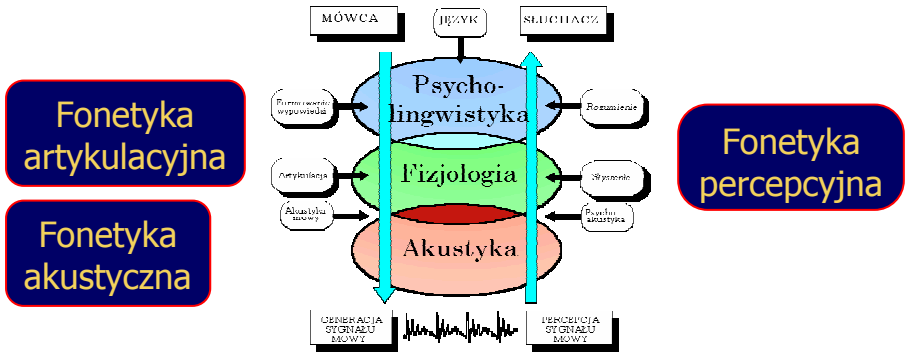
Dziedziny wiedzy obejmujące dwustronną komunikację werbalną



Podstawy opisu i klasyfikacji dźwięków mowy
- Opis artykulacyjny
- Opis akustyczny
- Opis percepcyjny

Fonetyka artykulacyjna
Przedmiotem fonetyki artykulacyjnej jest opisanie mechanizmu powstawania dźwięków mowy w narządzie artykulacyjnym człowieka.

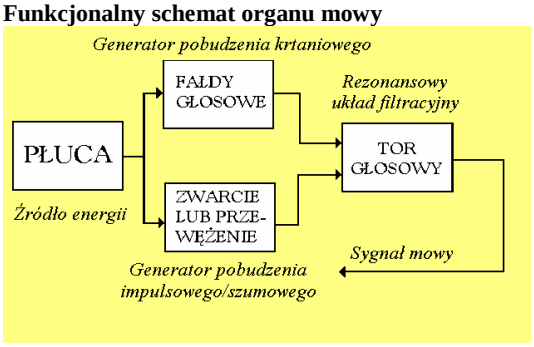
Fonetyka akustyczna
- Koncentruje się na analizie fizycznych własności dźwięków mowy promieniowanych wokół osoby mówiącej.
- Badanie dźwięków mowy odbywa się przy zastosowaniu fizycznych metod analizy sygnałów akustycznych.
- Jednocześnie poszukuje powiązań istniejących między czynnością artykulacyjną i wytworzonym sygnałem mowy

Fonetyka percepcyjna
- Bada percepcję dźwięków mowy, na poziomie układu centralnego.
- W badaniach stosowane są metody analizy subiektywnej oceny własności sygnałów akustycznych, zrozumiałości mowy itp.

Elementy narządu artykulacyjnego uczestniczące w formowaniu sygnału mowy
- Fałdy głosowe
- Podniebienie miękkie
- Podniebienie twarde
- Język
- Zęby
- Wargi

Źródłem energii promieniowanej podczas mówienia są płuca.
Podobnie jak ma to miejsce w instrumentach muzycznych dętych – źródłem energii niesionej przez dźwięk są płuca osoby grającej

Funkcjonalny schemat organu mowy



Cykle oddechowe: proporcje czasowe
Max pojemność płuc – ok. 7 litrów
Pojemność minimalna – 2 litry stale w płucach.
Objętość powietrza wymieniana podczas każdego cyklu oddechowego – 0.5 l
Częst. oddychania w stanie spoczynku – 12-20 cykli na minutę

Źródłem pobudzającym tor głosowy mogą być:

a) fałdy głosowe – modulują w sposób regularny przepływ powietrza wychodzącego z płuc,
b) szczelina utworzona w torze głosowym - powoduje powstanie zawirowań,
c) przeszkoda (zęby) – j.w.
d) krótkotrwały impuls powietrza – powstaje w wyniku nagłego otwarcia toru głosowego, po chwilowym zwarciu w określonym miejscu toru głosowego.

Instrumenty muzyczne stroikowe

Działają na podobnej zasadzie jak fałdy głosowe Np. Harmonijka ustna

Wzór na częstotliwość drgań fałdów głosowych

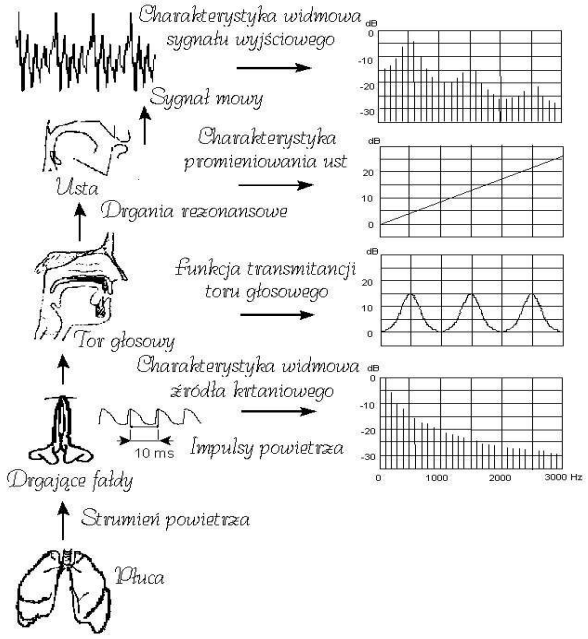
$$F_0 = \frac{1}{2\Pi}\sqrt{\frac{(K + K^*)}{m}}$$

m – masa fałdów
K – sztywność (napięcie) fałdów
K^* - sztywność aerodynamiczna

Narząd artykulacyjny jako układ akustyczny

Jest on swoistego rodzaju układem akustycznym, w którym można wyróżnić dwa podstawowe elementy:
a) źródło pobudzające
b) tor głosowy stanowiący w swej istocie rurę o zmiennym przekroju
wypełnioną powietrzem – w torze tym rozchodzi się fala płaska

Formowanie sygnału mowy



Stosunek powierzchni Ak/Ak+1 a charakterystyka częstotliwościowa

Nakładanie się fal padających i odbitych o różnym przesunięciu czasowym powoduje ich wielokrotne sumowanie (lub/i odejmowanie). Wielkość (amplituda) fal przenikających i odbitych zależy od stosunku powierzchni Ak/Ak+1. Stosunek tych powierzchni decyduje o charakterystyce częstotliwościowej układu cylindrów

Definicja formantu

Maksima w charakterystyce częstotliwościowej toru głosowego wpływające na różnicowanie dźwięków mowy danego języka nazywamy **formantami**. Oznacza to, że nie każde maksimum w widmie danego dźwięku mowy musi być formantem.

Rezonanse w falowodach cylindrycznych – fale stojące

Są dwa rodzaje falowodów cylindrycznych:

a) Rura zamknięta na jednym końcu, otwarta na drugim
b) Otwarta lub zamknięta na obu końcach – oba typy mają identyczne rezonanse

Falowody cylindryczne odgrywają podstawową rolę w instrumentach muzycznych (instrumenty dęte, organy itp.)

Konfiguracja toru głosowego, a częstotliwości formantowe

Między konfiguracją toru głosowego i częstotliwościami formantowymi istnieje związek, jednakże nie może być on jednoznacznie opisany. Różne konfiguracje geometryczne toru głosowego mogą mieć takie same częstotliwości formantowe, jak również różnym częstotliwościom formantowym mogą odpowiadać te same konfiguracje. Jednakże, zmiany w płaszczyźnie artykulacyjnej (miejsce i wysokość) powodują jednoznaczne zmiany w płaszczyźnie formantowej F1 i F2.

Charakterystyka aerodynamiczna spółgłosek
Podczas artykulacji spółgłosek w ponadkrtaniowej części toru głosowego powstaje zwężenie znacznie mniejsze, niż w przypadku artykulacji samogłoskowej. Wpływa ono na przepływ powietrza w tej części i może oddziaływać na pracę fałdów głosowych.
Zwężenie powoduje zmniejszenie amplitudy drgań fałdów głosowych, wskutek wzrostu ciśnienia ponadgłośniowego (różnica ciśnień pod- i ponad głośniowego jest mniejsza niż w przypadku artykulacji samogłoskowej). Może powodować też nieznaczne obniżenie częstotliwości drgań.

Efekty aerodynamiczne
Przy artykulacji spółgłosek powstają w zależności od stopnia zwężenia różne efekty aerodynamiczne i akustyczne. Zmniejszenie przekroju poprzecznego zwężenia powoduje zmniejszenie strumienia powietrza przepływającego w torze głosowym i wzrost ciśnienia ponadkrtaniowego. Gdy wzrost ten jest odpowiednio duży fałdy głosowe przestają poruszać się. Wzrost ciśnienia ponadkrtaniowego może nastąpić znacznie szybciej, gdy fałdy są rozwarte.

Stopień przewężenia
Sposób artykulacji spółgłosek określony jest przez wielkość zwężenia toru głosowego. Przy artykulacji spółgłosek przymkniętych ”j,l,ł” (approximants) powierzchnia przekroju poprzecznego zwężenia jest największa, natomiast przy spółgłoskach zwartych („p,t,k,b,d,g”) jest praktycznie równa zeru. Gwałtowne rozwarcie toru głosowego powoduje generację krótkiego impulsu szumowego.

Spółgłoski przymknięte
W tym przypadku zwężenie toru głosowego nie różni się w istotny sposób od zwężenia utworzonego dla samogłosek. Nie powoduje zaburzenia przepływu powietrza, dzięki czemu fałdy głosowe mogą swobodnie wykonywać ruchy drgające. Znamienne dla spółgłosek przymkniętych jest to, że zwężenie podczas ich artykulacji zmienia swoją wielkość. Można je wymówić tylko w sąsiedztwie samogłosek, stąd widoczne są często znaczne ruchy formantów. Obie komory przed i po zwężeniu uczestniczą w formowaniu dźwięku mowy.

Mechanizm powstawania turbulencji w szczelinie
Wypływ powietrza ze szczeliny przy osiągnięciu odpowiedniej prędkości przestaje być laminarny. Oddziaływanie ścian wskutek tarcia powoduje, że ruch cząsteczek w ich pobliżu jest bardziej hamowany, niż cząsteczki w środku strugi. Aby przepływ stał się turbulentny siły bezwładnościowe oddziaływujące na strugę przepływającego powietrza przekraczają siły wiążące ze sobą jego cząsteczek.

Warunki powstania turbulencji
Dla szczeliny określonych rozmiarów prędkość strugi powietrza musi przekroczyć pewną krytyczną wartość (określoną przez liczbę Reynoldsa), aby jej wypływ stał się turbulentny.

Liczba Reynoldsa

$$\mathrm{Re} = \frac{vh\rho}{\mu}$$

h-wymiar charakterystyczny (średnica)
m-współczynnik lepkości ośrodka

W przypadku przepływu powietrza przez cylindryczną rurę, liczba Reynoldsa zależy od gęstości ośrodka, rozmiarów przekroju rury, lepkości ośrodka i prędkości przepływu *v*. Dla rury przyjmuje się krytyczną wartość równą ~2300.

W przypadku przewężenia o powierzchni przekroju 0.6 cm2, i prędkości objętościowej przepływu 1000 cm3/s - Re=12000

Model równoważny (w układzie elektrycznym) źródła szumowego - szczelina
Lc=r*lc/Ac*, *lc* – długość szczeliny

$$R_c \approx \frac{k_c \rho V_c}{A_c^2}$$

k_c – współczynnik kształtu

Dla spółgłosek trących kc≈0.9
Funkcja transmitancji definiowana jako stosunek U0/Ps jest liniową funkcją powierzchni przekroju szczeliny Ac.

Miejsce artykulacji spółgłosek
Zwężenie toru głosowego przy artykulacji spółgłoskowej jest znacznie większe (może prowadzić nawet do chwilowego zamknięcia toru), niż w przypadku artykulacji samogłoskowej.
Tak więc w przypadku spółgłosek można mówić o miejscu artykulacji określającego np. położenia środka zwężenia lub miejsca chwilowego zamknięcia toru głosowego. Miejsce artykulacji ma wyraźny wpływ na strukturę akustyczną dźwięku mowy.
Źródło - filtr: spółgłoski trące
Widmo źródła szumowego jest formowane przez charakterystykę rezonansową przedniej komory znajdującą się między ustami i szczeliną. Na ogół wpływ tylnej komory jest pomijalnie mały, im mniejsza jest powierzchnia przekroju szczeliny, tym mniejszy jest jej wpływ.

Obwiednia widma spółgłosek trących
- Elementem formującym kształt widma spółgłosek trących jest komora utworzona z przodu szczeliny.
- Długość tej komory wyznacza najniższą jej częstotliwość rezonansową. Im jest dłuższa, tym ta częstotliwość jest mniejsza.

Klasyfikacja spółgłosek trących wg umiejscowienia zwężenia i/lub przeszkody w torze głosowym

Trące	/x/	/S/	/s'/	/s/	/f/
szczelina	głośnia	Palatalno-dziąsłowa	palatalna	dziąsłowa	Wargowo-zębowa
przeszkoda		dolne zęby	górne zęby	górne zęby	górne zęby
Przednia komora	Charakterystyka samogłoskowa	2-6 kHz	2-6 kHz	>4 kHz	b. mały wpływ

Źródło szumu dla głosek /S,s',s/ powstaje przede wszystkim na przeszkodzie i przy zachowaniu tej samej prędkości przepływu strugi powietrza ma największą energię w porównaniu z pozostałymi spółgłoskami trącymi (/x,f/).

Długość szczeliny
Szczelina przy artykulacji /s,s'/ jest stosunkowo krótka, dla /S/ - jest dłuższa.
Jeżeli długość przedniej komory jest bardzo mała, to jej najniższa częstotliwość rezonansowa jest tak wysoka, że jej udział w kształtowaniu widma dźwięku jest pomijalnie mały. Wówczas obwiednia widma promieniowanego dźwięku jest płaska. Tak jest np. w przypadku spółgłoski /f/.

Aerodynamika spółgłosek zwartych (wybuchowych)
- Tor głosowy podczas artykulacji tych głosek jest na chwilę zamknięty, a następnie szybko rozwarty.
- W pierwszej fazie następuje szybki wzrost ciśnienia ponadkrtaniowego i zamknięcie przepływu powietrza.
- W drugiej fazie – rozwarcie powoduje powstanie krótkiego impulsu szumowego.
- Źródło pobudzenia, podobnie jak w przypadku trących ma charakter turbulentny, ale czas pobudzenia jest znacznie krótszy (5-10 ms zamiast 100-200 ms).
- Szum jest formowany przez komorę utworzoną w torze głosowym z przodu, przed zwarciem.

Aspiracja
Niekiedy przy artykulacji spółgłosek zwartych, fałdy głosowe stosunkowo wolno przechodzą do pozycji, w której drgają. Powstaje przejściowa szczelina powodująca pojawienie się turbulencji.

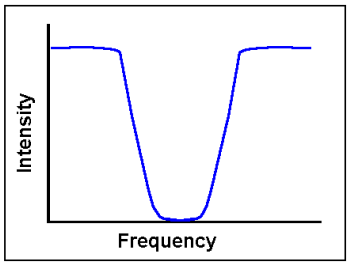
Spółgłoski zwarto-trące /ts, tS,ts'/
Już sama transkrypcja fonetyczna sygnalizuje, że artykulacja spółgłoski zwarto-trącej składa się z 2 faz: w pierwszej powstaje segment zwarcia (całkowite zamknięcie toru głosowego jak w przypadku głosek wybuchowych), w drugiej - utworzenie szczeliny (brak plozji), w wyniku czego zostaje wygenerowany krótki segment szumowy.

Udźwięcznianie spółgłosek
Uformowanie w torze głosowym szczeliny, czy nawet jego chwilowe zamknięcie nie musi spowodować zaprzestania ruchów fałdów głosowych. W języku polskim wszystkie spółgłoski bezdźwięczne (z wyjątkiem /x/) mają swoje dźwięczne odpowiedniki. Przy artykulacji spółgłosek bezdźwięcznych fałdy głosowe są rozwarte – przy dźwięcznych są do siebie zbliżone. Wówczas w formowaniu dźwięków mowy uczestniczą jednocześnie dwa źródła pobudzające różne części toru głosowego.

Analiza realizacji spółgłoski /r/
Koniuszek języka (apex) raz (najczęściej) lub dwa (niekiedy więcej) przywiera do wałka dziąsłowego. Zwarcie jest krótkotrwałe, na ogół niepełne. Realizacja tej spółgłoski silnie zależy od pozycji, kontekstu, często od nawyków osobniczych.

Artykulacja nosowa
Artykulacja nosowa powoduje opuszczenie podniebienia miękkiego i otwarcie wlotu do jamy nosowej. Od strony akustycznej powoduje to modyfikację charakterystyki przenoszenia toru głosowego. Przy artykulacji samogłosek nazalizowanych energia akustyczna jest promieniowana równolegle przez usta i nos. W przypadku samogłosek nosowych – przede wszystkim przez nos. Jednoczesne pobudzenie do drgań jamy ustnej i nosowej powoduje pojawienie się w charakterystyce toru tzw. antyformantów.

Antyformanty
W przeciwieństwie do samogłosek charakterystyka widmowa spółgłosek jest wyznaczona nie tylko przez formanty, ale również przez antyformanty.
Antyformant – przeciwieństwo formantu, charakterystyczne minimum w widmie dźwięku, tłumi składowe źródła w określonym zakresie częstotliwości.



Jakie elementy toru mogą powodują pojawianie się antyformantów
Częstotliwości antyformantów są określone przez wymiary tylnej komory i rozmiarów szczeliny (dla trących), wymiary komory ustnej ustnej (dla spółgłosek nosowych).

Kiedy mogą pojawiać się antyformanty ?
1) Gdy tor głosowy jest rozdzielony na dwie sprzężone ze sobą części np. w przypadku nazalizacji, czy artykulacji spółgłoski nosowej
2) Jama ustna zostaje rozdzielona na dwie równoległe do siebie części, jak to ma miejsce w przypadku artykulacji spółgłoski /l/
3) Szczelina przy artykulacji spółgłosek trących jest stosunkowo szeroka i występuje sprzężenie ze sobą tylnej i przedniej komory

Miejsce artykulacji spółgłosek – ruchy formantów
Ruchy formantów wskazują jakiego typu jest zmiana konfiguracji toru głosowego. Każdemu miejscu artykulacji spółgłoski odpowiadają odpowiednie ruchy formantów na przejściach od/do samogłoski. Największe ruchy formantów występują w pobliżu spółgłosek zwartych, najmniejsze dla przymkniętych.

Sposób artykulacji spółgłosek
1. Pobudzenie dźwięczne, bezdźwięczne, lub mieszane
2. Przepływ strugi powietrza zaburzony (szczelina, lub zwarcie lub ich kombinacja) lub nie
3. Konfiguracja toru głosowego stacjonarna lub nie w momencie artykulacji spółgłoski
4. Struktura jedno- lub polisegmentalna
5. Jama nosowa włączona lub nie

Wybrane cechy dystynktywne niektórych spółgłosek w płaszczyźnie miejsca artykulacji i typu pobudzenia

Cecha artyk.\głoska	b	d	g	p	t	k	s	z	m	n
wargowa	+	-	-	+	-	-	-	-	+	-
zębowe	-	+	-	-	+	-	+	+	-	+
tylno-językowa	-	-	+	-	-	+	-	-	-	-
pobudzenie krtaniowe	+	+	+	-	-	-	-	+	+	+

Efekty akustyczne spółgłoskowych ruchów artykulacyjnych
Artykulacji spółgłosek towarzyszą ruchy formantów spowodowane zmianami konfiguracji toru głosowego.
Gdy powstaje znaczne przewężenie w torze głosowym pojawia się źródło pobudzenia szumowego.
Chwilowemu zamknięciu toru głosowego towarzyszy niemal całkowity zanik sygnału (jeżeli wlot do jamy nosowej jest zamknięty), po którym może wystąpić pobudzenie impulsowe (głoski zwarte), bądź krótki segment pobudzenia szumowego (głoski zwarto-trące).

Cechy akustyczne dźwięków mowy
Akustyczny sygnał mowy niesie informacje umożliwiające rozpoznanie poszczególnych głosek wypowiedzianych w określonej sekwencji. Te elementy sygnału, które umożliwiają rozróżnienie lub identyfikację nazywamy cechami akustycznymi – obejmują one częstotliwości formantów, ich tranzjenty, widma plozji spółgłosek zwartych, widma szumu spółgłosek trących, obecność zwarcia – b. mała amplituda sygnału itp.

Cechy akustyczne sposobu artykulacji

Fazy wypowiedzi ustnej
Mowa jest procesem, podczas którego narządy artykulacyjne w sposób płynny następują przejścia między głoskami. Każda fraza (ograniczona obustronnie pauzami) stanowi pewną zorganizowaną całość, co przejawia się zarówno w jej strukturze segmentalnej (głoskowej i sylabicznej), jak i jej rozczłonowaniu rytmicznym i melodycznym.
Położenie głoski we frazie może wpływać na jej wymowę, bądź na jej ubezdźwięcznienie/udźwięcznienie

Charakterystyka wygłosu
W wygłosie wypowiedzi ruchy narządów mowy są wykonywane znacznie mniej dokładnie, z mniejszym nakładem energii, a także wolniej niż w nagłosie i śródgłosie. Przejawia się to przede wszystkim w:
- osłabianiu wygłosowych zwarć,
- w redukcji głosek otwartych,
- zmniejszaniu się (z wyjątkiem fraz pytających) częstotliwości F0,
- słabość wygłosu powoduje często ubezdźwięcznianie zwarto-wybuchowych, zwarto-trących i trących, a często i całej następującej po nich samogłoski.

Koartykulacja – jej źródło
- Ruchy artykulacyjne niezbędne do wypowiedzenia określonej głoski często uruchamiają tylko jeden (dwa) elementy układu artykulacyjnego, np. wargi, czubek języka itp. Np. przy artykulacji spółgłosek wargowych język ma swobodę do przyjęcia konfiguracji odpowiadającej następującej samogłosce.
- Innym czynnikiem jest tzw. ekonomizacja ruchów artykulacyjnych.
- Koartykulacja jest sprawnością wyuczoną. U małych dzieci jest znacznie słabsza.
- Koartykulacja jest czynnikiem, niekiedy bardzo silnie modyfikującym strukturę dźwiękową głosek

Przykład oddziaływania głosek na siebie - ubezdźwięcznianie
Sąsiadujące ze sobą dźwięki mowy w łańcuchu mowy wzajemnie na siebie mniej lub bardziej oddziaływują modyfikując artykulację głoski następującej lub poprzedzającej. Modyfikacja ta może pociągać za sobą zmianę typu głoski, zwłaszcza może to mieć miejsce na granicach między wyrazowych. Np. „wóz stoi” wymawia się „wus stoi”, choć w sekwencji wyrazów „wóz zatrzymał się” pierwszy wyraz jest wymawiany „wuz”.

Zalety koartykulacji
Informacja w segmencie odpowiadającym danej głosce jest nie tylko o głosce wymówionej, ale również o sąsiadujących z nią, np. dla sylaby /su/ w spółgłosce /s/ możemy ocenić jaka następuje po niej samogłoska.
Zjawisko to umożliwia rozumienie b. szybkiej mowy.

Wady koartykulacji z punktu widzenia analizy mowy
Brak wyraźnych, niezmiennych akustycznych „punktów" charakteryzujących daną głoskę. Ten sam fonem /s/ może zmienić się w inny. Por. „su" i „si". Również i w płaszczyźnie akustycznej ten sam dźwięk mowy może być interpretowany jako realizacja różnych fonemów, zależnie od kontekstu.

Uniwersalność koartykulacji
Cechy artykulacji, które nie są charakterystyczne dla danego języka, wynikają bowiem z ogólnych anatomicznych i fizjologicznych właściwości narządu mowy, mają charakter uniwersalny. Z tego powodu wartości parametrów fonetyczno-akustycznych (np. częstotliwości formantowe) nie są stałe w obrębie poszczególnych segmentów. Ta zmienność jest spowodowana przede wszystkim bezwładnością narządów artykulacyjnych. Nie mogą one w sposób skokowy zmieniać swojej konfiguracji z typowej dla jednej głoski na drugą konfiguracją, następującą przy kolejnej głosce.

Czynniki modyfikujące głoskę danej klasy
- Przypadkowe (dla tej samej osoby)
- Indywidualne zróżnicowania międzyosobnicze
- Zróżnicowania kontekstowe - koartykulacja
Istnieje naturalna tendencja do „ekonomizacji" ruchów artykulacyjnych, w wyniku czego granice między głoskami stają się mniej wyraźne, „przenikając" jedna w drugą. Stąd, każda głoska w mniejszym lub większym stopniu posiada niektóre cechy głoski poprzedzającej i następującej

Definicja koartykulacji
Koartykulacja jest zjawiskiem, podczas którego następuje nakładanie się ruchów artykulacyjnych właściwych dla sąsiadujących ze sobą głosek.

Rodzaje koartykulacji
- Antycypacja i przedłużenie
- Upodobnienia i uproszczenia w obrębie wyrazu
  Upodobnienia pod względem dźwięczności
  pod względem miejsca artykulacji
  pod względem stopnia zbliżenia narządów mowy
- Międzywyrazowe upodobnienia – na granicy wyrazów

Przykłady antycypacji
1) Zaokrąglenie warg typowe dla samogłoski /u/ może przenosić się na sąsiadujące z nią głoski, np. lukier.
2) Podobnie, jeśli nie ma sprzeczności w ruchach artykulacyjnych, układ masy języka typowy dla danej głoski może być już przygotowany podczas wymawiania głoski poprzedzającej, np. w fazie zwarcia por „tupać".
3) Podtrzymywanie (przedłużenie) np. bezdźwięczności:
„twardy" -> /tvardI/ -> /tfardI/

Przykład upodobnienia
Koartykulacja prowadzi do częściowego (niekiedy całkowitego) zacierania się różnic pomiędzy sąsiadującymi ze sobą dźwiękami i tym samym do tzw. upodobnień. Powodują one zmianę ich postaci dźwiękowej.
Upodobnienia obejmujące grupy głosek i połączone z redukcją (częściową, lub całkowitą) pewnych dźwięków tworzących te grupy nazywane są „uproszczeniami".
Np. „sześćset" -> /Ses'ts'set/-> /Ses'set/
Uproszczenia prowadzą niekiedy do „podstawień"
np. /Sejset/.
- Upodobnienia pod względem dźwięczności
Upodobnienie pod względem dźwięczności polega na zniesieniu różnicy między sąsiadującymi ze sobą głoskami: dźwięczną i bezdźwięczną. Np. „twarz" -> /tfaS/
- Upodobnienia pod względem miejsca artykulacji
Polegają na takim przesunięciu miejsca zwarcia lub szczeliny, by było ono takie same jak miejsce zwarcia lub szczeliny głoski sąsiedniej. Np. „ssie" -> /ss'e/->/s's'e/
- Upodobnienia pod względem zbliżenia
Np. „uszczelinowienie" głoski sąsiadującej w wyrazie „trzeba" -> /t_Seba/ -> /tSSeba/, „trzy" -> /tSSI/

Upodobnienia międzywyrazowe
- Na granicach form wyrazowych następują upodobnienia przede wszystkim pod względem dźwięczności.
- W wygłosie tzw. absolutnym (przed pauzą o dostatecznej długości) wszystkie spółgłoski dźwięczne z klas zwartych, zwarto-trących i trących są ubezdźwięczniane, ale jeżeli wyraz następny zaczyna się od spółgłoski dźwięcznej należącej do jednej z tych klas, wówczas końcowa spółgłoska poprzedniego wyrazu jest dźwięczna. W pozostałych przypadkach zachodzi ubezdźwięcznianie.

Segmentacja i koartykulacja
Ponieważ koartykulacja jest w sygnale mowy wszechobecna, trudno oczekiwać, by granice segmentów były zawsze jednoznaczne.
Z drugiej strony, jeżeli nie jesteśmy w stanie dokładnie określić w sygnale mowy początku i końca segmentów, to obszary nakładania się ruchów artykulacyjnych są wyznaczane jedynie w przybliżony sposób.

Pragmatyczna definicja granic segmentalnych sygnału mowy
Przyjmuje się założenie, że każda głoska jest reprezentowana przez quasistacjonarne widmo odpowiadające niezmiennej w czasie funkcji transmitancji toru głosowego, z wyjątkiem tych głosek charakteryzujących się przebiegiem tranzjentowym. Przebiegi te mogą dotyczyć zmian w funkcji źródła, bądź funkcji transmitancji toru głosowego. W pierwszym przypadku, gwałtowne zmiany częstotliwości podstawowej, zaś w drugim –zmiany w funkcji transmitancji wywołany szybkimi zmianami konfiguracji narządów artykulacyjnych, mogą być wykorzystane do określania granic segmentów.

Koartykulacja – podsumowanie

1) Koartykulacja jest wynikiem nakładania się ruchów artykulacyjnych
2) Elementy narządu artykulacyjnego o małej szybkości są bardziej podatne na efekt nakładania się
3) Między głoskami nie ma na ogół jednoznacznych, wyraźnych granic (z wyjątkiem pauz)
4) Mowa jest rozpoznawana w oparciu o obrazy akustyczne sylab
5) Koartykulacja jest najsilniejsza w obrębie sylaby
6) Samogłoski wpływają na artykulację sąsiedniej spółgłoski (również samogłoski)
7) Spółgłoski również wpływają na artykulację sąsiedniej samogłoski
8) Pewne dźwięki mowy są bardziej odporne na wpływ koartykulacji, inne mniej
9) Im większy jest konieczny ruch artykulacyjny przy przejściu z jednej głoski do następnej, tym większa jest koartykulacja
10) Samogłoski niskie są bardziej podatne na koartykulację w sąsiedztwie spółgłosek, niż samogłoski wysokie
11) Koartykulacja jest ograniczana w przypadku, gdy może powodować niejednoznaczną percepcję

Kod SAMPA

W transkrypcji fonetycznej tekstów ortograficznych stosowany jest kod SAMPA. Wersja polska:
http://www.phon.ucl.ac.uk/home/sampa/polish.htm
Umożliwia on bezpośrednie stosowanie w transkrypcji klawiatury QWERTY.

The vowel system comprises 8 phonemes, as follows. Those symbolized with ~ are nasalized.

SAMPA symbol	Orthography	Transcription	IPA
i	PIT	pit	pit
I	typ	tIp	tɨp or tɪp
e	test	test	test
a	pat	pat	pat
o	pot	pot	pot
u	puk	puk	puk
e~	gęś	ge~s'	gɛ̃ɕ or gɛ̃jɕ
o~	wąs	vo~s	võs or vow̃s

Consonants

The consonant system comprises 29 phonemes, as follows. The symbol ' indicates palatalization.

p	pik	pik
b	bit	bit
t	test	test
d	dym	dIm
k	kit	kit
g	gen	gen
f	fan	fan
v	wilk	vilk
s	syk	sIk
z	zbir	zbir
S	szyk	SIk
Z	żyto	ZIto
s'	świt	s'vit
z'	źle	z'le
x	hymn	xImn
ts	cyk	tsIk
dz	dzwon	dzvon
tS	czyn	tSIn
dZ	dżem	dZem
ts'	ćma	ts'ma
dz'	dźwig	dz'vik
m	mysz	mIS
n	nasz	naS
n'	koń	kon'
N	pęk	peNk
l	luk	luk
r	ryk	rIk
w	łyk	wIk
j	jak	jak

Tekst ortograficzny i jego transkrypcja fonetyczna

Fonem a litera

Te same znaki ortograficzne lub jednakowe ich sekwencje mogą odpowiadać różnym dźwiękom mowy: np. „wór” – /vur/, „wtórny” – /fturnI/
„marznąć” – /marznon’ts’/, „marzec” - /maZets/

Różne znaki ortograficzne mogą odpowiadać tym samym dźwiękom mowy
np. „auto” – /awto/, „dał” – /daw/

Różne sekwencje:
„dźwiga” – /dz’viga/, „dzień” – /dz’en’/

W transkrypcji fonetycznej uwzględnia się zjawisko koartykulacji !

Podstawowe reguły uproszczonej transkrypcji fonetycznej

- Literom samogłoskowym „y,e,a,o” odpowiadają fonemy /I,e,a,o/. Litery „u” i „ó” nie sygnalizują różnic w wymowie.
- Literę „i’ przed literą spółgłoskową wymawia się jako samogłoskę /i/
- Literę „i’ przed samogłoską wymawia się jako:
    - /j/ po zwartych, nosowej /m/, trących /f,v,x/, i głoskach /l,r/

/i/ na końcu wyrazu
    - podwójne „ii” po zwartych, nosowej /m/, trących /f,v/, głoskach /l,r/ i literze „ch” wymawia się jako /ji/
- Następujące grupy spółgłoska-samogłoska /i/ odpowiadają następującym fonemom:
- „si” – /s’/ „ci” - /ts’/
- „zi” – /z’/ „dzi” - /dz’/
    - „ni” - /n’/ wyjątek „Dania” –/dan’ja/, ale /dan’a/
- Samogłoski nosowe „ę,ą” wymawia się jako
    - /e~,o~/ na końcu wyrazu
- /em,om/ przed /p,b/
- /en,on/ przed /t,d,ts,tS,dz,dZ/
- /en’,on’/ przed /ts’,dz’/
- /eN,oN/ przed /k,g/
    - /e,o/ przed /l,w/ „wziąłem” – w czasie przeszłym
- Głoski zwarte (/b,d,g/), zwarto-trące (/dz,dz’,dZ/) i trące (/v,z,z’,Z/) wymówione przed głoskami bezdźwięcznymi, przerwą(w wygłosie) stają się bezdźwięcznymi i ich wymowa jest dokładna, jak ich bezdźwięcznych odpowiedników, tj. /p,t,k/, /ts,ts’,tS/ czy /f,s,s’,S/. To samo występuje u zbiegu wyrazów wymówionych bez przerwy
- O ubezdźwięcznieniu lub udźwięcznieniu całej sekwencji powyższych spółgłosek o różnym typie pobudzenia decyduje w zasadzie ostatnia w sekwencji głoska – np. „liczba” - /lidZba/, „rzadszy” - /Zat_SI/
- Od powyższej zasady jest wyjątek, gdy przed literą „w” lub sekwencją „rz” stoi głoska bezdźwięczna. Cała sekwencja staje się bezdźwięczna. np. „kwiat” – /kfjat/, „szwaczka” - /SfatSka/
- Nieregularności w wymowie „trz”, „drz”, „dż”, „dz” w obrębie wyrazu np. „trzech” - /tSSex/, ale „Czech” - /tSex/, „wodze” – /vodze/, „odzew” – /od_zef/
- Spółgłoski bezdźwięczne przed końcówką czasownikową „–my” pozostają bezdźwięczne np. „kupmy” - /kupmy/
- Grupy spółgłoskowe złożone ze spółgłosek zwartych, zwarto-trących i trących, które są wymówione w nagłosie lub śródgłosie form wyrazowych, są całkowicie dźwięczne lub bezdźwięczne – /fskotSIts’/, krufka/, /proz’ba/.
- Grupy mieszane – powyższe spółgłoski nie zmieniają dźwięczności spółgłosek przymkniętych - /kulka/, /puwka/, /krova/, zamknon’ts’/

Jednakże spółgłoski przymknięte wymówione w środku dłuższych sekwencji spółgłoskowych są najczęściej bezdźwięczne i wymawiane tak słabo, że często ulegają całkowitej redukcji – „jabłko” - /japko/, „rzemieślnik” - /Zemjes’n’ik/

Przykład transkrypcji fonetycznej (SAMPA) – mowa syntetyczna

Konwersja tekstu na mowę otwiera nowe możliwości, niedostępne w tradycyjnych systemach głosowych. Usługi katalogowe, informatory turystyczne, tematyczne serwisy informacyjne, czy portale głosowe, to tylko nieliczne zastosowania tej technologii.

Cechy prozodyczne mowy

Dotychczas przedmiotem naszych rozważań był opis dźwięków mowy (fonemów) języka polskiego, a więc jednostek, które są opisywane w płaszczyźnie artykulacyjnej, bądź akustycznej. Opis ten umożliwia nadanie z natury swej ciągłemu sygnałowi mowy struktury dyskretnej, przedstawianej w postaci sekwencji fonemów, głosek, sylab, wyrazów itp. Sekwencja ta jest wypowiadana, z określonym tempem (prędkością), rytmem, głośnością i melodią.

Cechy segmentalne vs. cechy suprasegmentalne mowy

Podział na segmenty – głoski, difony, sylaby, wyrazy, itp.
Cechy opisujące sekwencje (ciągi) segmentów – zmiany melodii (F0), intensywności, tempo wypowiedzi, rytm, akcenty, itp.

Cechy prozodyczne w automatycznym rozumieniu mowy

- ❑ Informacje prozodyczne są b. rzadko wykorzystywane w systemach rozumienia mowy
- ❑ Analiza prozodyczna może wspomagać wiele zadań :
    - ❖ automatyczna interpunkcja
    - ❖ rozpoznawanie wyrazów (np. zaimek pytajny – zaimek względny: *„czyj kapelusz? – powiedział czyj kapelusz nosi”*)
    - ❖ segmentacja składniowa wypowiedzi

Czynniki wpływające na czas i tempo wypowiedzi

<u>Iloczas</u> (czas trwania dźwięków mowy, a zwłaszcza samogłosek), sylab, wyrazów itp.
Parametry charakteryzujące tempo wypowiedzi – np. średni stosunek iloczasu dźwięków niesamogłoskowych/samogłoskowych,
Liczba samogłosek na jednostkę czasu

<u>Pauzy</u> (o czasie trwania większym od czasu trwania zwarć)
Średni czas trwania - średnia liczba pauz w obrębie wypowiedzi, wyznaczanych dla różnych progowych poziomów; średni czas trwania fraz do czasu wypowiedzi

Korelaty cech suprasegmentalnych sygnału mowy

Cechy suprasegmentalne sygnału mowy w płaszczyźnie percepcyjnej są następujące:

a) wysokość głosu
b) głośność
c) tempo, rytm, akcenty

Akustyczne korelaty cech suprasegmentalnych:

a) częstotliwość pobudzenia krtaniowego (wysokość)
b) poziom intensywności sygnału (głośność)
c) iloczas (długość segmentu)

Cechy suprasegmentalne kształtują prozodyczną strukturę języka – melodię, akcent i rytm

Relacje w płaszczyźnie percepcyjnej między wysokością, głośnością i długością (iloczasem)
Wrażenie wysokości głosu zależy głównie od częstotliwości drgań fałdów głosowych, jednakże pewien wpływ na percepcję wysokości mają również intensywność, jak i czas trwania danego segmentu.
W pierwszym przypadku, przy zwiększaniu poziomu sygnału o stałej częstotliwości towarzyszy wrażenie obniżania się jego wysokości, przy zmniejszaniu – podnoszenie się wysokości.

Rola iloczasu w percepcji wysokości
Minimalna długość segmentu, przy średnim poziomie natężenia, poniżej której nie można orzec, który z dwóch porównywanych ze sobą sygnałów jest wyższy lub niższy, wynosi nie mniej niż 60 ms (dla F0 ≈ 70 Hz). Zaś dla wyższych częstotliwości czas ten jest nieco krótszy.
Subiektywne względne różnice długości segmentów wypowiedzi, są określane na podstawie oceny iloczasu (np. głoska długa,
krótka itp.)

Barwa segmentów
Ze zmianami głośności i wysokości skorelowane są w sygnale mowy zmiany barwy, określone przede wszystkim przez sposób i miejsce artykulacji. Te dwa ostatnie czynniki decydują o postaci widma artykułowanego dźwięku. Jednakże modyfikacja głośności i wysokości może spowodować zmianę odczuwanej barwy głoski w kierunku jaśniejszej, bądź ciemniejszej, nie zmieniając przy tym znaczenia segmentu.

Rola cech prozodycznych w percepcji mowy
W percepcji łańcucha segmentów pierwszym poziomem analizy jest ich uporządkowanie według kryteriów stosowanych przy różnicowaniu wszelkiego typu dźwięków – więc segmenty długie - krótkie, głośne – ciche, wysokie – niskie, szumowe (bezdźwięczne) – dźwięczne, rozkład akcentów itp. Czynimy to również przy osłuchiwaniu się z językiem, którego zupełnie nie znamy.

Typy wypowiedzi rozróżnianych na podstawie intonacji
- pytania o rozstrzygnięcie (yes-no questions)
- pierwszy składnik wypowiedzi oznajmujących z uzupełnieniem
- końcowy składnik (uzupełnienie) wypowiedzi oznajmujących
- wypowiedzi oznajmujące
- wypowiedzi wykrzyknikowe (z podniesionym głosem)

Różnice w głośności głosek
Wśród czynników decydujących o dominacji danego segmentu w określonym łańcuchu głosek należy wymienić dźwięczność i głośność. Ta ostatnia jest proporcjonalna do stopnia otwarcia jamy ustnej. Najbardziej donośna spośród głosek języka polskiego (i nie tylko) jest samogłoska /a/, a następnie za nią idą pozostałe samogłoski wg stopnia otwarcia jamy ustnej /e,o,I,u,i/. Spółgłoski układają się w przybliżeniu w następującej kolejności:
Dźwięczne: /j,l,w/, nosowe, /r/, trące i zwarto-trące
Bezdźwięczne: trące (bez /f,x/), zwarto-trące i trące /f,x/)

Sylaby fonetyczne
Zmiany głośności między kolejnymi głoskami w strumieniu dźwięków mowy warunkują podział wypowiedzi na tzw. sylaby fonetyczne. Rdzeniem (ośrodkiem) sylaby fonetycznej jest segment głoskowy różniący się poziomem głośności od swego najbliższego otoczenia. Jego głośność jest niemal zawsze większa od głośności głoski występującej bezpośrednio przed nim i po nim.

Struktura sylabiczna wypowiedzi
Sylaba nie stanowi elementu funkcjonalnego jakim jest głoska. Jej jedyną funkcją jest segmentacja wypowiedzi, ułatwiająca artykulację i percepcję. Segmentacja ta dokonuje się poprzez rytmizację ciągu segmentów, spowodowaną podziałem tego ciągu na skutek chwilowych obniżeń poziomu emitowanego sygnału mowy.

Obniżenia te są wywoływane przez zwarcia, bądź szczeliny będącymi źródłem pobudzenia szumowego o niskim poziomie. Ośrodkami sylab są głoski o najwyższym poziomie (na ogół są to samogłoski).

Sylaby fonetyczne
Zmiany głośności między kolejnymi głoskami w strumieniu dźwięków mowy warunkują podział wypowiedzi na tzw. sylaby fonetyczne. Rdzeniem (ośrodkiem) sylaby fonetycznej jest segment głoskowy różniący się poziomem głośności od swego najbliższego otoczenia. Jego głośność jest niemal zawsze większa od głośności głoski występującej bezpośrednio przed nim i po nim.

Struktura sylabiczna wypowiedzi
Sylaba nie stanowi elementu funkcjonalnego jakim jest głoska. Jej jedyną funkcją jest segmentacja wypowiedzi, ułatwiająca artykulację i percepcję. Segmentacja ta dokonuje się poprzez rytmizację ciągu segmentów, spowodowaną podziałem tego ciągu na skutek chwilowych obniżeń poziomu emitowanego sygnału mowy.
Obniżenia te są wywoływane przez zwarcia, bądź szczeliny będącymi źródłem pobudzenia szumowego o niskim poziomie. Ośrodkami sylab są głoski o najwyższym poziomie (na ogół są to samogłoski).

Akcent wyrazowy
Definicja akcentu: Jest to to proces uwydatniający wybrane segmenty w sygnale mowy ciągłej, np. sylab w wyrazach lub wyrazów w zdaniach.
Uwydatnienie sylaby akcentowanej może polegać na silniejszym, a zarazem głośniejszym jej wypowiedzeniu, na bardziej precyzyjnym jej wymówieniu, co może spowodować jej wydłużenie czasu trwania.
Może też wystąpić tylko podwyższenie (niekiedy obniżenie) częstotliwości pobudzenia krtaniowego.

Akcent dynamiczny, rytmiczny i melodyczny

W zależności od tego, który z tych czynników przeważa, akcent jest określany jako:
dynamiczny – gdy czynnikiem dominującym w płaszczyźnie akustycznej są zmiany intensywności
rytmiczny – gdy o wrażeniu akcentu decydują zmiany iloczasów sylab, lub
melodyczny – gdy akcentowanie sylaby jest realizowane poprzez zmianę wysokości głosu
Dla języka polskiego przyjmuje się , że akcent jest zazwyczaj dynamiczny, choć jest to dyskusyjne.

Akcent wyrazowy
Definicja akcentu: Jest to to proces uwydatniający wybrane segmenty w sygnale mowy ciągłej, np. sylab w wyrazach lub wyrazów w zdaniach.
Uwydatnienie sylaby akcentowanej może polegać na silniejszym, a zarazem głośniejszym jej wypowiedzeniu, na bardziej precyzyjnym jej wymówieniu, co może spowodować jej wydłużenie czasu trwania.
Może też wystąpić tylko podwyższenie (niekiedy obniżenie) częstotliwości pobudzenia krtaniowego.

Położenie akcentu
Przyjmuje się, że w języku polskim akcent wyrazowy jest stały i spoczywa w zasadzie na przedostatniej sylabie formy wyrazowej. Są formy wyrazowe nie mające samodzielnego akcentu np. „się", „ci", „za", „mnie" itp. i dołączają się do wyrazu mającego swój akcent – np. „pod_lasem".
Akcent wyrazów zapożyczonych jest na ogół na 3-ej sylabie od końca – „logika". To samo może wystąpić w niektórych formach czasownikowych – „widzieliśmy".
Dłuższe formy wyrazowe obok akcentu na sylabie przedostatniej mają także akcent na pierwszej sylabie (akcent główny) – „prawdopodobnie" (o tym zadecydowały względy rytmiczne i melodyczne)

Realizacja akcentu w płaszczyźnie akustycznej
W zależności od języka mówca posługuje się jednym z akcentów jako dominującym dla danego języka.
W przykładzie dla języka angielskiego (z dominującym akcentem melodycznym), mówca niekiedy dodaje również akcent dynamiczny, a niekiedy obserwuje się wydłużenie sylaby, by uzyskać na niej słyszalne podniesienie melodii.

Funkcje melodii (intonacji) mowy
W języku polskim zmiany wysokości tonu krtaniowego, charakteryzują wraz z rozłożeniem akcentów, tempem wypowiedzi itp. dłuższe niż głoska odcinki wypowiedzi.
Zmiany F0 są nosicielami informacji o rozczłonowaniu składniowym tej wypowiedzi, o tym które jej fragmenty są szczególnie ważne, sygnalizują też koniec całej wypowiedzi, lub któregoś z jej członów.

Wzmocnienie sylaby
Wzmacnianie danej sylaby często odbywa się poprzez podniesienie częstotliwości F0 (w przykładzie na „O!", czy „Jak to..."). Takie uwydatnianie nazywa się akcentem logicznym (zdaniowym). Na ogół, wymaga to ponadto zwiększenia iloczasu uwydatnianej sylaby.
Obniżenie melodii jest zazwyczaj w wypowiedziach stanowiących zamkniętą całość. Podobnie jest w pozbawionych emocji poleceniach i rozkazach. Na końcu tych odcinków wypowiedzi, które wyodrębniają się, ale nie stanowią jeszcze zamkniętej całości, a więc takich, po których ma nastąpić ciąg dalszy melodia się wznosi. Podobnie melodia wznosi się na końcu zdania pytającego.

Rola cech prozodycznych w mowie
- porządkują i organizują strukturę czasową wypowiedzi
- są nosicielami informacji o jej podziale składniowym
- sygnalizują gramatyczną funkcję wypowiedzi (przede wszystkim melodia jest nosicielem tej informacji)
- sygnalizują stan emocjonalny

Muzyczna notacja dla mowy ?
- W dobie precyzyjnych pomiarów częstotliwości, czy ma jeszcze sens ?
- W muzyce podstawowym pojęciem jest interwał – różnica wysokości dwóch dźwięków wyrażona w jednostce miary, której podstawą jest oktawa i półton
- Muzyczny interwał jest muzyczną odległością między dźwiękami o różnej wysokości – ma bezpośredni związek z percepcją wysokości.
- Interwały są związane z częstotliwością, ale nie są identyczne (w różnych oktawach te same interwały są w skali częstotliwości różne)
- Tony 220 Hz i 440 Hz są muzycznie identyczne

Mowa a muzyka
Muzyczne interwały nie zależą od zakresu
– Oktawa może być dzielona muzycznie na wiele sposobów
– Melodia może wykorzystywać tylko jakąś część muzycznej przestrzeni dźwiękowej
– Mowa rozciąga lub zmniejsza całą przestrzeń dźwiękową. W zmienionej przestrzeni nadal dźwięk Wysoki pozostaje nadal Wysoki bez względu na to, czy przestrzeń ta została rozciągnięta, czy pomniejszona. W muzyce pomniejszony interwał jest różny od rozciągniętego
– Innymi słowy, muzyczna tonalność zmienia się w obrębie przestrzeni tonicznej, natomiast mowa tę przestrzeń sobie niemal dowolnie kształtuje

Nieadekwatność notacji muzycznej mowy
- Notacja nutowa sugeruje, że mowa jest „muzyczna".
- Muzyczna notacja może być myląca, sugerując strukturę tonalną melodii mowy, o czym nie ma przekonywujących danych.
- Jednakże badania neurologiczne wskazują na związek między percepcją konturu melodycznego i intonacją, ale nie między intonacją (w sensie lingwistycznym) i muzyczną tonalnością.

SOLA-Synchronized Overlap and Add

- Przetwarzanie segmentów czasowych

- Segmentacja na ciągi x[n] w zachodzących na siebie ramkach
  - Przesunięcie segmentów odpowiednio do wielkości współczynnika skalującego α
  - Wzajemne ustawienie, przedział nakładania/sumowania,
  - Obliczenie korelacji wzajemnej w przedziale nakładania się
  - Tak przesunąć względem siebie segmenty, by w tym przedziale współczynnik korelacji wzajemnej był maksymalny
  - wzmacnianie/tłumienie j.w.
  - Dowolne przesunięcie czasowe

Synteza sygnału mowy

- Skalowanie czasowe:
  - Skalowane segmenty muszą być dodane lub usunięte bez zmiany odległości między sąsiednimi impulsami krtaniowymi
    - Zmiana F0:
      - Po syntezie czas trwania segmentu nie ulega zmianie, natomiast konieczne jest przeskalowanie lokalnego okresu tonu krtaniowego
    - Segmenty mogą być pomijane (kompresja/obniżenie wysokości głosu)
    - Segmenty mogą być podwojone (rozciągnięcie/zwiększenie wysokości)
    - Artefakty:
      - „rozmazywanie tranzjentów", słyszalne „cięcia", zniekształcenia błędami fazowymi

Uniwersalizm niektórych sposobów wyrażania stanów emocjonalnych

Ekman wykazał , że niektóre stany emocjonalne są wyrażane w sposób niezależny od środowiska kulturowego:

- radość
- smutek
- złość, gniew
- strach, obawa
- odraza, wstręt (dla niektórych środowisk)
- zdziwienie, zaskoczenie (dla niektórych środowisk)

Pozostałe są kulturowo zmienne, w tym i „obojętność"

Multimodalna analiza twarzy

Oparta jest na analizie:

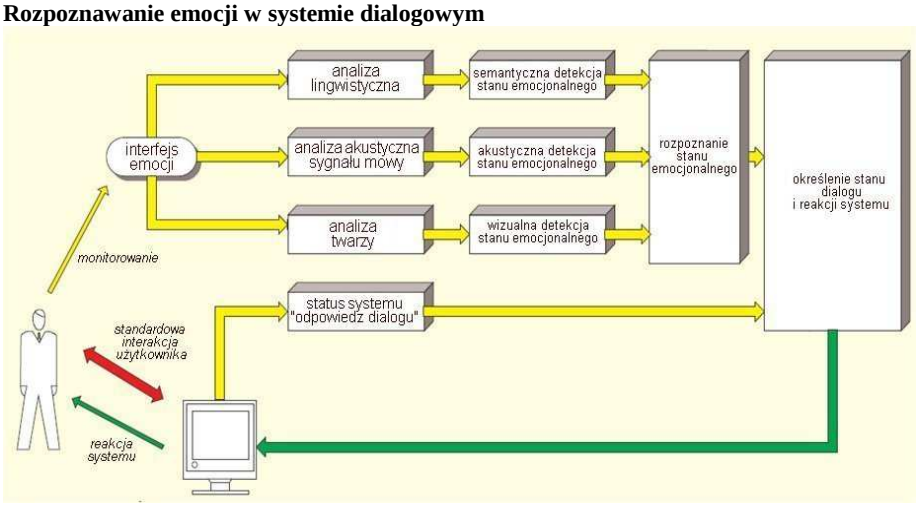
- Informacji o kolorze skóry
- Cechy elipsoidalne głowy
- Gradient luminancji/chrominancji
- Wstępny podział obszarów twarzy
- Określenie cech wyrazu twarzy
- Analiza sygnałów mikrofonowych
- …

Multimodalne środki emocji i jej rozpoznawanie

Obiekt analizy i rozpoznawania: twarz (wyraz, mimika) + mowa (głos, treść)

- Rozpoznawanie emocji -> systemy inteligentne (nadmiarowość, niepewność, niespójność informacji)
- Modelowanie emocji -> synteza emocji
- Interakcja -> rzeczywiste emocje -> baza danych

Rozpoznawanie emocji w systemie dialogowym



Etapy multimodalnej analizy i syntezy emocji

- Multimodalna analiza twarzy mówiącej osoby (tzw. Face Tracking)
- Ekstrakcja cech mimiki twarzy
- Ekstrakcja cech głosu
- Multimodalne rozpoznawanie emocji
- Multimodalna synteza emocji

Określenie cech wyrazu twarzy

Detekcja i śledzenie zmian cech

- Lokalizacja : w procesie uczenia i/lub poprzez heurystykę
- Ekstrakcja: wykorzystanie wiedzy *a priori*
- Informacje dotyczące kształtu/konturu
- Chwilowe zmarszczki
- …

Funkcje emocjonalne cech prozodycznych
Słuchacz na ogół kontroluje w wypowiedzi swój stan emocjonalny. W jego wyrażeniu posługuje się przede wszystkim tempem mówienia, głośnością, wprowadzaniem dodatkowych pauz, przedłużaniem niektórych dźwięków, a także modulowaniem melodii. W wypowiedziach nacechowanych emocjonalnie wahania melodii są znacznie większe, niż w wypowiedziach o charakterze neutralnym. Neutralne – 3-4 tony, z dużym ładunkiem emocjonalnym - > 1 oktawy.

Trudności w określaniu emocji
Nadanie wypowiedzi określonego typu emocji jest zadaniem bardzo złożonym. Osoby określające typ wypowiedzi pod względem emocji rzadko są zgodne w swych ocenach, z wyjątkiem krańcowych, lub wyraźnie kontrastowych typów emocji Słuchacze w swojej ocenie głównie opierają się na cechach prozodycznych, zwłaszcza na iloczasach i stylizowanym przebiegu F0.

Cechy emocji w sygnale mowy
- Prozodia nie uwzględnia jakości głosu, która może również nieść informację o stanie emocjonalnym osoby mówiącej (chrypka, krzyk, szept itp.) czy stylu mówienia (hyperartykulacja, wstawianie wydłużonych pauz…)
- Wydaje się, że cechy akustyczne emocji mogą być specyficzne dla języka
- Trudności w jednoznacznym określaniu emocji w sygnale mowy – często niesie równolegle szereg emocji jednocześnie, o podobnym charakterze

Emocje kontrastowe w płaszczyźnie akustycznej
<u>Strach/złość</u>
- zwiększona prędkość i głośność wypowiedzi
- podwyższone F0
- zwiększony zakres F0
- zaburzony rytm mowy
- dokładniejsza artykulacja
- zwiększona energia w zakresie wyższych częstotliwości

<u>Smutek/odprężenie</u>
- zmniejszona prędkość i głośność wypowiedzi
- obniżone F0
- zmniejszony zakres F0
- wyrównany rytm mowy, płynna mowa
- niedokładna artykulacja
- obniżona energia w zakresie wyższych częstotliwości

Miary akustyczne emocji
F0: zakres zmian, wartość średnia, nachylenie konturu (w górę/w dół), kształt konturu na sylabach akcentowanych
Struktura harmoniczna sygnału: udział szumów przydechowych, laryngalizacja (zwężone impulsy krtaniowe, duża zmienność okresu tonu krtaniowego)
Jasność brzmienia: stosunek energii w górnym zakresie częstotliwości do energii w dolnym zakresie
Głośność: zakres zmian, wartość średnia, kontur, plozji
Iloczasy: pauz, wyrazów, samogłoska/spółgłoska,

Narząd słuchu
W systemie percepcji dźwięków można wyróżnić 2 zasadnicze – układ peryferyjny słuchu i układ nerwowy tego narządu poprzez który dokonywane jest przetwarzanie bodźców na wyższych piętrach układu nerwowego (w mózgu). W narządzie słuchu dokonywane jest przetwarzanie zmian ciśnienia akustycznego na rozkład drgań na błonie podstawnej, który jest przekształcany na odpowiednie serie impulsów pobudzających nerw słuchowy. Informacje o odbieranych sygnałach docierających do narządu słuch są ekstrahowane na różnych poziomach układu nerwowego.

Funkcje kosteczek słuchowych
- swoistego rodzaju układ przekładni mechanicznej dopasowujący drgania w powietrzu do drgań w cieczy. Zamienia duży ruch tłoka o dużej powierzchni (błona bębenkowa) na mały ruch tłoka o małej powierzchni (podstawa strzemiączka w okienku owalnym). Wzmocnienie siły wynosi 27 razy. Transmisja dźwięków jest najskuteczniejsza w przedziale częstotliwości 500-4000 Hz.
- układ zabezpieczający – powyżej 90 dB(<1-2 kHz), następuje wzrost napięcia mięśni usztywniających układ kosteczek, w wyniku czego następuje ograniczenie przepływu energii akustycznej (odruch strzemiączkowy). Odruch ten jest zbyt wolny by chronić ucho przed hałasem impulsowym, np. wystrzał z broni palnej, gwałtowne pęknięcie ABS.

Funkcje transmitancji ucha zewnętrznego i środkowego
Zewnętrzny przewód słuchowy (o długości 2-3 cm, średnica 1 cm) ma skomplikowaną geometrię, co powoduje, że w jego charakterystyce transmitancji występuje szereg rezonansów (ok. 6) w zakresie od 3 do 12 kHz. Małżowina uszna wspomaga kierunkowe słyszenie dźwięków.
Funkcja transmitancji ucha środkowego ma jeden dominujący rezonans w pobliżu 1 kHz. Razem, obie części narządu słuchu kształtują częstotliwościową charakterystykę czułości słuchu z szerokim maksimum położonym w pobliżu 3 kHz.

Funkcje komórek rzęskowych
Komórki rzęskowe wewnętrzne są przymocowane do doprowadzających włókien nerwu ślimakowego i ich funkcją jako „rzeczywistych komórek słuchowych" jest zamiana informacji akustycznej na sygnały nerwowe. Komórki rzęskowe zewnętrzne są w przeważającym stopniu stymulowane przez włókna odprowadzające nerwu ślimakowego i często są opisywane jako „silnik" ślimakowego wzmacniacza. Ich zadaniem jest spowodowanie, aby maksymalne uwypuklenie błony podstawnej było bardziej wyraźne tak, aby komórki rzęskowe wewnętrzne to zarejestrowały. Tak więc komórki rzęskowe zewnętrzne służą jedynie do tego by wzmocnić wędrującą falę, podczas gdy komórki rzęskowe wewnętrzne zamieniają bodźce mechaniczne na potencjał bioelektryczny.

Efekt współdziałania zewnętrznych i wewnętrznych komórek rzęskowych
Tylko dzięki współdziałaniu i wzajemnym oddziaływaniu komórek rzęskowych wewnętrznych i zewnętrznych ucho posiada tak niski próg słyszenia (= podwyższenie amplitudy wędrującej fali) i taką czułość w rozróżnianiu częstotliwości(=strome przesunięcie wędrującej fali).

Mechaniczne i elektryczne własności komórek rzęskowych
Przy podstawie (bliżej okienka owalnego) komórki rzęskowe rozmieszczone wzdłuż błony podstawnej są odpowiednio dostrojone częstotliwościowo elektrycznie jak i mechanicznie. Rzęski przy okienku owalnym są krótsze i sztywniejsze, te bardziej oddalone są dłuższe i bardziej elastyczne. Jednocześnie własności komórek rzęskowych, decydujące o częstotliwości wyładowań elektrycznych własnych, są zgodne z rozmieszczeniem komórek wzdłuż membrany podstawnej. Częstotliwość wyładowań jest zgodna z rozkładem rezonansów błony podstawnej. A każdy neuron ma swoją „częstotliwość charakterystyczną".

Synchronizacja fazy z pobudzeniem sinusoidalnym
Dla częstotliwości < 5 kHz, impulsy nerwowe pojawiają się z określoną fazą zgodnie z cyklem sygnału pobudzającego. Wyładowania te nie pojawiają się w każdym cyklu pobudzenia. Jednakże odległość między pojedynczymi impulsami może wynosić 2,3 lub więcej cykli.

Przetwarzanie sygnału akustycznego na obraz wyładowań neuronowych
- Dokonuje się to w ślimaku – fala rozchodząca się wzdłuż membrany podstawnej pobudza określone jej miejsca do drgań.
- Percepcja częstotliwości sygnału odbywa się poprzez tzw. „pasma krytyczne", określające rozdzielczość częstotliwościową narządu słuchu.
- Można wyznaczyć ok. 24 pasm krytycznych rozmieszczonych na błonie podstawnej.
- Każde pasmo krytyczne na błonie zajmuje ok. 1,3 mm długości (ok. 1300 neuronów).

Zasadnicze punkty "teorii miejsca"
1. Istnieje korelacja miejsca położenia maksymalnej odpowiedzi (im wyższa częstotliwość miejsce to znajduje się bliżej okienka owalnego, przy podstawie ślimaka)
2. Zakres częstotliwości 20-5000 Hz rozkłada się na ponad 2/3 długości błony podstawnej (od 12 do 35 mm od okienka owalnego)
3. Wyższy zakres częstotliwości (5,000-20,000 Hz) przypada pozostałą część błony podstawnej (<1/3)
4. Stosunki częstotliwościowe bodźców są dokładnie odwzorowane przez stosunki odległości miejsc pobudzenia na błonie podstawnej

Zawodność teorii miejsca oceny wysokości dźwięku
Niezwykle małe rozmiary ślimaka i bardzo duża rozdzielczość w percepcji wysokości dźwięku wskazuje, że teoria miejsca nie wyjaśnia w pełni mechanizmu różnicowania dźwięków pod względem ich wysokości.
Podstawowe dane: długość błony podstawnej – ok. 3.2 cm
zdolność różnicowania ok. 1500 wysokości dźwięku, przy udziale 16000-20 000 komórek rzęskowych.
To sugerowałoby, że rozdzielczość drgań na długości błony podstawnej byłaby 0.002 cm. Tymczasem człowiek jest w stanie różnicować 2 jednoczesne dźwięki odległe od siebie o >7% (dla niskich częstotliwości) i >15% dla wysokich częstotliwości.

Krzywe strojenia
- Częstotliwościowa odpowiedź neuronu jest przedstawiana w postaci krzywej strojenia – określa jak głośny powinien być ton dla danej częstotliwości by pobudzić wyładowania w włóknie nerwu słuchowego
- Dla wysokich częstotliwości krzywa strojenia jest bardzo wąska zaś dla niskich częstotliwości – stosunkowo szeroka

Zjawisko „wyostrzania" w percepcji tonów
Teoria miejsca nie w pełni wyjaśnia obserwowanego zjawiska „wyostrzania", t.j. zdolności wyodrębniania bliskich w skali częstotliwości tonów. Jedna z prób wyjaśnienia opiera się na założeniu, że istnieje zjawisko tłumienia liczby wyładowań w neuronach sąsiadujących z miejscem maksymalnego szczytu drgań błony podstawnej. Wiadomo, że istnieje sprzężenie zwrotne z mózgu wspomagające to tłumienie.

Maskowanie
Maskowanie jest codziennie odczuwanym zjawiskiem, jedne dźwięki maskują.
Na przykład, dźwięki głośniejsze powodują, że cichsze stają się niesłyszalne.

Maskowanie = definicja
Maskowanie jest to zjawisko, w którym pojawienie się jednego dźwięku powoduje utratę słyszalności drugiego, lub zmniejszenie wrażenia jego głośności. Inaczej mówiąc następuje podniesienie progu słyszalności maskowanego dźwięku.

Wybrany dźwięk może maskować inne dźwięki, zwłaszcza te, które są dostatecznie blisko niego w skali częstotliwościowej (maskowanie częstotliwościowe) lub w skali czasowej (maskowanie czasowe).

Maskowanie częstotliwościowe
- Dźwięk o określonej częstotliwości maskuje dźwięki o innych częstotliwościach.
- Maskowanie przez dźwięk o niższej częstotliwości od maskowanego jest silniejsze, niż przez dźwięk o częstotliwości wyższej, zwłaszcza w przypadku dużych intensywności dźwięków.

Doświadczenie Fletchera
- Mierzył jak zmienia się próg słyszalności tonu w obecności szumu
- Szerokość pasma szumu, którego częstotliwość środkowa pokrywała się z częstotliwością maskowanego tonu była stopniowo zwiększana. Pociąga to wzrost energii szumu.

Przy stopniowym zwiększaniu pasma szumu próg słyszalności tonu rośnie do pewnego momentu. Dalszy wzrost pasma szumu nie powoduje istotnych zmian.

Pasmo krytyczne
Próg detekcji tonu sinusoidalnego wzrasta ze wzrostem szerokości pasma szumu maskującego. Po przekroczeniu pewnej wartości (pasma krytycznego filtru słuchowego) dalszy wzrost szerokości pasma szumu maskującego nie wpływa na wartość progu detekcji tonu (Fletcher, 1940)

Maskowanie a pasmo krytyczne

- aby usłyszeć określony ton człowiek musi skupić uwagę na sygnał wyjściowy z filtru, którego częstotliwość środkowa pokrywa się z częstotliwością tonu
- tylko w obrębie pasma krytycznego, stopniowy wzrost szerokości pasma szumu, zwiększa maskowanie tonu znajdującego w tym paśmie
- zwiększanie szerokości pasma szumu maskującego poza pasmo krytyczne, powoduje tylko pobudzanie sąsiednich filtrów słuchowych
- pobudzenie więcej niż jednego filtru słuchowego powoduje zwiększenie wrażenia głośności

Własności pasm krytycznych

- szerokość pasma krytycznego zależy od częstotliwości środkowej
- w mniejszym stopniu zależy od poziomu dźwięku
- dwa tony występujące w obrębie pasma krytycznego nie zwiększają słyszanej głośności w porównaniu z głośnością pojedynczego tonu.
- Dopiero gdy odległość między nimi jest większa od szerokości pasma krytycznego, wówczas wypadkowa głośność wzrasta.

Własności skali Bark

- Równe odległości w skali częstotliwości odpowiadają równym odległościom w skali percepcyjnej
- 1 bark = 1 szerokości pasma krytycznego
- Powyżej 500 Hz skala ta jest równoważna logarytmicznej skali częstotliwości
- Poniżej częstotliwości 500 Hz skala Bark jest funkcją liniową częstotliwości

Własności skali mel

- Punktem odniesienia jest ton 1000 Hz o poziomie 40 dB – 1000 meli = wysokość tonu o częstotliwości 1000 Hz
- Dla każdego tonu dobiera się drugi ton o częstotliwości odbieranej subiektywnie jako o dwukrotnie niższej (lub wyższej) wysokości, lub dokonuje się podziału danego zakresu częstotliwości na 4 percepcyjnie jednakowe interwały
- Do 500 Hz skala meli pokrywa się ze skalą częstotliwościową. Powyżej – zależność jest logarytmiczna
- 100 mel = 1 Bark

Pasma krytyczne mają wpływ na:

- Detekcję sygnału w ciszy
- Percepcję głośności
- Detekcję sygnału w szumie (maskowanie)
- Czułość na przesunięcie fazowe
- I wiele innych zjawisk …….

Czynniki wpływające na percepcję głośności

- Głośność dźwięku zależy od poziomu ciśnienia akustycznego
- Głośność dźwięku zależy od jego częstotliwości
- Głośność dźwięku zależy od jego zakresu częstotliwości
- Na wrażenie głośności dźwięku wpływają również czynniki czasowe

Pojęcie "rozdzielczości"

Określa dokładność z jaką można wyróżnić bodziec z pośród innych, o zbliżonych wartościach wybranego parametru

"Rozdzielczość częstotliwościowa"

Zdolność wyodrębnienia jednej składowej częstotliwościowej w dźwięku złożonym

Progowe badania wpływu zmian parametru fizycznego na percepcję dźwięku

W klasycznym ujęciu progiem nazywamy pewien punkt graniczny, w którym bodziec o zmieniającej się wartości określonego parametru (np. intensywności) lub wzrastająca różnica pomiędzy dwoma bodźcami stają się dostrzegalne (lub w którym bodziec lub malejąca różnica stają się niedostrzegalne).

Dwa progi w percepcji

- Progiem absolutnym nazywana jest wartość bodźca mierzona w warunkach eksperymentalnych, przy której zaczyna lub przestaje wywoływać reakcję.
- Progiem różnicowym (różnicy) nazywana jest minimalna (wzrastająca lub malejąca) różnica pomiędzy para bodźców, którą to różnicę można dostrzec w warunkach eksperymentalnych.

W postrzeganiu i wartościowaniu bodźców akustycznych przez człowieka udział biorą dwa niezależne mechanizmy; sensoryczny i decyzyjny

Zastosowanie badań progowych

Próg w ujęciu klasycznym, zarówno próg absolutny, jak i różnicowy, ma zastosowanie nie tylko w odniesieniu do badań prostych cech wrażeniowych takich jak głośność i wysokość.
Można go również określać przy badaniu innych zjawisk psychoakustycznych, na przykład takich jak lokalizacji źródeł dźwięku przez człowieka, czy percepcji zniekształceń nielinearnych.

Próg różnicowy częstotliwości

Jest to najmniejsza dostrzegalna różnica częstotliwości dwóch dźwięków. Oznacza się ją symbolem JND (ang. ***J**ust **N**oticeable **D**ifference)*. Ta zaledwie postrzegana różnica częstotliwości zależy od częstotliwości badanego dźwięku prostego, jego poziomu, czasu trwania oraz szybkości zmian jego częstotliwości.

zakres słyszalności dudnień

Dudnienia są wyraźnie słyszane, gdy różnica częstotliwości tonów pierwotnych jest < 15 Hz. Słyszy się tylko jeden ton o zmiennej amplitudzie.
Gdy różnica się powiększa nieznacznie powyżej tej granicy dźwięk staje się nieprzyjemny („chropowaty") bez wyraźnych dudnień. Do pewnej odległości ΔfD między tymi tonami, nie jest odczuwalna zmiana jakości dźwięku. Jest to granica różnicowania częstotliwościowego. Przy dalszym zwiększaniu różnicy częstotliwości między tymi tonami, zaczynają one być wyraźnie słyszalne jako 2 oddzielne tony. Ma to miejsce dla odległości większych od pasma krytycznego ΔfCB .

Pasmo krytyczne, a próg odczuwalnej minimalnej różnicy częstotliwości
Dla zadanej CZĘSTOTLIWOŚCI, pasmo krytyczne jest najmniejszym pasmem wokół której inne częstotliwości pobudzaja tę samą część błony podstawnej.
Natomiast, próg różnicy jest minimalną zauważalną różnicą (JND) pojedynczej częstotliwości, zaś pasmo krytyczne reprezentuje zdolność słuchającego do rozróżniania jednoczesnych tonów lub składowych dźwięków.

Źródło tonów kombinacyjnych
Różnicowe tony kombinacyjne nie są obecne w rzeczywistym sygnale.
Powstają one w wyniku pobudzenia membrany w miejscach odpowiadających tonom składowym (nie są one wynikiem złudzeń słuchowych !)
Są one wywołane „zniekształceniami" kształtu fali rozchodzącej się w płynie w kanale ślimakowym (powstają w nim turbulentne zawirowania).

Zniekształcenia obwiedni widma filtru słuchowego
Ma to miejsce w przypadku uszkodzeń słuchu.

- Szersze filtry słuchowe powodują powstanie „zamazanego" rozkładu pobudzenia, maksima stają się mniej wydatne, zmniejszony stosunek maksimów do minimów.
- Wprowadzenie szumu powoduje dodatkowo zacieranie różnic między wierzchołkami i minimami w widmie i zmniejsza cechy dystynktywne obwiedni widma

Maskowanie czasowe

- Maskowanie ma miejsce nawet, gdy sygnał maskujący i maskowany nie występują jednocześnie
- Maskowanie dźwięków wcześniejszych przez sygnał maskujący, tzw. maskowanie wsteczne (premaskowanie)
- Maskowanie dźwięków późniejszych, tzw. maskowanie resztkowe (postmaskowanie)

Charakterystyka maskowania czasowego
Maskowanie czasowe (nierównoczesne) polega na tym, że mózg nie jest w stanie przeanalizować dźwięków, które następują **tuż przed** (do 40 ms – zależnie od częstotliwości) oraz **tuż po** (do 200 ms, i więcej) dźwięku głośnym (**maskerze**).
Pierwszy typ maskowania , tzw. **wsteczne**, wynika z tego, że zanim dźwięk zostanie "zauważony" mija ok. 40 ms, a jeśli przed końcem tego czasu pojawi się dźwięk głośny, to proces analizowania tego cichego wariantu **zostaje przerwany**, a ucho i mózg **reagują tylko na sygnał maskujący**.

- **Maskowanie resztkowe** oprócz tego, że uwzględnia wspomniany czas na analizę dźwięku, to jeszcze czas potrzebny na tzw. **relaksację aparatu słuchu**, czyli powrót jego do stanu kiedy jest gotów odebrać z otoczenia kolejny dźwięk. Głośny dźwięk wymaga dłuższego po nim odpoczynku.

Maskowanie wsteczne
Wsteczne maskowanie jest związane z długością odpowiedzi impulsowej filtru słuchowego. Dla wysokich częstotliwości maskowanie wsteczne jest poniżej 1 ms dla wytrenowanych osób, przy jednousznym odsłuchiwaniu bodźców. Jednak zdolność wykrywania maskowanych wstecznie bodźców silnie zależy od predyspozycji słuchającego.

Maskowanie resztkowe (postmasking)
Maskowanie resztkowe sygnału testowego przez przebieg maskujący występuje zarówno, gdy sygnał zarówno znajduje się w obrębie odpowiedzi impulsowej filtru słuchowego, jak i neuronowej części systemu percepcyjnego.
Czas maskowania jest >20ms, a czasami stwierdza się, że czas ten może wynieść nawet kilkaset ms. W praktyce, w krzywej czasowej maskowania można wyróżnić dwie części – krótki obszar podtrzymywania maskowania oraz drugą część długiego zmniejszania maskowania. Im wyższy jest poziom sygnału maskowanego, tym krótszy jest czas postmaskingu.

Warunki amplitudowe w maskowaniu dźwięków

- Oczywiście jeśli w podanym przedziale czasu (-40 ms, +200 ms) pojawi się **dźwięk odpowiednio głośny**, on również zostanie "zauważony", te czasy pokazują maksymalny czas potrzebny w przypadku dźwięków **dużo cichszych od maskera** (o około 40 dB). Dzięki temu maskowaniu można z kodowanego dźwięku **wycinać ciche dźwięki** w odpowiednich miejscach, czyli tuż przed i po maskerze.

Prawo Hooke'a
Prawo Hooke'a stwierdza: odkształcenie rozchodzące się w ośrodku oddziaływuje na ścianki sześcianu z ciśnieniem liniowo proporcjonalnym do zmian jego objętości.
V=dx dy dz – objętość przed odkształceniem
du, dw, dv – zmiany wymiarów wzdłuż odpowiednio osi x, y, z
Ciśnienie P odnosi się jedynie do nadwyżki ciśnienia w stosunku do ciśnienia równowagi p0 w środowisku (ciśnienie atmosferyczne). Ciśnienie P nazywane jest ciśnieniem akustycznym.

Zmienne akustyczne
Podczas rozchodzenia się dźwięku w powietrzu (lub dowolnym ośrodku sprężystym), w każdym punkcie przestrzeni występują mierzalne fluktuacje ciśnienia, prędkości, temperatury i gęstości. Fizyczny stan ośrodka można opisać jako zmiany (stosunkowo małe) wokół pewnego stanu równowagi opisany przez wartości średnie powyższych parametrów. W akustyce obiektem analiz są właśnie zmiany wartości parametrów wokół pewnych wartości średnich.

Zależności fizyczne
Dla ośrodka idealnie sprężystego istnieje liniowa zależność między ciśnieniem akustycznym i zgęszczeniem lokalnym t.j. $p = K \cdot s$ gdzie zgęszczenie lokalne s jest definiowane jako stosunek przyrostu gęstości *s* do gęstości średniej w miejscu obserwacji
$s = \frac{d\rho}{\rho_0}$ zaś K - współczynnikiem sprężystości objętościowej

Ciśnienie fali akustycznej
Ciśnienie fali akustycznej odnosi się jedynie do nadwyżki ciśnienia w stosunku do ciśnienia równowagi w ośrodku rozchodzenia się fali (np. w powietrzu będzie to ciśnienie atmosferyczne). Ciśnienie P nazywane jest ciśnieniem akustycznym, czyli P = pa.

Ile energii niesie sygnał mowy?
" . . . 500 osób mówiących bez przerwy przez 12 miesięcy wytworzy energię wystarczającą do zaparzenia zaledwie 1 filiżanki herbaty."
Sygnał mowy generowany przez mężczyznę niesie energię 34 μW, przez kobietę – 18 μW (pomiar w odległości 1 m)

Zakres intensywności dźwięków słyszalnych
Minimalna intensywność dźwięku słyszalnego wynosi w przybliżeniu 10-12 W/m2. Intensywność dźwięku powodująca uszkodzenie słuchu – powyżej 1 W/m2.

Prawo Webera-Fechnera
Z badań psycho-akustycznych prowadzonych nad postrzeganiem różnic w głośności dźwięków wynika, że zgodnie z prawem Webera-Fechnera głośność dźwięku jest liniowo proporcjonalna do logarytmu z wartości bodźca.

Co wpływa na jakość brzmienia dźwięku stacjonarnego ?
1. Liczba i amplitudy harmonicznych
2. Składowe nieharmoniczne
3. Wysokość i zmiany tonu podstawowego
4. Tony różnicowe (zwiększają słyszalność tonu podstawowego)
5. Pasma krytyczne i maskowanie (formanty)

Przestrzenna lokalizacja źródła dźwięku
Przestrzenna lokalizacja - subiektywna ocena położenia źródła dźwięku w przestrzeni (kierunku i odległości) przez osobę znajdującą się w polu rozchodzącej się wokół niego fali akustycznej.
- percepcja w przestrzeni otwartej
- percepcja w przestrzeni zamkniętej (z odbiciami)

Czułość przestrzenna
Na współrzędne kierunku – lewo – prawo
Współrzędne podniesienia – góra – dół
Współrzędne odległości – od obserwatora
Słuchacze na ogół dość dobrze lokalizują położenie źródeł dźwięku znajdujących się na wprost nich, gorzej gdy są one z boku lub z tyłu głowy.
Lokalizacja dwuuszna - monouszna
W monousznej – decydujący jest fakt, że małżowina i głowa wpływają na charakterystykę częstotliwościową odbieranych dźwięków.

Czynniki wpływające na ocenę odległości od źródła
- ❑ Znajomość głośności znajomych źródeł
- ❑ Barwa dźwięku znanych źródeł (częstotliwości tonów wysokich są silniej tłumione w powietrzu, co powoduje zmianę barwy dźwięku przy oddalaniu się od jego źródła
- ❑ uwypuklenie czoła fali dźwiękowej
- ❑ stosunek natężenia dźwięku bezpośredniego do dźwięków odbitych
- ❑ doświadczenie słuchowe i wiązanie zjawisk akustycznych z obserwacjami wzrokowymi

Międzyuszna różnica poziomu (ILD)
Międzyuszna różnica poziomów zależy od kąta padania, i również od częstotliwości fali. Te o wysokiej częstotliwości ulegają mniejszemu ugięciu, a więc i cień akustyczny wokół głowy jest większy, niż w przypadku fal o niskiej częstotliwości. Dla głowy o średnicy ok. 17 cm, cień ten jest pomijalnie mały dla f< 500 Hz (λ=68 cm). Dla f>3000 Hz różnica jest istotna.

Jak obliczyć ITD ?
Różnica w czasie wynikająca z różnicy długości dróg d od źródła do lewego i prawego ucha (<1,3 kHz) : $d = r\cdot\theta + r\cdot\sin(\theta)$
r – promień głowy (8 cm)
θ – kąt ustawienia źródła, dla θ=300 (π/6), ITD=0.24 ms (dla prędkości fali 344 m/s)

Częstotliwość fali i IPD
Międzyuszne przesunięcie fazy dla fali o zadanej częstotliwości określa więc jednoznacznie opóźnienie w generowanych impulsach w narządzie słuchu. Dla ITD = 0.5 ms, w przypadku fali o częstotliwości f = 1 kHz, przesunięcie fazy IPD = 1800 . Dla f= 500 Hz, IPD =900 . W przypadku, gdy IPD wynosi więcej niż 3600 (co odpowiada maksymalnie 0.7 ms (dla głowy o średnicy = 8 cm) i częstotliwości 1430 Hz, fala dociera do obu uszu w tej samej fazie.

Nieoznaczoność fazy
W praktyce, nieoznaczoność fazy dla fali o zadanej częstotliwości jest w zakresie wyznaczonym przez odległość międzyuszną mniejszą od ½ długości fali. W praktyce nieoznaczoność jest pomijalnie mała, gdy odległość ta jest nie większa, niż ¼ długości fali.

Zależność kąta azymutalnego w przypadku dźwięków złożonych
Dźwięki złożone mają zmienną w czasie strukturę częstotliwościową i intensywność.
W dźwiękach złożonych są jednocześnie składowe nisko- i wysoko-częstotliwościowe. W tym przypadku, informacja azymutalna jest w przeważającym stopniu niesiona przez niskie częstotliwości, wpływających na percepcję ITD. Przy lateralizacji również i informacja niesiona przez ILD odgrywa pewną rolę.

Minimalna postrzegalna zmiana kąta obserwacji dla przebiegów sinusoidalnych
Zasadnicze punkty:
- Minimalna postrzegana różnica czasu ITD: 10 μs
- Minimalna postrzegana różnica poziomów ILD: 0.5-1 dB
- Różnice te są zależne od częstotliwości fali i kąta azymutalnego źródła
- Spadek dokładności postrzegania kąta azymutalnego źródła w obszarze 1.5 – 2 kHz sygnalizowany przez duplex theory w rzeczywistości nie ma miejsca. Opisywane przez nią mechanizmy nie działają skutecznie w tym obszarze.

Zależność kąta azymutalnego w przypadku dźwięków złożonych
Dźwięki złożone mają zmienną w czasie strukturę częstotliwościową i intensywność.
W dźwiękach złożonych są jednocześnie składowe nisko- i wysoko-częstotliwościowe. W tym przypadku, informacja azymutalna jest w przeważającym stopniu niesiona przez niskie częstotliwości, wpływających na percepcję ITD. Przy lateralizacji również i informacja niesiona przez ILD odgrywa pewną rolę.

Minimalna postrzegalna zmiana kąta obserwacji dla przebiegów sinusoidalnych
Zasadnicze punkty:
- Minimalna postrzegana różnica czasu ITD: 10 μs
- Minimalna postrzegana różnica poziomów ILD: 0.5-1 dB
- Różnice te są zależne od częstotliwości fali i kąta azymutalnego źródła
- Spadek dokładności postrzegania kąta azymutalnego źródła w obszarze 1.5 – 2 kHz sygnalizowany przez duplex theory w rzeczywistości nie ma miejsca. Opisywane przez nią mechanizmy nie działają skutecznie w tym obszarze.

Podsumowanie (dla przebiegów sinusoidalnych)
- Lokalizacja jest oparta na ocenie ILD i ITD
- ILD jest miarą międzyusznej różnicy poziomów w danym momencie czasu
- ITD jest miarą różnicy czasu fali dźwiękowej docierającej do lewego i prawego ucha
- ILD jest skuteczną miarą kąta azymutalnego dla częstotliwości > 2000 - 3000 Hz
- ITD jest skuteczną miarą dla częstotliwości< 1000 Hz
- Istnieje nieostrość w lokalizacji przód – tył w oparciu tylko o parametry ITD i ILD, która jest likwidowana poprzez ruchy głowy

Teoria Batteau (1967, 1968)
- odbicia powstające w małżowinie usznej niosą dane pomocne w ocenie lateralizacji i stopnia podniesienia źródła.
- w odlewach małżowin pomierzył zakresy zmian opóźnień dla kątów azymutalnych (2 – 80 μs) i podniesienia (100 – 300 μs)
- eksperymentalny odsłuch przez protezy małżowin dawał wrażenie eksternalizacji dźwięku

Charakterystyka przenoszenia głowy – Head Related Transfer Function
Charakterystyka przenoszenia głowy HRTF jest stosunkiem widma sygnału docierającego do ucha do widma sygnału docierającego do punktu przestrzeni zajmowanego przez środek głowy (czyli gdy nie ma w tym miejscu obserwatora). Para tych funkcji uwzględnia wszystkie statyczne parametry lokalizacji: ITD, ILD i charakterystyki częstotliwościowe małżowin.
HRTF dotyczy filtracji przestrzennej (anatomiczne funkcje przenoszenia).

Własności funkcji HRTF
- Jest w rzeczywistości asymetryczna z powodu kształtu małżowiny usznej oraz odbić od głowy i ramion
- HRTF określa w jakim stopniu różne składowe częstotliwościowe są wzmacniane/tłumione przez głowę dla różnych położeń źródła
- Funkcja ta odgrywa rolę tylko dla dźwięków szerokopasmowych

Funkcja transmitancji głowy HRTF – cechy widmowe lokalizacji źródła
- Funkcja HRTF jest głównie wyznaczona przez charakterystykę muszli usznej
- W mniejszym stopniu (i w zakresie niskich częstotliwości) przez głowę i tułów (ramiona, klatka piersiowa, kolana)
- Funkcja HRTF niesie informacje umożliwiające lokalizację położenia źródła
- W przypadku niemożności poruszania głową, niosą jedyne informacje umożliwiające lokalizację źródła, gdy znajduje się ono na stożku nieostrości

Założenia funkcji HRTF
Funkcja transmitancji ludzkiej głowy HRTF wykorzystuje założenia teorii Batteau, według której ucho pełni rolę sumatora, do którego wpadają sygnały odbite z różnym opóźnieniem i różnym tłumieniem od różnych fragmentów małżowiny, a odbijające zewnętrzne elementy małżowiny grają rolę zarówno przy detekcji kąta wzniesienia, jak i odległości, czy azymutu źródła.

Małżowina uszna jako swoistego rodzaju filtr
- **Teoria Blauerta** utożsamia natomiast małżowinę uszną z filtrem.

W zależności od kierunku czoła fali małżowina uszna wzmacnia niektóre części widma częstotliwości, a inne tłumi. W płaszczyźnie środkowej wg Blauerta wrażenie położenia źródła zależy nie od jego rzeczywistego kierunku, a od częstotliwości dźwięku.

Pomiar funkcji HRTF dla danego obserwatora
Pomiar HRTF może być wykonany w dwojaki sposób:
Monousznie - różnica funkcji źródła i funkcji pomierzonej w przewodzie słuchowym
Dwuusznie – przez wyznaczenie różnicy w odpowiednich punktach przewodów słuchowych tych funkcji.
(zakłada się przy tym, że tłumienie wysokich częstotliwości w powietrzu jest pomijalne)

Własności funkcji HRTF
Pojedyncza funkcja HRTF składa się z dwóch filtrów, po jednym dla każdego ucha, które zawierają wszystkie informacje o dźwięku (np. IID, ITD, widmo) istotne dla lokalizacji źródła przez obserwatora. Charakterystyka filtrów zmienia się w zależności od miejsca, z którego dochodzą dźwięki do obserwatora. Kompletna funkcja HRTF zawiera zestaw wielu filtrów, opisujących sferyczne środowisko dźwiękowe - 360 stopni, we wszystkich kierunkach dla wszystkich odległości. Filtry te zmieniają się w zależności od miejsca, z którego dochodzą dźwięki do obserwatora.

Przestrzenne słyszenie dźwięku
Dlaczego człowiek słyszy trójwymiarowo?
Są na to 3 teorie i każda z nich wydaje się być słuszna:
1) małżowina + kanał uszny stanowią układ rezonansowy; wzbudzenie określonych rezonansów zależy od kierunku i odległości źródła dźwięku od obserwatora
2) wrażenie położenia źródła zależy nie tylko od jego rzeczywistego kierunku ale od widma dźwięku, gdyż w zależności od kierunku czoła fali małżowina uszna wzmacnia niektóre częstotliwości, a inne tłumi
3) ucho pełni rolę sumatora do którego wpadają sygnały odbite z różnym opóźnieniem i różnym tłumieniem od różnych fragmentów małżowiny, a odbijające zewnętrzne elementy małżowiny grają rolę zarówno przy detekcji kąta wzniesienia, jak i odległości czy azymutu źródła

Efekt 3D przy odsłuchu słuchawkowym
Model ludzkiej głowy skonstruowany z materiałów o impedancji akustycznej odpowiadającej impedancjom tkanki kostnej czaszki, tkanki mięśniowej, skórnej i nerwowej mózgu jest bardzo kosztowny
Inny i tańszy (sztuczna głowa kosztuje bardzo dużo) sposób uzyskania efektu 3D w nagraniu jest użycie mikrofonów binauralnych, których membrany znajdują się w pobliżu błon bębenkowych. Realizator dźwięku umieszcza np. małe przetworniki w swoich uszach, we wlotach kanałów usznych.
Efekt 3D jest słyszalny wyłącznie przy odsłuchu na słuchawkach, gdyż membrany słuchawek znajdują się wówczas w przybliżeniu w miejscu membran mikrofonów użytych w nagraniu.

Lokalizacja w pomieszczeniu z odbiciami
Na wielkość ITD wpływa pogłos i odbicia, gdyż zależy ona od zgodności sygnałów docierających do uszu.
Natomiast na ILD mogą wpływać fale stojące, ale ogólnie biorąc, pomieszczenie ma mniejszy wpływ na ten parametr. Przy lokalizacji słuchacz głównie wykorzystuje informacje niesione przez składowe w zakresie wysokich częstotliwości.
Efekt precedensu – słuchacze lokalizują w oparciu o ocenę, z której strony dochodzi wcześniejsza fala bezpośrednia.

Odsłuch w przestrzeni z odbiciami: tłumienie echa i zjawisko precedensu
W wielu otoczeniach, bezpośrednia fala dźwiękowa docierająca do obserwatora jest jedną z wielu. Na ogół słyszy on obecność tylko jednego źródła zlokalizowanego przez niego w miejscu, w pobliżu którego znajduje się.

Percepcja odległości
- W otwartej przestrzeni i w komorze bezechowej**:**

Znajomość źródła dźwięku (np. mowa) znacznie ułatwia ocenę odległości. Malenie intensywności z odległością wskutek rozpraszania sferycznego mocy dźwięku, zaczyna być postrzegane dla odległości >3m
Własności widmowe – absorpcja w powietrzu wzrasta w powietrzu z odległością szczególnie silnie dla wysokich częstotliwości, wpływ jest zauważalny dla odległości >15m
- W pomieszczeniu z odbiciami:

Lepsza jest ocena odległości – porównuje się dźwięk bezpośredni z dźwiękami odbitymi. Błąd – 15-30%, a w określonych przypadkach większy

Percepcja dźwięku w przestrzeni z pierwszym odbiciem
Można wyróżnić trzy zakresy czasu :
1) lokalizacja sumacyjna (opóźnienie < 1 ms): dwa przebiegi są ze sobą łączone: postrzegana lokalizacja jest sumą ważoną parametrów lokalizacyjnych (ILD, ITD i charakterystyk częstotliwościowych).
2) Zjawisko precedensu (dla opóźnień ok.1-5 ms): postrzegany jest tylko jeden dźwięk - ten, który dociera pierwszy do obserwatora jest dominujący.
3) Próg percepcji echa (dla opóźnień > 5 ms): słyszane są dwa oddzielne dźwięki.

Zjawisko dominacji (pierwszeństwa)
Najbardziej istotna dla lokalizacji jest fala, która do obserwatora dociera pierwsza. Aby ten efekt wystąpił w przypadku odbić, maksymalne czasy opóźnień nie powinny być większe od kilkudziesięciu milisekund (powyżej - słyszalne echo)

Zjawisko Haasa (precedensu)
Zjawisko to uwidacznia fakt, że w percepcji kierunku położenia źródła opóźnienie fali docierającej do obserwatora ma znacznie większy wpływ, niż różnica poziomów. W przypadku dwóch identycznych źródeł promieniujących falę dźwiękowe o tym samym natężeniu odbiorca lokalizuje źródło pozorne dokładnie pośrodku między nimi. Dla opóźnień 0.1 – 1 ms jednego z sygnałów następuje przesunięcie źródła pozornego w kierunku źródła promieniującego bez opóźnienia. Aby uzyskać ponownie centralne położenie źródła pozornego należy zwiększyć poziom opóźnionej fali o 10 dB.

Lokalizacja dźwięków złożonych
- Dźwięki złożone mają zmienną w czasie strukturę częstotliwościową i intensywność
- Poprzez filtrację składowych o wysokiej częstotliwości i niskiej częstotliwości można stwierdzić, że pierwsze składowe też są w pewnym stopniu skuteczne w lokalizacji źródła, choć przesunięcia fazowe w percepcji obuusznej nie są jednoznaczne (tego teoria duplex nie wyjaśnia). Badania te wykonuje się stosując krótkie impulsy nisko- i wysoko-częstotliwościowe

Słuch a wzrok
1) Dźwięk zawiera zupełnie inną informację o otoczeniu, niż światło.
2) Informacja wizualna towarzyszy nam zwykle przez cały czas, natomiast dźwięk (słyszalny dla człowieka) powstaje wtedy, gdy coś się zmienia, np. gdy obiekty materialne wibrują, przemieszczają, zderzają się, ulegają deformacji itp.
3) Słuch to zmysł dotyczący zdarzeń, a nie scen. W związku z tym układ słuchowy przetwarza dane dźwiękowe w inny sposób, niż robi to układ wzrokowy z danymi wizualnymi.
4) Zasadniczym zadaniem wzroku jest informowanie nas, gdzie co się znajduje, natomiast głównym zadaniem słuchu jest zwracanie uwagi, że coś się dzieje.

Funkcjonowanie słuchu poniżej/na progu świadomości
1) Słuch jest ostatnim zmysłem, który przestaje funkcjonować, gdy tracimy świadomość.
2) Przy drzemce, odbiór pozostałych bodźców zmysłowych słabnie, natomiast dźwięków, staje się intensywniejszy, i jednocześnie pierwszym zmysłem, który zaczyna funkcjonować, gdy odzyskuje się przytomność.

Dominacja percepcji wzrokowej

1) Człowiek jest wzrokowcem, a jednak nieustannie używany jest słuch, aby kontrolować, co dzieje się w otoczeniu, w obrębie 3600 . Słuch uzupełnia naszą percepcję wzrokową - choć zwykle nie zwraca się na to większej uwagi, z wyjątkiem specyficznych sytuacji, w których brak dźwięku odbierany jest jako silnie nienaturalny.

2) Oczywiście, słuch jest także zmysłem mowy, wówczas informacja wizualna jest na ogół tylko jej uzupełnieniem.

3) Przedmiotem percepcji słuchowej są nie tylko same fizyczne dźwięki, ale również znaczenia, jakie niosą, nawet na progu świadomości.

Choć w sytuacjach, kiedy informacje napływające od różnych zmysłów są sprzeczne, dominuje wzrok, nic dziwnego, biorąc pod uwagę wrażliwość naszych uszu, że to słuch dominuje nad wzrokiem, jeśli chodzi o określanie czasu występowania zdarzeń.

Rozdzielczość czasowa

Rozdzielczość czasowa słuchu jest bez porównania lepsza od rozdzielczości czasowej wzroku. Obraz kinowy składający się z 24 klatek na sekundę odbieramy jako coś jednolitego, a nie jako 24 pojedyncze obrazy. Natomiast 24 stuknięcia w ciągu sekundy usłyszymy jako serię stuknięć — nie zlewają się one bowiem w jeden ciągły dźwięk.

Skąd różnice w prędkości działania receptorów wzrokowych i słuchowych

Wiele elementów układu słuchowego jest wyraźnie wyspecjalizowanych w pomiarze czasu. Niemniej jednak zasadnicze znaczenie ma tutaj budowa narządu odbierającego dźwięki.

W przypadku wzroku, światło jest przekształcane na impulsy nerwowe w stosunkowo powolnym procesie chemicznym zachodzącym w komórkach receptorowych. Natomiast w uchu, dźwięk przekształcany jest na impulsy nerwowe na szybkiej drodze mechanicznej, a następnie bioelektrycznej.

Minimalne czasy postrzegania zmian bodźców

Wrażliwość układu słuchowego na różnice czasowe jest wyjątkowa — wykrywa on okresy ciszy między dźwiękami, które trwają jedynie 1 ms. Układ wzrokowy musi widzieć dany obraz przez około 30 ms, aby informacja o nim dotarła do świadomości.

Akustyka pomieszczeń

- Wiele zjawisk akustycznych jest przedstawianych w uproszczeniu, bowiem często przyjmuje się, że warunki otoczenia spełniają warunki pola swobodnego .
- W polu swobodnym poziom natężenia dźwięku maleje co 6 dB przy każdorazowym dwukrotnym zwiększaniu odległości od źródła
- Jednak obecność powierzchni odbijających powoduje zniekształcenie warunków pola swobodnego:
  - Występowanie wielokrotnych odbić powoduje pojawienie się pogłosu (dla niezbyt niskich częstotliwości)
  - Odbicia między równoległymi do siebie powierzchniami mogą prowadzić do powstania rezonansów fal stojących (mody pomieszczenia - dla stosunkowo niskich częstotliwości)

Zasadnicze problemy w pomieszczeniach

- Obniżenie poziomu hałasów
- Zrozumiałość mowy
- Jakość mowy
- Jakość brzmienia muzyki

Podstawowe cechy akustyczne pomieszczeń

- Szum tła
- Czas pogłosu
- Poziom poglosu
- Echa (duże pomieszczenia)
- Obecność fal stojących (małe pomieszczenia)

Co wpływa na akustykę pomieszczeń

- Miejsca i kąty odbić
- Rozkład czasowy odbić
- Jakość odbić:
  - W funkcji częstotliwości
  - Współczynniki pochłaniania
- Liczba odbić
- Mody pomieszczenia, które ulegają wzmocnieniu

Jak opisywać zanikanie dźwięku w pomieszczeniu ?

- Problem miary w skali czasu.
- W połowie czasu wybrzmiewania?

po t 1/2 moc dźwięku jest równa ½ mocy początkowej

- Można zastosować funkcję opisującą zanikanie mocy dźwięku np. $P(t) = P0\text{-}t/t\_1/2$
- lub w postaci wykładniczej

$P(t) = P0 \exp(-t/tzanikanie)$

- lub w poniższy sposób
- $P(t) = P010\text{-}t/t\_z.$
- Przy odpowiednim dobraniu *t_z lub tzanikanie* powyższe funkcje są równoważne

Odbicia i pogłos

- Do słuchacza po bezpośredniej fali dźwiękowej docierają fale odbite od ścian pomieszczenia
- Nakładające się na nią fale odbite o odpowiednim opóźnieniu dają wrażenie *pogłosu*
- Stosunek energii niesionej przez falę bezpośrednią do energii fal odbitych stanowi wskazówkę, co do rozmiarów pomieszczenia, wykładzin na powierzchniach ograniczających i odległości od źródła.

Wczesne odbicia

- Czas pojawienia się pierwszego wczesnego odbicia jest ważnym parametrem w ocenie estetycznej akustyki sal. Dlaczego ? Nie ma fizycznych podstaw wyjaśnienia tego faktu !
- Wiadomo (z symulacji), że jeżeli pierwsze odbicie jest opóźnione o więcej niż ok. 65 ms, wówczas słyszy się echo – niepożądany efekt.

Rola odbić w ocenie nagrań

- Odbicia między 50ms and 150ms wpływają na wrażenie odległości, ale odbywa się to kosztem zmniejszonej zrozumiałości
- Odbicia z tego zakresu brzmią "ciemno". Dobierając odpowiednio amplitudę i opóźnienie wczesnych odbić można uzyskać nagrania o dużej przestrzeni, głębi i „obszernym" planie dźwiękowym
- Nagranie ze zbyt niskim poziomem wczesnych odbić brzmi jako zbyt bliskie i o sztucznym brzmieniu.
- Istnieje optymalny poziom wczesnych odbić od -4 do -6 dB w stosunku do poziomu dźwięku bezpośredniego
- Poziom dźwięku w zakresie >150ms jest krytyczny – zmiana w tym zakresie o 3 dB pociąga za sobą zmianę o ok. 1 dB pola pogłosowego
- Słyszalność pola pogłosowego silnie zależy od czasu pogłosu

Czas pogłosu a akustyka pomieszczenia

- Uznaje się, że najważniejszym parametrem charakteryzującym akustykę pomieszczenia jest czas pogłosu.
- Jest to parametr czasowy charakteryzujący zanik dźwięku w pomieszczeniu lub zmalenie jego poziomu do określonej wartości.
- Na przebieg czasowy zanikania dźwięku w pomieszczeniu wpływa nie tylko jego wielkość, lecz również rodzaj wykładzin ścian.
- Duże pomieszczenia mają stosunkowo długi czas pogłosu.
- Pomieszczenia o bardziej wytłumionych ścianach mają zmniejszony czas pogłosu.

Pojęcie czasu pogłosu

- Powszechnie stosowana definicja czasu pogłosu, *RT*60, jest czasem, w którym energia dźwięku w pomieszczeniu zmniejszy się o 60 dB w stosunku do energii początkowej.
- Pomiar czasu pogłosu może być wykonany poprzez wytworzenie krótkiego impulsu dźwiękowego za pomocą strzału, pęknięcia balonika, czy klaśnięcia.
- Dlaczego spadek o 60 dB ? Poziom orkiestry w crescendo dla większości utworów osiąga ok. 100 dB, zaś poziom szumów tła w przeciętnej sali koncertowej wynosi ok. 40 dB.
- W praktyce pomiar ten jest trudny do zmierzenia . Z powodu nieliniowej charakterystyki zanikania dźwięku trudno ograniczyć zakres pomiaru poziomów.

Definicja czasu pogłosu

Fala odbita pod kątem Qi dociera do obserwatora w chwili Ti niosąc energię Ei . Średni czas, w którym docierają odbicia do obserwatora wynosi:

$$TS = \frac{\sum E_i T_i}{\sum E_i}$$

Czasu pogłosu jest czasem, po upływie którego poziom energii dźwięku w pomieszczeniu zmniejszy się 106 razy, to jest

$$\frac{E(T_{60})}{E_0} = 10^{-6}$$

Wzór Sabine'a (1900)

$$T_{60} = 0.163(s/m)\frac{V}{S_e}$$

V – objętość pomieszczenia [m3]
Se – chłonność ścian pomieszczenia [m2]
Se = a1S1 + a2S2 + a3S3 + ….
ai - współczynnik pochłaniania ściany i
ai = 1 – bi
bi – współczynnik odbicia

Charakterystyka pomieszczenia

Czas pogłosu – czas potrzebny do stłumienia dźwięku o 60 dB. Zależy od:

- wymiarów i kształtu pomieszczenia (objętość pomieszczenia)
- materiałów pokrywających ściany (współczynnika pochłaniania wykładzin α
- chłonności akustycznej całego pomieszczenia Se określonej przez ważoną sumę współczynników absorpcji poszczególnych powierzchni
- obiektów znajdujących się w pomieszczeniu (dodatkowe odbicia i pochłanianie)

Konieczność kompresji dźwięku

- Inne techniki i inne wymagania, niż w przypadku obrazów video
- Szybkość transmisji dla danych CD audio jest znacznie mniejsza niż dla video, ale jednak przekracza możliwości połączenia dial-up (modemowego)
- Wymagana szybkość transmisji dla CD:
  44100*2*2 bajty/s = 176400 B/s = 1,41 Mbit/s
- Zajętość pamięci 3 minuty zapisu stereo = 31 MB

Dlaczego kompresja sygnałów audio jest możliwa ?

- Rozkład funkcji gęstości prawdopodobieństw próbek nie jest równomierny
- Próbki nie są od siebie niezależne, zarówno w dziedzinie czasu, jak i częstotliwościowej (istnieje redundancja)
- Ograniczenia narządu słuchu powodują, że są cechy czy zmiany w sygnale percepcyjnie nie różnicowane (zakres nieistotnych różnic)

Trudności w kompresji dźwięku

- Złożoność fal dźwiękowych, ich trudno przewidywalny charakter utrudnia stosowanie efektywnych bezstratnych metod kompresji
- Różnego typu dźwięki stawiają różne wymagania wobec systemów kompresji
- ✓ Muzyka
- ✓ Mowa
- ✓ Dźwięki otoczenia i zależnie od przeznaczenia

Kwantyzacja liniowa - nieliniowa

- Percepcja głośności dźwięku jest proporcjonalna do logarytmu jego amplitudy
- Nieliniowe techniki kwantyzacji ograniczają rozmiary próbek (wymagana jest mniejsza ilość bitów)
- W liniowej kwantyzacji poziomy kwantyzacji są jednakowo odległe od siebie, w logarytmicznej – blisko siebie dla małych wartości, coraz bardziej odległe dla większych

Zalety nieliniowej kwantyzacji sygnału

Sygnał telefoniczny jest próbkowany z częstotliwością 8 kHz. Kompresja mu-law (stosowana również w dyktafonach) koduje w 8 bitach zakres zmian dynamiki, który przy liniowej konwersji wymagałby 12 bitów. Czyli redukcja danych jest o 1/3.

Kompresja mowy – liniowe kodowanie predykcyjne (LPC – linear prediction coding)

Właściwości LPC

- Znaczna kompresja mowy
- Zastosowany jest matematyczny model toru głosowego
- Zamiast transmisji próbek sygnału wysyłane są parametry modelu toru głosowego
- Osiągane są b. małe wymagania co do prędkości transmisji danych – 2,4 kbps (takie jak w b. kiepskich liniach telefonicznych)
- Brzmienie mowy nieco „maszynowe”, choć zrozumiała

Liniowe kodowanie predykcyjne

- Wartość danej próbki (o rozmiarze k-bitów) prognozuje się jedynie na podstawie wartości poprzedzających ją M próbek.
- Rząd predykcji równa się liczbie próbek po której uśredniamy współczynniki.

Błąd predykcji

Błąd między próbką aktualną i prognozowaną:
Suma błędów kwadratowych w analizowanym segmencie sygnału, która może być zminimalizowana (za n próbek):
Przyrównując pochodne cząstkowe *E* względem a*i* otrzymujemy zbiór równań minimalizujących błąd

Struktura kodera LPC

1. Sygnał mowy jest segmentowany na nie zachodzące na siebie ramki
2. Sygnał jest poddawany preemfazie, by wyrównać obwiednię widma w zakresie wyższych częstotliwości
3. Detektor dźwięczności dokonuje klasyfikacji ramek (1 bit)
4. Sygnał poddawany jest analizie LPC – wyznaczonych zostaje 10 współczynników
5. Współczynniki te poddawane są kwantyzacji i wraz z indeksami są przesyłane w bloku informacji o danej ramce
6. Skwantowane współczynniki są stosowane do zbudowania filtru błędu predykcji realizującego filtrację preemfazowanego sygnału mowy w celu wyznaczenia na wyjściu błędu predykcji
7. Okres tonu podstawowego jest estymatą realizowaną z sygnału błędu predykcji (dla ramek uznanych za dźwięczne)

Kodowanie LPC i mu-law

Ramka w LPC – około 22,5 ms, co odpowiada 180 próbkom, przy częstotliwości próbkowania of 8000 kHz
Liczba współczynników predykcji = 10 (42 bity)
F0 i informacja o dźwięczności – 7 bitów
Wzmocnienie G – 5 bitów
Globalna suma bitów dla jednej ramki- 54 bit (2400 bps)

Model toru głosowego złożony z wielu odcinków cylindrycznych

W torze głosowym funkcja przekroju jest zmienna w czasie podczas mówienia. Dla wielu dźwięków mowy źródło pobudzające jest takie same. Sygnał pobudzający rozchodzi od głośni do ust ulegając kolejnym odbiciom na złączach odcinków cylindrycznych (bez strat energii)

Zalety i wady wokodera LPC

- o Zalety
  - rozdzielone parametry F0, wzmocnienie, dźwięczność/bezdźwięczność mogą być oddzielnie modyfikowane (np. do zmiany głosu męski/żeński)
  - mały strumień danych – 2400 bps
- o Wady
  - stosunkowo niska jakość syntezy mowy dla głosów żeńskich
  - nie osiąga jakości mowy telefonicznej

Podstawy kompresji percepcyjnej

- W sygnale istnieją składowe, których narząd słuchu nie postrzega
- Niektóre dźwięki mogą być poniżej progu słyszalności
- Niektóre dźwięki mogą być maskowane przez inne

Próg słyszalności

- Próg słyszalności:
  - wartość poziomu powyżej którego dźwięk jest słyszalny
  - Zmienia nieliniowo z częstotliwością
  - Dźwięki niskoczęstotliwościowe i wysokoczęstotliwościowe muszą być o znacznie wyższym poziomie, niż te ze środkowego pasma częstotliwościowego
  - Największa czułość słuchu jest w zakresie pasma częstotliwości sygnału mowy

Model psychoakustyczny

- W algorytmie kompresji stosowany jest model psycho-akustyczny opisujący zmiany czułości słuchu z częstotliwością oraz wynikające ze zjawiska maskowania
- Maskowanie – głośne dźwięki mogą spowodować, że cichsze stają się niesłyszalne. Zależność ta wynika bezpośrednio ze stosunku ich poziomów, ale również ze stosunku ich częstotliwości

- maskowanie powoduje, że w obrębie głośnego tonu następuje podniesienie krzywej progowej czułości słuchu (również i szumy mogą stać się niesłyszalne)
- w obrębie głośnych tonów kwantyzacja może być o mniejszej rozdzielczości (stąd mniejsza ilość bitów do kodowania głośnych składowych – tym samym maskowany jest szum kwantyzacji)

Kodowanie percepcyjne

- wykorzystuje się własności i ograniczenia w percepcji dźwięków przez narząd słuchu
- w modelu uwzględnione są:
  - zmienna z częstotliwością czułość słuchu
  - maskowanie częstotliwościowe
  - maskowanie czasowe
- kompromis między kodowaniem wysokiej i niższej jakości jest wynikiem doboru odpowiedniego rozmiaru strumienia danych

Kodowanie podpasmowe w MPEG- Audio

Po przejściu przez filtr pasmowy, wskutek decymacji z podpasm usuwane są próbki, w taki sposób, że każdy filtr wyznacza tylko co 32 próbkę (filtr jest polifazowy). Zdecymowane sekwencje próbek sa poddawane zmodyfikowanej transformacie cosinusowej typu IV (MDCT). Fizycznie zwiększa to ilość pasm analizy do 192 lub 576 (długość transformaty jest dobierana przez blok modelu psychoakustycznego). Modyfikacja transformaty polega na tym, że obejmuje ona dwa bloki próbek (12 lub 36), nakładających się w połowie długości ramek.

Dodatkowa informacja poboczna

- Sygnał audio jest przetwarzany w ciągi dyskretnych bloków próbek – bloki te są nazywane ramkami
- Każda ramka (24 ms = 1152 bitów) na wyjściu z każdego podpasma jest:
  - Skalowana w celu normalizacji szczytowego poziomu sygnału
  - Kwantyzacja jest dobrana odpowiednio do bieżącego stosunku sygnału do poziomu maskowania
- Dekoder musi znać bieżący współczynnik skali oraz zastosowane poziomy kwantyzacji
- Informacja ta musi być dołączona do strumienia danych
- Ten dodatkowy wzrost strumienia jest bardzo mały w porównaniu z zyskami przeprowadzonej kompresjii

Teoria z wykładów DSM

Komputer jest jedynym urządzeniem umożliwiającym tzw. przekaz multimedialny

odtwarzacz CD nie jest urządzeniem multimedialnym

Cechy bodźców rozróżniane przez słuch

Zakres częstotliwości – 20 Hz-16000 kHz (l =17,2 m- 2,15 cm)

***Odszumianie** – usuwanie z nagrań niepożądanych dźwięków*

***Język** – system znaków i określonych reguł fonologicznych, syntaktycznych i semantycznych rządzących kombinacją tych znaków*

***Morfologia** – budowa i odmiana wyrazów*

***Działanie modułu fonetycznego** ma na celu dokonanie konwersji wyrazów przedstawionych w postaci kodu ortograficznego na kod fonetyczny z dodatkowymi informacjami (np. dotyczącymi akcentu), określającymi ich wymowę*

***Analiza morfologiczna** umożliwia określenie wymowy deklinacyjnych i koniugacyjnych form wyrazów znajdujących się w słowniku, a przede wszystkim zmianę znaczenia spowodowaną zmianą dźwięku mowy lub intonacją*

***Moduł syntezy mowy** generuje akustyczny sygnał mowy, na podstawie sekwencji określonych fonemów uzyskanych na podstawie przetwarzania tekstu, wzorców iloczasowych, konturu melodycznego i obwiedni amplitudy*

***Difon** – element zawierający w całości przejście między głoskami, poprzedzone częścią głoski poprzedzającej i zakończone częścią głoski następującej*

***fonetyki artykulacyjnej**- jest opisanie mechanizmu powstawania dźwięków mowy w narządzie artykulacyjnym człowieka*

Fonetyka akustyczna
- Koncentruje się na analizie fizycznych własności dźwięków mowy promieniowanych wokół osoby mówiącej
- Badanie dźwięków mowy odbywa się przy zastosowaniu fizycznych metod analizy sygnałów akustycznych
- Jednocześnie poszukuje powiązań istniejących między czynnością artykulacyjną i wytworzonym sygnałem mowy

Fonetyka percepcyjna *- Bada percepcję dźwięków mowy, na poziomie układu centralnego*

Narzady artykulacyjne człowieka: jama(nosowa, ustna, gardlowa), podniebienie miękkie twarde, wargi, jezyk, szpara głośni, tchawica
Elementy narządu artykulacyjnego uczestniczące w formowaniu sygnału mowy: wargi, jezyk, żeby, podniebienie, faldy glosowe

Źródłem energii promieniowanej podczas mówienia są płuca

źródłem energii niesionej przez dźwięk są płuca osoby grającej

Max pojemność płuc – ok. 7 litrów **Pojemność minimalna** – 2 litry stale w płucach.
Objętość powietrza wymieniana podczas każdego cyklu oddechowego – 0.5 l **Częst. oddychania w stanie spoczynku** – 12-20 cykli na minutę

Źródłem pobudzającym tor głosowy mogą być:
a) fałdy głosowe – modulują w sposób regularny przepływ powietrza wychodzącego z płuc,
b) szczelina utworzona w torze głosowym - powoduje powstanie zawirowań,
c) przeszkoda (zęby) – j.w.
d) krótkotrwały impuls powietrza – powstaje w wyniku nagłego otwarcia toru głosowego, po chwilowym zwarciu w określonym miejscu toru głosowego.

Fonacja może się rozpoczynac przy przy cisnieniu podgłośniowym większym niż 3 cm H2O od ciśnienia atmosferycznego a nie może przy mniejszym

Wzór na częstotliwość drgań fałdów głosowych

$$F_0 = \frac{1}{2\Pi}\sqrt{\frac{(K + K^*)}{m}}$$

<u>Średnia długość fałdów</u>:
noworodki – 5 mm
dzieci – 10-13 mm
kobiety –11-15 mm
mężczyźni – ok. 20 mm

Narząd artykulacyjny jako układ akustyczny
a) źródło pobudzające
b) tor głosowy

długość toru głosowego - 17 cm
długość odcinka cylindrycznego - 1 cm

Liczba rezonansów w torze głosowym istotnych dla percepcji dźwięku samogłoskowego jest ograniczona i nie przekracza zazwyczaj 5-7

Maksima w charakterystyce częstotliwościowej toru głosowego wpływające na różnicowanie dźwięków mowy danego języka nazywamy **<u>formantami</u>**. Oznacza to, że nie każde maksimum w widmie danego dźwięku mowy musi być formantem

Są dwa rodzaje falowodów cylindrycznych:

a) Rura zamknięta na jednym końcu, otwarta na drugim
b) Otwarta lub zamknięta na obu końcach – oba typy mają identyczne rezonanse

Częstotliwości formantowe samogłosek

Samogloska	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
/i/	188-275	2078-2836	2670-3432	3316-4144
/y/	262-391	1689-2362	2424-3146	3124-4226
/e/	524-630	1580-2228	2468-3146	3064-4034
/a/	683-1021	1132-1566	2328-2860	3098-4088
/o/	493-679	788-1100	2410-3026	3194-3954
/u/	242-338	558-789	2266-3188	2942-4058

Dwa źródła pobudzenia toru głosowego

- Źródło krtaniowe - *pobudzenie periodyczne (harmoniczne)* powstające w wyniku drgań fałdów głosowych
- Źródło szumowe - *szum* powstający w wyniku gwałtownej zmiany ciśnienia lub przewężenia w torze głosowym.

Elementem formującym kształt widma spółgłosek trących jest komora utworzona z przodu szczeliny.

Długość tej komory wyznacza najniższą jej częstotliwość rezonansową. Im jest dłuższa, tym ta częstotliwość jest mniejsza

W przeciwieństwie do samogłosek charakterystyka widmowa spółgłosek jest wyznaczona nie tylko przez formanty, ale również przez antyformanty

Kiedy mogą pojawiać się antyformanty
1) Gdy tor głosowy jest rozdzielony na dwie sprzężone ze sobą części np. w przypadku nazalizacji, czy artykulacji spółgłoski nosowej
2) Jama ustna zostaje rozdzielona na dwie równoległe do siebie części, jak to ma miejsce w przypadku artykulacji spółgłoski /l/
3) Szczelina przy artykulacji spółgłosek trących jest stosunkowo szeroka i występuje sprzężenie ze sobą tylnej i przedniej komory

Położenie głoski we frazie może wpływać na jej wymowę, bądź na ubezdźwięcznienie/udźwięcznienie

W wygłosie wypowiedzi ruchy narządów mowy są wykonywane znacznie mniej dokładnie, z mniejszym nakładem energii, a także wolniej niż w nagłosie i śródgłosie

Koartykulacja jest zjawiskiem, podczas którego następuje nakładanie się ruchów artykulacyjnych właściwych dla sąsiadujących ze sobą głosek.

Rodzaje koartykulacji
Antycypacja i przedłużenie
Upodobnienia i uproszczenia w obrębie wyrazu
Upodobnienia pod względem dźwięczności
pod względem miejsca artykulacji
pod względem stopnia zbliżenia narządów mowy
Międzywyrazowe upodobnienia – na granicy wyrazów

Fonem a litera

Te same znaki ortograficzne lub jednakowe ich sekwencje mogą odpowiadać różnym dźwiękom mowy: np. „**wór" – /vur/, „wtórny" – /fturnI/**
„marznąć" – /marznon'ts'/, „marzec" - /maZets/
Różne znaki ortograficzne mogą odpowiadać tym samym dźwiękom mowy
np. **„auto" – /awto/, „dał" – /daw/**

Różne sekwencje:
„dźwiga" – /dz'viga/, „dzień" – /dz'en'/
W transkrypcji fonetycznej uwzględnia się zjawisko koartykulacji !

Literom samogłoskowym **„y,e,a,o"** odpowiadają fonemy **/I,e,a,o/**. Litery „u" i „ó" nie sygnalizują różnic w wymowie.

Literę „i' przed literą spółgłoskową wymawia się jako samogłoskę /i/

Literę „i' przed samogłoską wymawia się jako:
- /j/ po zwartych, nosowej /m/, trących /f,v,x/, i głoskach /l,r/

/i/ na końcu wyrazu
- podwójne „ii" po zwartych, nosowej /m/, trących /f,v/, głoskach /l,r/ i literze „ch" wymawia się jako /ji/

Następujące grupy spółgłoska-samogłoska /i/ odpowiadają następującym fonemom:
- „si" – /s'/ „ci" - /ts'/
- „zi" – /z'/ „dzi" - /dz'/
- **„ni" - /n'/ wyjątek „Dania" –/dan'ja/, ale /dan'a/**

Samogłoski nosowe „**ę,ą"** wymawia się jako
- **/e~,o~/** na końcu wyrazu

- /em,om/ przed /p,b/
- /en,on/ przed /t,d,ts,tS,dz,dZ/
- /en',on'/ przed /ts',dz'/
- /eN,oN/ przed /k,g/
- **/e,o/ przed /l,w/ „wziąłem" – w czasie przeszłym**

Głoski zwarte (/b,d,g/), zwarto-trące (/dz,dz',dZ/) i trące (/v,z,z',Z/) wymówione przed głoskami bezdźwięcznymi, przerwą(w wygłosie) stają się bezdźwięcznymi i ich wymowa jest dokładna, jak ich bezdźwięcznych odpowiedników, tj. /p,t,k/, /ts,ts',tS/ czy /f,s,s',S/. To samo występuje u zbiegu wyrazów wymówionych bez przerwy

O ubezdźwięcznieniu lub udźwięcznieniu całej sekwencji powyższych spółgłosek o różnym typie pobudzenia decyduje w zasadzie ostatnia w sekwencji głoska – **np. „liczba" - /lidZba/, „rzadszy" - /Zat_SI/**

Od powyższej zasady jest wyjątek, gdy przed literą „w" lub sekwencją „rz" stoi głoska bezdźwięczna. Cała sekwencja staje się bezdźwięczna. **np. „kwiat" – /kfjat/, „szwaczka" - /SfatSka/**

Nieregularności w wymowie „trz", „drz", „dż", „dz" w obrębie wyrazu np. **„trzech" - /tSSex/,** ale **„Czech" - /tSex/, „wodze" – /vodze/, „odzew" – /od_zef/**

Spółgłoski bezdźwięczne przed końcówką czasownikową „–my" pozostają bezdźwięczne np. „kupmy" - /kupmy/

Grupy spółgłoskowe złożone ze spółgłosek zwartych, zwarto-trących i trących, które są wymówione w nagłosie lub śródgłosie form wyrazowych, są całkowicie dźwięczne lub bezdźwięczne – /fskotSIts'/, krufka/, /proz'ba/.

Grupy mieszane – powyższe spółgłoski nie zmieniają dźwięczności spółgłosek przymkniętych - /kulka/, /puwka/, /krova/, zamknon'ts'/
Jednakże spółgłoski przymknięte wymówione w środku dłuższych sekwencji spółgłoskowych są najczęściej bezdźwięczne i wymawiane tak słabo, że często ulegają całkowitej redukcji – „jabłko" - /japko/, „rzemieślnik" - /Zemjes'n'ik/

Przykład SAMPA
Konwersja tekstu na mowę otwiera nowe możliwości, niedostępne w tradycyjnych systemach głosowych. Usługi katalogowe, informatory turystyczne, tematyczne serwisy informacyjne, czy portale głosowe, to tylko nieliczne zastosowania tej technologii.

konversja tekstu na move otfjera nove moZlivos'ts'i n'edostempne f tradItsIjnIx sIstemax gwosovIx | uswugi katalogove informatorI turIstItSne tematItSne servisI informatsIjne tSI portale gwosove to tIlko n'elitSne zastosovan'a tej texnologji

Definicja akcentu: Jest to to proces uwydatniający wybrane segmenty w sygnale mowy ciągłej, np. sylab w wyrazach lub wyrazów w zdaniach

Akcent dynamiczny, rytmiczny i melodyczny

Przyjmuje się, że w języku polskim akcent wyrazowy jest stały i spoczywa w zasadzie na przedostatniej sylabie formy wyrazowej. Są formy wyrazowe nie mające samodzielnego akcentu np. „się", „ci", „za", „mnie" itp. i dołączają się do wyrazu mającego swój akcent – np. „pod_lasem".

Wzmacnianie danej sylaby często odbywa się poprzez podniesienie częstotliwości F0

Narząd słuchu

W systemie percepcji dźwięków można wyróżnić 2 zasadnicze – układ peryferyjny słuchu i układ nerwowy tego narządu poprzez który dokonywane jest przetwarzanie bodźców na wyższych piętrach układu nerwowego (w mózgu). W narządzie słuchu dokonywane jest przetwarzanie zmian ciśnienia akustycznego na rozkład drgań na błonie podstawnej, który jest przekształcany na odpowiednie serie impulsów pobudzających nerw słuchowy. Informacje o odbieranych sygnałach docierających do narządu słuch są ekstrahowane na różnych poziomach układu nerwowego.

Zasadnicze elementy narządu słuchu
Ucho zewnętrzne : Małżowina, zewnętrzny kanał słuchowy
Ucho środkowe: Młoteczek, Kowadełko, Błona bębenkowa, półkolisty kanał poziomy, strzemiączko
Ucho wewnętrzne: Nerw słuchowy, ślimak, Okienko okrągłe, Kanał Eustachiusza

Schemat funkcjonalny organu słuchu
Ucho zewnętrzne: Małżowina uszna, fala dzwiekowa, zewnętrzny kanał słuchowy
Ucho środkowe: błona bębe, młoteczek, kowadełko, okienko owalne, strzemiączko, okienko okragłe
Ucho wewnętrzne: schody przedsionka, organ Cortiego, membrana podstawna, schody bębenka, ślim

Charakterystyka częstotliwościowa ucha zewnętrznego

"**Czy w uchu srodkowym dokonuje sie analiza czestotliwosciowa?" TAK**

Ucho wewnętrzne działa jak swoistego rodzaju detektor poziomu o stałym poziomie detekcji

Przetwarzanie sygnału akustycznego na obraz wyładowań neuronowych Dokonuje się to w ślimaku
Dwukrotnej zmianie częstotliwości (czyli o oktawę), niezależnie od zakresu, towarzyszy zmiana miejsca pobudzenia błony podstawnej o 3.5 – 5 mm

Trzy percepcyjne skale częstotliwości Bark Mel ERB

Własności skali Bark

- Równe odległości w skali częstotliwości odpowiadają równym odległościom w skali percepcyjnej
- 1 bark = 1 szerokości pasma krytycznego
- Powyżej 500 Hz skala ta jest równoważna logarytmicznej skali częstotliwości
- Poniżej częstotliwości 500 Hz skala Bark jest funkcją liniową częstotliwości
- Zakres zmian skali od 1 do 24, czyli obejmuje pierwsze 24 pasma krytyczne

Własności skali Mel

- Punktem odniesienia jest ton 1000 Hz o poziomie 40 dB – 1000 meli = wysokość tonu o częstotliwości 1000 Hz
- Dla każdego tonu dobiera się drugi ton o częstotliwości odbieranej subiektywnie jako o dwukrotnie niższej (lub wyższej) wysokości, lub dokonuje się podziału danego zakresu częstotliwości na 4 percepcyjnie jednakowe interwały
- Do 500 Hz skala meli pokrywa się ze skalą częstotliwościową. Powyżej – zależność jest logarytmiczna
- 100 mel = 1 Bark
- Filtry melowe znalazły zastosowanie w przetwarzaniu sygnału mowy

Własności skali ERB

- Skala ERB jest wyrażana w Hz
- Zakres 16 000 Hz dzieli się na 40 pasm
- Szerokość pasma również zależy od częstotliwości środkowej

9

Pasma krytyczne mają wpływ na:

- Detekcję sygnału w ciszy
- Percepcję głośności
- Detekcję sygnału w szumie (maskowanie)
- Czułość na przesunięcie fazowe
- I wiele innych zjawisk …….

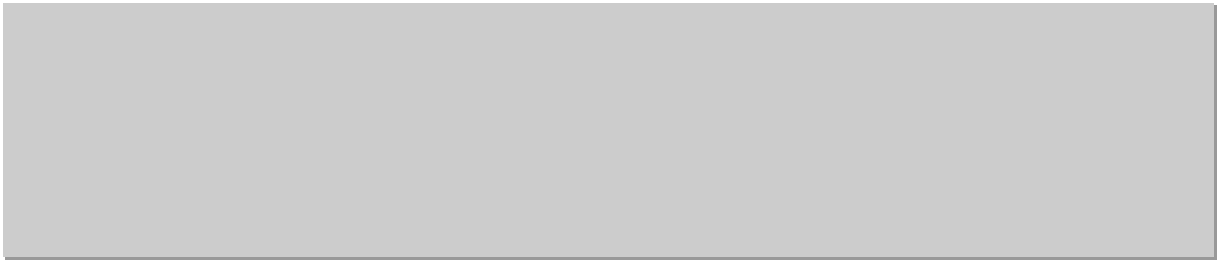
Czynniki wpływające na percepcję głośności

- Głośność dźwięku zależy od poziomu ciśnienia akustycznego
- Głośność dźwięku zależy od jego częstotliwości
- Głośność dźwięku zależy od jego zakresu częstotliwości
- Na wrażenie głośności dźwięku wpływają również czynniki czasowe

Pojęcie “rozdzielczości”
Określa dokładność z jaką można wyróżnić bodziec z pośród innych, o zbliżonych wartościach wybranego parametru

“**Rozdzielczość częstotliwościowa”**
Zdolność wyodrębnienia jednej składowej częstotliwościowej w dźwięku złożonym

Przeciętne wartości progów różnicy częstotliwości dla różnych zakresów



Energia niesiona przez dźwięk:

W polu idealnie rozproszonym intensywność dźwięku maleje odwrotnie proporcjonalnie do kwadratu odległości od źródła

Intensywność jest proporcjonalna do kwadratu ciśnienia skutecznego.
Im większa jest intensywność dźwięku, tym jest odbierany jako głośniejszy

Ile energii niesie sygnał mowy?
Sygnał mowy generowany przez mężczyznę niesie energię 34 μW, przez kobietę – 18 μW (pomiar w odległości 1 m)

Trzy współrzędne słyszenia przestrzennego: odległość, podniesienie, kat azymutalny (horyzontalny)

Międzyuszna różnica poziomów zależy od kąta padania, i również od częstotliwości fali

Różnica czasu ITD jest równoważna przesunięciu fazy. Minimalna postrzegana różnica kąta azymutalnego odpowiada minimalnej (10-20 μs) postrzegalnej różnicy czasu ITD.

Małżowina uszna ma określoną częstotliwościowo zależną charakterystykę kierunkową
Małżowina uszna wspomaga ocenę podniesienia źródła

Charakterystyka częstotliwościowa małżowiny jest bardziej czuła na kierunek góra – dół, niż lewo - prawo.

W ocenie wysokości położenia źródła, międzyuszne różnice intensywności (ILD) i czasu (ITD) nie odgrywają istotnej roli

Logarytmiczna skala kwantyzacji daje lepsze odwzorowanie cichszych dźwięków, niż liniowa