40 tzw. rzęsek formujących proste rzędy. Do każdej z tych komórek dochodzi około 20 neuronów aferentnych, przekazujących impulsy elektryczne na wyższe piętra drogi słuchowej. Rzęski tych komórek nie mają bezpośredniej styczności z błoną pokrywkową, znajdująca się nad narządem Cortiego. Ruchy błony podstawnej w górę i w dół powodują cykliczne zbliżanie narządu Cortiego i błony pokrywkowej, prowadząc do przeginięcia rzęsek raz w jedną, raz w drugą stronę. Przeginięcie to połączone jest z otwieraniem i zamykaniem kanałów znajdujących się w rzęskach, którymi do ujemnie spolaryzowanych komórek rzęskowych mogą napływać dodatnie jony potasu z endolimfy (Pickles, 1988; Yates i in., 1992; 1995). Chwilowe podwyższenie potencjału tych komórek odbierane jest przez synapsy włókien neuronowych i transmitowane do wyższych pięter układu słuchowego. Znakomita większość informacji o bodźcu akustycznym, jaka dostępna jest na różnych piętrach drogi słuchowej, pochodzi z wewnętrznych komórek rzęskowych.

Zewnętrzne komórki rzęskowe uporządkowane są u ludzi w pięciu rzędach. Jest ich około 25 000, a każda z nich ma około 140 rzęsek. Do komórek tych dochodzi około 1800 neuronów eferentnych, tj. takich, które przekazują sygnały z mózgu. Ponadto komórki te charakteryzują się kurczliwością pod wpływem zmiany potencjału: jeśli zwiększymy ich potencjał, to ulegają one skróceniu (Neely, 1993; Santos-Sacchi, 1993; 1997). Kurczliwość oraz specyficzny sposób unerwienia tych komórek są odpowiedzialne za dużą czułość i dobrą selektywność częstotliwościową słuchu.

Przeginięcie rzęsek zewnętrznych komórek rzęskowych powoduje otwieranie i zamykanie kanałów znajdujących się w rzęskach: ruch błony podstawnej w kierunku błony pokrywkowej połączony jest z otwarciem kanałów jonowych, którymi napływają jony potasu. Powodują one wzrost potencjału tych komórek, a więc również ich skrócenie. Skrócenie to jest największe w szczytowym wychyleniu błony podstawnej, kiedy kanały jonowe są całkowicie otwarte. Skrócenie zewnętrznych komórek rzęskowych prowadzi do lepszego zbliżenia błony pokrywkowej i narządu Cortiego oraz do lepszej stymulacji wewnętrznych komórek rzęskowych. W konsekwencji obserwuje się też zwiększenie liczby impulsów czynnościowych neuronów (aferentnych) unerwiających te komórki. Brak motorycznych cech zewnętrznych komórek rzęskowych (kurczliwości) wymagałby znacznie większego wychylenia błony podstawnej (większego natężenia dźwięku), by można było zaobserwować aktywność we włóknach dochodzących do wewnętrznych komórek rzęskowych. Zewnętrzne komórki rzęskowe odgrywają w słyszeniu bardzo istotną rolę, a ich uszkodzenie prowadzi do upośledzenia czułości słuchu. Ich funkcja motoryczna stanowi podstawę funkcjonowania wzmacniacza ślimakowego.

Jednym z dowodów na istnienie wzmacniacza ślimakowego jest przebieg krzywej strojenia przedstawiony na rys. 3. Gdy błona podstawna jest w dobrym stanie fizjologicznym, jej krzywa strojenia jest znacznie bardziej selektywna (kółka) niż w przypadku ustania funkcji motorycznej zewnętrznych komórek rzęskowych (kwadrat). Innym argumentem na korzyść funkcjonowania takiego wzmacniającego procesu jest szereg nieliniowych właściwości słuchu, a w szczególności istnienie tzw. tonów kombinacyjnych (Tyler i in., 1979). Jedną z najistotniejszych cech układów liniowych jest to, że w sygnale wyjściowym z takiego układu mogą występować tylko sygnały o częstotliwościach równych częstotli-

wościom sygnałów podanych na wejściu. Jeśli w sygnale wyjściowym obecne będą takie sygnały, których nie ma w sygnale wejściowym, to układ jest nieliniowy. Jeśli słuchamy dwutonu o częstotliwościach np. 1 i 1.2 kHz, to na skutek nieliniowości sluchu, we wrażeniu obecny jest także sygnał o częstotliwości około 0.8 kHz. Nie jest to bynajmniej efekt psychologiczny: ton o częstotliwości 0.8 kHz (tzw. ton kombinacyjny) został wygenerowany przez ślimak. Jego obecność nie jest jednak oczywista, bowiem dopiero zdudnienie go poprzez dodanie do dwutonu (1 i 1.2. kHz) tonu o częstotliwości nieznacznie różniącej się od 800 Hz (np. 804 Hz) pozwala stwierdzić jego obecność (Smooenburg, 1972a; 1972b).

Innym przejawem istnienia aktywnych procesów biologicznych w ślimaku jest generowanie bardzo cichych dźwięków przez ucho. W wielu sytuacjach, za pomocą mikrofonu sondującego umieszczonego blisko błony bębenkowej, można rejestrować słabe sygnały generowane spontanicznie przez ucho. Ponadto jeśli nasz układ słuchowy zostanie pobudzony impulsem (trzaskiem), to można zarejestrować dźwięk będący reakcją ucha na sygnał pobudzający. Początkowa część tego dźwięku jest wynikiem odbicia dźwięku od ucha środkowego. Jednak dalsza część tego sygnału pojawia się z opóźnieniem 54-60 ms i jest rezultatem biologicznej aktywności ślimaka. Dźwięki te nazywane są odpowiednio spontaniczną i wywołaną emisją otoakustyczną (Kemp, 1978), a ich analiza znajduje coraz szersze zastosowanie w diagnostyce sluchu.

Jedną z dróg wyjaśnienia istnienia emisji otoakustycznej jest uwzględnienie motorycznych cech zewnętrznych komórek rzęskowych. Każdy neuron eferentny doprowadzony jest jednocześnie do wielu komórek rzęskowych zewnętrznych. Jeśli spontaniczne wyładowanie jednego neuronu eferentnego dotrze do pewnej grupy tych komórek, to wywoła ich chwilową depolaryzację, a w ślad za tym skrócenie ich długości. Skrócenie to spowoduje zmniejszenie odległości pomiędzy błoną podstawną a błoną pokrywkową, a więc wytworzy zaburzenie ciśnienia w schodach bębenka i w schodach przedsionka. Zaburzenie to dotrze do okienka owalnego i może wprowadzić w drgania strzemiączko, które za pośrednictwem kowadełka i młoteczka przekaże je na błonę bębenkową.

1.3. Nerw słuchowy

Zasadnicza część informacji o sygnale akustycznym dostępna w nerwie słuchowym pochodzi z komórek rzęskowych wewnętrznych. Do każdej z tych komórek dochodzi około 20 neuronów aferentnych i każdy z nich połączony jest wyłącznie z jedną komórką. W odpowiedzi na ton, a więc najprostszy sygnał akustyczny, wychyleniom podlega pewien (ściśle uzależniony od częstotliwości tonu) obszar błony podstawnej, pobudzając znaczną liczbę komórek rzęskowych, i jeszcze większą liczbę neuronów. Informacja o częstotliwości dźwięku kodowana jest poprzez pobudzenie określonej grupy neuronów dochodzących do komórek rzęskowych, znajdujących się w określonych obszarach błony podstawnej (tzw. teoria miejsca). Informacja o natężeniu dźwięku jest natomiast kodowana przede wszystkim poprzez liczbę impulsów czynnościowych neuronów na sekundę. Nie ma jednak prostej zależności pomiędzy natężeniem a liczbą tych impulsów w jednostce czasu: wzrost natężenia prowadzi w ogólności do wzrostu częstotliwości wyładowań neuronów, ale również do poszerzenia pobudzonego obszaru błony podstawnej oraz do wzrostu liczby aktywnych neuronów. Dla średnich wartości poziomu dźwięku znaczna część neuronów słucho-

wych osiąga stan nasycenia: dalszy wzrost poziomu dźwięku (powyżej ok. 50 dB) nie wywołuje zmiany ich aktywności. Wzrost poziomu natężenia jest wówczas sygnalizowany przez tzw. neurony nienasycone oraz przez wzrost liczby aktywnych neuronów.

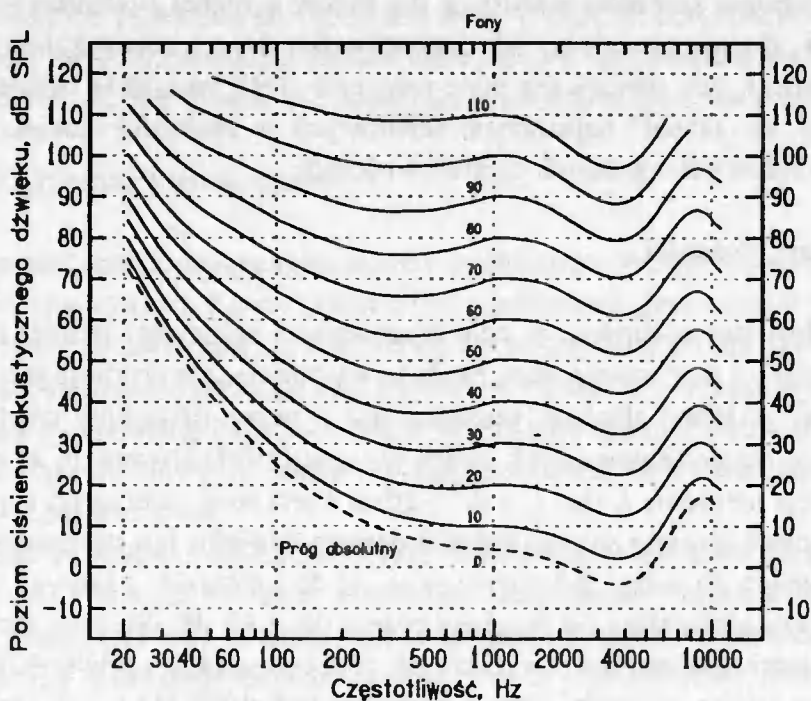
Częstotliwość impulsów neuronowych nie jest jedyną informacją o bodźcu dostępną w nerwie słuchowym. Dla częstotliwości bodźca do 5 kHz przebieg wyładowań neuronów odzwierciedla strukturę czasową bodźca (Pickles, 1988). Wyładowania te mają tendencję do występowania dla ściśle określonej fazy bodźca, choć nie muszą występować w każdym jego okresie. Zjawisko to nazywane jest synchronicznością fazową. Interwały czasowe między kolejnymi wyładowaniami są więc całkowitymi wielokrotnościami okresu bodźca i zawierają jednoznaczną informację o częstotliwości sygnału. Określając częstotliwość sygnału układ słuchowy może bazować na informacji o tym, które z neuronów są aktywne oraz na odstępach czasowych pomiędzy impulsami. W przypadku bodźca złożonego synchroniczność fazowa jest zdeterminowana przez najefektywniejsze składowe. Przebiegi wyładowań różnych neuronów zawierają informacje o względnych amplitudach składowych bodźca i o różnych częstotliwościach składowych.

2. PERCEPCYJNA ANALIZA DŹWIĘKU W UKŁADZIE SŁUCHOWYM

2.1. Progi słyszalności i percepcja głośności

2.1.1. Progi słyszalności

Próg słyszalności (próg absolutny, próg detekcji sygnału) jest najmniejszym poziomem natężenia dźwięku, który wywołuje zaledwie spostrzegane wrażenie słuchowe. Wyznaczając taki próg, dźwięki prezentuje się zwykle za pomocą słuchawek, a pomiaru natężenia dokonuje się za pomocą mikrofonu umieszczonego blisko błony bębenkowej. Przykładowy próg słyszalności przedstawiono na rys. 5 (najniżej położona krzywa). Dźwięki słyszalne leżą w przedziale od 20 do 20 000 Hz, choć wrażliwość słuchu zmienia się znacząco ze zmianą częstotliwości. Jesteśmy stosunkowo mało wrażliwi na dźwięki o bardzo małych i bardzo dużych częstotliwościach. Największa czułość słuchu przypada na zakres częstotliwości od 500 do 5000 Hz i wynika w pewnym stopniu z charakterystyki przeniesienia ucha środkowego. Przedstawiona krzywa ilustruje średnie rezultaty dużej grupy słuchaczy o słuchu normalnym. Dlatego indywidualne progi wyższe lub niższe o 20 dB w dalszym ciągu są uznawane za mieszczące się w normie.



Rys. 5. Próg absolutny i krzywe jednakowej głośności. Parametrem jest wartość poziomu głośności w fonach.

W audiologii progi słyszalności przedstawia się zazwyczaj na tzw. audiogramie, tj. wykresie ilustrującym ubytki słuchu względem słuchu normalnego.

2.1.2. Krzywe jednakowej głośności

Głośność jest tym atrybutem wrażenia słuchowego, która pozwala na uszeregowanie dźwięków od cichych do głośnych. Głośność jest cechą subiektywną dźwięku i nie można jej zmierzyć bezpośrednio. Omawiając tę cechę wrażenia dźwiękowego najwygodniej jest posługiwać się skalą, która nie tylko przyporządkowuje wielkości fizyczne (np. natężenie) charakteryzujące dźwięki ich odpowiednikom subiektywnym ale i taką, która umożliwia porównanie głośności dźwięków o różnej częstotliwości. Jedną z takich skal jest rodzina tzw. krzywych jednakowej głośności, które przedstawiono na rys. 5 (Robinson i Dadson, 1956). Utworzono je dzięki zrównaniu głośności analizowanego dźwięku (o dowolnej częstotliwości) z głośnością tonu o częstotliwości 1000 Hz (tzw. standardu) i różnych poziomach. Wyznaczona w ten sposób miara głośności nazywana jest *poziomą głośnością*. Poziom głośności dowolnego dźwięku jest równy poziomowi ciśnienia akustycznego tonu standardowego (o częstotliwości 1000 Hz), z którym jest on jednakowo głośny.

Poziom głośności tonu standardowego o częstotliwości 1000 Hz, z definicji, jest liczbowo równy jego poziomowi ciśnienia akustycznego. Jednostką poziomu głośności jest *fon*: dany dźwięk ma tyle fonów, ile decybeli (SPL) ma ton o częstotliwości 1 kHz, który jest jednakowo głośny z analizowanym dźwiękiem. Dla niskich poziomów głośności krzywe jednakowej głośności są zbliżone kształtem do krzywej progowej, a dla wysokich po-

ziomów głośności stają się bardziej płaskie. Oznacza to, że tempo wzrostu głośności wraz ze wzrostem poziomu natężenia jest różne dla tonów o różnej częstotliwości. Zatem głośność dźwięku złożonego zależy od częstotliwości jego poszczególnych składowych zwłaszcza wówczas, gdy dźwięk ma małe natężenie. Jeśli wszystkie składowe mają takie same amplitudy, to „udział” najniższych składowych w głośności dźwięku jest znacznie mniejszy niż składowych o średnich częstotliwościach.

2.1.3. Skalowanie głośności

Skale głośności stworzono w celu wyznaczenia zależności pomiędzy subiektywną głośnością dźwięku i jego natężeniem. Skale te wyznaczono wykorzystując metodę szacowania wielkości, w której słuchacz proszony jest o ocenę dźwięków względem dźwięku porównawczego. Na podstawie takich badań Stevens (1957) stwierdził, że głośność L jest potęgową funkcją natężenia I , tzn. $L = kI^{0.3}$, gdzie k jest stałą zależną od słuchacza i zastosowanych jednostek. Innymi słowy głośność danego dźwięku jest proporcjonalna do natężenia podniesionego do potęgi 0.3. Oznacza to, że dwa dźwięki, z których jeden oceniany jest jako dwa razy głośniejszy od drugiego różnią się o 10 dB. Stevens zaproponował, by jednostką głośności dźwięku był *son*. Głośność jednego *sona* jest wielkością przyjętą arbitralnie i jest ona równa głośności tonu o częstotliwości 1000 Hz o poziomie 40 dB SPL. Ton o częstotliwości 1000 Hz i poziomie 30 dB SPL jest zazwyczaj oceniany jako dwukrotnie cichszy od tonu 40 dB i ma głośność 0.5 *sona*, a dźwięk o poziomie 50 dB SPL ma głośność 4 *sonów*. Ta prosta zależność nie obowiązuje dla poziomów niższych od 40 dB.

Związek pomiędzy natężeniem i głośnością sygnałów opisany prawem potęgowym został wielokrotnie potwierdzony eksperymentalnie. Jednak mimo to skalowanie głośności poddawano krytyce, bowiem żadna ze stosowanych metod badawczych nie wydaje się być wolna od efektów ubocznych (Warren, 1981).

2.2. Selektywność częstotliwościowa

2.2.1. Istota selektywności częstotliwościowej

Sygnal sinusoidalny wywołuje wrażenie tonalne o czystej, wyraźnej barwie. Wysokość tonu jest związana z jego częstotliwością, a głośność z natężeniem. Gdy zaprezentuje się jednocześnie dwa tony, to ich percepcyjne rozdzielenie (usłyszenie ich jako dwóch oddzielnych dźwięków) zależy od odległości w dziedzinie częstotliwości. Jeśli tony te są znacząco rozseparowane, to słyszymy dwa oddzielne dźwięki: ucho ludzkie zachowuje się jak analizator rozkładający dźwięk złożony (dwuton) na składowe tonalne. Zdolność układu słuchowego do percepcyjnego rozłożenia dźwięku na składowe nazywa się rozdzielczością (selektywnością) częstotliwościową. Jeśli jednak składowe dwutonu mają zbliżone częstotliwości to słyszymy pewien dźwięk pośredni, będący mieszaniną dwóch tonów składowych: rozdzielczość częstotliwościowa jest niewystarczająca do rozseparowania tych składowych. Najmniejszy odstęp częstotliwości dwóch tonów, przy którym stają się one słyszalne oddzielnie, zależy od ich częstotliwości, gdyż rozdzielczość częstotliwościowa nie jest stała i pogarsza się ze wzrostem częstotliwości (Plomp i Mimpen, 1968).

Rozdzielczość częstotliwościowa jest fundamentalną właściwością słuchu i przejawia się we wszystkich aspektach percepcji dźwięku. Wzrok, w odróżnieniu od słuchu, nie posiada zdolności do analizy bodźca: światło białe jest dla nas światłem białym, a nie zbiorem barw, z których się ono składa. Nie potrafimy w nim dostrzec barw składowych.

2.2.2. Wstęga krytyczna, filtry słuchowe

Selektywność częstotliwościową układu słuchowego opisuje się zwykle w kontekście tzw. filtrów słuchowych. Każdy punkt błony podstawnej „jest wrażliwy” na inną częstotliwość, co pozwala stwierdzić, że zachowuje się w przybliżeniu tak jak filtr pasmowo-przepustowy¹ o określonej częstotliwości środkowej i paśmie przepustowym. Cała błona podstawna może być potraktowana jako zbiór filtrów tego typu. Właściwości tych filtrów, nazywanych filrami słuchowymi, można wyznaczać w badaniach fizjologicznych bądź też w badaniach psychofizycznych.

Wyznaczeniu kształtu (charakterystyki) filtru słuchowego poświęcono jak dotychczas wiele uwagi. Pierwszą hipotezę, postulującą prostokątny kształt filtru słuchowego, zaproponował Fletcher (1940). Mierzył ona próg detekcji sygnału sinusoidalnego w obecności pasma szumu o częstotliwości środkowej równej częstotliwości tonu. Gęstość widmowa mocy szumu (moc szumu w każdym jego jednostkowym paśmie) była jednakowa, co oznacza, że zwiększanie szerokości pasma szumu powodowało wzrost całkowitej mocy szumu. Początkowemu zwiększaniu szerokości pasma szumu towarzyszył wzrost progu detekcji tonu. Jednak począwszy od pewnej szerokości pasma próg ten ustalał się na pewnym poziomie i dalsze zwiększanie szerokości pasma nie zwiększało progu detekcji tonu. Na tej podstawie Fletcher stwierdził, że: (i) słuchacz dokonuje detekcji tonu na tle szumu, wykorzystując filtr o częstotliwości środkowej zbliżonej do częstotliwości tonu, (ii) dopóki szerokość pasma szumu maskującego mieści się w filtrze słuchowym, dopóty szum ten wpływa na próg detekcji sygnału. Gdy szerokość pasma szumu wzrasta, lecz w dalszym ciągu nie przekracza szerokości filtru słuchowego, wzrasta również próg detekcji sygnału. Jednak gdy szerokość pasma szumu przekroczy szerokość filtru słuchowego, to część szumu, która znalazła się poza filtrem, przestaje wpływać na detekcję tonu i próg przyjmuje stałą wartość. Szerokość pasma szumu o częstotliwości środkowej równej częstotliwości tonu, przy której próg detekcji tonu przestaje wzrastać, nazywa się *wstęgą krytyczną* (Fletcher, 1940).

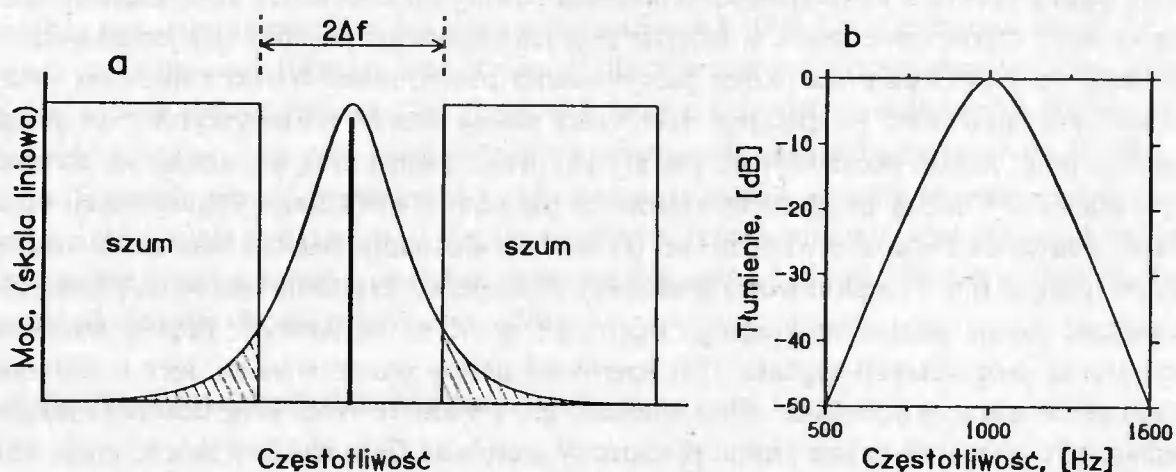
Koncepcja prostokątnych filtrów słuchowych pozwala zinterpretować rezultaty eksperymentów dotyczących maskowania tonów przez sygnały szerokopasmowe (Zwicker i Feldtkeller, 1976), choć w odniesieniu do innych typów sygnałów maskujących nie sprawdza się.

Niezależnie od znacznego uproszczenia kształtu filtru słuchowego (tj. założenia, że jest on prostokątny) i związanych z tym konsekwencji, koncepcja wstęg krytycznych jest w

¹ Filtr jest urządzeniem elektrycznym umożliwiającym zmniejszenie amplitud niektórych składowych częstotliwościowych sygnału. Filtr pasmowoprzepustowy redukuje znacząco amplitudy najniższych i najwyższych składowych, pozostawiając bez zmian składowe przypadające na jego tzw. pasmo przepustowe.

dalszym ciągu szeroko stosowana, a wyznaczeniu szerokości wstęg krytycznych poświęcono w latach 1950-1970 wiele uwagi. Ich podsumowaniem jest zależność zaproponowana przez Scharfa (1970), przedstawiona na rys. 7 linią przerywaną. Wstęgi krytyczne mają szerokość stałą (równą 100 Hz) dla częstotliwości środkowych mniejszych od 500 Hz. Dla częstotliwości większych od 500 Hz szerokość wstęgi krytycznej wzrasta ze wzrostem częstotliwości: przyjmuje się, że szerokość ta jest równa około 17 % częstotliwości środkowej.

Jedną z najczęściej stosowanych obecnie metod wyznaczania charakterystyki filtru słuchowego (Patterson, 1974; 1976) oparta jest na wyznaczeniu progów maskowania tonu, co schematycznie przedstawiono na rys. 6a. Ton maskowany (zaznaczony grubą pionową linią) ma stałą częstotliwość, a sygnałem maskującym jest szum pasmowozaporowy, o częstotliwości środkowej pasma zaporowego równej częstotliwości sygnału. Zmieniając szerokość pasma zaporowego ($2\Delta f$) wyznacza się próg detekcji sygnału jako funkcję szerokości pasma zaporowego. Dla sygnału umieszczonego w środku symetrii pasma zaporowego szumu optymalny stosunek sygnału do szumu uzyskuje się na wyjściu filtru o częstotliwości środkowej równej częstotliwości sygnału, który pokazano schematycznie na rys. 6a.



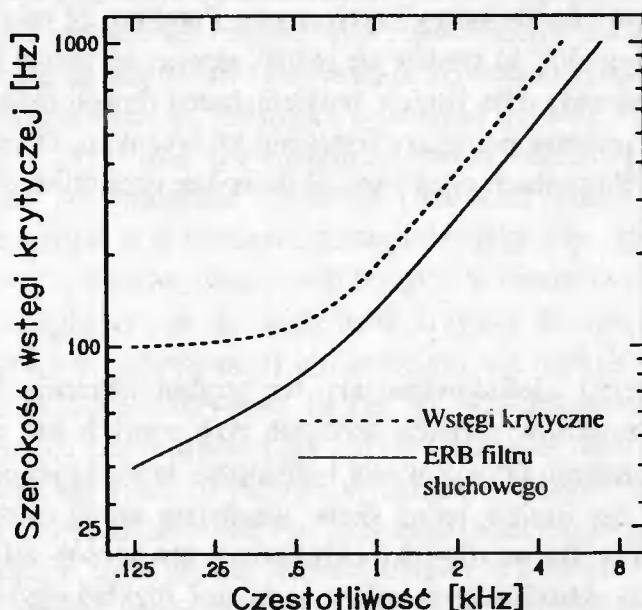
Rys. 6. Schemat metody Pattersona wyznaczania kształtu filtru słuchowego oraz przykładowa charakterystyka takiego filtru (Patterson, 1976).

Gdy szerokość zaporowego pasma szumu wzrasta, to przez filtr słuchowy przejdzie coraz mniej szumu i dlatego próg detekcji sygnału maleje. Przy założeniu, że próg odpowiada stałemu ilorazowi sygnału do szumu na wyjściu filtru, zmiana progu detekcji sygnału ze zmianą szerokości pasma zaporowego umożliwia wyznaczenie kształtu filtru słuchowego, którego przykład przedstawiono na rys. 6b. Filtr ten ma zaokrąglony wierzcho-

lek, dość strome zbocza i nie można opisać go za pomocą liczby, tak jak wstęgi krytycznej. Dlatego też szerokość filtru słuchowego wyraża się dość często za pomocą tzw. ekwiwalentnej szerokości prostokątnej, ERB. Jest ona liczbowo równa szerokości idealnego prostokątnego filtru o wartości transmitancji równej maksymalnej transmitancji filtru słuchowego, przy czym moc przechodzącego szumu przez ten filtr jest równa mocy szumu przechodzącego przez filtr słuchowy. Przebieg zależności ekwiwalentnej szerokości filtrów słuchowych od częstotliwości jest opisany za pomocą następującej zależności (Glasberg i Moore, 1990):

$$ERB = 24.7(4.37F + 1)$$

gdzie F jest częstotliwością środkową filtru wyrażoną w kHz. Przebieg tej zależności jest w pełni udokumentowany w przedziale częstotliwości od 80 do 12 000 Hz i zilustrowano ją na rys. 7 linią ciągłą.



Rys. 7. Porównanie szerokości wstęg krytycznych i ekwiwalentnych szerokości prostokątnych (ERB) filtru słuchowego (Moore, 1997).

Zasadnicza różnica pomiędzy szerokościami wstęg krytycznych a ekwiwalentnymi szerokościami filtrów słuchowych obserwowana jest dla częstotliwości mniejszych od 500 Hz. Wstęgi krytyczne są około trzykrotnie szersze niż ekwiwalentna szerokość filtrów słuchowych dla częstotliwości 100-120 Hz. Różne przebiegi szerokości wstęg krytycznych i filtrów słuchowych w zależności od częstotliwości sugerują, że w tym zakresie częstotliwości modele te nie są zgodne (Sęk, 1994; Sęk i Moore, 1994). Dla częstotliwości większych niż 500 Hz szerokości wstęg krytycznych są generalnie większe niż ekwiwalentna szerokość filtrów słuchowych. Jednak w przedziale częstotliwości 500-5000 Hz iloraz tych miar szerokości filtru słuchowego jest w przybliżeniu stały, co pozwala stwierdzić, że w tym przedziale częstotliwości oba modele są zgodne.

Przedstawione dotychczas rozważania dotyczyły analizy funkcjonowania słuchu na przykładzie wykorzystania jednego filtru. Choć zakładano, że decyzja słuchacza oparta jest na jednym filtrze, to w rzeczywistości korzystamy zazwyczaj z wielu filtrów słuchowych. Przykładem bodźców, które wywołują aktywność wielu filtrów słuchowych może być np. muzyka czy mowa, które są sygnałami o paśmie znacznie szerszym niż jeden filtr słuchowy. Ponadto percepcja barwy dźwięku zależy, przynajmniej częściowo, od rozkładu aktywności różnych filtrów słuchowych, a detekcja sygnału w warunkach maskowania często zależy od wyniku porównania sygnałów wyjściowych z różnych filtrów słuchowych (Moore, 1992).

Innym nasuwającym się pytaniem jest to, czy istnieje skończona liczba sąsiadujących ze sobą wstęg krytycznych, czy też ciąg nakładających się na siebie takich wstęg. Scharf (1970) np. przedstawił tablicę, w której zawarł 24 następujące po sobie wstęgi krytyczne. Górna częstotliwość graniczna każdej wstęgi była jednocześnie dolną częstotliwością graniczną następnej wstęgi krytycznej o większej częstotliwości środkowej, co sugerować może istnienie skończonej liczby wstęg krytycznych. Pomimo, że taka metoda prezentacji wyników jest bardzo wygodna, to wydaje się jednak pewne, że wstęgi krytyczne mają charakter ciągły, a nie dyskretny. Nie istnieje bowiem żaden dowód eksperymentalny na istnienie nieciągłości lub przerwy pomiędzy wstęgami krytycznymi. Dlatego możemy mówić o wstędze krytycznej (filtrze słuchowym) wokół dowolnej częstotliwości z zakresu słyszalnego.

2.2.3. Percepcja barwy

Barwa dźwięku jest zdefiniowana jako ten atrybut wrażenia dźwiękowego, który umożliwia rozróżnienie głosów różnych osób lub tych samych nut granych na różnych instrumentach. W odróżnieniu od wysokości i głośności, które są jednowymiarowe, barwa jest wielowymiarowa: nie istnieje jedna skala, względem której można by porównywać barwę różnych dźwięków. Barwa dźwięku określona w ten sposób zależy od kilku fizycznych cech tego dźwięku, wśród których należy wymienić rozkład jego energii w dziedzinie częstotliwości, zmiany tego rozkładu w czasie oraz charakter dźwięku, tzn. czy jest periodyczny i wytwarza wrażenie tonu o częstotliwości z zakresu słyszalnego, lub czy jest to dźwięk nieregularny wytwarzający wrażenie szumu. Zasadnicze znaczenie dla barwy dźwięku, zwłaszcza w odniesieniu do instrumentów muzycznych, ma sposób prezentacji dźwięku, tzn. czy jest on prezentowany w sposób ciągły, czy z przerwami oraz sposób włączania i wyłączania dźwięku: szybkie narastanie dźwięku w początkowej jego fazie ma zupełnie inną barwę niż narastanie powolne.

Dla ciągłych dźwięków periodycznych można przytoczyć uściśloną definicję barwy podaną przez Amerykańską Organizację Normalizacyjną: „barwa dźwięku jest tym atrybutem wrażenia słuchowego, względem którego słuchacz może ocenić dwa stałe w czasie dźwięki złożone o tej samej głośności, wysokości i czasie trwania jako niepodobne”. Barwa zdefiniowana w ten sposób zależy głównie od względnych amplitud poszczególnych składowych widmowych dźwięków złożonych. Np. dźwięki zawierające głównie duże częstotliwości mają „ostrą” barwę, a dźwięki, w których dominują częstotliwości małe mają barwę „łagodną”. Rozróżnianie barwy jest kolejnym dowodem na to, że ucho działa jak anali-

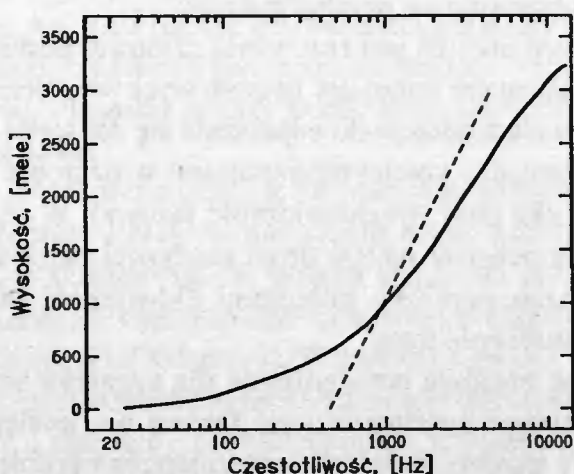
zator widmowy. Składowe dźwięku złożonego są częściowo rozdzielane przez filtry słuchowe, a rozkład aktywności tych filtrów w funkcji ich częstotliwości środkowej determinuje barwę.

2.3. Percepcja wysokości

2.3.1. Wysokość tonów. Skalowanie wysokości

Wysokość jest tym atrybutem wrażenia słuchowego, względem którego dźwięki można uporządkować na skali muzycznej. Innymi słowy – zmiany wysokości wywołują wrażenie melodii. Ponieważ wysokość jest atrybutem subiektywnym, to nie można jej mierzyć bezpośrednio. Przyporządkowanie pewnemu dźwiękowi wartości wysokości polega na wskazaniu częstotliwości tonu mającego subiektywnie taką samą wysokość, co rozpatrywany dźwięk. Dla sygnałów sinusoidalnych (tonów) wysokość jest ściśle związana z ich częstotliwością: im większa jest częstotliwość, tym większa jest wysokość.

Poprzez analogię do skalowania głośności podejmowano próby skalowania wysokości dźwięków i zaproponowano, by jednostką wysokości był *mel*. Pierwszą skalę tego typu wprowadził Stevens w 1937 roku, przyjmując za dźwięk odniesienia ton o częstotliwości 1000 Hz i przypisując mu arbitralnie wysokość 1000 meli (Hartmann, 1997). Wysokość innych dźwięków określano w odniesieniu do tego dźwięku. Np. dźwięk dwa razy wyższy ma 2000 meli, dźwięk o połowę niższy – 500 meli. Skonstruowaną w ten sposób skalę przedstawiono linią ciągłą na rys. 8. Nachylenie krzywej wyrażającej wysokość tonu w melach do osi odciętych (częstotliwości) jest mniejsze niż nachylenie powszechnie stosowanej skali muzycznej (linia przerywana na rys. 8). Np. pasmo obejmujące trzy oktawy 200-1600 Hz (o stosunku częstotliwości jak 1:8) ma stosunek najmniejszej do największej wysokości (w melach) jak 1:4,3. Wysokość wyrażona w melach wzrasta w ogólności ze wzrostem częstotliwości do około 5000 Hz, przy czym interwał oktawy dla małych częstotliwości jest percepcyjnie mniejszy niż interwał oktawy dla dużych częstotliwości. Np. dla oktawy 100-200 Hz różnica wysokości wynosi $300 - 160 = 140$ meli, a dla oktawy 1-2 kHz różnica wysokości wynosi $1550 - 1000 = 550$ meli.



Rys. 8. Porównanie skali meli (linia ciągła) i skali muzycznej (linia przerywana).

Zwicker (1990) zaproponował inną skalę meli, w której przyjął, że tonowi o częstotliwości 125 Hz odpowiada 125 meli. Zakres tej skali wynosi 2400 meli, tak że każdym 100 melom odpowiada jedna wstęga krytyczna. Przyjęcie tych wartości miało na celu powiązanie wysokości z położeniem maksimum drgań błony podstawnej. Ponieważ każda wstęga krytyczna odpowiada stałemu odcinkowi błony podstawnej (ok. 0.9 mm), to w ten sposób próbowano połączyć wysokość dźwięku z odległością miejsca maksymalnego pobudzenia od okienka owalnego.

Wyrażanie wysokości w melach poddawano krytyce głównie ze względu na to, że jednostki te są nieprzydatne w określaniu wysokości dźwięków muzycznych. Z jednej strony różnym oktaвам przypisuje różne wartości wysokości w melach, podczas gdy wiadomo z codziennego doświadczenia, że relacje między interwałami muzycznymi nie zależą od bezwzględnych wysokości oktawy. Wysokość dźwięku rozpatruje się zwykle w dwóch wymiarach. Jednym z nich jest wysokość tonalna, a drugim wysokość muzyczna. Opis wysokości dźwięków za pomocą meli znajduje zastosowanie tylko w odniesieniu do wysokości tonalnej i nie może być stosowany do wysokości muzycznej.

2.3.2. Teorie percepcji wysokości

Jedną z klasycznych teorii percepcji dźwięku jest tzw. teoria miejsca (lub lokalizacyjna), która postuluje, że wysokość jest związana z rozkładem aktywności różnych punktów błony podstawnej i rozkładem aktywności różnych neuronów słuchowych. Ton wywołuje maksymalną aktywność tych włókien nerwowych, które unerwiają komórki rzęskowe w najaktywniejszych miejscach błony podstawnej. Zakłada się, że „położenie” tego maksimum określa wysokość całego dźwięku. Zmiany częstotliwości są wykrywane na podstawie tych miejsc, w których aktywność zmienia się najbardziej. Teoria ta napotyka znaczne trudności w wyjaśnieniu wysokości bardzo głośnych dźwięków. Choć wysokość nieznacznie zależy od poziomu dźwięku, to bezpośrednie obserwacje wychylenia błony podstawnej świadczą o tym, że położenie punktu maksymalnego jej wychylenia przesuwają się dla bardzo głośnych dźwięków aż o pół oktawy (Pickles, 1988). Przyjmując więc teorię miejsca można by oczekiwać znaczącej zmiany wysokości dźwięków wywołanej wzrostem ich natężenia, czego nie obserwuje się psychofizycznie.

Alternatywą do teorii miejsca jest tzw. teoria czasowa, postulująca że wysokość ma związek z czasowym przebiegiem impulsów neuronowych wywołanych przez ten bodziec. Wyładowania neuronów mają tendencję do pojawiania się dla ściśle określonej fazy dźwięku. Interwały czasowe pomiędzy kolejnymi impulsami w neuronie są całkowitymi wielokrotnościami okresu dźwięku (tzw. synchroniczność fazowa). W przypadku gdy sygnałem jest ton zakłada się, że na pewnym piętrze drogi słuchowej określany jest najczęściej powtarzający się przedział czasu pomiędzy impulsami. Odwrotność najkrótszego z tych przedziałów wskazuje na częstotliwość tonu.

Teoria czasowa nie znajduje potwierdzenia dla sygnałów sinusoidalnych o bardzo dużej częstotliwości, ponieważ synchroniczność fazowa nie występuje dla częstotliwości większych od 5 kHz. Nie stanowi to jednak poważniejszego problemu, ponieważ dźwięki wytwarzane przez instrumenty muzyczne, ludzki głos oraz większość dźwięków spotykanych w codziennym życiu mają częstotliwości podstawowe mniejsze od 5 kHz. Za pomocą

tej teorii można też interpretować wysoką czułość słuchu na zmiany częstotliwości (Sęk i Moore, 1995), co przemawia na rzecz jej przyjęcia.

Teoria czasowa i teoria miejsca nie wykluczają się wzajemnie. Mechanizm lokalizacyjny obowiązuje w całym zakresie częstotliwości słyszalnych, a mechanizm synchronizacji fazowej (teoria czasowa) „wspomaga” go znacząco w zakresie częstotliwości do 4-5 kHz.

2.2.3. Wysokość dźwięków złożonych

Każdy dźwięk złożony, którego częstotliwość powtarzania leży w zakresie częstotliwości 20-20 000 Hz może mieć wysokość. O ile w przypadku tonów wysokość związana jest z ich częstotliwością, to w przypadku dźwięków złożonych wysokość związana jest z częstotliwością powtarzania fali dźwiękowej. W przypadku periodycznego dźwięku złożonego wysokość odpowiada jego częstotliwości podstawowej. Określenie wysokości dźwięku złożonego polega na przyporządkowaniu mu tonu mającego tę samą co dźwięk złożony wysokość.

Dźwięki periodyczne mogą być rozłożone na składowe sinusoidalne zawierające składową podstawową i szereg składowych harmonicznnych, których częstotliwości są wielokrotnościami składowej podstawowej. Np. krótki impuls powtarzający się 200 razy w ciągu jednej sekundy ma składową podstawową o częstotliwości 200 Hz i harmoniczne o częstotliwościach 400, 600, 800, ..., itd. Hz. Wysokość takiego dźwięku jest podobna do wysokości tonu o częstotliwości 200 Hz, choć obecność składowej o częstotliwości 200 Hz nie decyduje o jego wysokości. Można bowiem tę składową usunąć (odfiltrować), a wysokość takiego dźwięku pozostanie niezmienną. Zjawisko to nazywa się słyszeniem nieobecnej składowej podstawowej, ponieważ pomimo braku składowej o częstotliwości 200 Hz (składowej podstawowej) wysokość dźwięku w dalszym ciągu jest taka jak tonu o tej częstotliwości. Wrażenie wysokości pozostaje prawie niezmienną, gdy dźwięk zostanie odfiltrowany w ten sposób, że pozostanie w nim tylko kilka składowych harmonicznnych o dużych częstotliwościach np. 800, 1000 i 1200 Hz. Koncepcja, według której wysokość zależy od odstępu częstotliwościowego składowych harmonicznnych jest również niepoprawna. Można to wykazać przesuując częstotliwości składowe, które pozostały po odfiltrowaniu, o stałą wartość, np. o 10 Hz, tak aby harmoniczne miały częstotliwości 810, 1010 i 1210 Hz. Wysokość takiego dźwięku jest inna (wyższa) niż wysokość dźwięku rozpatrywanego poprzednio, choć odległości między składowymi obu dźwięków w dziedzinie częstotliwości są nadal takie same.

Teorie percepcji wysokości zakładają, że tworzenie wrażenia wysokości jest procesem dwustopniowym (Moore, 1997). W pierwszym etapie analizowane są niskie harmoniczne za pomocą filtrów słuchowych. W drugim etapie ma miejsce analiza pobudzenia filtrów słuchowych i określenie częstotliwości składowej podstawowej. Odczuwana wysokość odpowiada częstotliwości tej składowej. W jednym z modeli tego typu wysokość jest określana na podstawie podharmonicznnych składowych obecnych w dźwięku (Terhardt, 1979). Układ słuchowy „znajduje” podharmoniczne wszystkich składowych obecnych w dźwięku złożonym. Wysokość dźwięku złożonego jest równa częstotliwości tej podharmonicznej, która jest podharmoniczną największej liczby składowych dźwięku. Podharmo-

niczne dźwięku złożonego z trzech tonów o częstotliwościach 800, 1000 i 1200 Hz przedstawiono w tabeli 1. Każda z tych składowych ma swoje „własne” podharmoniczne: podharmoniczna o częstotliwości 200 Hz występuje najwięcej razy, tj. dla każdego z analizowanych tonów. To właśnie ona określa wysokość rozważanego dźwięku złożonego.

Tabela 1. Częstotliwości podharmonicznych tonów o częstotliwościach 800, 1000 i 1200 Hz.

Częstotliwość składowej, Hz		
800	1000	1200
400	500	600
266.7	333.3	400
200	250	300
160	200	240
133.3	166.7	200
114.29	142.86	171.4

Znajdowanie podharmonicznych każdego tonu w wielotonie jest możliwe na podstawie informacji zawartych w synchroniczności fazowej: wyładowania neuronowe przypadają zawsze dla tej samej fazy sygnału pobudzającego. Jednak odstępy pomiędzy nimi odpowiadają zawsze całkowitej wielokrotności okresu tego tonu. Zatem najmniejszy odstęp pomiędzy impulsami w danym neuronie odpowiada częstotliwości tonu, a każda kolejna całkowita wielokrotność tego odstępu – kolejnym jego podharmonicznym. Tonom o częstotliwościach 800, 1000 i 1200 Hz odpowiadają odstępy pomiędzy impulsami, które przedstawiono w tabeli 2. Tylko odstęp 5 ms występuje dla wszystkich trzech składowych i odpowiada on częstotliwości 200 Hz, czyli wysokości analizowanego dźwięku.

Tabela 2. Odstępy czasu (ms) pomiędzy impulsami w neuronach dla tonów 800, 1000 i 1200 Hz.

<i>n</i>	Częstotliwość składowej, Hz		
	800	1000	1200
1	1.25	1	0.83333
2	2.5	2	1.66666
3	3.75	3	2.5
4	5	4	3.33333
5	6.25	5	4.16666
6	7.5	6	5

Najnowsze teorie zakładają, że percepcja wysokości dźwięków złożonych jest procesem analizy pobudzenia opartym na wstępnej analizie składowych dźwięku. Analiza ta może zależeć zarówno od informacji dotyczących „miejsca”, jak i informacji czasowej.

Gdy w dźwięku obecne są tylko jego wysokie harmoniczne, wysokość jest w dalszym ciągu percypowana, lecz wrażenie jest słabsze.

2.4. Rozdzielczość czasowa ucha

Znakomita większość informacji, jaką niesie dźwięk, zawarta jest w zmienności jego parametrów fizycznych. Stąd też układ słuchowy jest dobrze przystosowany do detekcji zmian dźwięków w czasie. Ograniczenia tej zdolności odzwierciedlają tzw. czasową rozdzielczość ucha. Pomiar rozdzielczości czasowej ucha polega na wyznaczeniu progów detekcji przerw (interwałów ciszy) w szumie. Rezultaty tych eksperymentów wykazały, że najkrótszy interwał ciszy, jaki potrafimy odebrać trwa 2-3 ms (Penner, 1977).

Wykorzystanie szumu szerokopasmowego do określenia czasowej zdolności rozdzielczej nie pozwala na uchwycenie wpływu filtrowania peryferyjnego na ten proces. Ponieważ szerokość filtrów słuchowych wzrasta wraz ze wzrostem częstotliwości środkowej, to można się spodziewać, że rozdzielczość czasowa układu słuchowego powinna zmieniać się wraz z częstotliwością: dla małych częstotliwości rozdzielczość ta powinna być gorsza.

Próg detekcji interwału ciszy w sygnale wąskopasmowym wzrasta monotonicznie wraz ze zmniejszaniem się częstotliwości środkowej pasma szumu. Dla dużych częstotliwości środkowych wartość progu detekcji interwału ciszy jest zbliżona do wartości tego progu uzyskiwanych dla sygnałów szerokopasmowych, co sugeruje, że podczas detekcji przerw w sygnale szerokopasmowym słuchacze wykorzystują przede wszystkim składowe o dużych częstotliwościach. Zwiększenie progu detekcji dla małych częstotliwości środkowych może być związane z odpowiedzią czasową filtru słuchowego.

2.5. Lokalizacja dźwięków

Międzyuszne różnice czasu i natężenia są dwoma zasadniczymi czynnikami umożliwiającymi lokalizację źródeł dźwięku. Dźwięk ze źródła usytuowanego po lewej stronie dociera najpierw do ucha lewego, a po chwili do ucha prawego. Jego natężenie w uchu lewym jest większe niż w uchu prawym. Dla sygnałów sinusoidalnych różnica czasów ich dotarcia do obu uszu jest równoznaczna różnicy ich faz, co wykorzystuje układ słuchowy. Wykorzystanie tej różnicy w całym zakresie częstotliwości słyszalnych nie jest możliwe: dla dźwięków o częstotliwości powyżej 1500 Hz długość fali staje się porównywalna z wymiarami głowy i słuchacz nie może określić, który cykl fali z lewego ucha odpowiada danemu cyklowi fali z ucha prawego. Dla tych częstotliwości różnice fazy niosą niejednoznaczne informacje o lokalizacji źródeł dźwięku. Dla małych częstotliwości (do 1500 Hz) zdolność detekcji międzyusznych różnic czasu jest bardzo dobra i wynosi 10-20 μ s, co odpowiada zmianie kąta padania dźwięku w płaszczyźnie horyzontalnej o około 1-2°.

Międzyuszne różnice natężenia umożliwiają lokalizację źródeł dźwięku o dużych częstotliwościach, większych od 1500 Hz. Dla tych częstotliwości głowa jest źródłem cienia akustycznego i powstająca w ten sposób międzyuszna różnica natężenia jest wykorzystywana przez układ słuchowy do lokalizacji.

Koncepcja zakładająca, że lokalizacja źródła dźwięku o małych częstotliwościach odbywa się na podstawie międzyuszných różnic czasu (fazy), a źródeł dźwięku o dużych częstotliwościach na podstawie międzyuszných różnic natężenia, nosi nazwę teorii dupleksowej. I choć sprawdza się ona dość dobrze dla tonów, to nie jest wystarczająco ścisła dla dźwięków złożonych. Większość dźwięków, jakie napotykamy w codziennym życiu ma częstotliwości powtarzania obwiedni mniejsze od 600 Hz i dlatego międzyuszne różnice czasu są wykorzystywane do ich lokalizacji w większości sytuacji.

Istnienie różnic pomiędzy sygnałami docierającymi do lewego i prawego ucha nie wyjaśnia wyczerpująco naszej zdolności do lokalizowania dźwięków. Np. różnice czasu lub natężenia nie istnieją, gdy dźwięk dociera do nas dokładnie z przodu lub z tyłu albo z dołu lub z góry, a mimo to potrafimy dość precyzyjnie określić kierunek położenia źródła dźwięku. Lokalizacja tego typu jest możliwa dzięki małżowinom usznym. Stwierdzono bowiem (Gardner i Gardner, 1973; Blauert, 1983), że małżowiny uszne dokonują znacznych zmian dźwięków dochodzących do błon bębenkowych. Zmiany te zależą od kierunku padania dźwięku i określa się je jako tzw. filtrowanie przestrzenne. Małżowina uszna nie ogranicza się jedynie do filtrowania tego typu, lecz odgrywa również rolę w ocenie, czy źródło dźwięku usytuowane jest wewnątrz czy na zewnątrz głowy. Dźwięk jest oceniany jako docierający do nas z zewnątrz jedynie wtedy, gdy zostanie on przetransformowany widmowo przez małżowiny. Dlatego źródło dźwięku odtwarzanego przez słuchawki jest oceniane jako leżące wewnątrz głowy, bowiem małżowiny uszne, a także głowa, nie wywierają na taki dźwięk żadnego wpływu. Jednak dźwięki odtwarzane za pomocą słuchawek można spreparować tak, aby brzmiały jak dźwięki dochodzące z zewnątrz. W tym celu należy je zarejestrować przy użyciu tzw. sztucznej głowy (wiernego modelu głowy ludzkiej) lub przetworzyć, opierając się na tzw. funkcji przeniesienia głowy. Funkcja ta stanowi charakterystykę przeniesienia układu, jaki stanowią głowa i małżowiny uszne, a wyznacza się ją eksperymentalnie dla każdego słuchacza, jako że jej przebiegi charakteryzują się znaczną zmiennością międzysobniczą (Blauert, 1983).

Dźwięk ze źródła dochodzi do uszu wieloma drogami. Część dźwięku dociera drogą bezpośrednią, a część po jednym lub wielu odbiciach. Jednak zazwyczaj nie zdajemy sobie sprawy z tych odbić i wydaje się, że mają one niewielki wpływ na ocenę kierunku źródła dźwięku. Powodem tego jest tzw. efekt precedensu (zjawisko Haasa, lub tzw. efekt pierwszeństwa). Jeśli interwał czasu pomiędzy kilkoma następującymi po sobie dźwiękami jest wystarczająco krótki, to są one słyszane jako pojedynczy dźwięk. Położenie źródła jest całkowicie zdeterminowane przez położenie źródła dźwięku bezpośredniego, tj. tego, którego dźwięk dociera pierwszy. Wynika stąd, że odbicia mają niewielki wpływ na percepcję kierunku, choć oczywiście mogą wpływać na barwę, głośność i ogólną jakość dźwięku.

Zjawisko precedensu występuje jedynie dla dźwięków nieciągłych, takich jak mowa czy muzyka i zanika, gdy odbicia są znacznie opóźnione lub mają natężenie porównywalne z natężeniem dźwięku bezpośredniego. W normalnych warunkach efekt ten odgrywa jednak ważną rolę w lokalizacji i identyfikacji dźwięków. Jest to zjawisko słyszenia dwuosznego. Gdy jedno z uszu zostanie zatkane, odbicia od ścian pomieszczenia są łatwiej słyszalne, a dźwięki wydają się bardziej huczące.

Podsumowanie

Głównym zagadnieniem omawianym w niniejszym artykule jest mechanizm filtrowania peryferyjnego i selektywność częstotliwościowa, które pełnią zasadniczą rolę w naszym słuchu. Selektywność częstotliwościowa umożliwia percepcyjne rozseparowanie dźwięków o różnej częstotliwości, detekcję sygnału na tle szumu, identyfikację barwy dźwięków, percepcję wysokości dźwięków złożonych oraz spostrzeganie zmian dźwięków w czasie. Podstawowe cechy selektywności częstotliwościowej można zinterpretować zakładając, że peryferyjny układ słuchowy funkcjonuje tak jak zbiór nakładających się na siebie pasmowoprzepustowych filtrów o określonych parametrach. Podstawą tego mechanizmu jest błona podstawna ślimaka, która dokonuje konwersji częstotliwości dźwięku na drganie ściśle określonego jej miejsca, a następnie na zmiany potencjałów neuronów dochodzących do tego miejsca. Proces ten jest wspierany przez zewnętrzne komórki rzęskowe, przyczyniające się do znacznego wzrostu selektywności częstotliwościowej błony podstawnej i czułości układu słuchowego.

Mechanizm analizy częstotliwościowej działa również na poziomie nerwu słuchowego. Informacje o częstotliwości, natężeniu i składzie widmowym bodźca przenoszone są zarówno w postaci rozkładu aktywności włókien nerwowych, jak i w czasowym przebiegu ich wyładowań odgrywających ważną rolę w percepcji wysokości.

Mechanizm analizy dźwięku jest wrażliwy fizjologicznie. Jego czasowe uszkodzenie mogą wywołać niektóre środki farmakologiczne czy też przebywanie w środowisku nadmiernie głośnych dźwięków. Nasz organizm nie posiada zdolności zregenerowania tego mechanizmu, co w konsekwencji oznacza trwałą, bezpowrotną utratę czułości słuchu.

LITERATURA CYTOWANA

- Blauert, J. (1983). *Spatial Hearing*. Cambridge: Mass. MIT Press.
- Durrant, J.D. i Lovrinic, J.H. (1995). *Basis of hearing science*. Baltimore: Williams & Wilkins.
- Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys.* 12, 47-65.
- Gardner, M.B. i Gardner, R.S. (1973). Problem of localization in the median plane: effect of pinnae cavity occlusion. *J. Acoust. Soc. Am.* 53, 400-408.
- Glasberg, B.R. i Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103-138.
- Hartmann, W.M. (1997). *Signals, sound, and sensation*. Woodbury, New York: American Institute of Physics.
- Kemp, D.T. (1978). Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* 64, 1386-1391.
- Moore, B.C.J. (1992). Across-channel processes in auditory masking. *J. Acoust. Soc. Jpn. (E)* 13, 25-37.
- Nelly, S.T. (1993). A model of cochlear mechanics with outer hair cell motility. *J. Acoust. Soc. Am.* 94, 137-146.

- Palmer, A.R. (1995). Neural signal processing. W: B.C.J. Moore (red.), *Hearing*. San Diego: Academic Press.
- Patterson, R.D. (1974). Auditory filter shape. *J. Acoust. Soc. Am.* 55, 802-809.
- Patterson, R.D. (1976). Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.* 59, 640-654.
- Penner, M.J. (1977). Detection of temporal gaps in noise as a measure of the decay of auditory sensation. *J. Acoust. Soc. Am.* 64, 552-557.
- Pickles, J.O. (1988). *An introduction to the physiology of hearing*, 2nd Ed. London: Academic Press.
- Plomp, R. i Mimpen, A.M. (1968). The ear as a frequency analyzer II. *J. Acoust. Soc. Am.* 43, 764-767.
- Robinson, D.W. i Dadson, R.S. (1956). A re-determination of the equal-loudness relations for pure tones. *Brit. J. Appl. Phys.* 7, 166-181.
- Santos-Sacchi, J. (1993). Harmonics of outer hair cell motility. *Biophysical Journal* 65, 2217-2227.
- Santos-Sacchi, J. (1997). Fast outer hair cell motility. W: C.D.G.P. Dallos, J.W. Matthews, M.A. Ruggero, C.R. Steele (red.), *The mechanics and biophysics of hearing*. New York: Springer-Verlag.
- Scharf, B. (1970). Critical band. W: J.V. Tobias (red.), *Foundations of modern auditory theory*. New York: Academic Press.
- Sek, A. (1994). Modulation thresholds and critical modulation frequency based on random amplitude and frequency changes. *J. Acoust. Soc. Jpn.* 15, 67-75.
- Sek, A. i Moore, B.C.J. (1994). The critical modulation frequency and its relationship to auditory filtering at low frequencies. *J. Acoust. Soc. Am* 95, 2606-2615.
- Sek, A. i Moore, B.C.J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *J. Acoust. Soc. Am.* 97, 2479-2486.
- Smooenburg, G.F. (1972a). Audibility region of combination tones. *J. Acoust. Soc. Am.* 52, 603-614.
- Smooenburg, G.F. (1972b). Combination tones and their origin. *J. Acoust. Soc. Am.* 52, 615-632.
- Stevens, S.S. (1957). On the psychophysical law. *Psychol. Rev.* 64, 153-181.
- Terhardt, E. (1979). Calculating virtual pitch. *Hear. Res.* 1, 152-182.
- Tyler, R.S., Small, A.M. i Abbas, P.J. (1979). Unmasking produced by combination tones. *J. Acoust. Soc. Am.* 66, 379-387.
- Warren, R.M. (1981). Measurement of sensory intensity. *Behav. Brain. Sci.* 4, 175-189.
- Yates, G.K. (1995). Cochlear structure and function. W: B.C.J. Moore (red.), *Hearing*. San Diego: Academic Press.
- Yates, G.K., Johnstone, B.M., Patuzzi, R.B. i Robertson, D. (1992). Mechanical preprocessing in the mammalian cochlea. *Trends in Neuroscience* 15, 57-61.
- Zwicker, E. i Fastl, H. (1990). *Psychoacoustics – facts and models*. Berlin: Springer-Verlag.
- Zwicker, E. i Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel-Verlag.