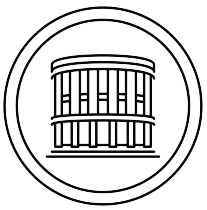
UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



# ÚLOHA MATEMATIKY PRI MODELOVANÍ FUNGOVANIA ČASTÍ AUDITÓRNEHO SYSTÉMU ČLOVEKA DÔLEŽITÝCH PRI VNÍMANÍ HUDBY

DIPLOMOVÁ PRÁCA

2025 Bc. Zuzana TOMČÍKOVÁ

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

# ÚLOHA MATEMATIKY PRI MODELOVANÍ FUNGOVANIA ČASTÍ AUDITÓRNEHO SYSTÉMU ČLOVEKA DÔLEŽITÝCH PRI VNÍMANÍ HUDBY

## DIPLOMOVÁ PRÁCA

| | |
|---|---|
| Študijný program: | Ekonomicko-finančná matematika a modelovanie |
| Študijný odbor: | Matematika |
| Školiace pracovisko: | Katedra aplikovanej matematiky a štatistiky |
| Vedúci práce: | doc. RNDr. Peter Mederly, CSc. |

Bratislava 2025 **Bc. Zuzana TOMČÍKOVÁ**






Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

# ZADANIE ZÁVEREČNEJ PRÁCE

**Meno a priezvisko študenta:** Bc. Zuzana Tomčíková
**Študijný program:** ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
**Študijný odbor:** matematika
**Typ záverečnej práce:** diplomová
**Jazyk záverečnej práce:** slovenský
**Sekundárny jazyk:** anglický

**Názov:** Úloha matematiky pri modelovaní fungovania častí auditórneho systému človeka dôležitých pri vnímaní hudby
*The role of mathematics in modeling of functioning of the parts of the human auditory system important in music*

**Anotácia:** Existencia vzťahu medzi hudbou a matematikou je známa už staročia. Hudba inšpirovala k formulácii a riešeniu problémov v rôznych oblastiach matematiky, na druhej strane matematika pomohla popísať, pochopiť a vyriešiť viaceré problémy v oblasti hudby. V súčasnosti sa matematika zásadným spôsobom uplatňuje aj v oblasti neurovedy zameranej na problematiku fungovania auditórneho systému človeka vo vzťahu k vnímaniu hudby.

**Vedúci:** doc. RNDr. Peter Mederly, CSc.
**Katedra:** FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
**Vedúci katedry:** doc. Mgr. Igor Melicherčík, PhD.

**Dátum zadania:** 11.01.2024

**Dátum schválenia:** 25.01.2024 prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

**Poďakovanie** Chcela by som sa poďakovať vedúcemu mojej diplomovej práce doc. RNDr. Petrovi Mederlymu, CSc. za poskytnutú literatúru, ochotu a odbornú pomoc pri vypracovaní práce ako aj pri konzultáciách. Ďalej chcem poďakovať aj svojej rodine za trpezlivosť a podporu.

# Abstrakt v štátnom jazyku

TOMČÍKOVÁ, Zuzana: Úloha matematiky pri modelovaní fungovania častí auditórneho systému človeka dôležitých pri vnímaní hudby [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: doc. RNDr. Peter Mederly, CSc., Bratislava, 2025, 89 s.

Diplomová práca sa zaoberá problematikou matematického modelovania vnímania výšky tónu, ktoré je základným predpokladom pre vnímanie hudby. Po stručnom úvode do problematiky vzťahov matematiky a hudby v prvej kapitole a predstavení potrebných nástrojov z oblasti matematiky, fyziky, anatómie a fyziológie auditórneho systému človeka v druhej a tretej kapitole sa práca sústreďuje na dve hlavné témy. Prvou z nich je charakteristika a klasifikácia existujúcich modelov vnímania výšky tónu obsiahnutá v štvrtej kapitole. Druhou hlavnou témou je podrobná analýza Meddis–Hewittovho časového modelu vnímania výšky tónu, ktorá je obsahom piatej kapitolu a vlastná programová implementácia tohto modelu vrátane jej otestovania na synteticky generovaných aj reálnych zvukových signáloch popísaná v šiestej kapitole. V prílohe práce je uvedený kód programu v jazyku Python.

**Kľúčové slová:** zvukový signál, sluchový vnem, výška tónu, gamatónový filter, auditórny systém človeka, autokorelácia, Meddis–Hewittov model

# Abstract

TOMČÍKOVÁ, Zuzana: The role of mathematics in modeling of functioning of the parts of the human auditory system important in music [Diploma Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: doc. RNDr. Peter Mederly, CSc., Bratislava, 2025, 89 p.

This diploma thesis focuses on the problem of mathematical modeling of pitch perception, which is a fundamental prerequisite for music perception. After a brief introduction to the relationship between mathematics and music in the first chapter and the presentation of essential tools from mathematics, physics, anatomy and physiology of the human auditory system in the second and third chapters, the thesis focuses on two main topics. The first is the characterization and classification of existing models of pitch perception, covered in the fourth chapter. The second key topic is a detailed analysis of the Meddis–Hewitt temporal model of pitch perception, presented in the fifth chapter, along with our own program implementation of this model, including its testing on both synthetically generated and real audio signals, described in the sixth chapter. The appendix includes the Python source code of the program.

**Keywords:** auditory signal, auditory perception, pitch, gammatone filter, human auditory system, autocorrelation, Meddis–Hewitt model

# Obsah

**Úvod** **9**

**1 Matematika a hudba** **11**

**2 Fyzikálne základy vnímania hudby** **14**

2.1 Prvá fáza zvukového javu . . . . . 14

2.2 Druhá fáza zvukového javu . . . . . 16

2.3 Tóny . . . . . 17

2.4 Fourierova analýza . . . . . 18

2.5 Autokorelácia . . . . . 18

2.6 Digitálne filtre . . . . . 19

2.6.1 Gamatónový filter . . . . . 20

2.6.2 ERB . . . . . 21

**3 Fyziologické základy vnímania hudby** **22**

3.1 Kódovanie akustických informácií v nervovom systéme . . . . . 22

3.2 Periférna časť auditórneho systému . . . . . 23

3.3 Centrálna časť auditórneho systému . . . . . 28

**4 Modelovanie vnímania výšky tónu** **31**

4.1 Všeobecná charakteristika modelov vnímania výšky tónu . . . . . 31

4.2 Spektrálne modely vnímania výšky tónu . . . . . 34

4.2.1 História vývoja spektrálnych modelov . . . . . 35

4.3 Časové modely vnímania výšky tónu . . . . . 36

4.3.1 História vývoja časových modelov . . . . . 37

**5 Meddis–Hewittov model** **39**

5.1 Popis modelu . . . . . 39

5.1.1 Vplyvy vonkajšieho a stredného ucha . . . . . 40

5.1.2 Filtrácia bazilárnou membránou . . . . . 40

5.1.3 Mechanicko-nervová transdukcia . . . . . 41

5.1.4 Refraktérne efekty . . . . . 44

5.1.5 Distribúcia časových intervalov . . . . . 45
5.1.6 Sčítanie naprieč kanálmi . . . . . 46
5.1.7 Extrakcia výšky tónu . . . . . 46

**6 Programová implementácia Meddis–Hewittovho modelu 48**
6.1 Popis programu MH . . . . . 48
6.2 Použitie programu MH na signál s chýbajúcou fundamentálnou frekvenciou 59
6.3 Použitie programu MH na tón trúbky . . . . . 63
6.4 Použitie programu MH na zvuk tibetskej spievajúcej misky . . . . . 67

**Záver 71**

**Literatúra 73**

**Príloha A 76**

# Úvod

Matematika a hudba sa na prvý pohľad môžu javiť ako dve úplne odlišné oblasti – jedna založená na presných výpočtoch a formálnych pravidlách, druhá na emóciách, tvorivosti a estetickom vnímaní. Pri hlbšom pohľade však odhalíme pozoruhodné spojenie medzi týmito dvoma svetmi. Už v antickom Grécku si Pythagoras všimol súvislosť medzi dĺžkami strún a výškou tónov [14], čím položil základy matematického chápania hudobných intervalov.

Odvtedy sa vzťah medzi matematikou a hudbou neustále prehlboval. Matematika dnes predstavuje nevyhnutný nástroj na skúmanie akustických javov a spracovanie zvukových signálov. Umožňuje presne modelovať šírenie zvuku a jeho interakciu s ľudským uchom.

Diplomová práca sa venuje matematickému modelovaniu vnímania výšky tónu, ktoré je základným prvkom v procese vnímania hudby. V tejto súvislosti sa v nej sústreďujeme na dve hlavné témy. Prvou je charakteristika a klasifikácia existujúcich modelov vnímania výšky tónu a druhou je podrobná analýza a programová implementácia Meddis–Hewittovho časového modelu vnímania výšky tónu, ktorý zohral kľúčovú úlohu pri vývoji v tejto oblasti. Problematikou matematického modelovanie sluchového vnímania hudby sa zaoberala aj diplomová práca [5] z roku 2023. Súčasťou jej výsledkov bola programová implementácia časti Goldsteinovho spektrálneho modelu, o ktorom hovoríme v 4.2.1.

Prístup, ktorý sme v práci použili, spája matematické modelovanie, digitálne spracovanie signálu a poznatky z anatómie a neurofyziológie auditórneho systému človeka. Budeme vychádzať najmä z [9], [11], [14] a [18]. Diplomová práca je rozdelená do šiestich kapitol.

V prvých troch si pripravujeme potrebné nástroje na splnenie cieľov práce a v druhých tieto ciele napĺňame. Prvá kapitola je úvodom do problematiky vzťahov medzi hudbou a matematikou.

V druhej kapitole sa venujeme matematickým a fyzikálnym nástrojom, ktoré budeme pri modelovaní procesu vnímania výšky tónu potrebovať. Tieto zahŕňajú základný model zvukových javov a charakterizáciu vlastností zvukových signálov a ich šírenia. Z matematických nástrojov načrtneme použitie Fourierovej analýzy ako nástroja pre frekvenčnú dekompozíciu komplexných tónov, použitie autokorelácie ako nástroja pre zisťovanie periodicity a na záver kapitoly sa venujeme digitálnym filtrom.

V tretej kapitole charakterizujeme auditórny systém človeka. Popíšeme anatomickú štruktúru periférnej časti auditórneho systému a fyziológiu jej fungovania. Následne stručne popíšeme aj centrálnu časť auditórneho systému.

Štvrtá kapitola je venovaná problematike modelov vnímania výšky tónu. Obsahuje charakteristiku, klasifikáciu a históriu týchto modelov.

Piata a šiesta kapitola je venovaná Meddis–Hewittovmu časovému modelu vnímania výšky tónu, ktorý je jedným z najvýznamnejších modelov tohto druhu a podstatne ovplyvnil vývoj v tejto oblasti. V piatej kapitole sa venujeme jeho popisu a analýze a v šiestej predstavujeme vlastnú programovú implementáciu tohto modelu v jazyku Python a jej použitie na 4 rôzne typy zvukových signálov, a to tak počítačovo syntetizovaných, ako aj generovaných hudobnými nástrojmi.

Príloha obsahuje kód programu.

# 1 Matematika a hudba

Hudba je neodmysliteľnou súčasťou ľudského života už tisícročia. Svojou schopnosťou ovplyvňovať emócie výrazne formuje každodenné prežívanie človeka. Už od staroveku sa ľudia pokúšali pochopiť, prečo má hudba taký silný vplyv na naše emócie. Táto otázka sa stala predmetom rozsiahleho vedeckého skúmania, ktoré sa zaoberá nielen fyzikálnou podstatou hudby, ale aj tým, ako ju človek vníma. Pri hľadaní odpovedí hrala od nepamäti kľúčovú úlohu matematika, ktorá pomáha odhaľovať zákonitosti skryté v hudbe a jej vplyve na emócie. Aj keď na prvý pohľad ide o dve rozdielne oblasti, matematiku, považovanú za exaktnú vedu a hudbu, ktorá zasahuje najmä emócie, ich prepojenie ľudí fascinuje už viac ako 2500 rokov. Táto téma prechádzala rôznymi obdobiami záujmu, pričom základná otázka, ako zvuk vzniká, šíri sa a je vnímaný, ostáva stále rovnako aktuálna.

Aby sme dosiahli cieľ nášho skúmania, teda čo najviac porozumieť tomu, ako hudba funguje a ako nám matematika pomáha pri jej pochopení, musíme vytvoriť model prostredia, v ktorom sa hudba vyskytuje a v ktorom pôsobí na človeka. Keďže hudba je formou zvukového javu, tento model sa bude týkať práve tohto druhu javov a interakcie medzi nimi a človekom. Tento model by mal zahŕňať nielen fyzikálnu podstatu zvukových javov, ale aj mechanizmus ich vplyvu na človeka, vrátane anatomických štruktúr a fyziologických procesov, ktoré sú nevyhnutné na vnímanie zvuku.

Zvukový jav možno podľa [16] vo všeobecnosti opísať ako sériu procesov zahŕňajúcich tri základné zložky alebo fázy, ktorými sú: zdroj zvukového signálu, prenosové médium a prijímateľ. Zdroj vytvára zvukové vibrácie generujúce zvukový signál, médium tento signál prenáša a prijímateľ ho zachytáva a spracováva. Medzi týmito zložkami dochádza k prenosu energie vo forme zvukových vĺn tvorených šíriacim sa zvukových signálom. Zvukový signál je tvorený osciláciami tlaku vo forme, rýchleho striedania pozdĺžneho stláčania a rozpínania vzduchu, na ktoré je naše ucho citlivé.

Na to, aby sme lepšie pochopili túto sériu procesov v kontexte hudby, môžeme sa pozrieť na konkrétny príklad. Predpokladajme, že zdrojom zvukového signálu je nejaký hudobný nástroj, ktorý pozostáva z niekoľkých základných fyzikálnych komponentov:

- Primárny excitačný mechanizmus, pod ktorým si môžeme predstaviť napríklad ťah sláčika po strune na husliach, fúkanie vzduchu na hranu flauty alebo plátok klarinetu, prúdenie vzduchu medzi perami hráča na nátrubkovom dychovom nástroji alebo

medzi hlasivkami speváka.

- Základný vibrujúci element, ktorý pri aktivácií primárnym mechanizmom dokáže udržiavať presne definované vibračné módy na konkrétnych frekvenciách, ako napríklad struny huslí, vzduchový stĺpec v dychových nástrojoch alebo vzduchový stĺpec v hlasovom trakte speváka. Tento element určuje výšku tónu a poskytuje harmonické zložky, ktoré ovplyvňujú farbu tónu.

- Rezonátor, ako je rezonančná doska klavíra či telo sláčikového nástroja, ktorý zosilňuje oscilácie vibrujúceho elementu a transformuje ich na zvukové vlny v okolí (v prípade dychových nástrojov a spevákov plní vzduchový stĺpec v tomto modeli úlohu tak vibrujúceho elementu ako aj rezonátora).

Prenosové médium predstavuje vzduch, ktorý prenáša zvukové vlny od zdroja k poslucháčovi. Zvukové vlny sú ovplyvňované hranicami priestoru – stenami, stropom, podlahou či ľuďmi, ktorí spôsobujú odrazy a absorpciu zvuku.

Za prijímateľa zvukového signálu budeme pokladať poslucháča, ktorého auditórny systém zachytáva zvukové vlny, prevádza ich na elektrické signály v sluchovom nerve, ktorých výsledkom po ďalšom spracovaní v mozgu je sluchový vnem. Proces spracovania zvukového signálu auditórnym systémom človeka podrobnejšie popíšeme v kapitole 3.

V predchádzajúcom popise sme namiesto jednoduchého slova zvuk používali výrazy „zvukový signál“ a „sluchový vnem“. Pre pochopenie zvukových javov je totiž dôležité zdôrazniť rozdiel medzi fyzikálnou podstatou zvuku a jeho psychologickým vnímaním. Zvukový signál ako fyzikálny jav predstavuje merateľné vibrácie šíriace sa prostredím, zatiaľ čo sluchový vnem je subjektívnym psychologickým zážitkom, ktorý vzniká spracovaním týchto vibrácií v ľudskom mozgu. V dôsledku toho je potrebné rozlišovať objektívne fyzikálne vlastnosti zvuku ako zvukového signálu a subjektívne vnímané atribúty zvuku ako sluchového vnemu. V našej práci budeme z tohto dôvodu štandardne používať presné označenia „zvukový signál“ a „sluchový vnem“. Jednoduché označenie „zvuk“ budeme v tejto práci používať len v prípadoch, keď nemôže prísť k nedorozumeniu.

Osobitným prípadom zvukových signálov, ktorý má kľúčový význam pre hudbu, sú periodické zvukové signály, ktoré sa nazývajú hudobnými tónmi, skrátene tónmi. Hlavnými vnímanými atribútmi hudobného tónu sú výška, hlasitosť a farba tónu. Výška tónu (angl.

*pitch*) sa definuje ako vlastnosť umožňujúca posúdiť vzájomnú polohu tónov na stupnici „nižší“ / „vyšší“. Hlasitosť predstavuje atribút odrážajúci silu sluchového vnemu. Farba tónu je atribút, ktorý umožňuje rozlišovať medzi tónmi rôznych hudobných nástrojov aj vtedy, keď majú rovnakú výšku a hlasitosť.

V ďalšej kapitole sa zameriame na podrobnejší popis prvých dvoch fáz zvukového javu vrátane charakterizácie niektorých fyzikálnych vlastností zvuku, ktoré vytvárajú základ pre naše sluchové vnímanie. Tieto fyzikálne charakteristiky zohrávajú kľúčovú úlohu v tvorbe hudby a jej interpretácii.

# 2 Fyzikálne základy vnímania hudby

Modelovanie prvej a druhej fázy zvukového javu, teda zdroja zvukového signálu a jeho šírenia prostredím, patrí do oblasti fyziky, ktorá úzko spolupracuje s matematikou. Fyzika skúma mechanizmy, ktorými vznikajú zvukové javy a matematika poskytuje nástroje na ich presný opis a analýzu. V tejto kapitole budeme vychádzať z [16] a [4].

## 2.1 Prvá fáza zvukového javu

Zdrojom zvukového signálu je vždy nejaký druh vibrácie alebo kmitania. Typickými príkladmi sú kmitanie struny hudobného nástroja, blany bubna či chvenie hlasiviek v prípade ľudského hlasu. Preto je prirodzené, že skúmanie zvukových javov sa začína štúdiom kmitavého pohybu, ktorý predstavuje základný fyzikálny proces vedúci k vytváraniu zvukových vĺn. Tieto vibrácie sú kľúčové nielen pre pochopenie samotného vzniku zvuku, ale aj pre jeho šírenie a spracovanie v ďalších fázach.

Kmitanie je druh mechanického pohybu, pri ktorom sa pohybujúci predmet alebo časť predmetu striedavo vzďaľuje a vracia do svojej rovnovážnej polohy. Klasickým príkladom je kyvadlo, ktoré sa pravidelne vracia do rovnovážnej polohy po tom, čo bolo vychýlené. Aby bol pohyb považovaný za periodický, musí sa pohybujúci predmet alebo časť predmetu vracať do rovnovážnej polohy v rovnakých časových intervaloch a medzi návratmi vykonávať rovnaký vzor pohybu. Časový interval, po ktorom sa tento vzor pohybu opakuje, nazývame perióda $T$. Základný vzor pohybu, ktorý sa opakuje počas jednej periódy, označujeme ako cyklus.

Najjednoduchším typom kmitavého pohybu je jednoduchý harmonický pohyb, pri ktorom sa objekt vychyľuje z rovnovážnej polohy a cyklicky sa do nej vracia pod vplyvom sily úmernej veľkosti vychýlenia. Príkladom takéhoto pohybu je závažie s hmotnosťou $m$ zavesené na pružine. V rovnovážnej polohe je súčet všetkých síl pôsobiacich na závažie nulový. Ak závažie vychýlime z tejto polohy, pružina začne na závažie pôsobiť silou, ktorá ho smeruje späť do rovnovážnej polohy. Táto sila je priamo úmerná vzdialenosti od rovnovážnej polohy a je vyjadrená vzťahom:

$$F = -ky, \tag{1}$$

kde $k$ je tuhosť pružiny a $y$ je výchylka z rovnovážnej polohy.

Po uvoľnení závažia z vychýlenej polohy sa začne pohybovať smerom k rovnovážnej polohe. Pri dosiahnutí rovnovážnej polohy má závažie maximálnu rýchlosť a pokračuje zotrvačnosťou ďalej. Po vzdialení sa od rovnovážnej polohy na opačnú stranu začne pružina opäť pôsobiť silou, ktorá pohyb spomaľuje, až závažie zastaví a začne sa vracať späť. Tento cyklus sa opakuje, pričom pohyb má periodický charakter.

Podľa druhého Newtonovho pohybového zákona bude táto sila vyjadrená:

$$F = ma, \tag{2}$$

kde $a$ je zrýchlenie, ktoré môžeme vyjadriť ako druhú deriváciu výchylky $y$ podľa času $t$:

$$a = \frac{d^2y}{dt^2}. \tag{3}$$

Kombináciou vzťahov (1), (2) a (3) dostávame lineárnu diferenciálnu rovnicu druhého rádu:

$$\frac{d^2y}{dt^2} + \frac{ky}{m} = 0. \tag{4}$$

Všeobecné riešenie tejto rovnice má tvar:

$$y = Acos(\sqrt{\frac{k}{m}}t) + Bsin(\sqrt{\frac{k}{m}}t).$$

Toto riešenie môžeme prepísať do tvaru:

$$y(t) = c\,sin(\sqrt{\frac{k}{m}}t + \phi).$$

Konštanta $c$ predstavuje maximálnu výchylku závažia do kladných aj záporných hodnôt a jej absolútna hodnota sa nazýva amplitúda. Závažie môže dosiahnuť maximálnu výchylku v čase $t = 0$ alebo aj v iných časoch. Pre tieto prípady bol zavedený parameter počiatočnej fázy $\phi$, ktorý umožňuje posúvať vrcholy sínusoidy popisujúcej pohyb závažia pozdĺž časovej osi podľa potreby.

Jednoduchý harmonický pohyb je periodický, pričom jeho perióda $T$ je definovaná vzťahom:

$$T = 2\pi\sqrt{\frac{m}{k}}.$$

Z toho pre frekvenciu $f$ vyplýva vzťah:

$$f = \frac{1}{T} = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

Ak je perióda $T$ vyjadrená v sekundách, potom sa frekvencia $f$, ako počet cyklov za sekundu, vyjadruje v hertzoch (Hz). Túto frekvenciu možno označiť za preferovanú alebo rezonančnú frekvenciu príslušného systému vykonávajúceho jednoduchý harmonický pohyb. Ak táto spadá do rozsahu počuteľného pre človeka, čo je približne 20 – 20 000 Hz, tak zvukové signály generované systémom kmitajúcim s touto frekvenciou je človek schopný vnímať.

## 2.2 Druhá fáza zvukového javu

V tejto časti sa budeme venovať prenosu zvukového signálu, konkrétne jeho prenosu vzduchom. Zdrojom zvukového signálu je vždy nejaký druh vibrácie alebo kmitania, napríklad kmitanie struny. Tieto vibrácie sa vzduchom prenášajú vo forme zvukových vĺn.

Z hľadiska fyziky sa dá vlnenie charakterizovať ako fyzikálny jav, pri ktorom prichádza k časovej a priestorovej periodickej zmene stavu prostredia, ktorá má formu zmeny hodnoty nejakej fyzikálnej veličiny, pričom táto zmena sa šíri v priestore. Vzhľadom na úzku väzbu vlnenia na kmitanie, sú viaceré základné pojmy týchto dvoch javov spoločné, ako napríklad perióda $T$, frekvenica $f$ a uhlová rýchlosť $\omega$, pričom $\omega = 2\pi f$. Pretože na rozdiel od kmitania, vlnenie prebieha nielen v čase, ale aj v priestore, pre jeho popis sa zavádza aj vlnová dĺžka, ktorou sa označuje vzdialenosť dvoch vrcholov $\lambda$, $\lambda = vT = \frac{v}{f} = \frac{2\pi v}{\omega}$, kde $v$ je rýchlosť šírenia vlnenia.

Z dôvodu existencie gravitácie je vzduch okolo nás pod tlakom, čo znamená, že pôsobí silou na objekty, ktoré sa v ňom nachádzajú. Táto sila sa nazýva atmosferický tlak. Zvuková vlna je v podstate malá opakovaná zmena tlaku vzduchu v mieste vzniku zvukového signálu, ktorá sa šíri v priestore. Vo vrcholoch zvukovej vlny je tlak vzduchu vyšší ako bežný atmosferický tlak a v údoliach zvukovej vlny je tlak vzduchu nižší ako tento tlak. Tieto rozdiely sú však malé oproti absolútnej veľkosti bežného atmosferického tlaku.

S každým vlnením sú spojené dva smery: smer šírenia a smer oscilácií, ktoré vlnenie spôsobujú. Podľa vzájomného vzťahu týchto dvoch smerov sa vlnenie delí na priečne a pozdĺžne. Pri priečnom vlnení je smer oscilácií kolmý na smer šírenia vlny . Priečne vlnenie môže existovať v médiu, ktoré kladie nejaký odpor voči ohýbaniu, to znamená, že pri ohybe vzniká spätná sila pôsobiaca proti tej, ktorá ohyb vyvolala. Príkladom takéhoto prostredia je napnutá struna. Pri pozdĺžnom vlnení je smer oscilácií rovnobežný so smerom šírenia

vlny. Pozdĺžne vlnenie vzniká v médiu, ktoré kladie odpor voči stláčaniu a v ktorom vzniká spätná sila pôsobiaca proti tej, ktorá stláčanie vyvolala napríklad vo vzduchu. Zvukové vlnenie širiace sa vo vzduchu má pozdĺžny charakter, pretože pri ňom prichádza k osciláciam molekúl vzduchu rovnobežne so smerom šírenia vlny.

## 2.3 Tóny

Zvuky okolo nás sú vo všeobecnosti veľmi bohaté a komplexné. Niektoré zo zvukov, ktoré sa využívajú pre vedecké účely, musia však byť jednoduché a stabilné, nesmú vykazovať zmeny amplitúdy ani frekvencie. Z uvedených dôvodov je najpoužívanejším zvukom používaným pre výskumné účely jednoduchý tón. Ako už sme uviedli, slovom tón označujeme zvukový signál s periodickým priebehom. Potom jednoduchý tón môžeme definovať ako zvukový signál so sínusovým priebehom. Len zriedkavo sa s ním stretneme v bežných podmienkach, ale vyjasnenie jeho vzťahu so zvukmi, s ktorými sa stretávame v reálnom živote, je dôležité pri pochopení fungovania sluchu.

Keď brnkneme na strunu gitary, neprodukuje iba jednoduchý tón zodpovedajúci rezonančnej frekvencii celej struny, ale aj zmes, ktorá obsahuje vyššie harmonické s frekvenciami dvojnásobku, trojnásobku, štvornásobku alebo $n$-násobku rezonančnej frekvencie. Harmonický rad tónov je postupnosť zvukových signálov so sínusovým priebehom, ktorých frekvencie tvoria harmonický rad frekvencií.

Najnižšia frekvenčná zložka, ktorá zodpovedá rezonančnej frekvencii celej struny, sa nazýva fundamentálna frekvencia. Frekvenčné zložky, ktoré vznikajú rezonanciou polovice, tretiny, štvrtiny, $n$-tiny struny, sa nazývajú vyššie harmonické. Fundamentálnu frekvenciu $f_0$ považujeme zároveň za prvú harmonickú, takže $n$-tá harmonická má hodnotu $n$-násobku fundamentálnej frekvencie $f_0$:

$$f_n = nf_0.$$

Ako už vieme, periodické javy v prírode majú často formu jednoduchých harmonických pohybov, ktoré môžeme matematicky opísať pomocou sínusoidy. Tieto jednoduché pohyby však v skutočnom svete často neexistujú osamotene. Väčšina tónov, ktoré počujeme, je zložená z viacerých harmonicky súvisiacich zložiek, ktoré vytvárajú zložený zvukový signál. Pri analýze takýchto zvukov a pochopení ich štruktúry nám výrazne pomáha Fourierova analýza.

## 2.4 Fourierova analýza

Fourierova analýza a syntéza sú kľúčové matematické nástroje na pochopenie toho, čo sa v hudbe deje. Založené sú na Fourierovej vete, ktorá tvrdí, že každý periodický signál, vrátane zvukových vĺn, môže byť rozložený na súčet jednoduchých sinusových vĺn s rôznymi amplitúdami, frekvenciami a fázami. Tieto nástroje sa stali základom modernej akustiky a sú nevyhnutné pre analýzu a spracovanie signálov v rôznych oblastiach, vrátane hudby.

Na základe Fourierovej vety môže byť podľa [14] každý periodický signál vyjadrený ako funkcia času vo forme súčtu nekonečného radu, ktorý môžeme symbolicky zapísať v tvare:

$$x(t) = \sum A_k \cdot \cos(2\pi f_k t + \phi_k), \tag{5}$$

pričom frekvencie všetkých zložiek sa dajú vyjadriť ako celočíselné násobky fundamentálnej frekvencie $f_0$. Praktickou výhodou tejto metódy je, že aj keď teoreticky môžeme na opis signálu potrebovať nekonečný počet sínusoíd, v praxi je možné vybrať konečnú podmnožinu, ktorá signál dostatočne presne priblíži.

Fourierova analýza priraďuje signálu amplitúdové spektrum a fázové spektrum, ktoré vyjadrujú intenzitu a fázové posuny jednotlivých frekvencií. Fourierova syntéza naopak priradí týmto spektrám funkciu, ktorá sa skladá z týchto zložiek a vytvára výsledný signál.

## 2.5 Autokorelácia

Všeobecný algoritmus na zisťovanie výšky tónu v časových modeloch je založený na autokorelácii. Autokorelačná funkcia umožňuje porovnávať signál so samým sebou, ale posunutý v čase. Porovnávanie sa vykonáva pomocou súčinu hodnôt signálu v časoch $x(t)$ a $x(t+\tau)$. Pre spojitý signál $x(t)$ definovaný pre čas od $t_0$ až $T$ je autokorelačná funkcia definovaná ako:

$$r(\tau) = \int_{r_0}^{T} x(t)x(t-\tau)dt. \tag{6}$$

Táto funkcia dosahuje maximálne hodnoty pre rôzne časové posuny, ktoré sú násobkami základnej periódy zvukového signálu. Napríklad pri jednoduchom tóne bude autokorelačná funkcia dosahovať maximum pri $\tau = 0$ a následne pri intervaloch zodpovedajúcich perióde tónu. Prvé maximum, ktoré sa objaví pri nenulovom posune, nám poskytuje informáciu o perióde zvukového signálu, čo je kľúčové pre určenie výšky tónu. [14]

## 2.6 Digitálne filtre

Podľa [19] pod digitálnym filterom chápeme taký filter, ktorý pracuje s digitálnymi signálmi, ako je napríklad zvuk reprezentovaný v počítači. Ide o výpočtovú operáciu, ktorá berie jednu postupnosť čísel (vstupný signál) a vytvára novú postupnosť čísel (filtrovaný výstupný signál).

Filtre majú dve využitia: separácia signálu a obnovu signálu. Separácia signálu je potrebná vtedy, keď je signál znečistený šumom alebo inými signálmi. Obnova signálu sa používa vtedy, keď bol signál nejakým spôsobom zdeformovaný. Každý lineárny filter má impulznú odozvu, skokovú odozvu a frekvenčnú odozvu. Každá z týchto odoziev obsahuje úplné informácie o filtri, ale v inej forme. Ak je daná jedna z týchto troch charakteristík, zvyšné dve sú tiež pevne dané a dajú sa priamo vypočítať.

Najpriamejší spôsob, ako zrealizovať digitálny filter, je vykonať konvolúciu medzi vstupným signálom a impulznou odozvou filtra. Takto možno vytvoriť všetky možné lineárne filtre. Ďalším spôsobom, ako vytvoriť digitálne filtre je rekurzia. Keď je filter zrealizovaný pomocou konvolúcie, každý vzorkovaný bod na výstupe sa vypočíta tak, že sú vstupné vzorky vhodne váhované a spočítané. Rekurzívne filtre sú rozšírením tohto postupu, používajú okrem vzoriek zo vstupu aj predchádzajúce výstupné hodnoty.

Impulzné odozvy rekurzívnych filtrov sú tvorené sinusoidami, ktoré exponenciálne klesajú v amplitúde. Teoreticky to znamená, že ich impulzné odozvy sú nekonečne dlhé. V praxi však amplitúda nakoniec klesne pod prahovú hodnotu systému a zvyšné hodnoty sa zanedbajú. Práve kvôli tejto vlastnosti sa rekurzívnym filtrom hovorí aj filtre s nekonečnou impulznou odozvou (angl. *infinite impulse response*, IIR). Na porovnanie filtre realizované konvolúciou sa nazývajú filtre s konečnou impulznou odozvou (angl. *finite impulse response*, FIR). [20], [17]

FIR filtre sú lineárne systémy [18]. Lineárne filtre nemôžu zavádzať nové frekvenčné zložky, môžu len zosilniť alebo stlmiť jednotlivé frekvenčné zložky vstupu a zmeniť ich fázu zavedením oneskorenia. Ak určitý systém reaguje na sínusový vstup sínusovým výstupom s rovnakou frekvenciou, považujeme to za znak toho, že systém je lineárny. To platí aj v prípade, že vstup je zmesou viacerých sínusových vĺn. V takom prípade bude výstup lineárneho filtra tiež zmesou frekvencií, pričom relatívne amplitúdy jednotlivých frekvenčných zložiek sa môžu dramaticky zmeniť, ale neobjavia sa frekvencie, ktoré vo

vstupe neboli prítomné.

### 2.6.1 Gamatónový filter

Gamatónový filter (angl. *gammatone filter*) predstavuje matematický model impulznej odozvy sluchového systému, ktorý vznikol na základe časovej analýzy spikov [6], [13]. Gamatónový filter je definovaný impulznou odozvou:

$$g(t) = t^{n-1} exp(-2\pi bt) cos(2\pi f_0 t + \phi), \quad t \geq 0 \tag{7}$$

Názov gamatónový vychádza z toho, že obálka impulznej odozvy má tvar gama funkcie a štruktúru impulznej odozvy predstavuje sínusová vlna s frekvenciou $f_0$. Parametre gamatónového filtra sú:

- $n$ je rád filtra a pri pevnom $b$ určuje tvar obálky. Čím je $n$ väčšie, tým je obálka menej asymetrická.
- $b$ (v Hz) je šírka pásma filtra a pri pevnom $n$ ovplyvňuje trvanie impulznej odozvy. Väčšie $b$ znamená kratšie trvanie odozvy.
- $f_0$ (v Hz) je frekvencia tónu.
- $\Phi$ (v radiánoch) je fáza tónu a určuje relatívne umiestnenie jemnej štruktúry tónu vzhľadom na obálku.

Dôvod pre použitie gamatónového filtra ako modelu sluchového filtra vychádza z toho, že fyziologické merania zo slimáka a psychofyzikálne merania u ľudí vykazujú veľmi podobné vlastnosti frekvenčnej selektivity. Teda to, ako rozlišujeme rôzne frekvencie zvuku na základe subjektívneho vnímania, sa zhoduje s tým, čo pozorujeme priamo na úrovni mechaniky vnútorného ucha.

Keby sme mali súbor takýchto gamatónových filtrov, pričom každý z nich by bol naladený na trochu inú frekvenciu a boli by usporiadané podľa frekvencie, mohli by sme pomocou takéhoto súboru gamatónových filtrov vykonávať podrobnú frekvenčnú analýzu prichádzajúcich zvukových signálov a v reálnom čase počítať niečo veľmi podobné spektrogramu. A to takým spôsobom, že by sme nechali zvuky prechádzať všetkými filtrami paralelne, ako prichádzajú. [18]

### 2.6.2 ERB

Periférna časť auditórneho systému sa často modeluje ako sústava prekrývajúcich sa pásmových filtrov (angl. *bandpass filters*), ktoré prepúšťajú signály v určitom frekvenčnom pásme medzi dolnou a hornou prahovou frekvenciou a potláča frekvencie mimo tohto pásma [17], čo úzko súvisí s pojmom kritického pásma. Mnohí autori sa zhodujú, že šírka kritického pásma súvisí so šírkou sluchového filtra pri danej centrálnej frekvencii [12]. Konkrétny tvar amplitúdovej charakteristiky gamatónového filtra štvrtého rádu sa výrazne podobá funkcii, ktorá sa štandardne používa na opis vnímaného sluchového filtra u ľudí. Táto podobnosť naznačuje, že šírka pásma filtra korešponduje s približne konštantnou vzdialenosťou na bazilárnej membráne.

Glasberg a Moore [2] odhadli, že ekvivalentná obdĺžniková šírka pásma (angl. *equivalent rectangular bandwidth*, ERB) sluchového filtra v Hz sa dá vyjadriť nasledovne:

$$ERB = 24.7(\frac{4.37 f_c}{1000} + 1), \tag{8}$$

kde $f_c$ predstavuje centrálnu frekvenciu filtra v Hz. S rastúcou frekvenciou sa sluchové filtre v slimáku rozširujú, ich šírka (ERB) rastie približne úmerne s $f_c$ [18]. Naopak, vzdialenosť medzi harmonickými zložkami komplexného tónu, ktorá je daná fundamentálnou frekvenciou $f_0$ , zostáva konštantná. To znamená, že vo vyšších frekvenčných oblastiach sa harmonické zložky stávajú spektrálne hustejšími vzhľadom na šírku filtrov. Výsledkom je, že nižšie harmonické sú v sluchovom systéme dobre rozlíšiteľné, každá z nich stimuluje iné miesto na bazilárnej membráne. Naopak, vyššie harmonické sa v rámci jedného filtra prekrývajú, a preto nie sú samostatne rozlíšiteľné.

Niekedy je užitočné škálovať frekvenciu pomocou jednotiek ERB [2]. Rovnicu, ktorá vyjadruje vzťah medzi počtom ERB k frekvencii, možno získať ako:

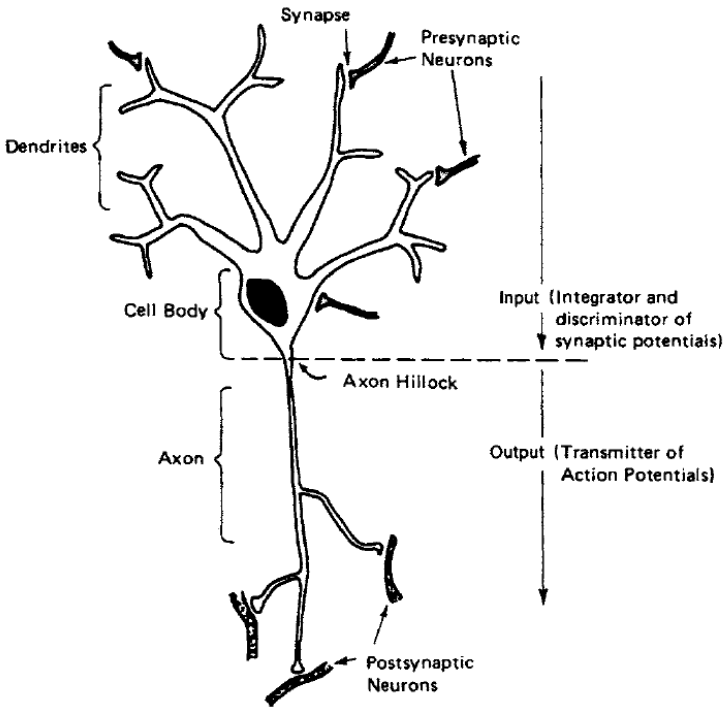
$$ERB(f_c) = 21.4\, log_{10}(\frac{4.37 f_c}{1000} + 1). \tag{9}$$

# 3 Fyziologické základy vnímania hudby

V tejto kapitole sa budeme venovať spôsobu, akým sa zvukový signál, ako mechanické vlnenie šíriace sa prostredím stáva sluchovým vnemom. Fyziologické procesy zapojené do tejto transformácie začínajú, keď je zvuk zakódovaný a spracovaný ako nervová aktivita štruktúrami ucha. Auditórny systém, ktorý je zodpovedný za vnímanie zvuku, môžeme rozdeliť na dve hlavné časti: periférnu časť a centrálnu časť. Každá z nich zohráva špecifickú úlohu v procese transformácie zvukového signálu na vnímaný zvuk. Vychádzať budeme najmä z [18].

## 3.1 Kódovanie akustických informácií v nervovom systéme

Procesy v tejto podkapitole, v ktorej čerpáme z [16], sú kľúčové pre pochopenie, ako nervový systém kóduje akustické informácie a akým spôsobom tieto informácie putujú cez nervový systém až do mozgu, kde sa vykonáva ich konečné spracovanie a interpretácia.




**Obr. 1:** Štruktúra neurónu. Dendrity prijímajú vstupné signály od presynaptických neurónov, ktoré sa následne spracúvajú a vyhodnocujú v oblasti axónu. Táto oblasť rozhoduje, či sa signál dostatočne zosilní na vytvorenie akčného potenciálu. (Zdroj [16]).

Neurón je základná bunka nervového systému, ktorá je zodpovedná za prijímanie,

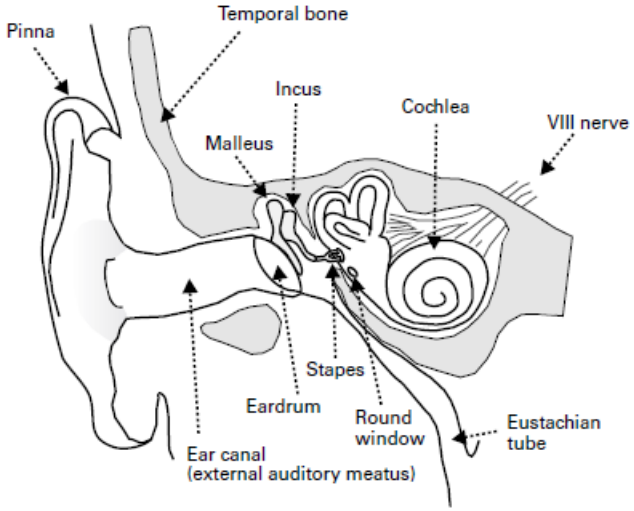
spracovanie a prenášanie elektrických signálov v tele. Tri základné časti neurónu, ktoré môžeme vidieť na Obr. 1 sú telo neurónu, dendrity a axón. Telo neurónu je zodpovedné za spracovanie signálov, ktoré prichádzajú z dendritov a rozhoduje, či sa vygeneruje akčný potenciál, ktorý sa ďalej prenáša axónom. Axón je dlhý výbežok neurónu, ktorý slúži na prenos elektrických signálov od tela neurónu k ďalším neurónom. Dendrity sú rozvetvené výbežky, ktoré zachytávajú signály z iných neurónov a odovzdávajú ich do tela neurónu na ďalšie spracovanie.

Neuróny generujú elektrické impulzy – akčné potenciály (ďalej len spiky), ktoré sa prenášajú cez synapsie, čo je štruktúra, ktorá umožňuje komunikáciu medzi dvoma neurónmi. Na jednej strane synapsie sa nachádza presynaptická bunka, ktorá uvoľňuje neurotransmitery a na druhej strane postsynaptická bunka, ktorá tieto signály prijíma. Keď spike dosiahne koniec axónu presynaptického neurónu, spôsobí uvoľnenie neurotransmiterov do synaptickej štrbiny. Tieto neurotransmitery sa viažu na špecifické receptory na postsynaptickej bunke, čo vedie k zmene jej elektrického potenciálu. Tento proces môže byť buď excitačný, čo zvyšuje pravdepodobnosť vyvolania akčného potenciálu v postsynaptickej bunke alebo inhibičný, čo znižuje pravdepodobnosť aktivácie postsynaptickej bunky. V tomto procese je kľúčový neurotransmiter ako glutamát, ktorý je najbežnejším excitačným neurotransmiterom.

## 3.2 Periférna časť auditórneho systému

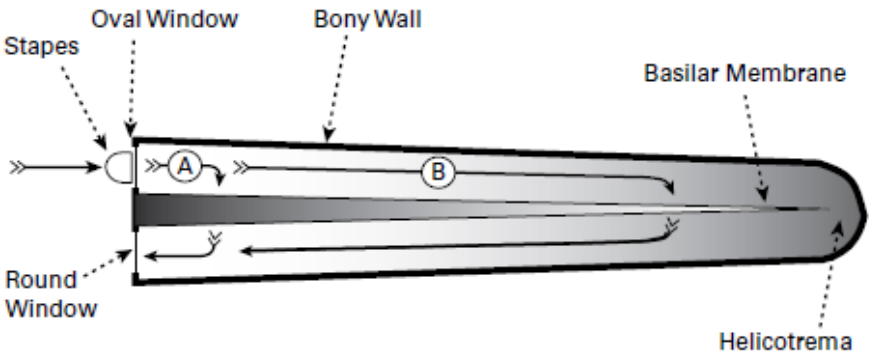
Sluchové vnímanie začína, keď zvukové vlny vstupujú do zvukovodu a tlačia na bubienok. Bubienok oddeľuje vonkajšie ucho od stredného ucha. Stredné ucho obsahuje systém troch malých kostičiek, ktoré sú známe ako kladivko, nákovka a strmienok. Účelom stredného ucha je sústreďovať tlak z bubienka na oveľa menšiu plochu strmienka a tým prenášať drobné zvukové vibrácie do slimáka, ktorý je súčasťou vnútorného ucha. Obr. 2 znázorňuje základnú stavbu periférneho auditórneho systému

Slimák je v podstate stočená trubica, uzavretá v tvrdej kostnej škrupine, vnútorne pozdĺžne rozdelená dvomi membránami do troch častí, ktoré sú naplnené fyziologickými tekutinami, nazývanými perilymfa a endolymfa. Vonkajšia stena slimáka pozostáva z pevnej kosti s membránovým obložením. Jediné otvory v tvrdej kostnej škrupine slimáka sú oválne okienko, ktoré sa nachádza priamo pod strmienkom a okrúhle okienko, ktoré sa




**Obr. 2:** Vonkajšie, stredné a vnútorné ucho. Vonkajšie ucho tvorí ušnica a zvukovod. Stredné ucho obsahuje tri kostičky (kladivko, nákovka, strmienok), ktoré zosilňujú vibrácie a prenášajú ich k slimáku vo vnútornom uchu. Vnútorné ucho zahŕňa slimáka, kde sa zvukové vibrácie premieňajú na spiky a začiatok sluchového nervu, ktorý tieto spiky prenáša do mozgu. (Zdroj [18])

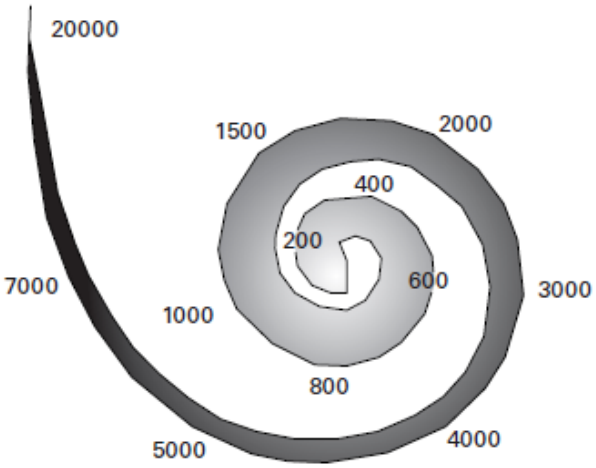
nachádza nižšie, ako môžeme vidieť na Obr. 3 . Keď strmienok vibruje v rytme zvukového signálu, prenášajú sa tieto vibrácie prostredníctvom jemnej membrány uzatvárajúcej oválne okienko do tekutinou naplnených priestorov slimáka a zvyšuje sa tak v týchto priestoroch tlak. Keďže tekutiny v slimáku sú nestlačiteľné a takmer úplne obklopené tvrdou kostnou škrupinou, tlak na oválne okienko spôsobí pohyb tekutiny v slimáku, čo spôsobí pohyb membrány uzatvárajúcej okrúhle okienko smerom von.




**Obr. 3:** Šírenie zvukových vibrácií v slimáku, ktorý je znázornený v rozvinutej forme. Rôzne frekvencie zvuku rozkmitajú bazilárnu membránu na rôznych miestach, pričom vyššie frekvencie sú bližšie k oválnemu okienku a nižšie frekvencie sú bližšie k vrcholu slimáka. (Zdroj [18]).

Takmer po celej dĺžke slimáka sa nachádza štruktúra známa ako bazilárna membrána, ktorá je jednou z dvoch membrán rozdeľujúcich vnútro slimáka tri komory: scala vestibuli, scala media a scala tympani. Bazilárna membrána má zaujímavé vlastnosti, je úzka a pevná na bazálnom konci slimáka, teda blízko oválneho a okrúhleho okienka, ale široká a pružná na vzdialenom, apikálnom konci, teda pri vrchole slimáka.

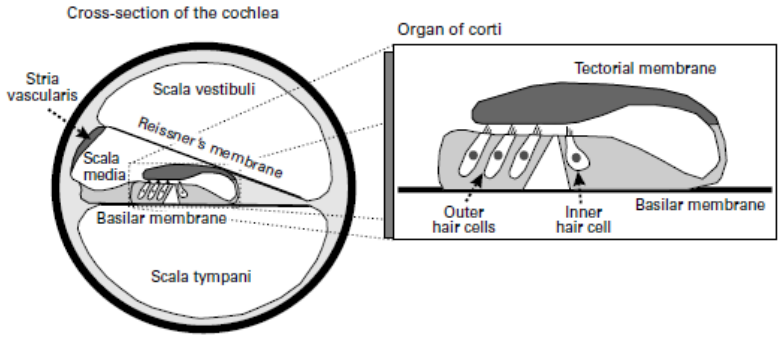
Slimák je vybavený dvoma zdrojmi mechanického odporu proti vibráciam prenášaným do jeho vnútra zo stredného ucha. Jedným je tuhosť bazilárnej membrány a druhým je zotrvačnosť tekutín v slimáku. Veľkosť týchto odporov sa pozdĺž slimáka mení, pričom táto zmena prebieha v navzájom opačných smeroch, čo môžeme vidieť na Obr. 4. Gradient tuhosti bazilárnej membrány klesá, keď sa vzďaľujeme od bazálneho konca slimáka, naopak zotrvačný gradient tekutín v slimáku v tomto smere stúpa. Prichádzajúca vibrácia bude teda hľadať kompromisnú cestu, ktorá je dostatočne dlhá na to, aby tuhosť klesla, ale nie taká dlhá, aby zotrvačný odpor výrazne vzrástol. Keďže zotrvačný odpor závisí od frekvencie, optimálna cesta s celkovým najnižším odporom závisí od frekvencie. Týmto spôsobom má každý bod bazilárnej membrány svoju vlastnú optimálnu frekvenciu, ktorá spôsobí, že tento bod bude vibrovať viac než ktorýkoľvek iný.




**Obr. 4:** Rozloženie frekvencií v Hz na bazilárnej membráne v slimáku. Vysoké frekvencie rozkmitajú bazilárnu membránu na začiatku slimáka pri oválnom okienku, zatiaľ čo nízke frekvencie na jeho vrchole. Toto rozloženie umožňuje slimáku analyzovať zvuky podľa ich výšky. (Zdroj [18]).

Pozdĺž celej dĺžky bazilárnej membrány sa na jej hornej strane nachádza Cortiho orgán.

Keď sa časti bazilárnej membrány rozvibrujú v reakcii na prichádzajúci zvukový signál, zodpovedajúce časti Cortiho orgánu sa dvíhajú a klesajú spolu s membránou. Cortiho orgán má zvláštny preložený tvar, ako môžeme vidieť na Obr. 5, kde spodná časť priamo sedí na bazilárnej membráne a nachádzajú sa na nej v radoch uložené senzorické vláskové bunky. Vláskové bunky sú špeciálnym typom receptorových buniek, ktoré na rozdiel od neurónov nevytvárajú akčné potenciály a nemajú axóny ani dendrity. Na strane ležiacej bližšie k stredu špirály slimáka sa Cortiho orgán preklápa a vytvára štruktúru pripomínajúcu strechu, ktorá sa nazýva tektoriálna membrána. Táto membrána prichádza do kontaktu so stereocíliami, čo sú vlásky, ktoré sa nachádzajú na senzorických vláskových bunkách. Keď sa Cortiho orgán pohybuje nahor a nadol, tektoriálna membrána kĺže po vrcholoch vláskových buniek. Pri pohybe nahor tlačí stereocílie smerom od stredu špirály slimáka, zatiaľ čo pri pohybe nadol ich tlačí opačným smerom.

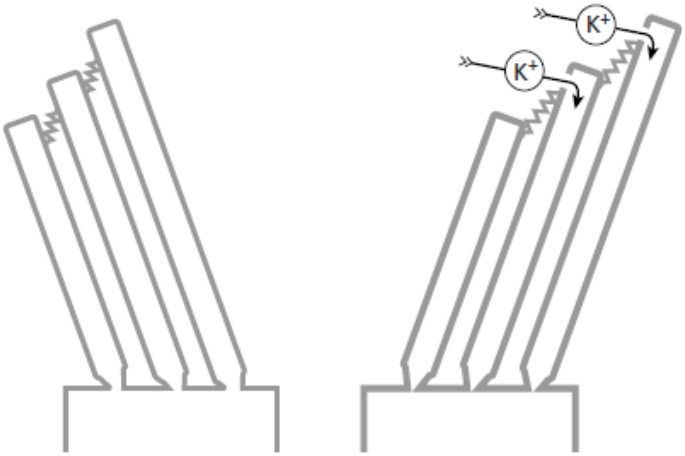



**Obr. 5:** Prierez slimákom a detailná štruktúra Cortiho orgánu. Slimák je rozdelený na tri komory: scala vestibuli, scala media a scala tympani, ktoré sú oddelené Reissnerovou a bazilárnou membránou. Na bazilárnej membráne je uložený Cortiho orgán, ktorý obsahuje vnútorné a vonkajšie vláskové bunky, pričom nad nimi sa nachádza tektoriálna membrána. Pohyb tektoriálnej membrány spôsobuje ohýbanie stereocílií vláskových buniek, čo vedie k ich depolarizácii a následnej premene mechanického signálu na spike.(Zdroj [18]).

Z Obr. 5 môžeme vidieť, že vláskové bunky existujú v dvoch druhoch a to vnútorné a vonkajšie vláskové bunky. Pre účely nášho modelovania sú dôležité len vnútorné vláskové bunky, preto sa v ďalšom budeme zaoberať len nimi.

Každá vláskova bunka obsahuje zväzok niekoľkých desiatok stereocílií, ktorých špičky sú spojené jemnými bielkovinovými vláknami. Posunutie zväzku vláskovej bunky smerom k najdlhšej stereocií spôsobí napätie na bielkovinových vláknach, zatiaľ čo posunutie

opačným smerom napätie uvoľní. Každý cyklus vibrácie bazilárnej membrány spôsobuje zodpovedajúci cyklus zvyšujúceho a klesajúceho napätia. Bielkovinové vlákna sú pripojené k iónovým kanálom, ktoré sa otvárajú vzhľadom na vytvorené napätie, čo umožňuje, aby ióny draslíka $K^+$ prúdili elektrickým a koncentračným gradientom z endolymfy do vláskovej bunky. Tento proces môžeme vidieť na Obr. 6.

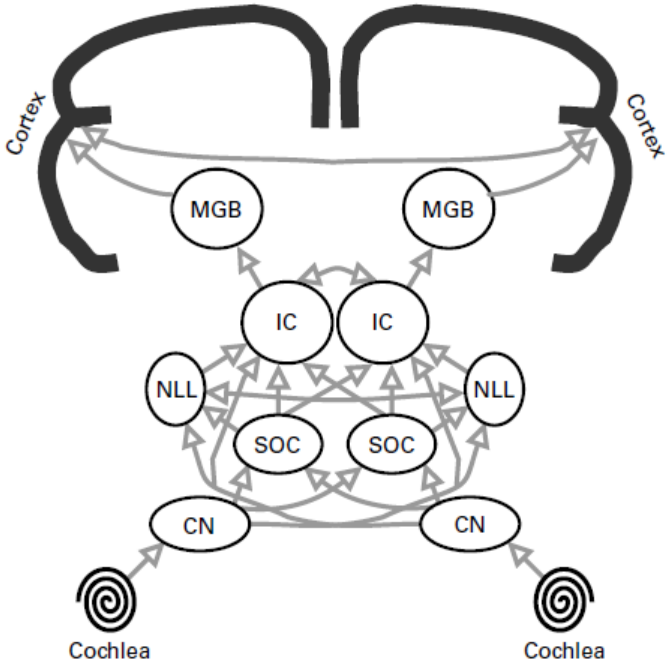



**Obr. 6:** Znázornenie pohybu stereocílií smerom k najdlhšej stereocií, vlákna sa napínajú, čo otvára iónové kanály a umožňuje prúdenie iónov draslíka $K^+$ z endolymfy do vláskovej bunky. Tento proces mení elektrické napätie vo vláskovej bunke a opakuje sa pri každom cykle vibrácie bazilárnej membrány. Táto mechanika umožňuje vláskovým bunkám prevádzať mechanické pohyby na elektrické signály. (Zdroj [18]).

Hlavnou úlohou vláskových buniek je prenášať zvukové signály pomocou glutamáterických synapsií prostredníctvom sluchového nervu. Depolarizáciou vláskových buniek, pri ktorej dochádza k otvoreniu iónových kanálov, sa uvoľňuje neurotransmiter glutamát, čo vedie ku generovaniu spiku vo vlákne sluchového nervu. Intenzita uvoľňovania glutamátú závisí od miery depolarizácie vláskovej bunky, ktorá priamo odráža amplitúdu vibrácií bazilárnej membrány. Čím silnejšie vibruje konkrétna časť bazilárnej membrány, tým viac glutamátu sa uvoľňuje a tým vyššia je frekvenica spikov. Tieto spiky sú systematicky usporiadané podľa tonotopie bazilárnej membrány.

## 3.3 Centrálna časť auditórneho systému

V tejto časti stručne popíšeme vzostupnú sluchovú dráhu (angl. *ascending auditory pathway*), to znamená cestu, ktorou postupujú v centrálnej časti auditórneho systému akustické informácie od výstupu zo slimáka až do najvyšších centier v sluchovej kôre (angl. *auditory cortex*, AC), ktorou sa rozumejú špecifické časti mozgovej kôry spojené primárne s fungovaním auditórneho systému. Budeme vychádzať z [15] a [18]. Anatómia sluchovej dráhy je veľmi komplikovaná a preto sa tu zameriame len na základnú charakteristiku jej hlavných bodov (angl. *processing centers, processing stations*), v ktorých prebieha spracovávanie akustických informácií, a prepojení medzi nimi. Pri popise budeme postupovať od výstupu zo slimáka až po najvyššiu úroveň auditórneho systému, ktorou je sluchová kôra.




**Obr. 7:** Centrálna časť auditórneho systému. Zvuk sa prenáša zo slimáka cez kochleárne jadrá (CN), ďalej cez superior olivary complex (SOC), lateral lemniscal nucleus (NLL), a nakoniec do inferior colliculus (IC). Signály sú následne smerované do medial geniculate body (MGB) v talame, ktoré ich prenášajú do sluchovej kôry v mozgu. (Zdroj [18]).

Po výstupe zo slimáka sa vlákna sluchového nervu stávajú súčasťou VIII hlavového nervu, ktorý sa nazýva vestibulokochleárny. Vlákna sluchového nervu následne vstupujú do kochleárneho jadra (angl. *cochlear nucleus*, CN) nachádzajúceho v mozgovom kmeni na

rozhraní predĺženej miechy a mostu. CN sa člení na ventrálnu časť (angl. *ventral cochlear nucleus*, VCN) a dorzálnu časť (angl. *dorsal cochlear nucleus*, DCN). VCN sa ďalej člení na prednú časť (angl. *anterior VCN*, AVCN) a zadnú časť (angl. *posterior VCN*, PVCN). Každé vlákno sluchového nervu sa po vstupe do CN rozvetvuje a vysiela svoje vetvy do všetkých troch častí CN. Každá z týchto vetiev vytvára početné synapsie s rôznymi typmi neurónov nachádzajúcich sa v troch častiach CN. Výstupy neurónov nachádzajúcich sa v CN smerujú ďalej do rôznych častí sluchovej dráhy. Hlavné body na tejto dráhe, v ktorých sa spracovávajú akustické informácie, sú schematicky zobrazené na Obr. 7.

Takmer všetky výstupy z CN sa nakoniec dostanú do inferior colliculus (IC), ktoré je prvým veľkým centrom spracovávajúcim akustické informácie v najvyššej časti mozgového kmeňa, v strednom mozgu. Časť z týchto výstupov sa do IC dostáva priamo a časť cez superior olivary complex (SOC) nachádzajúci sa v dolnej časti mostu v mozgovom kmeni. V SOC sa stretávajú a interagujú zvukové informácie z pravého a ľavého ucha, prichádza tu teda k binaurálnej interakcii. Toto miesto podstatne prispieva k priestorovému vnímaniu zvuku.

Axóny z CN a SOC potom idú vo zväzku vlákien známom ako lateral lemniscus do IC. Cestou môžu, ale nemusia, vysielať postranné vetvy do ventrálnych, intermediálnych a dorzálnych jadier v lateral lemniscus (angl. *nuclei of lateral lemniscus*, NLL). Pritom dráha z CN do IC je z väčšej časti prekrížená, takže neuróny v strednom mozgu a mozgovej kôre sú prevažne excitované zvukmi prichádzajúcimi z ucha na opačnej (kontralaterálnej) strane tela.

Aj samotné IC má zložitú organizáciu s prepojeniami medzi IC nachádzajúcim sa v pravej časti a IC nachádzajúcim sa v ľavej časti stredného mozgu. Tieto prepojenia umožňujú ďalšiu binaurálnu interakciu v rámci vzostupnej sluchovej dráhy. IC sa delí na viacero podjadier. Najväčším z nich je centrálne jadro (angl. *central nucleus of the IC*, ICC), do ktorého prichádza najviac vstupov z nižších úrovní mozgového kmeňa. Väčšina výstupov z IC smeruje do ďalšieho kľúčového bodu sluchovej dráhy nachádzajúceho sa v talame, ktorý je súčasťou medzimozgu, a to do jadra medial geniculate body (MGB), ktoré slúži ako relé (teda spínacie zariadenie) do mozgovej kôry.

Vo všetkých uvedených kľúčových bodoch vzostupnej sluchovej dráhy sa zachováva tonotopická organizácia, čo znamená, že neuróny v rámci týchto jadier sú viac-menej

naladené na konkrétne frekvencie (nazývané charakteristickými frekvenciami neurónu) a anatomicky usporiadané podľa týchto frekvencií. Výstupy z MGB smeruje prevažne do primárnych časti sluchovej kôry (angl. *auditory cortex*, AC) nachádzajúcich sa v spánkových lalokoch oboch mozgových hemisfér. Časti sluchovej kôry v oboch hemisférach sú prepojené tzv. komisurálnymi spojeniami (angl. *commissural connections*) cez svorové teleso (angl. *corpus callosum*). Svorové teleso všeobecne zabezpečuje vzájomné prepojenie mozgových hemisfér. Tieto prepojenia poskytujú ďalšiu možnosť výmeny informácií v auditórnom systéme iniciovaných signálmi prichádzajúcimi z oboch uší.

# 4 Modelovanie vnímania výšky tónu

V tejto kapitole uvedieme základné informácie o modeloch vnímania výšky tónu, ktorá sa dá pokladať za najdôležitejšiu vlastnosť sluchového vnemu. Bez tejto vlastnosti by sa hudba zredukovala len na rytmické zvuky, ľudská reč by stratila veľkú časť možností vyjadrenia emócií a značne by sa obmedzili možnosti človeka pri identifikácii zdrojov zvuku, ktorá je v mnohých prípadoch dôležitá aj z hľadiska jeho prežitia. Pri písaní tejto kapitoly sme vychádzali najmä zo zdrojov [14], [22].

## 4.1 Všeobecná charakteristika modelov vnímania výšky tónu

Skúmanie atribútu výšky sluchového vnemu je neoddeliteľnou súčasťou histórie skúmania zvuku a siaha až do jej zdokumentovaných začiatkov spojených s Pytagorovým skúmaním vzťahu medzi výškou tónu a dĺžkou struny, ktorá tón vydáva.

Ako sme už uviedli v kapitole 1, výškou tónu (angl. *pitch*) sa označuje jedna z vlastností sluchového vnemu. Ide o subjektívne vnímanú charakteristiku, ktorá odráža to, ako „vysoko“ alebo „nízko“ daný zvuk poslucháč vníma. Sluchové vnemy vznikajú u poslucháča na základe zvukových signálov. Nie všetky sluchové vnemy však majú vlastnosť výšky. Ukazuje sa, že predpokladom, aby poslucháč pri sluchovom vneme vnímal atribút výšky, je istá časová pravidelnosť zvukového signálu, ktorý sluchový vnem vyvolal. Vnem výšky štandardne vzniká pri tónoch, ktorými sa označujú periodické (alebo takmer periodické) zvukové signály. Vnem výšky však môže sprevádzať aj niektoré typy hlukov alebo šumov, teda nepravidelných zvukových signálov. Niekedy sa preto vyskytuje aj označenie „výška zvuku“, ale štandardne sa na označenie atribútu výšky sluchového vnemu používa spojenie „výška tónu“. Túto konvenciu dodržiavame aj v tejto práci. Výnimočne, keď to pokladáme za adekvátne, osobitne v tejto kapitole, používame aj dlhšie označenie „výška sluchového vnemu“ alebo jednoducho len „vnem výšky“. V angličtine tento terminologický problém nevzniká, pretože atribút výšky sluchového vnemu sa označuje jedným slovom „pitch“.

Na vysvetlenie fenoménu vnímania výšky tónu vzniklo množstvo modelov. Možno ich rozdeliť na dve základné skupiny, ktorými sú spektrálne modely vnímania výšky tónu (angl. *spectral pitch perception models*) a časové modely vnímania výšky tónu (angl. *tem-*

*poral pitch perception models*).

Zvukový signál ako časový sled tlakových zmien v nejakom prostredí, ktorým je štandardne vzduch, je obvykle charakterizovaný troma fyzikálnymi vlastnosťami: frekvenciou, amplitúdou a fázou. Periférna časť auditórneho systému transformuje každú z týchto vlastností do neurálneho kódu vo forme časovej postupnosti akčných potenciálov alebo spikov (angl. *spike trains*) vo vláknach sluchového nervu. Tieto následne vstupujú do centrálnej časti auditórneho systému v mozgu, kde po ďalšom spracovaní na jeho jednotlivých úrovniach nakoniec vznikne sluchový vnem, ktorý pre niektoré druhy zvukových signálov má aj atribút výšky. Aj keby sa mohlo zdať, že by nemalo byť náročné nájsť súvislosť medzi fyzikálnymi vlastnosťami zvukového signálu a výškou sluchového vnemu, ktorú vyvoláva, nie je to tak. Ukazuje sa, že charakterizovať jednotným spôsobom širokú škálu podmienok vedúcich k vnímaniu výšky sluchového vnemu je náročne a nedarí sa ho zatiaľ dosiahnuť.

Štúdium výšky sluchového vnemu je skúmaním vzťahu medzi fyzikálnymi vlastnosťami zvukového signálu, jeho transformácie v auditórnom systéme a samotným vznikom vnemu výšky. Hľadanie modelov vysvetľujúcich tento psychofyzikálny vzťah trvá stáročia a stále prebieha debata o tom, ktoré vlastnosti zvukového signálu sú rozhodujúce pre výšku sluchového vnemu, ktorý vyvoláva. Hľadané modely by mali fungovať v širokej škále kontextov, v rámci ktorých je výška dôležitou vlastnosťou zvukového vnemu.

Ukazuje sa, že niektoré z vyvinutých modelov sú vhodné pre niektoré kontexty, a iné modely pre iné kontexty. Univerzálny model, ktorý by vnem výšky vysvetlil v každej situácii, zatiaľ neexistuje. V tejto súvislosti si treba tiež uvedomiť, že pociťovaná intenzita vnemu výšky sluchového vnemu nie je vždy rovnaká a závisí od okolností, za ktorých vnem vzniká. Napríklad tón, v ktorom sú zastúpené len vysoké harmonické, spôsobuje menej intenzívne vnímanie výšky ako tón obsahujúci aj nízke harmonické.

V elementárnom prípade jednoduchého zvukového signálu, teda tónu so sínusovým priebehom, zodpovedá výška príslušného sluchového vnemu frekvencii tohto tónu. Aj takýto jednoduchý signál sa dá matematicky reprezentovať dvojakým spôsobom: pomocou jeho spektra (spektrálna reprezentácia) alebo prostredníctvom jeho časového priebehu, teda ako funkcia času (časová reprezentácia).

V prípade spektrálnej reprezentácie je takýto signál charakterizovaný jednoduchým spektrom priraďujúcim jeho amplitúdu a fázu jeho frekvencii. Výška sluchového vnemu

zodpovedá frekvencii spektrálneho komponentu. V prípade časovej reprezentácie je tento zvukový signál reprezentovaný funkciou sínus vyjadrujúcou závislosť tlakových zmien, ktoré ho tvoria (akustický tlak), od času. Výška sluchového vnemu v tomto prípade zodpovedá obrátenej hodnote periódy takejto funkcie. Uvedené dve reprezentácie zvukového signálu sú navzájom inverzné a možno medzi nimi prechádzať pomocou Fourierovej transformácie. Neplatí to len pre jednoduchý tón, ale aj pre zložitejšie zvukové signály.

Na fyzikálnej úrovni sú teda spektrálna a časová reprezentácia zvukového signálu ekvivalentné. Ak však uvažujeme o súvislosti fyzikálnych vlastností zvukového signálu a výšky sluchového vnemu, ktorý vyvoláva a ktorý vzniká na vyšších úrovniach auditórneho systému v mozgu, musíme si uvedomiť, že pri prechode auditórnym systémom je zvukový signál podstatne transformovaný. Zmeny, ktorými signál prechádza, osobitne pri prechode periférnou časťou auditórneho systému, ovplyvňujú tak jeho spektrálnu, ako aj časovú reprezentáciu, a to rôznym spôsobom. Po prechode periférnou časťou auditórneho systému strácajú spektrálna a časová reprezentácia transformovaného signálu svoju ekvivalentnosť, a tým sa mení aj to, akým spôsobom prispieva spektrum a časový priebeh zvukového signálu k výške príslušného sluchového vnemu.

Povedané inými slovami, ak vychádzame z experimentálne overeného poznatku, že výška sluchového vnemu súvisí s frekvenciou (a ekvivalentne s periódou) príslušného zvukového signálu, tak centrálny auditórny systém dostane po prechode zvukového signálu periférnou časťou auditórneho systému jeden typ informácií odvodený z jeho spektrálnej reprezentácie a druhý typ informácií odvodený z jeho časovej reprezentácie.

Takéto tvrdenie je možné urobiť na základe skutočnosti, že vnútorné ucho je schopné zabezpečiť tak istý druh frekvenčnej analýzy, ako aj analýzy časového priebehu zvukového signálu a výsledky týchto analýz transformovať do neurálneho kódu. Treba si však uvedomiť, že výsledky týchto dvoch analýz už nemusia byť navzájom ekvivalentné a úplná informácia o frekvencii vstupného zvukového signálu, ktorá bola obsiahnutá ekvivalentne v jeho spektrálnej aj časovej reprezentácii, sa po prechode periférnou časťou auditórneho systému rozpadla na dve neekvivalentné skupiny informácií, z ktorých by mal centrálny auditórny systém byť schopný „poskladať“ informáciu o frekvencii vstupného signálu ako podklad pre vygenerovanie výšky sluchového vnemu.

Modely vnímania výšky sluchového vnemu musia brať túto skutočnosť do úvahy, a

teda vychádzať skôr z neurálnych reprezentácií zvukového signálu na výstupe periférnej časti auditórneho systému ako len z fyzikálnych vlastností vstupnému zvukového signálu.

Spôsob, ktorým jednotlivé modely spracovávajú spektrálne a časové informácie, ktoré vznikli transformáciou vstupného zvukového signálu pri prechode periférnou časťou auditórneho systému, umožňuje ich rozdelenie na dve vyššie uvedené základné skupiny: spektrálne modely a časové modely. Spektrálne modely berú vo všeobecnosti ako východisko spektrálne infomácie, teda informácie o frekvenciách, ktoré sa dajú extrahovať z informácií na výstupe periférnej časti auditórneho systému a časové modely vychádzajú vo všeobecnosti z časových informácií z tohto zdroja.

## 4.2 Spektrálne modely vnímania výšky tónu

Spektrálne modely vychádzajú prioritne z frekvenčného obsahu zvukového signálu, ktorý sa po po prechode periférnou časťou auditórneho systému premieta do tonotopickej štruktúry informácií v sluchovom nerve a následne aj vo vyšších úrovniach auditórneho systému. Kľúčovú úlohu pritom zohráva skutočnosť, že miesta maximálnej citlivosti na jednotlivé frekvenčné zložky zvukového signálu sú distribuované pozdĺž celej bazilárnej membrány, čo umožňuje uvedenú tonotopickú štruktúru, to znamená väzbu jednotlivých frekvencií na miesta na bazilárnej membráne a následne na vlákna sluchového nervu, ktoré sú na tieto miesta naviazané. Rôzne frekvencie teda excitujú rôzne oblasti v mozgu.

Spektrálne modely sa preto zvyknú nazývať aj miestnou teóriou (angl. *place theory*) vnímania výšky tónu, podľa ktorej rôzne výšky tónu počujeme preto, lebo zvukové signály s rôznou frekvenciou excitujú rôzne neuróny, a tým aj rôzne miesta v mozgu. Na základe dlhoročných výskumov je známe, že miestnou teóriou je možné vysvetliť viaceré stránky vnímania výšky tónov. Podstatnou súčasťou spektrálnych modelov je tzv. porovnávanie vzorov (angl. *pattern matching, template matching*) vychádzajúce z predpokladu, že v auditórnom systéme prebieha porovnávanie tonotopicky zakódovanej frekvenčnej štruktúry vnímaného tónu so šablónami (angl. *template*) uloženými v našej pamäti, zodpovedajúcimi rôznym harmonickým spektrám, z ktorých každá je spojená s nejakou výškou tónu. Tá zo šablón, ktorá sa zhoduje s vnímaným spektrom najlepšie, potom poskytne vnem výšky tónu, ktorá je s touto šablónou spojená. Uloženie šablón do pamäte vzniká na základe učenia, prostredníctvom expozície rôznym zvukom s bohatou harmonickou štruktúrou.

Uplatňuje sa pritom Hebbov mechanizmus učenia, ktorý sa dá stručne charakterizovať pravidlom tvrdiacim, že ak sa dva neuróny aktivujú súčasne, zosilňujú sa medzi nimi vzájomné synaptické väzby (angl. *neurons that fire together, wire together*).

### 4.2.1 História vývoja spektrálnych modelov

Gréci ako prví porozumeli kvalitatívnej súvislosti medzi výškou tónu a frekvenciou, no presný kvantitatívny vzťah pre výpočet frekvencie bol stanovený až oveľa neskôr Marinom Mersennom (1636) a Galileom Galileim (1638), ktorí ho určili pre kmitajúcu strunu. Mersenne si všimol, že v zvukovom signáli generovanom strunou alebo hlasom, možno počuť aj výšky tónov zodpovedajúce prvým piatim harmonickým. Nerozumel však, ako môže struna vibrovať fundamentálnou frekvenciou a zároveň aj viacerými jej násobkami. Až Joseph Sauveur v roku 1701 experimentálne dokázal, že struna môže skutočne vibrovať súčasne na viacerých harmonických a zaviedol pojmy fundamentálna frekvencia a harmonická. Fourierova veta (5) dovtedy známu teóriu ešte viac skomplikovala, keďže ukázala, že periodický zvukový signál možno rozložiť na jednoduché sínusové vlny. Ak má každá z nich vlastnú periódu, vzťah medzi periódou a výškou tónu už nie je jednoznačný.

Helmholtz (1877) svoju teóriu zhrnul do štyroch hlavných bodov. Iba periodické signály obsahujúce fundamentálnu frekvenciu evokujú tón odpovedajúci tejto perióde. Harmonické môžu vyvolávať aj ďalšie tóny. Fázy harmonických neovplyvňujú výšku ani farbu tónu. Súčasťou teórie bola aj fyziologická zložka. Helmholtz videl bazilárnu membránu ako súbor pružných strún, pričom každá bola naladená na inú frekvenciu, od vysokých tónov pri báze, po nízke pri vrchole slimáka. Tento model interpretoval činnosť slimáka ako približnú Fourierovu transformáciu. Helmholtzova teória bola elegantná a prepojená s Fourierovou transformáciou, ktorá, rozkladala zvuk na sínusoidy. Avšak pri pozorovaní niektorých experimentov a teoretických modelov sa ukázalo, že tento prístup nie je úplný. Aj keď sa zdalo, že výška tónu je závislá na prítomnosti fundamentálnej frekvencie, niektoré experimenty ukázali, že aj v prípade, že fundamentálna frekvencia chýba, môže byť vnímaná rovnaká výška tónu, ak je podnet periodický.

Teórie, ktoré nasledovali po Helmholtzovej teórii začali riešiť problémy, ktoré Helmholtzova teória nedokázala plne vysvetliť. Tieto teórie a na ich základe vyvinuté modely lepšie reagovali na problém chýbajúcej fundamentálnej frekvencie tým, že predpokladali,

že mozog je schopný identifikovať fundamentálny tón aj vtedy, keď nie je prítomný v samotnom podnete. Tieto moderné teórie sa viac zameriavajú na schopnosť nervového systému doplniť chýbajúce komponenty a analyzovať zvukové vlny ako celok, namiesto toho, aby sa zameriavali na jednoduchú analýzu fundamentálneho tónu a harmonických.

Jedným z takýchto nových modelov bol model založený na porovnávaní vzorov. Myšlienka porovnávania vzorov sa po prvýkrát objavila v práci Terharda v roku 1974 [21] Viaceré matematické modely realizujúce túto myšlienku boli publikované v rokoch 1980–1990. Za najvýznamnejšiu z týchto prác sa pokladá Goldsteinov článok [3] zavádzajúci pojem centrálneho procesora pre spracovanie výšky tónu.

Pokračujúc v evolúcii modelov vnímania výšky tónu sa dnes využívajú prístupy, ktoré už nepredpokladajú, že vnútorné ucho je striktne obmedzené na rezonančné frekvencie, ktoré reagujú len na jeden harmonický signál. Mnohé moderné teórie predpokladajú schopnosť ucha analyzovať komplexné zvuky a detegovať fundamentálnu frekvenciu aj pri absencii jej harmonických. Tieto pokroky ukázali, že náš zmyslový systém, konkrétne ucho a mozog, sú schopné sa prispôsobiť a riešiť zložitejšie akustické signály, než pôvodne predpokladal Helmholtz. Tento pokrok priniesol nový pohľad na spôsob, akým vnímame a spracúvame zvuk v každodennom živote.

## 4.3 Časové modely vnímania výšky tónu

Časové modely predpokladajú, že základom pre vnímanie výšky tónu sú časové postupnosti akčných potenciálov (angl. *spike trains*) vznikajúce pri počúvaní zvukového signálu v sluchovom nerve a ich periodické vlastnosti. Vychádzajú teda z transformácie časového priebehu zvukového signálu do postupnosti spikov vo vláknach sluchového nervu iniciovanej vláskovými bunkami v slimáku na základe pohybov bazilánej membrány vyvolanej zvukovým signálom, a nie z jeho frekvenčného obsahu. Na rozdiel od spektrálnych modelov sa teda nezameriavajú na frekvenčný obsah zvukového signálu , ale na jeho periodicitu v čase.

Tieto modely predpokladajú, že mozog zisťuje periodicitu resp. opakujúce sa časové vzory v postupnostiach spikov mechanizmom založenom na rovnakom princípe ako autokorelačná analýza 2.5. Tak, ako pri spektrálnych modeloch je ich fyziologickou základom tonotopický charakter vibrácií bazilárnej membrány, v prípade časových modelov je fyzi-

ologickým základom tzv. fázová synchronizácia (angl. *phase locking*) spočívajúca v tom, že generovanie spiku v sluchovom nerve sa s najväčšou pravdepodobnosťou uskutočňuje vždy v rovnakej fáze zvukového signálu. Na základe získaných informácií o periodicite dokážu časové modely určiť výšku tónu aj v prípadoch, kde spektrálne modely zlyhávajú. Ide napríklad o prípad, keď spracovávaný zvukový signál neobsahuje žiadne rozlíšené harmonické zložky. Osobitne zreteľne sa to prejavuje v prípadoch chýbajúcej fundamentálnej frekvencie skombinovanej s chýbajúcimi rozlíšenými harmonickými zložkami. Tieto modely vo všeobecnosti fungujú aj pri nedokonalom frekvenčnom rozlíšení, ktoré spektrálne modely nedokážu pokryť z princípu.

### 4.3.1 História vývoja časových modelov

Historické teórie vnímania výšky tónu prechádzali významným vývojom od starovekých filozofických úvah až po moderné fyzikálne a neurobiologické modely. Už Pytagorejci vnímali zvuk ako sériu rýchlo sa opakujúcich pulzov, ktoré sa prekrývajú tak, že vnímame jeden tón. Táto predstava predznamenáva neskoršie koncepty frekvenčnej analýzy zvuku.

V 17. storočí Marin Mersenne a René Descartes rozlíšili dva dôležité aspekty zvukových vibrácií. Frekvenciu, ktorá určuje výšku tónu a rýchlosť šírenia zvuku, ktorá na výšku nemá vplyv. Toto rozlíšenie položilo základy pre neskoršie štúdium vzťahu medzi frekvenciou a vnímanou výškou tónu. V 19. storočí Rutherford navrhol tzv. telefónnu teóriu (angl. *telephone theory*), podľa ktorej ucho funguje ako pasívny prijímač vibrácií, ktoré len sprostredkováva mozgu. Tento model elegantne vysvetľoval transformáciu zvukových vibrácií na spiky, ale nedokázal odpovedať na otázku, ako sú tieto vibrácie analyzované a vnímané ako rôzne výšky tónov.

Volley teória (angl. *volley theory*) Wevera a Braya (1930) riešila problém, že jednotlivé nervové vlákna nedokážu dostatočne rýchlo reagovať na vysoké frekvencie. Podľa tejto teórie vlákna v sluchovom nerve nefungujú samostatne, ale striedajú sa v aktivácii, čím spoločne dokážu reprezentovať aj vyššie frekvencie, než zvládne jedno vlákno samo. Neskoršie výskumy ukázali, že generovanie akčných potenciálov má skôr stochastický charakter a nie je dokonale synchronizované s frekvenciami zvukových signálov.

Jednoduché tóny spôsobujú pravidelný periodický pohyb bazilárnej membrány, čo umožňuje aplikovať volley teóriu. Pri zložitejších signáloch však výskumy Schoutena (1962)

ukázali, že vnímaná výška tónu sa môže meniť, aj keď obálka signálu odrážajúca zmenu jeho amplitúd v čase zostáva rovnaká. To vyvrátilo hypotézu, že výška tónu je určovaná len obálkou. Ďalšie modely sa preto sústredili na jemnú časovú štruktúru signálu, konkrétne na intervaly medzi vrcholmi v signáli. Tento prístup však predpokladal väčšiu citlivosť na fázu, než aká bola experimentálne pozorovaná, čo viedlo k potrebe ešte presnejšieho modelu.

Riešením sa stal autokorelačný model výšky tónu, ktorý analyzuje opakujúce sa vzory v signáli a nezávisí priamo od fázy jednotlivých parciálnych tónov. Tento model umožňuje spoľahlivo určovať periodicitu aj pri zložitejších zvukoch a dnes tvorí základ moderných teórií výšky tónu. Skúmanie týchto mechanizmov je dôležité nielen pre pochopenie fungovania sluchového systému, ale aj pre aplikácie v digitálnom spracovaní zvuku, hudobnej akustike a neurovede.

Ako prvý navrhol model vnímania výšky tónu založený na procese autokorelácie Licklider (1951). Autokoreláciu možno chápať ako spôsob analýzy zvuku, pri ktorom sa zisťuje podobnosť signálu so sebou samým pri rôznych časových oneskoreniach. Výsledkom tohto procesu je dvojrozmerný vzor, v ktorom jedna os predstavuje charakteristickú frekvenciu a druhá os časové oneskorenie. Ak je zvuk periodický, v tomto vzore sa objaví pravidelný hrebeň na oneskorení zodpovedajúcom perióde zvuku.

Licklider však nekonkretizoval, akým spôsobom mozog spracúva tento vzor na extrakciu výšky tónu. Tento nedostatok riešili Meddis a Hewitt, ktorí navrhli sčítať autokorelačné vzory naprieč všetkými frekvenčnými kanálmi, čím vznikla súhrnná autokorelačná funkcia (SACF). Táto funkcia umožňuje jednoznačne určiť periodicitu signálu, a tým aj vnímanú výšku tónu. Meddis–Hewitov model zohral v histórii časových modelov mimoriadne významnú úlohu. Tomuto modelu sa podrobne venujeme v ďalších kapitolách.

# 5 Meddis–Hewittov model

Táto kapitola sa venuje Meddis–Hewittovmu modelu, ktorého miesto vo vývoji modelov vnímania výšky tónu sme uviedli v časti 2.5 a ktorý patrí k najznámejším časovým modelom v tejto oblasti.

Identifikácia výšky tónu je v tomto modeli založená na extrahovaní časových informácií z aktivít sluchového nervu. Na tento účel model používa autokorelačnú analýzu založenú na myšlienke, že výška tónu sa dá extrahovať z priebehu generovania akčných potenciálov vo vláknach sluchového nervu s využitím autokorelačnej funkcie na hodnotenie priebehu aktivít v jednotlivých vláknach, resp. v skupinách vlákien citlivých na rovnakú frekvenciu. Túto myšlienku ako prvý prezentoval Licklider v práci z roku 1951 [7]. Meddis–Hewittov model rozširuje túto myšlienku tak, že kombinuje výsledky autokorelačnej analýzy uskutočnenej na uvedených skupinách vlákien do súhrnnej autokorelačnej funkcie (SACF). Rozhodnutia o vnímanej výške tónu sú v tomto modeli založené výlučne na tejto súhrnnej autokorelačnej funkcii (SACF).

Model vychádza z experimentálne overeného predpokladu, že vnem výšky tónu je potrebné odvodzovať z rozlíšených aj nerozlíšených harmonických zložiek tónu v slimáku, pretože oba tieto druhy zložiek obsahujú informácie o periodicite zvukového signálu. Model kombinuje tieto informácie z rozlíšenej aj nerozlíšenej oblasti jednoduchým agregačným procesom. Autori modelu tiež porovnali jeho výsledky s psychofyzikálnymi výsledkami vychádzajúcimi z výsledkov klasických štúdií s ľudskými poslucháčmi. Pri popise tohto modelu budeme vychádzať z [9].

## 5.1 Popis modelu

Model pozostáva z viacerých fáz, ktoré zodpovedajú časovej následnosti v procese spracovávania vstupného zvukového signálu od jeho vstupu do vonkajšieho ucha až po získanie informácie o periodicite obsiahnutej v nervových impulzoch vo vláknach sluchového nervu, ktoré tento signál po spracovaní vo vnútornom uchu vyvolal. Táto informácia je potom podkladom pre vnímanú výšku tónu predpokladanú modelom. V ďalšom popíšeme každú z uvedených fáz modelu individuálne. Pre všetky fázy platí, že vstupné zvukové signály sú modelované ako postupnosti čísiel bez fyzikálnej dimenzie. Okrem explicitne uvedených

výnimiek, signály trvajú 100 ms a výsledky predstavujú stav modelu na konci tohto času.

### 5.1.1 Vplyvy vonkajšieho a stredného ucha

Tlakové zmeny tvoriace zvukový signál vstupujúci do vonkajšieho ucha vykazujú na jeho konci, teda pri bubienku, zosilnenie v porovnaní so vstupom do zvukovodu. Toto zosilnenie je najvýraznejšie v oblasti frekvencií medzi $2-5\,\text{kHz}$ a vzniká vďaka rezonančným vlastnostiam vonkajšieho zvukovodu, ktorý funguje ako akustický rezonátor a prirodzene zosilňuje určité frekvenčné pásmo. Stredné ucho, ktoré zabezpečuje prenos zvukového signálu z vonkajšieho zvukovodu do vnútorného ucha, plní okrem samotného prenosu zvuku aj funkciu pásmového filtra, pričom jeho maximálna účinnosť sa nachádza pri frekvencii približne 1 kHz a na nižších frekvenciách dochádza k strmému poklesu prenosu. Tento pokles je spôsobený obmedzenou schopnosťou mechanického systému kostičiek v strednom uchu efektívne prenášať veľmi nízkofrekvenčné signály. Na dosiahnutie realistickej simulácie týchto procesov sa v modeli skombinovala funkcia zosilnenia vo vonkajšom uchu a filtrovania v strednom uchu do jednej pásmovej priepuste, ktorá je implementovaná ako digitálny pásmový filter na vstupe modelu.

### 5.1.2 Filtrácia bazilárnou membránou

Bazilárna membrána vo vnútornom uchu funguje ako súbor pásmových filtrov, pričom rôzne časti membrány reagujú na rôzne frekvencie zložiek zvukového signálu. Ako sme uviedli v časti 3.2, vysoké frekvencie spôsobujú vibrácie bazilárnej membrány bližšie ku vchodu do slimáka, zatiaľ čo nízke frekvencie aktivujú jej vzdialenejšie časti. Tento proces je v modeli simulovaný pomocou 128 digitálnych gamatónových filtrov, ktoré predstavujú pásmové filtre s kritickou šírkou pásma. Gamatónové filtre sú matematické modely navrhnuté tak, aby simulovali spôsob, akým ľudské ucho rozkladá zvukový signál na jednotlivé frekvenčné zložky. Sú založené na gamatónovej funkcii, ktorá veľmi presne opisuje odozvu sluchového systému. Každý filter má svoju centrálnu frekvenciu, pričom v našom modeli sú filtre rovnomerne rozmiestnené s rozostupom približne o $0,25$ ekvivalentnej šírky pásma (ERB) od seba. Bližšie informácie o ERB sú spomenuté v odseku 2.6.2. Celý súbor filtrov pokrýva frekvenčný rozsah od 80 Hz do 8 kHz, čím zahŕňa väčšinu sluchového spektra človeka. ERB, ktorá je podobná kritickému pásmu sluchového systému, predstavuje

vzdialenosť približne $0,9$ mm na bazilárnej membráne a udáva, ako citlivo vnútorné ucho rozlišuje jednotlivé frekvencie.

Použitie gamatónových filtrov je motivované psychofyzikálnymi experimentmi, ktoré skúmajú, ako ľudia subjektívne vnímajú zvuk. Tieto filtre presne modelujú spracovanie zvuku v sluchovom systéme, pričom poskytujú dobrú aproximáciu kritických pásiem. Napriek tomu existujú určité rozdiely oproti fyziologickej realite. Gamatónové filtre v našom modeli sú lineárne, čo znamená, že spracovávajú všetky intenzity signálu rovnako. V skutočnosti však sluchové vlákna vykazujú nelineárnu odpoveď, pri vyšších intenzitách sa mení ich citlivosť a odozva v dôsledku mechanizmov aktívneho zosilnenia v slimáku.

### 5.1.3 Mechanicko-nervová transdukcia

Výstup z každého filtra je vstupom do simulátora vláskových buniek (Obr. 9), ktorý konvertuje mechanický pohyb bazilárnej membrány na pravdepodobnosť výskytu akčného potenciálu v príslušnom vlákne sluchového nervu. Model predpokladá, že pravdepodobnosť výskytu akčného potenciálu je lineárne závislá od množstva neurotransmitera v synaptickej štrbine medzi vnútornou vláskovou bunkou a jej zodpovedajúcim sluchovým nervovým vláknom:

$$p(t) = hc(t)dt \tag{10}$$

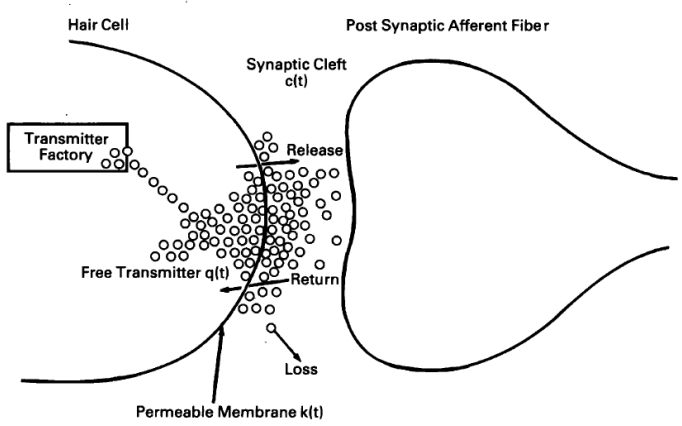
kde $p(t)$ je pravdepodobnosť výskytu akčného potenciálu počas intervalu $t$ až $t+dt$, $c(t)$ je množstvo neurotransmitera v synaptickej štrbine, $h$ je konštanta modelu, $dt$ je časový krok simulácie (v uvedenom modeli je dt $= 0,00005$ s). Neurotransmiter sa uvoľňuje do synaptickej štrbiny z vláskových buniek v množstvách, ktoré závisia od permeability bunkovej membrány.

Permeabilita membrány je ovplyvňovaná amplitúdou signálu nasledujúcim spôsobom:

$$\begin{aligned} k(t) &= \frac{g(x(t)+A)}{x(t)+A+B}dt, \quad && x(t)+A > 0, \\ k(t) &= 0, && x(t)+A < 0, \end{aligned} \tag{11}$$

kde $x(t)$ je okamžitá amplitúda signálu po filtrovaní a $A, B, g$ sú parametre modelu.

Popis dynamiky prenosu neurotransmitera medzi vnútornou vláskovou bunkou a synaptickou štrbinou smerom k sluchovému nervovému vláknu je modelovaný sústavou troch diferenciálnych rovníc:

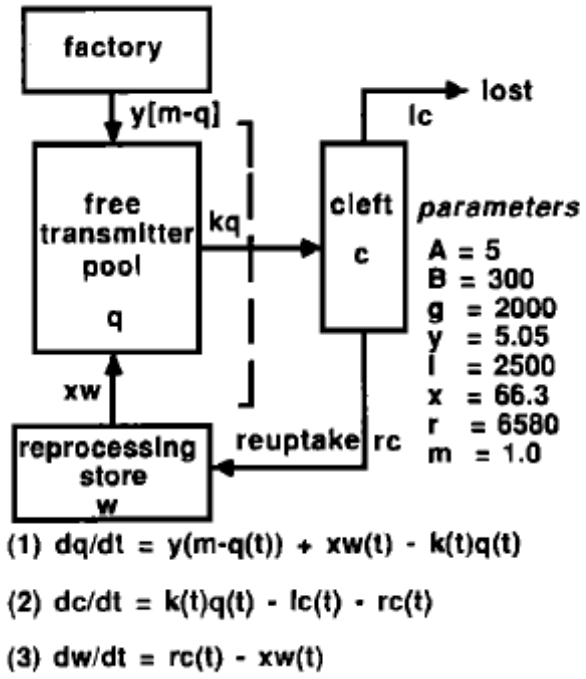



**Obr. 8:** Schematické znázornenie synaptického prenosu použitého v modeli. Neurotransmiter je syntetizovaný vo vláskovej bunke a hromadí sa v zásobníku voľného neurotransmitera, ktorého množstvo v čase označujeme ako $q(t)$. Z tohto zásobníka je následne cez polopriepustnú membránu s časovo závislou permeabilitou $k(t)$ uvoľňovaný do synaptickej štrbiny, kde jeho aktuálne množstvo označujeme ako $c(t)$. Zo štrbiny je neurotransmiter buď reabsorbovaný späť do bunky alebo sa stratí. (Zdroj [10])

$$\begin{aligned} \frac{dq}{dt} &= y(m - q(t)) + xw(t) - k(t)q(t), \\ \frac{dc}{dt} &= k(t)q(t) - lc(t) - rc(t), \\ \frac{dw}{dt} &= rc(t) - xw(t), \end{aligned} \tag{12}$$

kde $k(t)$ je permeabilita membrány a $x, y$ sú parametre modelu. Pričom $x$ je parameter ovplyvňujúci rýchlosť uvoľňovania neurotransmitera z recyklačného zásobníka a $y$ je parameter ovplyvňujúci rýchlosť dopĺňania neurotransmitera do zásobníka. Množstvo neurotransmitera uvoľneného do synaptickej štrbiny v čase $t$ je $k(t)q(t)dt$, kde $q(t)$ predstavuje množstvo neurotransmitera v zásobníku voľného neurotransmitera (ďalej len množstvo neurotransmitera v zásobníku) vo vláskovej bunke.

Neurotransmiter sa zo synaptickej štrbiny späť reabsorbuje do vláskovej bunky rýchlosťou $rc(t)$. Časť neurotransmitera sa však stráca zo systému rýchlosťou $lc(t)$. Tento úbytok spôsobuje adaptáciu frekvencie vzniku akčných potenciálov, pri ktorej sa počiatočne vysoká aktivita neurvového vlákna postupne znižuje, aj keď stimulácia pretrváva. Parameter $l$ ovplyvňuje rýchlosť strát neurotransmitera zo synaptickej štrbiny a parameter $r$ určuje rýchlosť návratu neurotransmitera do recyklačného zásobníka.




**Obr. 9:** Model synaptického prenosu medzi vnútornou vláskovou bunkou a sluchovým nervovým vláknom. Neurotransmiter je uvoľňovaný do synaptickej štrbiny v závislosti od permeability membrány $k(t)$, odkiaľ sa môže byť buď stratiť rýchlosťou $lc(t)$, alebo reabsorbovať späť rýchlosťou $rc(t)$. Model je definovaný sústavou troch diferenciálnych rovníc, ktoré popisujú dynamiku synaptického prenosu a recyklácie. (Zdroj [9]).

Reabsorbovaný neurotransmiter je dočasne uložený v recyklačnom zásobníku v množstve označenom $w(t)$, kde prebieha jeho krátkodobá recyklácia. Následne je rýchlosťou $xw(t)$ presunutý späť do zásobníka voľného neurotransmitera, čím sa obnovuje jeho disponibilita na ďalšie uvoľnenie do synaptickej štrbiny. Prevažná časť neurotransmitera neustále cirkuluje medzi zásobníkom voľného neurotransmitera, synaptickou štrbinou a recyklačným zásobníkom. Ak však dôjde k poklesu dostupného voľného neurotransmitera v dôsledku jeho strát zo synaptickej štrbiny, tento deficit sa postupne kompenzuje syntézou nového neurotransmitera v biosyntetickom systéme vláskovej bunky rýchlosťou $y(m - q(t))$, kde $m$ je maximálne dostupné množstvo neurotransmitera.

Fungovanie modelu je popísané pomocou rovnice pre permeabilitu (11) a troch diferenciálnych rovníc (12), ktoré sú tiež uvedené na Obr. 9. Parametre modelu, taktiež zobrazené na Obr. 9, boli zvolené tak, aby zabezpečili vysokú spontánnu rýchlosť generovania spikov v sluchovom nervovom vlákne 60 spike/s.

V tomto modeli je implementovaná len jeden typ nervových vlákien. V skutočnosti však nervové vlákna vykazujú rôzne vlastnosti, najmä čo sa týka spontánnej frekvencie a dynamického rozsahu. Táto rozmanitosť nebola v tomto modeli implementovaná.

### 5.1.4 Refraktérne efekty

Pravdepodobnosť generovania spiku vo vlákne závisí aj od nedávnej histórie spikov v tomto vlákne. Tento efekt je v modeli aproximovaný nasledujúcim vzťahom:

$$p'(t) = p(t)(1 - \sum_{i=1}^{N} p'(t-T)W(t-T)), \tag{13}$$

kde $T = idt$, $p'(t)$ je pravdepodobnosť výskytu spiku po zohľadnení refraktérnych efektov, tie predstavujú obdobie po spiku, počas ktorého nervové vlákno nie je schopné generovať ďalší spike, aj keď je stimul dostatočne silný, $p(t)$ je neupravená pravdepodobnosť (založená na množstve neurotransmitera v synaptickej štrbine) a $W(t)$ je empiricky odvodená váhová funkcia. Hodnota $N$ vyjadruje počet predchádzajúcich časových krokov ovplyvňujúcich aktuálnu pravdepodobnosť spiku $p'(t)$ a závisí od dĺžky refraktérneho efektu a časového kroku $dt$ . Hodnoty $p'(t-T)$ musia byť vypočítané skôr, ako je možné vyčísliť $p'(t)$.

V modeli je váhová funkcia $W(t)$ nastavená nasledovne:

$$\begin{aligned} W(t-T) &= 1, \quad t - T < 0.001s \\ W(t-T) &= 0, \quad t - T > 0.001s, \end{aligned} \tag{14}$$

čo zodpovedá absolútnej refrektérnej perióde 1 ms.

Pri simulácii spikov musia byť podľa [11] dodržané dve základné podmienky. Počas refraktérneho obdobia nie je povolený ďalší spike v tom istom vlákne. Tento stav je zabezpečený znížením pravdepodobnosti spiku prostredníctvom váhovej funkcie $W(t)$. Ďalej musí byť spike vygenerovaný len vtedy, keď hodnota náhodného čísla z rovnomerného rozdelenia na intervale $[0, 1)$ neprekročí okamžitú upravenú pravdepodobnosť spiku $p'(t)$. Týmto spôsobom sa model riadi fyziologickou požiadavkou, že spiky nevznikajú úplne náhodne, ale pravdepodobnosť ich vzniku závisí od hladiny neurotransmitera a histórie aktivácie konkrétneho vlákna.

### 5.1.5 Distribúcia časových intervalov

V nasledujúcej fáze modelu sa počítajú histogramy časových intervalov medzi spikmi, pomocou ktorých bude možné určiť autokorelačnú funkciu. Histogram časových intervalov medzi spikami je vytvorený osobitne pre každý filtrovaný kanál. Použité boli intervaly medzi každými dvoma spikami, teda nielen medzi bezprostredne po sebe nasledujúcimi. Pravdepodobnosť výskytu časového intervalu dĺžky $\delta t$ medzi dvoma spikami, z ktorých jeden nastal v čase $t$, je daná vzťahom:

$$h(t, \delta t) = p'(t)p'(t - \delta t)dt. \tag{15}$$

Autori popisovaného modelu uvádzajú, že myšlienka využitia autokorelácie v časových modeloch vnímania výšky tónu pochádza od Licklidera, ktorý ju aj formálne demonštroval vo svojej práci z roku 1959 [8]. Licklider ďalej navrhol systém založený na výpočte histogramov časových intervalov medzi spikami, ktorý interpretoval ako kĺzavú autokorelačnú funkciu (angl. *running autocorrelation function*). Každá dvojica spikov s časovým odstupom $\delta t$ prispieva do histogramu podľa svojej pozície v čase a je vážená exponenciálnym členom, ktorý znižuje vplyv starších spikov. Táto formulácia oprávňuje označovať takéto histogramy ako autokorelačné funkcie (ACF). Ich grafické znázornenie sa nazýva autokorelogram (angl. *autocorrelogram*).

Matematicky je autokorelačná funkcia pre interval $\delta t$ v čase $t$ pre kanál $k$ definovaná ako:

$$h(t, \delta t, k) = \sum_{i=1}^{T_max} p'(t - T)p'(t - T - \delta t)e^{-\frac{T}{\tau}}dt, \tag{16}$$

kde $T = idt$.

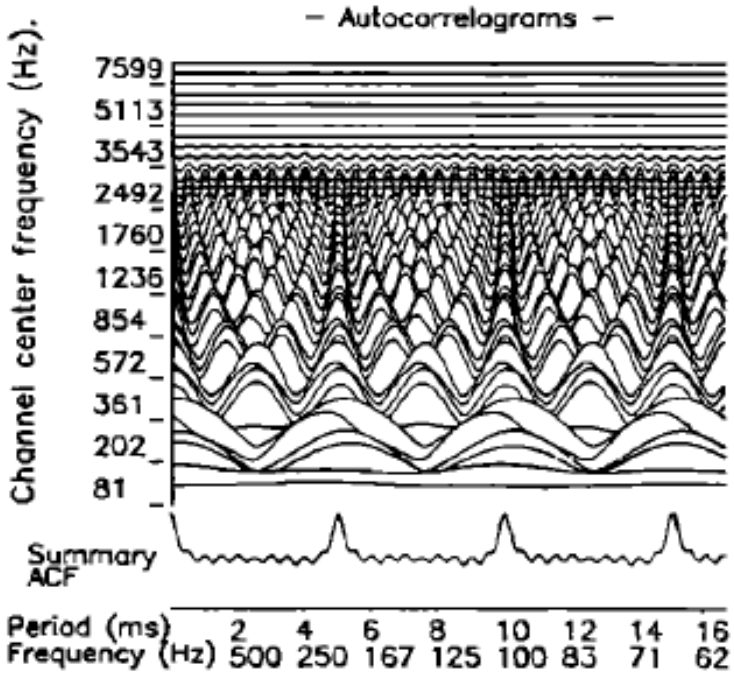
Táto rovnica počíta príspevky všetkých dvojíc spikov, ktoré sa vyskytli v minulosti vzhľadom na aktuálny čas $t$, pričom časový odstup $T$ rastie s krokom . Váhovací člen $e^{-\frac{T}{\tau}}$ má funkciu zabúdania, teda spiky staršie ako niekoľko milisekúnd majú zanedbateľný vplyv. Na základe odporúčaní Licklidera, ktorý navrhol časovú konštantu $\tau \approx 2,5$ ms, sa za postačujúci rozsah výpočtu hodnôt do minulosti považuje maximálne $T_max = 3\tau = 7,5$ ms. Príspevky po tomto čase sú už natoľko malé, že ich možno zanedbať.

### 5.1.6 Sčítanie naprieč kanálmi

Autokorelačné funkcie kanálov vypočítané pre jednotlivé kanály sa nevyužívajú bezprostredne na odhad vnímanej výšky tónu, ale ich spriemerovaním cez všetky kanály sa určí súhrnná autokorelačná funkcia (SACF) daná vzťahom:

$$s(t, \delta t) = \sum_{k=1}^{128} \frac{h(t, \delta t, k)}{128}, \tag{17}$$

kde $h(t, \delta t, k)$ je kĺzavá ACF $h(t, \delta t)$ pre $k$-ty kanál vzorkovaný v čase $t$. Graf súhrnnej autokorelačnej funkcie sa nazýva súhrnný korelogram (angl. *summary ACF*).




**Obr. 10:** Výstup modelu v reakcii na stimul pozostávajúci z prvých desiatich harmonických s fundamentálnou frekvenciou 200 Hz. Zobrazený je len posledný úsek simulácie v trvaní 7, 5 ms. Horná časť grafu ukazuje autokorelogramy pre jednotlivé frekvenčné kanály, kde sú viditeľné periodické štruktúry súvisiace s harmonickými zložkami podnetu. Spodná časť grafu zobrazuje súhrnnú autokorelačnú funkciu (SACF), ktorá vykazuje jasné maximá pre periódy zodpovedajúce fundamentálnej frekvencii podnetu. (Zdroj [9]).

### 5.1.7 Extrakcia výšky tónu

Maximum súhrnnej autokorelačnej funkcie sa používa ako jednoduchý indikátor vnímanej výšky tónu. Takto určený indikátor poskytuje v mnohých prípadoch uspokojivý výsledok.

Napríklad na Obr. 10 je vidno, že SACF nadobúda maximum pri frekvencii opakovania spikov 200 Hz (oneskorenie 5 ms), čo zodpovedá vnímaniu výšky tónu na tejto frekvencii. Ak SACF nadobúda maximum vo viacerých bodoch, vyberie sa ten s najkratším časovým oneskorením.

# 6 Programová implementácia Meddis–Hewittovho modelu

V tejto kapitole predstavíme nami vytvorený program MH implementujúci Meddis–Hewittov model v programovacom jazyku Python a jeho použitie na vybrané umelo vytvorené aj prirodzené zvukové signály. Celý kód programu je možné nájsť v Príloha A.

## 6.1 Popis programu MH

Pri tvorbe programu MH sme vychádzali z informácií v [9], [10] a [11], ale využili sme aj poznatky popísané v podkapitole 2.6 o digitálnych filtroch. Použili sme vzorkovací krok $dt = 0.00005$ s a analyzovali sme vstupný signál s dĺžkou 1 s. V grafických výstupoch programu vykresľujeme kvôli prehľadnosti spravidla posledných 25 ms priebehu zobrazovanej veličiny.

Na ilustráciu fungovania programu sme v jeho nasledujúcom popise využili umelo generovaný vstupný signál s predpisom:

$$y(t) = sin(2\pi 200t) + sin(2\pi 400t) + sin(2\pi 600t), \tag{18}$$

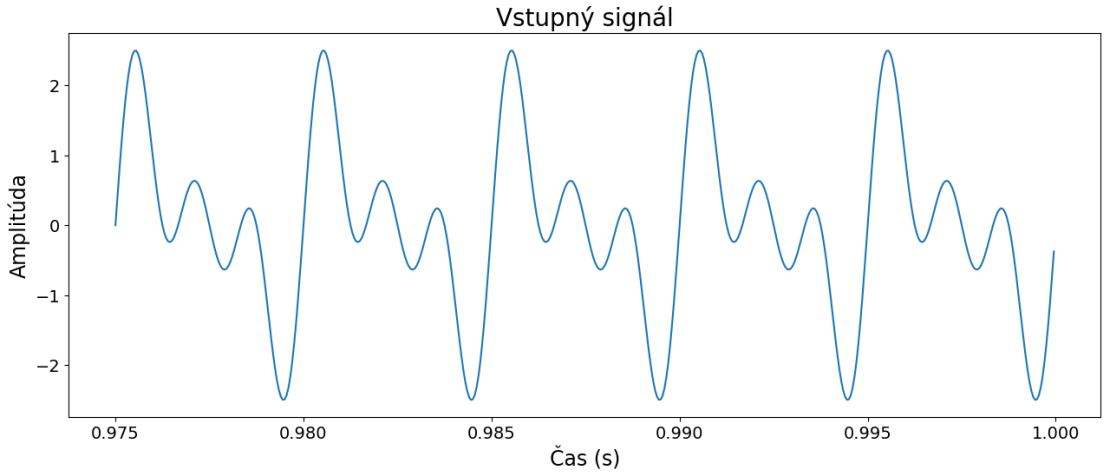
kde $t$ je čas v sekundách. Ide o signál zložený z jednoduchých signálov s frekvenciami 200 Hz, 400 Hz a 600 Hz. Na nasledujúcom Obr. 11 môžeme vidieť vykreslenie tohto vstupného signálu.

Prvá fáza programu MH simuluje vplyvy vonkajšieho a stredného ucha na priebeh vstupného signálu popísané v odseku 5.1.1. Na dosiahnutie realistickej simulácie týchto procesov boli skombinované funkcie zosilnenia vo vonkajšom uchu a filtrovania v strednom uchu do jednej pásmovej priepuste, ktorá je implementovaná ako digitálny pásmový filter na vstupe modelu.

Pre tento pásmový digitálny filter boli podľa [9] použité nasledujúce koeficienty :

$$y_i = 0.8878x_i - 0.8878x_{i-2} - 0.2243y_{i-1} + 0.7757y_{i-2}. \tag{19}$$

Výstupný filtrovaný signál môžeme vidieť na Obr. 12. Jednoduché zložky vstupného signálu s frekvenciami 200 Hz, 400 Hz a 600 Hz ležia pod „aktívnymi“ pásmami vonkajšieho ucha (2 – 5 kHz) aj stredného ucha (okolo 1 kHz). Z tohto dôvodu dochádza po

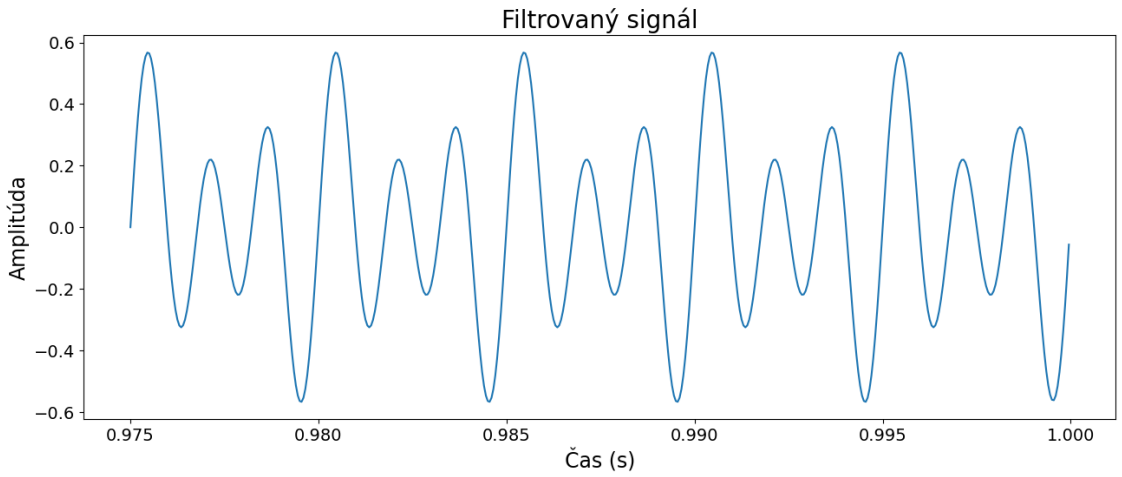



**Obr. 11:** Vstupný signál použitý na ilustráciu fungovania programu na implementáciu Meddis–Hewitovho modelu. (Zdroj: vlastné spracovanie)

prechode signálu cez tento filter k výraznému poklesu jeho amplitúdy. Tento jav zodpovedá fyziologickej realite, kde mechanický systém stredného ucha obmedzene prenáša nízkofrekvenčné komponenty.

Ďalším krokom v programe je aplikácia súboru gamatónových filtrov, ako je opísané v odseku 2.6.1. Platí, že každý filter má svoju centrálnu frekvenciu. Pomocou prevodného vzťahu (9) medzi frekvenciou a vyjadrením frekvencie v ERB škále sme rovnomerne rozmiestnili body s centrálnymi frekvenciami filtrov tak, aby boli vzájomne posunuté o 0.25 ERB. V našej programovej implementácií modelu sme týmto postupom dostali 124 digitálnych filtrov, ktoré modelujú spôsob, akým ľudské ucho rozkladá zvukový signál na jednotlivé frekvencie. Každý z týchto filtrov predstavuje samostatný kanál, ktorý reaguje najcitlivejšie na svoju centrálnu frekvenciu a poskytuje výstup zodpovedajúci danej frekvenčnej zložke vstupného signálu. Tieto filtre pokrývajú rozsah medzi 80 Hz a 8000 Hz. Výstupy súboru gamatónových filtrov, nazývané tiež kochleogram môžeme vidieť vykreslený na Obr. 13.

Digitálnu implementáciu súboru gamatónových filtrov sme realizovali pomocou nasledujúceho algoritmu [6]. Tento algoritmus umožňuje efektívne aproximovať gamatónové filtre $n$-tého rádu pomocou kaskády identických IIR filtrov prvého rádu, čím sa výrazne znižuje výpočtová náročnosť v porovnaní s FIR implementáciou. V našej implementácii




**Obr. 12:** Výstupný filtrovaný signál po prechode vonkajším a stredným uchom. Vzhľadom na to, že frekvenčné zložky pôvodného signálu pôvodného signálu sa nachádzajú pod 1 kHz, amplitúda filtrovaného signálu je výrazne zoslabená. Zodpovedá to fyziologickému faktu, že nízke frekvencie sú v sluchovom systéme prenášané menej efektívne. (Zdroj: vlastné spracovanie)

sme použili filtre štvrtej rádu $n = 4$, ako sme vysvetlili v odseku 2.6.2.

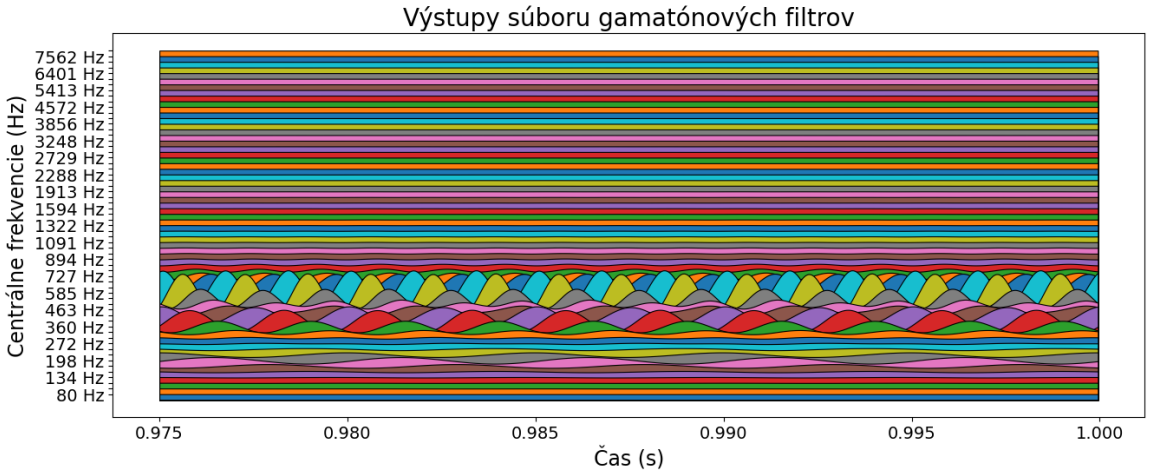
Ako vstup bol využitý výstupný signál $y(k)$ z pásmového filtra reprezentujúceho pôsobenie vonkajšieho a stredného ucha. Signál $y(k)$ sa najskôr frekvenčne posunie o $-f_c$ Hz, čím sa komponenty spektra v okolí frekvencie $f_c$ presunú do oblasti okolo 0 Hz. Touto transformáciou vznikne z pásmového signálu nízkofrekvenčný signál, ktorý je možné efektívne filtrovať dolnopriepustným IIR filtrom (2.6). Matematicky sa tento posun realizuje pomocou komplexnej exponenciály, čím vznikne komplexný signál $z(k)$ s imaginárnou jednotkou $j$ v nasledujúcom tvare:

$$z(k) = e^{-2\pi j f_c k dt} y(k). \tag{20}$$

Následne prechádza tento posunutý signál $z(k)$ cez rekurzívny filter prvého rádu s výstupom $w(k)$,

$$w(k) = w(k-1) + (1 - e^{-2\pi b dt}(z(k-1) - w(k-1))), \tag{21}$$

kde $b$ predstavuje ekvivalentnú obdĺžnikovú šírku pásma (ERB) pre príslušnú centrálnu frekvenciu. Na výpočet ERB sme využili vzťah (8).




**Obr. 13:** Výstupy súboru gamatónových filtrov. Na vertikálnej osi sú zobrazené centrálne frekvencie jednotlivých filtrov, ktoré pokrývajú rozsah od 80 Hz do 8000 Hz. Kvôli prehľadnosti sú zobrazené centrálne frekvencie len pre podmnožinu zo 124 filtrov. Každá farebná oblasť reprezentuje výstup jedného filtra. Spektrálne zloženie použitého vstupného signálu sa výrazne prejavuje na priebehu grafického znázornenia zodpovedajúceho frekvenciám okolo 200 Hz až 600 Hz. (Zdroj: vlastné spracovanie)

Aby sa dosiahol vyšší rád filtra $n = 4$, tento krok sa opakuje celkovo štyrikrát ako kaskáda identických filtrov, pričom výstup jedného sa stáva vstupom ďalšieho. Nakoniec sa výstup $w(k)$ frekvenčne posunie o $+f_c$ Hz a vezme sa z neho len reálna časť, čím vznikne výstup $x(k)$ :

$$x(k) = \Re(e^{2\pi j f_c k dt} w(k)) \tag{22}$$

Tento algoritmus bol aplikovaný na každý z 124 gamatónových filtrov s príslušnými centrálnymi frekvenciami $f_c$, $c = 1, \ldots, 124$. K stručnej charakteristike zobrazenia výstupov súboru gamatónových filtrov na Obr. 13 ešte dodávame, že zobrazené jednotlivé výstupy filtrov sú posunuté vertikálne podľa ich centrálnej frekvencie. Vodorovná os reprezentuje čas a farby jednotlivých oblastí zobrazujú amplitúdu odpovede daného filtra v danom čase. Podobne ako bazilárna membrána reaguje na $f = 200$ Hz tým, že sa najviac prehne v mieste citlivom na túto frekvenciu, aj náš výstup vykazuje zvýšenú citlivosť v okolí $f = 200$ Hz. Aktivita filtrov v blízkosti tejto frekvencie vzniká v dôsledku prekryvu

jednotlivých digitálnych filtrov. Hoci sú ich centrálne frekvencie v ERB škále rozmiestnené s krokom 0.25 ERB, vo frekvenčnej oblasti sa ich pásma navzájom prekrývajú. Tento jav realisticky simuluje správanie bazilárnej membrány, kde frekvenčné rozsahy jednotlivých miest nie sú úplne oddelené, ale sa plynule prekrývajú. Všeobecne vidíme, že najvýraznejšia aktivita sa sústreďuje okolo frekvencií približne 200 Hz, 400 Hz a 600 Hz, čo korešponduje s frekvenčným zložením vstupného signálu.

Nasledujúca fáza programu realizuje simuláciu procesu mechanicko-nervovej transdukcie vo vláskových bunkách vnútorného ucha. Ide o kľúčový krok v spracovaní zvukového signálu, pri ktorom sa akustické mechanické zvukové signály transformujú na spiky v nervových vláknach sluchového nervu. Model popisuje časové zmeny množstva neurotransmitera v troch priestoroch a to množstvo neurotransmitera $q(t)$ v zásobníku vo vnútri vláskovej bunky, množstvo neurotransmitera $c(t)$ v synaptickej štrbine a množstvo neurotransmitera $w(t)$ v recyklačnom zásobníku. Meddis a Hewitt v článku [11] zverejnili implementáciu tejto simulácie v programovacom jazyku BASIC, z ktorej vychádza implemetácia tohto procesu v našom programe. Simulácia prebieha samostatne pre každý frekvenčný kanál, pričom vstupom je signál spracovaný príslušným filtrom zo súboru gamatónových filtrov. Pre každý kanál sa na začiatku nastavujú počiatočné podmienky, ktoré zodpovedajú stavu pri spontánnej aktivite nervového vlákna:

- m = 1.0 je maximálne množstvo voľného neurotransmitera v zásobníku zo systému diferenciálnych rovníc (12).
- A = 5 je konštanta pre permeabilitu membrány zo vzorca (11).
- B = 300 je konštanta pre permeabilitu membrány zo vzorca (11).
- g = 2000 je konštanta pre permeabilitu membrány zo vzorca (11).
- x = 66.31 vyjadruje rýchlosť uvoľňovania neurotransmitera z recyklačného zásobníka zo systému diferenciálnych rovníc (12).
- y = 5.05 určuje rýchlosť dopĺňania neurotransmitera zo systému diferenciálnych rovníc (12).
- l = 2500 určuje rýchlosť strát neurotransmitera zo synaptickej štrbiny zo systému diferenciálnych rovníc (12).

- r = 6580 vyjadruje rýchlosť návratu neurotransmitera do recyklačného zásobníka zo systému diferenciálnych rovníc (12).

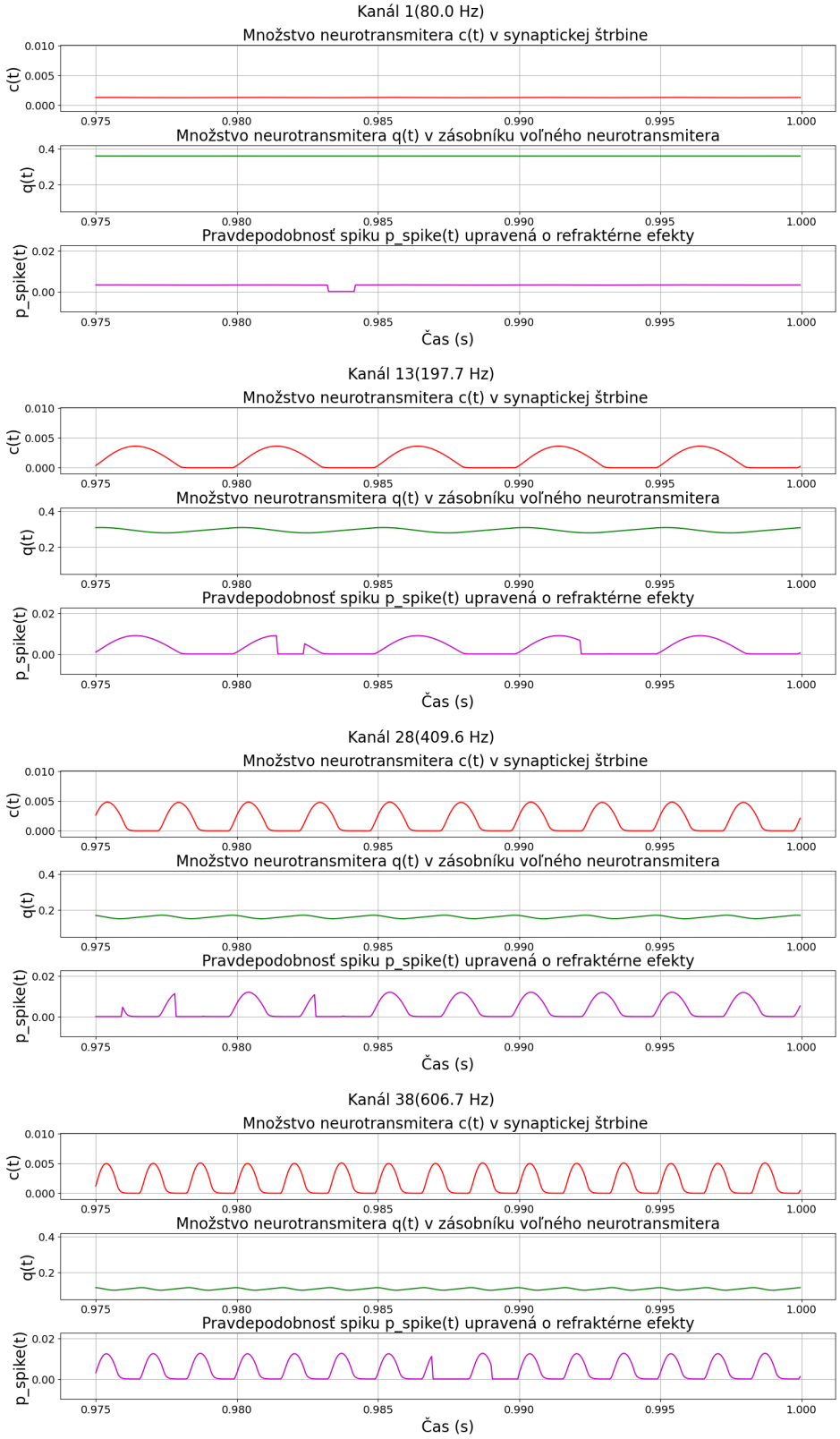
- h = 50000 je konštanta pre pravdepodobnosť spiku zo vzorca (10)

Po úvodnej fáze ticha sa začína spracovanie samotného vstupného signálu. V tomto bode sa amplitúda vstupného signálu upraví podľa zvolenej intenzity 80 dB. Výsledná hodnota signálu následne ovplyvňuje permeabilitu synaptickej membrány $k(t)$, ktorá určuje mieru uvoľňovania neurotransmitera zo zásobníka voľného neurotransmitera do synaptickej štrbiny. Ďalej sa počítajú jednotlivé príspevky ku zmenám v množstvách neurotransmitera: dopĺňanie zásobníka voľného neurotransmitera, jeho uvoľňovanie do synaptickej štrbiny, strata zo synaptickej štrbiny, spätná reabsorbcia a presun späť do zásobníka voľného neurotransmitera. Tieto zmeny sa prejavujú v aktuálnych hodnotách premenných $q(t)$, $c(t)$ a $w(t)$.

Od množstva neurotransmitera v synaptickej štrbine sa potom odvíja pravdepodobnosť spiku (10), pričom sa zohľadňuje aj citlivosť synapsie $h$ a veľkosť časového kroku $dt$. Spike vznikne, ak náhodne vygenerované číslo z rovnomerného rozdelenia na intervale $[0, 1)$ neprekročí túto pravdepodobnosť a zároveň od posledného spiku uplynulo viac než trvanie refraktérneho obdobia, ako opisujeme v odseku 5.1.4. Tým sa zohľadňuje biologický fakt, že po vzniku spiku nasleduje krátke obdobie, počas ktorého je opätovný vznik spiku veľmi nepravdepodobný.

Výsledné premenné sa uchovávajú samostatne pre každý kanál. Najdôležitejšie z nich, množstvo neurotransmitera v synaptickej štrbine $c(t)$, množstvo neurotransmitera v zásobníku $q(t)$ a pravdepodobnosť spiku upravená o refraktérne efekty $p_spike(t)$ môžeme vidieť vykreslené na Obr. 14. Zobrazené kanály reprezentujú rôzne časti spektra vstupného signálu, pričom sledujeme, ako model reaguje v závislosti od blízkosti centrálnej frekvencie kanála k frekvenčným zložkám vstupného signálu. Na Obr. 14 sú zobrazené výsledky simulácie získané našim programom MH pre štyri kanály s rôznymi centrálnymi frekvenciami: 80 Hz, 197.7 Hz, 409.6 Hz a 606.7 Hz.

V prípade kanála 1 s centrálnou frekvenciou $80,0$ Hz pozorujeme, že všetky tri veličiny zostávajú takmer konštantné. Množstvo neurotransmitera $c(t)$ v synaptickej štrbine je na veľmi nízkej hodnote a nevykazuje žiadne výraznejšie oscilácie. Množstvo neurotransmitera $q(t)$ v zásobníku je prakticky stabilné, bez významných zmien. Pravdepodobnosť




**Obr. 14:** Ukážka výsledku simulácie pre parametre : množstvo neurotransmitera $c(t)$ v synaptickej štrbine, množstvo neurotransmitera $q(t)$ v zásobníku a pravdepodobnosť spiku $p_spike(t)$ kanálov s centrálnymi frekvenciami 80 Hz, 197.7 Hz, 409.6 Hz a 606.7 Hz. (Zdroj: vlastné spracovanie)

spiku $p_spike(t)$ je v celom intervale prakticky nulová. Tento priebeh naznačuje, že pre tento kanál je vstupný signál veľmi slabý, čím je aktivita spikov minimálna.

Takéto správanie je typické pre kanály, ktorých centrálne frekvencie nerezonujú so žiadnou zložkou vstupného signálu. V týchto prípadoch sa do synaptickej štrbiny neuvoľňuje dostatok neurotransmitera na to, aby vyvolal aktiváciu v postsynaptickom nervovom vlákne. Výsledkom je, že generovanie spikov má len spontánny charakter. Tieto kanály teda nereflektujú periodicitu vstupu a ich príspevok k detekcii výšky tónu je zanedbateľný.

Naopak, v kanáli 13 s centrálnou frekvenciou $197,7$ Hz už vidíme výraznejšiu dynamiku. Množstvo neurotransmitera $c(t)$ v synaptickej štrbine vykazuje pravidelné oscilácie, čo odráža cyklické uvoľňovanie neurotransmitera. Množstvo neurotransmitera $q(t)$ v zásobníku ukazuje obdobné oscilácie v dôsledku jeho uvoľňovania zo zásobníka do synaptickej štrbiny. Pravdepodobnosť spiku $p_spike(t)$ je vyššia než v prvom kanáli. Vidíme tu aj známky refraktérnych efektov: po prudkom zvýšení pravdepodobnosti spiku nasleduje krátke obdobie utlmenia, kedy pravdepodobnosť spiku opäť klesá. Ako sme už uviedli, toto správanie je dôsledkom fyziologického mechanizmu, kde po spiku nasleduje fáza zníženej aktivity nervového vlákna.

V kanáli 28 s frekvenciou $409,6$ Hz je dynamika ešte výraznejšia. Množstvo neurotransmitera $c(t)$ v synaptickej štrbine ukazuje rýchlejšie a ostrejšie oscilácie, čo reflektuje vyššiu frekvenciu spracovávaného signálu. Množstvo neurotransmitera $q(t)$ v zásobníku sa pravidelne mení. Spiky sa v tomto kanáli vyskytujú s vysokou pravdepodobnosťou. Refraktérne efekty sú opäť zreteľné: po zvýšení pravdepodobnosti spiku vidíme mierne potlačenie ďalšej pravdepodobnosti na veľmi krátky čas.

V prípade kanála 38 s frekvenciou $606,7$ Hz vidíme najvýraznejšiu dynamiku zo všetkých analyzovaných kanálov. Množstvo neurotransmitera $c(t)$ v synaptickej štrbine osciluje rýchlo a s vysokou amplitúdou, čo naznačuje silnú synchronizáciu medzi vstupným signálom a uvoľňovaním neurotransmitera. Množstvo neurotransmitera $q(t)$ v zásobníku ukazuje sústavné oscilácie. Pravdepodobnosť spiku $p_spike(t)$ má v tomto kanáli veľmi výrazný priebeh: prakticky každá oscilácia neurotransmitera vedie k prudkému nárastu pravdepodobnosti spiku. Refraktérne efekty sa prejavujú veľmi typickým spôsobom, po spiku nasleduje krátkodobý pokles aktivity, čím sa zabezpečuje biologická obnova nervového vlákna pred ďalším spikom.

Maximálna hodnota pravdepodobnosti spiku po zohľadnení refraktérných efektov bola na úrovni približne $p_spike \approx 0.015$. Táto hodnota predstavuje pravdepodobnosť spiku v jednom časovom kroku $dt$, čo možno prepočítať na očakávaný počet spikov za sekundu (angl. *firing rate*) nasledovne:

$$\text{Firing rate} = \text{spike} \cdot \text{vzorkovacia frekvencia} = 0.015 \cdot 20\,000 = 300 \text{ spike/s}. \tag{23}$$

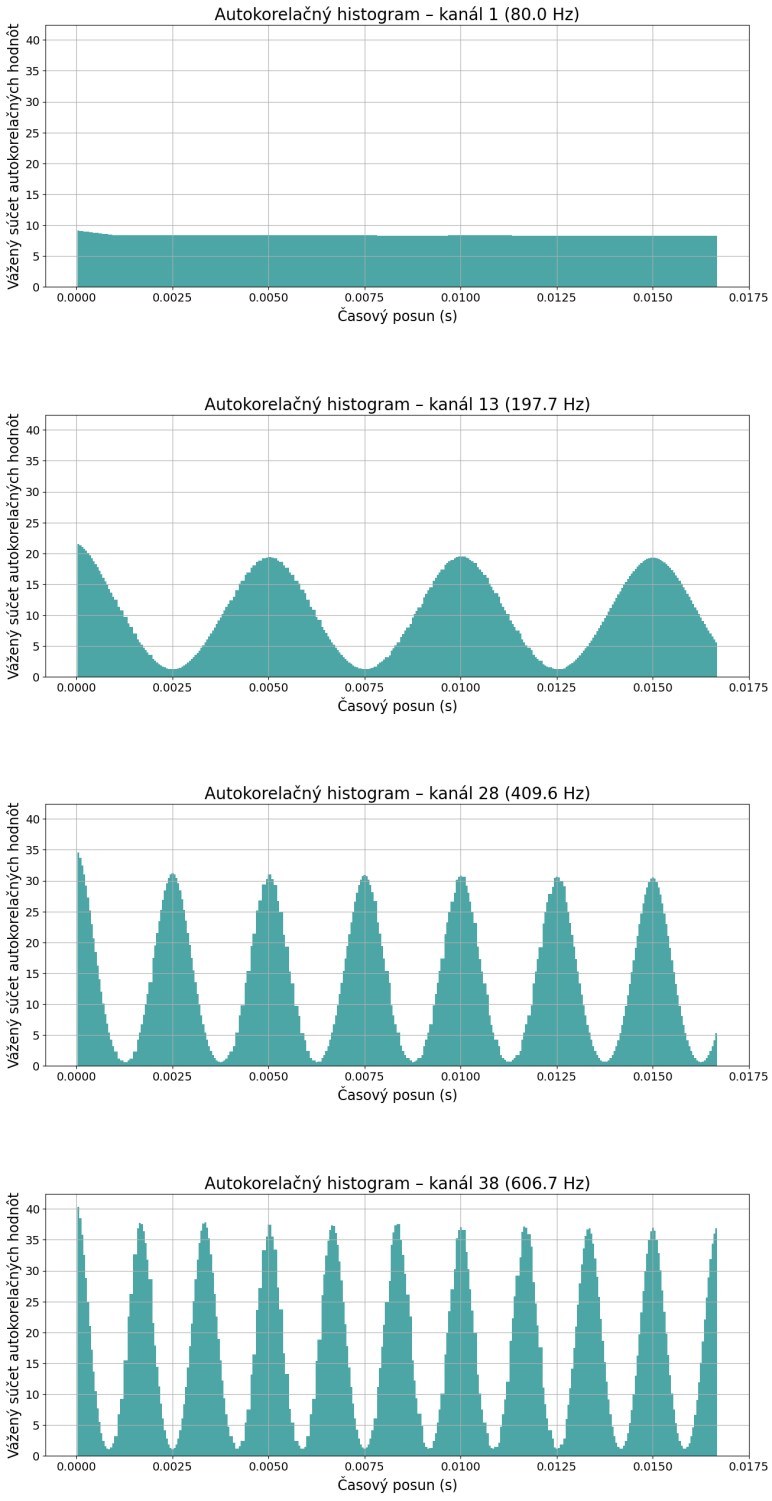
V nasledujúcom kroku program vytvára pre každý kanál zvlášť histogram časových intervalov medzi medzi spikmi. Namiesto toho, aby boli sledované len intervaly medzi po sebe idúcimi spikmi, do úvahy boli brané všetky možné kombinácie dvojíc spikov v danom kanáli. Ako sme uviedli v odseku 5.1.5 pravdepodobnosť výskytu určitého časového intervalu medzi dvoma spikmi je vyjadrená pomocou jednoduchého súčinu ich pravdepodobností výskytu v čase $t$ a $t - \delta t$ (15).

Intervaly $\delta t$ sme obmedzili na rozsah od $0,00005$ s do $0,01667$ s. Spodná hranica zodpovedá minimálnej vzorkovacej perióde modelu, vrchná hranica zodpovedá frekvencii 60 Hz, čo je spodná hranica výšky tónu, ktorú by sme mohli potencionálne brať ako kandidáta na fundamentálnu frekvenciu. Aby sme zohľadnili aj časové rozostupy medzi spikmi a ich význam v čase, aplikovali sme exponenciálne váženie s časovou konštantou $\tau = 2,5$ms. Pre výpočet histogramu sme využili vzťah (16) s hore uvedenými obmedzeniami hodnoty $\delta t$.

Týmto spôsobom bol realizovaný výpočet autokorelačnej funkcie (ACF) popísaný v odseku 5.1.5, čo zodpovedá diskrétnej aproximácií spojitého vzorca (6). Výsledné autokorelačné funkcie potom slúžili ako základ pri hľadaní fundamentálnej frekvencie. Keďže exponenciálne váženie spôsobuje, že hodnoty rýchlo klesajú s rastúcim časovým odstupom, histogram bol počítaný len v rozsahu $7,5$ ms do minulosti (t. j. od aktuálneho času $t$ s po $t - 0,0075$ s). Tento interval bol dostatočný na zachytenie relevantných periodicít v signáli.

Na Obr. 15 sú zobrazené autokorelačné histogramy pre štyri rôzne kanály signálu s centrálnymi frekvenciami 80 Hz, 197.7 Hz, 409.6 Hz a 606.7 Hz. V kanáli 1 s centrálnou frekvenciou 80 Hz je histogram plochý a bez výrazných vrcholov. Nezaznamenáva sa výrazná periodicita spikov.

Kanál 13 s centrálnou frekvenciou 197.7 Hz už v histograme obsahuje periodickú štruktúru, s maximálnymi hodnotami približne každých 5 ms. Tento interval zodpovedá frek-




**Obr. 15:** Autokorelačné histogramy pre štyri vybrané kanály s rôznymi centrálnymi frekvenciami 80 Hz, 197, 7 Hz, 409, 6 Hz a 606, 7 Hz. Na horizontálnej osi je znázornený časový posun v sekundách a na vertikálnej osi počet výskytov autokorelovaných hodnôt. (Zdroj: vlastné spracovanie)

vencii 200 Hz, čo korešponduje s centrálnou frekvenciou tohto kanála. To naznačuje, že práve tento kanál obsahuje spiky dobre synchronizované s periódou vstupného signálu, a teda pravdepodobne významne prispieva k odhadu výšky tónu.

V pásme kanála 28 (409.6 Hz) je periodicita spikov veľmi zreteľná. Histogram vykazuje pravidelné maximá každých približne 2.5 ms, čo zodpovedá frekvencii 400 Hz. Táto zložka signálu je teda v spikoch daného kanála veľmi výrazne zastúpená a spiky prebiehajú so silnou časovou synchronizáciou. Jasne definované maximá v histograme potvrdzujú, že tento kanál zachytáva harmonickú zložku vstupného signálu. Histogram spikov v kanáli 38 (606.7 Hz) vykazuje ešte ostrejšiu a pravidelnejšiu periodicitu. Maximá sa objavujú približne každých 1.65 ms, čo zodpovedá frekvencii 606 Hz, teda presne centrálnej frekvencii kanála.

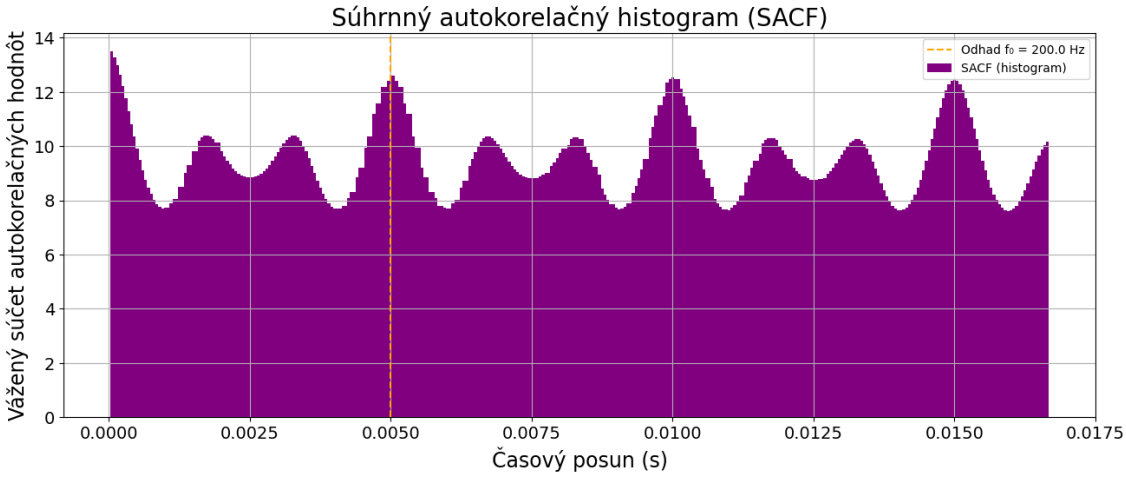
V kanáloch, ktorých centrálne frekvencie nerezonujú so žiadnou frekvenčnou zložkou vstupného signálu, napríklad ako v kanáli 1, autokorelačný histogram nevykazuje výrazné periodicity. V takýchto prípadoch spiky nevykazujú žiadnu jasnú časovú štruktúru ani opakujúci sa vzor, čo sa prejavuje ako plochý histogram bez dominantných vrcholov. Dôvodom je, že vstupný signál v týchto pásmach neobsahuje žiadne zložky, takže mechanicko-neuronová transdukcia v príslušnom kanáli generuje spiky len sponténne. Tieto kanály preto významne neprispievajú k výpočtu periodicity ani pri odhade fundamentálnej frekvencie. Takéto správanie zároveň potvrdzuje selektivitu modelu, keďže periodicita je zachytená len v tých frekvenčných pásmach, ktoré skutočne obsahujú relevantné harmonické zložky.

Ako sme uviedli v odseku 5.1.6, s použitím individuálych ACF sa spriemerovaním cez všetky kanály podľa vzťahu (17) určí súhrnná autokorelačná funkcia (SACF), ktorá slúži ako základ pre extrakciu fundamentálnej frekvencie $f_0$. Na jej určenie sa použije pozícia maxima v SACF. Toto maximum zodpovedá najpravdepodobnejšiemu periodickému intervalu vstupného signálu, a teda aj fundamentálnej frekvencii.

Na Obr. 16 je zobrazený súhrnný autokorelačný histogram pre použitý vstupný signál. V grafe je možné pozorovať výrazné periodické maximá s časovým rozostupom približne 0.005 s, čo zodpovedá fundamentálnej frekvencii 200 Hz:

$$f_0 = \frac{1}{0.005} = 200 \text{ Hz}$$

Tento výsledok zodpovedá fundamentálnej frekvencii vstupného signálu, ktorá je v




**Obr. 16:** Graf súhrnnej autokorelačnej funkcie (SACF) vypočítanej programom zodpovedajúcim použitému vstupnému signálu. SACF má pravidelne sa opakujúce maximá s rozostupom približne 0.005 sekundy, čo potvrdzuje periodicitu zodpovedajúcu 200 Hz, teda fundamentálnu frekvenciu vstupného tónu. (Zdroj: vlastné spracovanie)

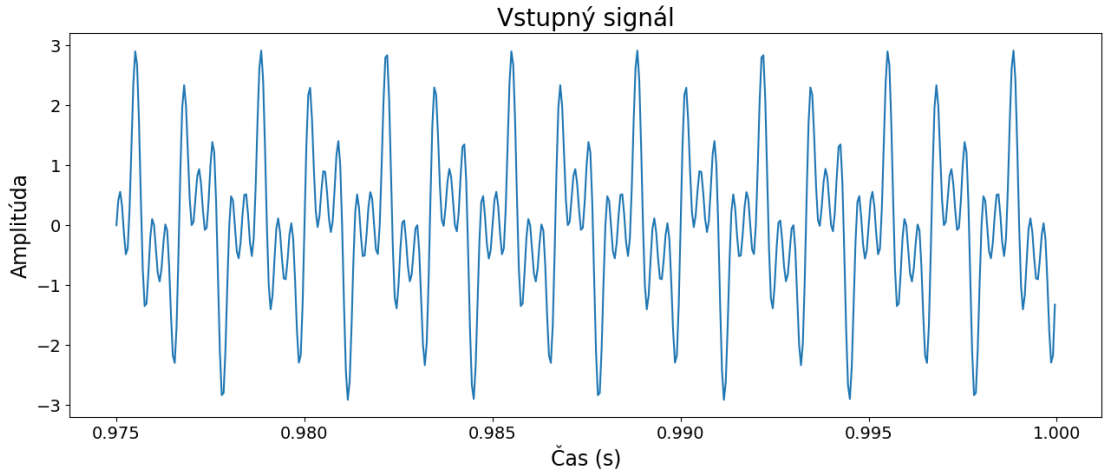
grafe vyznačená zvislou čiarou.

## 6.2 Použitie programu MH na signál s chýbajúcou fundamentálnou frekvenciou

V tejto podkapitole budeme ilustrovať použitie nášho programu MH na spracovanie zloženého periodického signálu, medzi zložkami ktorého sa nenachádza zložka s jeho fundamentálnou frekvenciou. Ide o prípad označovaný ako *missing fundamental*. Je významný tým, že vyvoláva u poslucháča vnem výšky tónu zodpovedajúcej fundamentálnej frekvencii, hoci zložka s touto frekvenciou v signáli chýba. Použitie programu budeme ilustrovať na vstupnom signáli danom predpisom:

$$y(t) = sin(2\pi 600t) + sin(2\pi 1500\ t) + sin(2\pi 2400t). \tag{24}$$

Ako prvé signál prechádza pásmovým filtrom, ktorého výsledok je znázornený na Obr. 18. Ako je zrejmé z (24), vstupný testovací signál je zložený z troch sínusových zložiek s frekvenciami 600 Hz, 1500 Hz a 2400 Hz. Po prechode signálu cez pásmový filter, ktorý modeluje vlastnosti vonkajšieho a stredného ucha, dochádza k najväčšiemu tlmeniu zložky



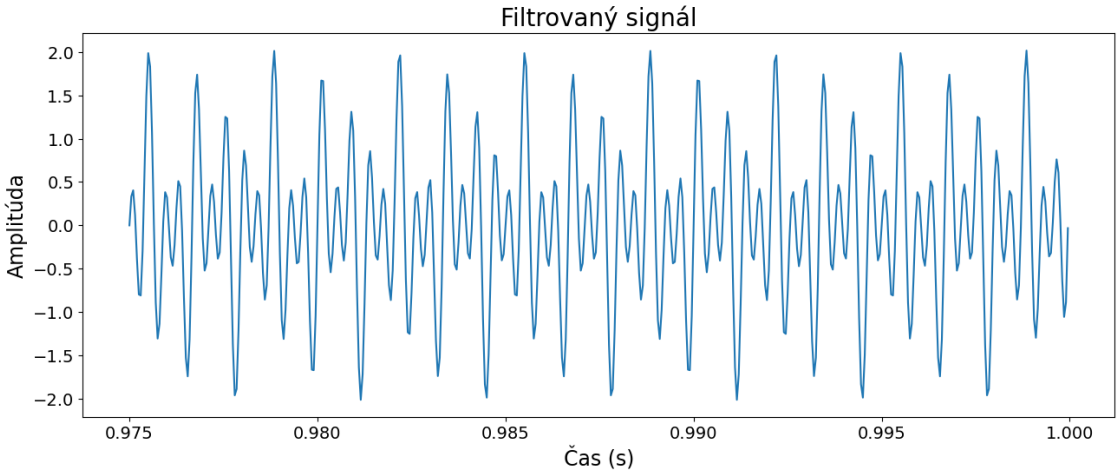

**Obr. 17:** Vstupný signál, ktorý je tvorený superpozíciou troch sínusových zložiek s frekvenciami 600 Hz, 1500 Hz a 2400 Hz. Tento signál slúži ako vstup pre ďalšie spracovanie (Zdroj: vlastné spracovanie)

s frekvenciou 600 Hz, pretože leží na okraji priepustného pásma filtra. Vo výslednom filtrovanom signáli zostávajú všetky frekvenčné komponenty, ale s menšou amplitúdou. Tento efekt sa prejavuje aj zreteľným poklesom celkovej amplitúdy signálu, čo potvrdzuje účinnosť simulovaného filtra a jeho schopnosť tlmiť nízkofrekvenčné zložky tak, ako to robí aj ľudské ucho v reálnych podmienkach.

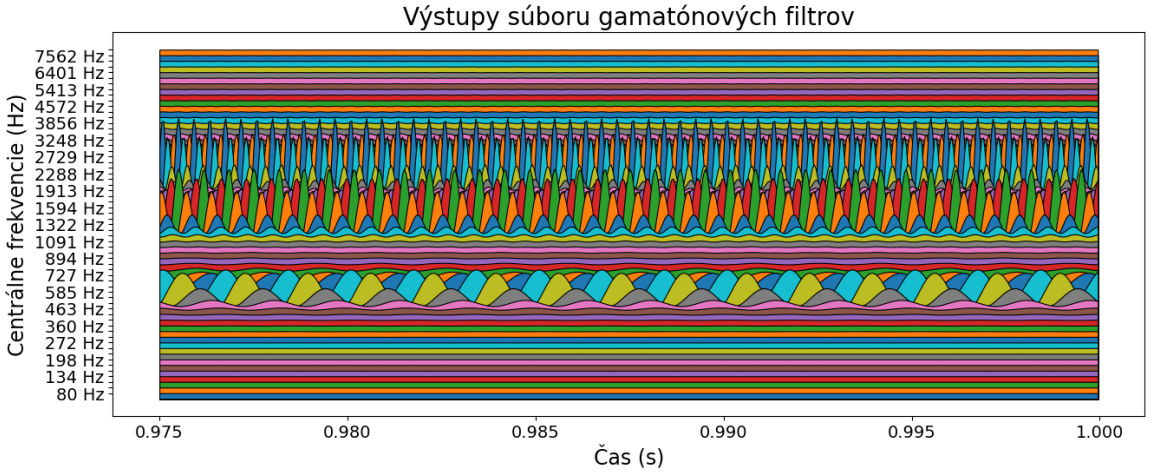
Po spracovaní vstupného signálu simuláciou vlastností vonkajšieho a stredného ucha pomocou pásmového filtra, nasleduje ďalší krok, modelovanie procesov v slimáku pomocou súboru gamatónových filtrov. Na Obr. 19 sú znázornené výstupy súboru gamatónových filtrov aplikované na filtrovaný signál. Najsilnejšia odpoveď sa objavuje okolo 1500 Hz a 2400 Hz, čo potvrdzuje predchádzajúce pozorovanie, že 600 Hz zložka bola v predchádzajúcom kroku najviac tlmená.

Ďalej nasleduje simulácia mechanicko-neurónovej transdukcie. Tento proces opisuje prechod mechanických vibrácií bazilárnej membrány na neurónovú odpoveď v sluchovom nerve. Obraz o časových štruktúrach odpovedí nervového systému poskytujú autokorelačné histogramy. Tieto histogramy sú počítané pre jednotlivé výstupy gamatónových filtrov.

Na výsledných autokorelačných histogramoch (Obr. 20) možno pozorovať výrazné pe-

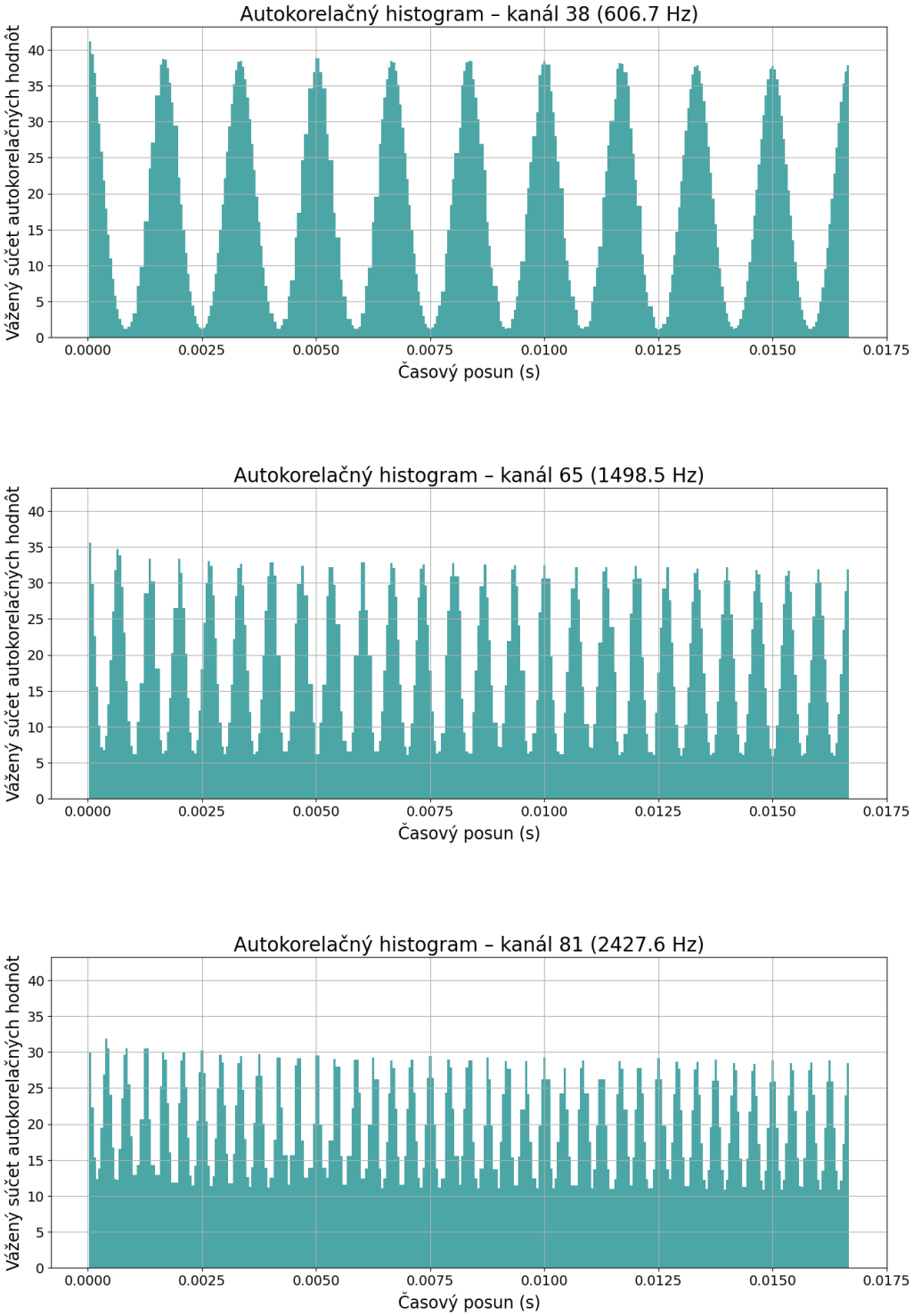



**Obr. 18:** Výstup filtrovaného signálu po čiastočnom utlmení zložiek vstupného signálu, osobitne zložky s najnižšou frekvenciou 600 Hz. (Zdroj: vlastné spracovanie)



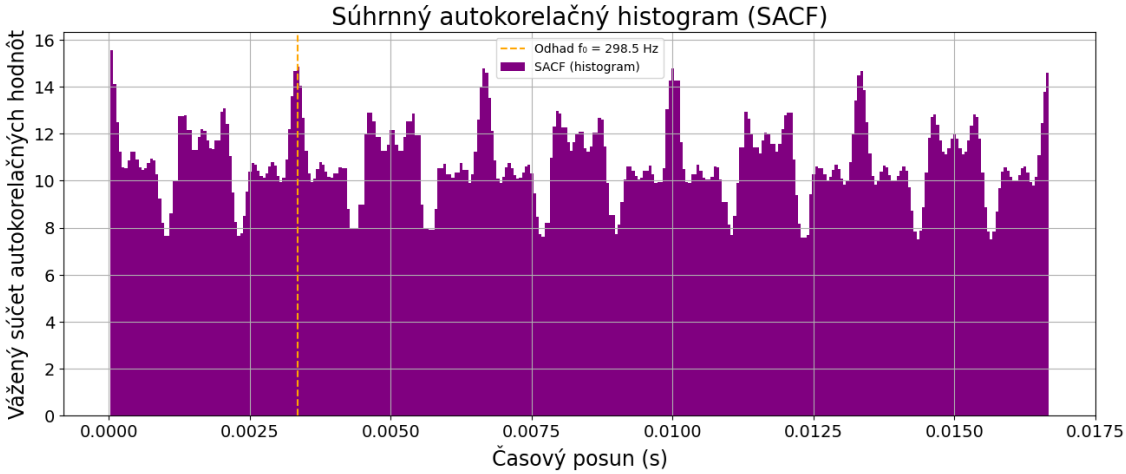

**Obr. 19:** Výstupy zo súboru gamatónových filtrov aplikované na vstupný signál. Na obrázku je možné pozorovať, ako jednotlivé frekvenčné zložky vstupného signálu (s frekvenciami 600 Hz, 1500 Hz a 2400 Hz) aktivujú filtre s im blízkymi centrálnymi frekvenciami. Najvýraznejšia aktivita sa nachádza v pásmach zodpovedajúcich týmto frekvenciám, čo odráža spektrálne zloženie vstupného signálu. (Zdroj: vlastné spracovanie)

riodicity v odpovediach filtrov, ktorých centrálne frekvencie sú blízke frekvenčným zložkám vstupného signálu, teda približne 600 Hz, 1500 Hz a 2400 Hz. V príslušných filtroch sa objavujú výrazné vrcholy v krátkych časových intervaloch zodpovedajúcich periódam




**Obr. 20:** Autokorelačné histogramy pre tri kanály so stúpajúcou centrálnou frekvenciou: 606, 7 Hz, 1498, 5 Hz a 2427, 6 Hz. Každý histogram zobrazuje periodické opakovanie štruktúr v signáli v závislosti od časového posunu, pričom vyššia frekvencia vedie k hustejšiemu rozloženiu vrcholov. (Zdroj: vlastné spracovanie)

príslušným k týmto frekvenciám. Histogramy tak poskytujú dôležité informácie pre ďalší krok spracovania, ktorým je odhad fundamentálnej frekvencie signálu.



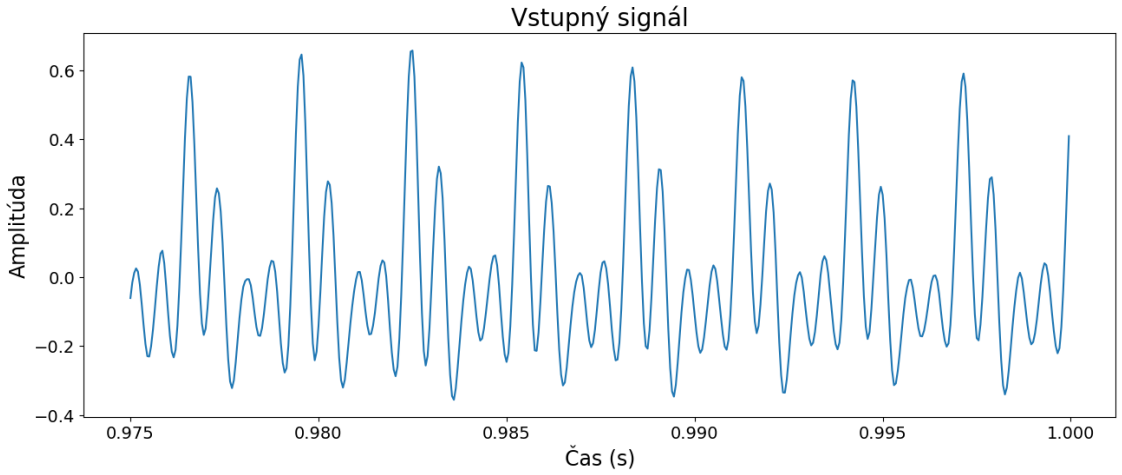

**Obr. 21:** Súhrnný autokorelačný histogram. Predstavuje spriemerovaný súčet autokorelačných histogramov jednotlivých kanálov. Vyznačená je aj pozícia maxima, zodpovedajúca odhadu fundamentálnej frekvencie $f_0 = 298,5$ Hz. (Zdroj: vlastné spracovanie)

Na získanie celkového obrazu periodicity naprieč celým spektrom bol zostavený súhrnný autokorelačný histogram (SACF) ako spriemerovaný súčet histogramov jednotlivých kanálov. Výsledný SACF odhalil fundamentálnu frekvenciu $f_0 = 298,5$ Hz, ktorá síce nebola priamo prítomná vo vstupnom signáli, ale zodpovedá približne najväčšiemu spoločnému deliteľovi harmonických 600 Hz, 1500 Hz, 2400 Hz, teda fundamentálnej frekvencií 300 Hz.

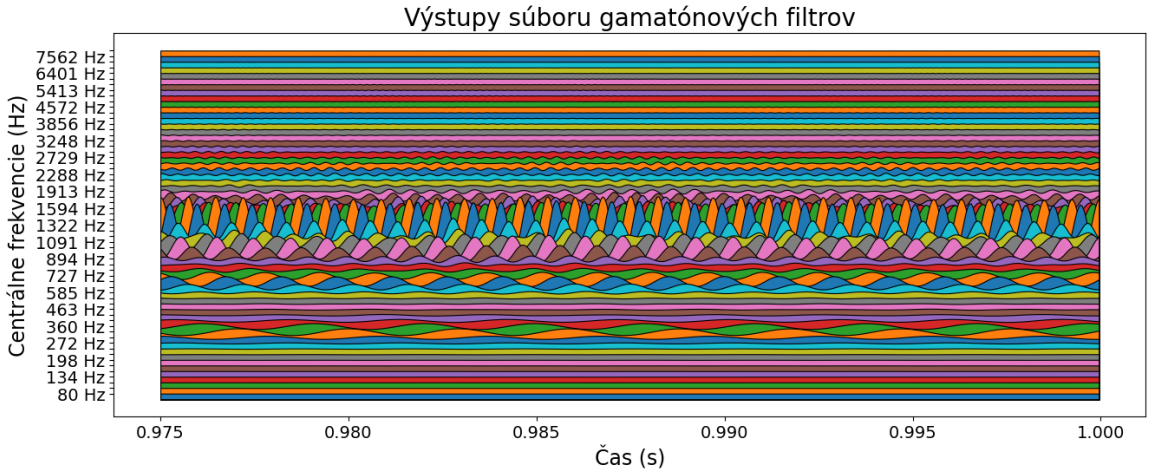
## 6.3 Použitie programu MH na tón trúbky

Náš program implementujúci Meddis–Hewittov model sme sa rozhodli otestovať aj na prirodzene generovanom zvukovom signáli a tak overiť jeho schopnosť zachytiť relevantné črty zložitejších zvukových signálov. Ako prvý prirodzene generovaný signál sme zvolili nahrávku jedného tónu zahraného na štandardnej trúbke. Jednalo sa konkrétne o tón $F4$ (jednočiarkované f). Nahrávku sme realizovali programom Audacity a jej spektrálna analýza realizovaná týmto programom vykázala ako fundamentálnu frekvenciu zahraného tónu hodnotu 341 Hz. Pre úplnosť uvádzame, že v modernom rovnomerne temperovanm

ladení zodpovedá tónu $F4$ hodnota frekvencie 349 Hz, čo znamená, že nami použitá trúbka bola v čase záznamu naladená o trochu nižšie. Výstupy z jednotlivých fáz nášho programu použitého na vyššie uvedený zvukový signál sú zachytené na Obr. 22 – Obr. 25.

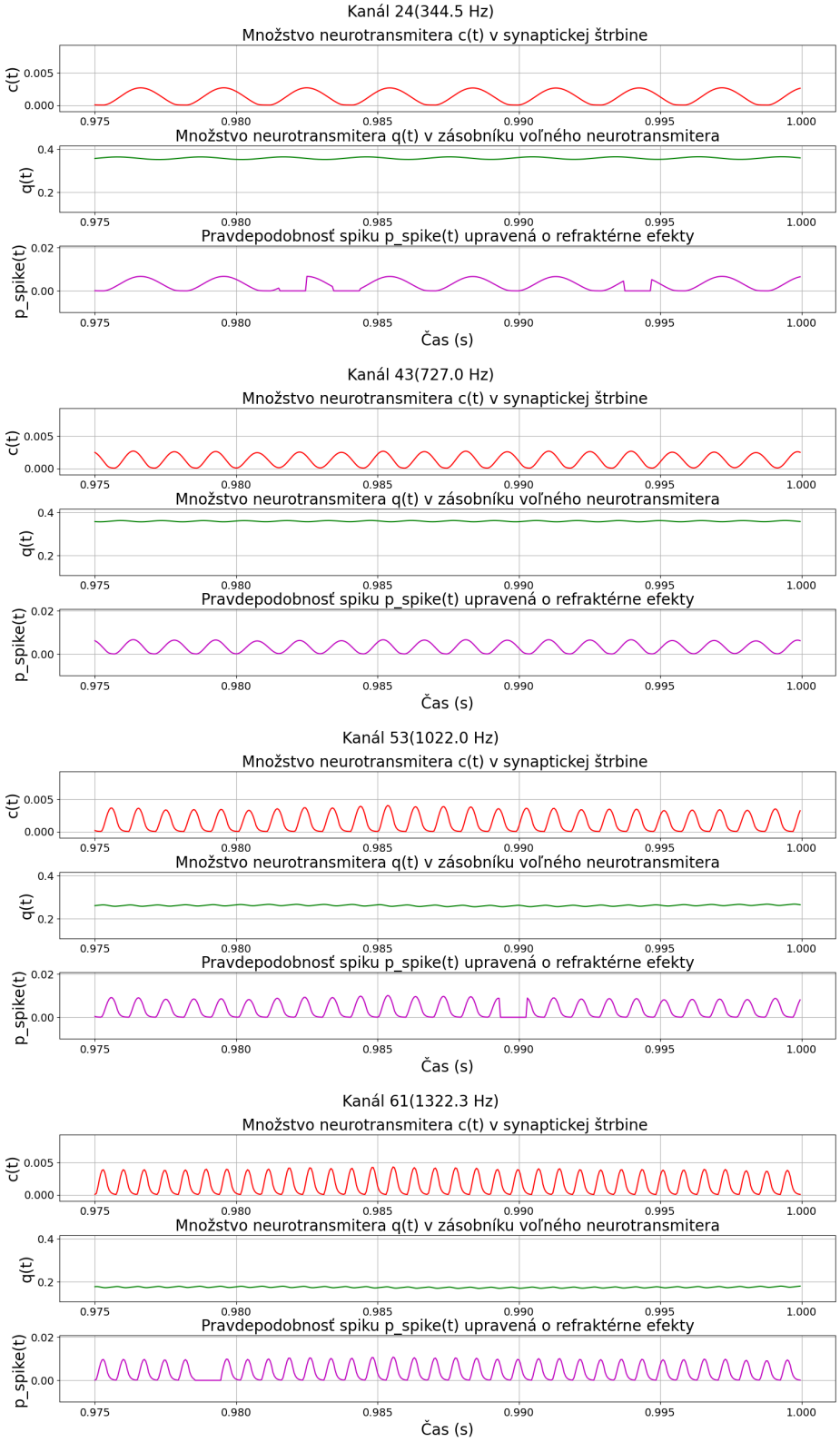



**Obr. 22:** Časový priebeh vstupného zvukového signálu trúbky, tón $F4$. (Zdroj: vlastné spracovanie)




**Obr. 23:** Výstupy súboru gamatónových filtrov pre signál trúbky. Zreteľná aktivita v kanáloch okolo 350 Hz, 700 Hz, 1100 Hz a 1350 Hz poukazuje na prítomnosť harmonických zložiek v blízkosti týchto frekvencií. (Zdroj: vlastné spracovanie)

Vo výstupoch kanáloch zo súboru gammatónových filtrov na Obr. 23, ktoré vykazovali



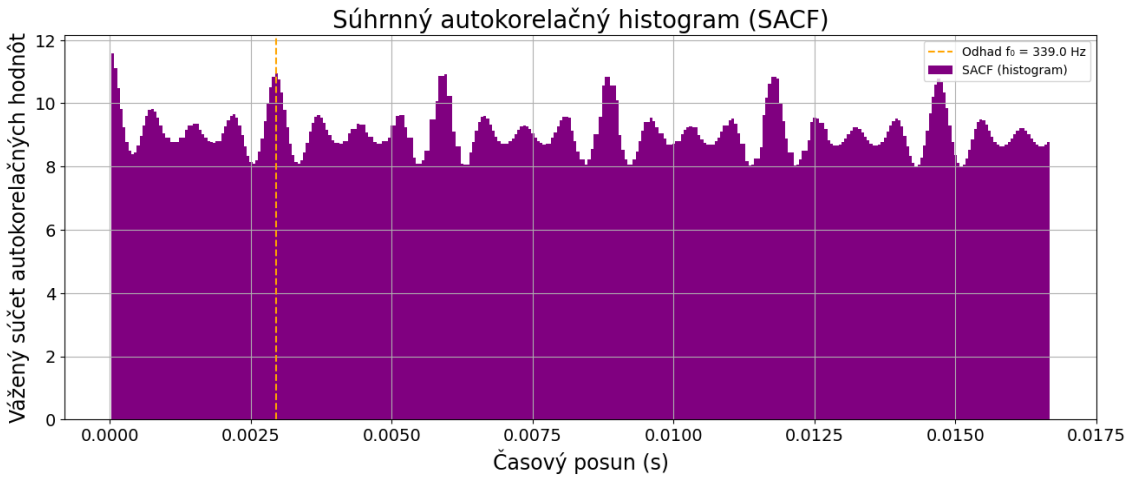

**Obr. 24:** Dynamika synaptických premenných $c(t), q(t)$ a $p_spike(t)$ charakterizujúcich mechanicko- nervovú transdukciu v štyroch vybraných kanáloch gamatónového súboru filtrov s centrálnymi frekvenciami 344.5 Hz, 727.0 Hz, 1022.0 Hz a 1322.3 Hz pre zvukový signál trúbky. (Zdroj: vlastné spracovanie)

zvýšenú aktivitu, je možné pozorovať výraznejšie dynamické zmeny premenných $c(t)$, $q(t)$ a $p_spike(t)$. V týchto kanáloch dochádza k intenzívnejšiemu vyčerpávaniu neurotransmitera zo zásobníka a jeho rýchlejšiemu presunu do synaptickej štrbiny. Množstvo neurotransmitera v synaptickej štrbine kolíše podľa amplitúdy výstupného signálu v danom kanáli. Zásobník sa dynamicky vyčerpáva a dopĺňa, pričom model zachytáva oneskorené dopĺňanie neurotransmitera počas silnej stimulácie.

Ako môžeme vidieť na Obr. 24 ,pravdepodobnosť spiku v nervovom vlákne odvodená od aktuálneho množstva neurotransmitera v synaptickej štrbine a následne upravená s prihliadnutím na refraktérne efekty dosahuje maximálnu hodnotu $p_spike \approx 0.01$. Z tejto hodnoty možno odvodiť očakávanú frekvenciu spikov (angl. *firing rate*):

$$\text{Firing rate} = \text{spike} \cdot \text{vzorkovacia frekvencia} = 0.01 \cdot 20\,000 = 200 \text{ spike/s}. \tag{25}$$

Táto hodnota je v rozsahu hodnôt uvádzaných v experimentálnych štúdiách v tejto oblasti pre zvukové signály podobného charakteru ako nami použitý zvukový signál.




**Obr. 25:** Súhrnný autokorelačný histogram (SACF) pre signál trúbky. SACF dosahuje prvé maximum pri časovom oneskorení zodpovedajúcom frekvencii približne 339 Hz, čo takmer presne zodpovedá frekvencii vstupného tónu $F4$, ktorá bola 341 Hz (Zdroj: vlastné spracovanie)

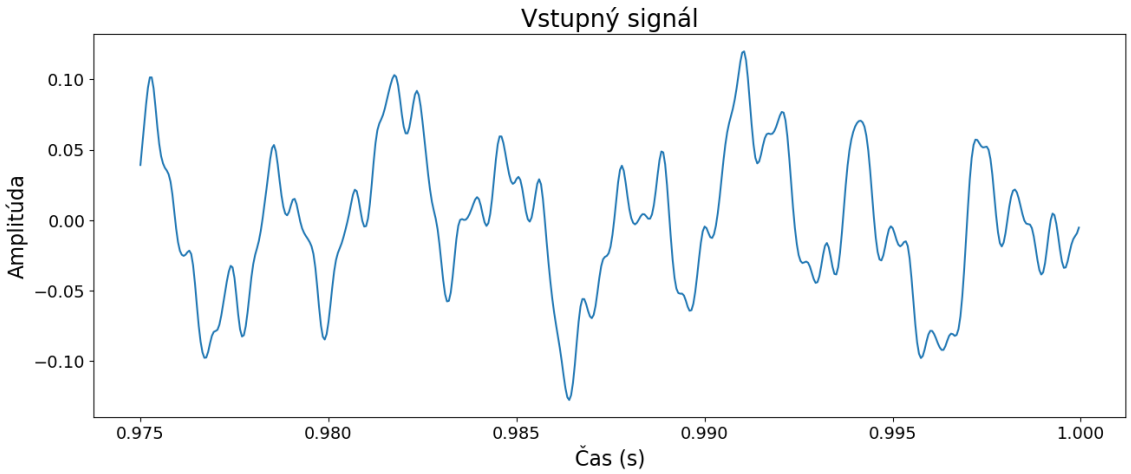
Maximum v SACF priamo súvisí fundamentálnou frekvenciou $f_0$. Z Obr. 25 možno určiť frekvenciu ako $f_0 = 339$ Hz, čo je takmer totožné s frekvenciou vstupného tónu $F4$, 341 Hz. Na základe vyššie uvedeného možno konštatovať, že výsledok potvrdzuje schopnosť

nášho programu implementujúceho Meddis–Hewittov model presne identifikovať výšku tónu aj pri spracovaní zvukového signálu generované reálnym hudobným nástrojom.

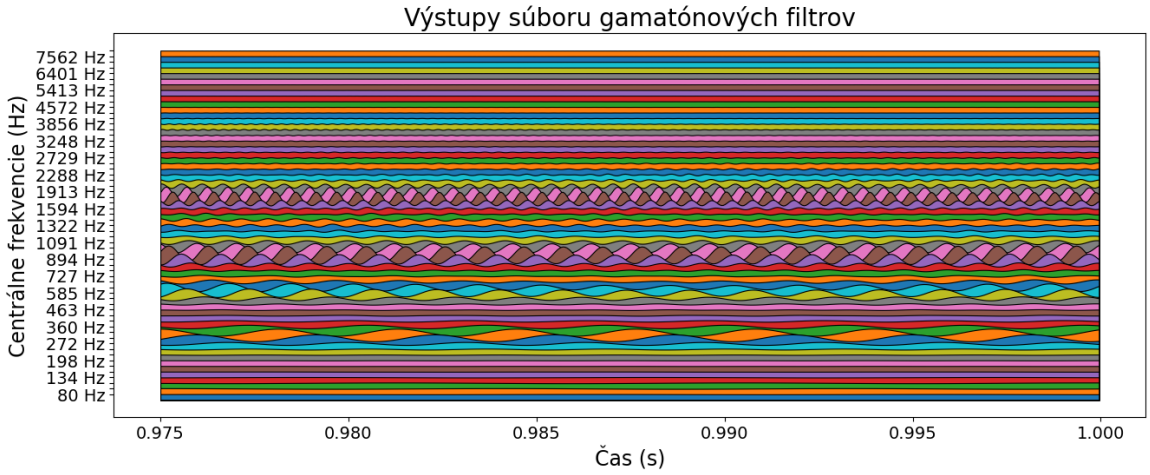
## 6.4 Použitie programu MH na zvuk tibetskej spievajúcej misky

Náš program sme sa rozhodli otestovať aj na analýze zvuku generovaného tibetskou spievajúcou miskou (angl *tibetan singing bowl*). Spievajúce misky sú misky vyrobené obvykle z bronzu. Možno ich rozozvučať úderom drevených paličiek s rôznymi hlavicami (napríklad z plste alebo kože) alebo trením ich okraja krúživým pohybom napríklad paličkou s koženým poťahom.

Z akustického hľadiska sa spievajúce misky chovajú podobne ako zvony [1]. Ich zvuk vzniká excitáciou ich rezonančných frekvencií. Výsledný zvuk pozostáva z viacerých blízkych frekvenčných zložiek, ktoré navzájom interferujú. Ich interferencia spôsobuje vznik rázov, teda pomalej zmeny amplitúdy signálu. Tento jav môžeme vnímať ako pulzovanie zvuku a je dôsledkom superpozície dvoch alebo viacerých frekvencií s malým rozdielom. Frekvenčné zložky nie sú, na rozdiel od klasických hudobných nástrojov, napríklad strunových alebo dychových, časťou žiadneho harmonického radu. Z toho vyplýva, že zvukový signál generovaný spievajúcou miskou nie je periodický. Z hľadiska poslucháča to znamená, že takýto zvukový signál vníma nie ako tón s jednou výškou, ale ako viacero tónov




**Obr. 26:** Priebeh zvukového signálu spievajúcej tibetskej misky. Signál má neperiodický charakter. (Zdroj: vlastné spracovanie)




**Obr. 27:** Výstupy zo súboru gamatónových filtrov pre signál tibetskej spievajúcej misky. Dominantné odozvy sa nachádzajú v kanáloch s centrálnymi frekvenciami približne 300 Hz, 600 Hz, 900 Hz a 1800 Hz. (Zdroj: vlastné spracovanie)

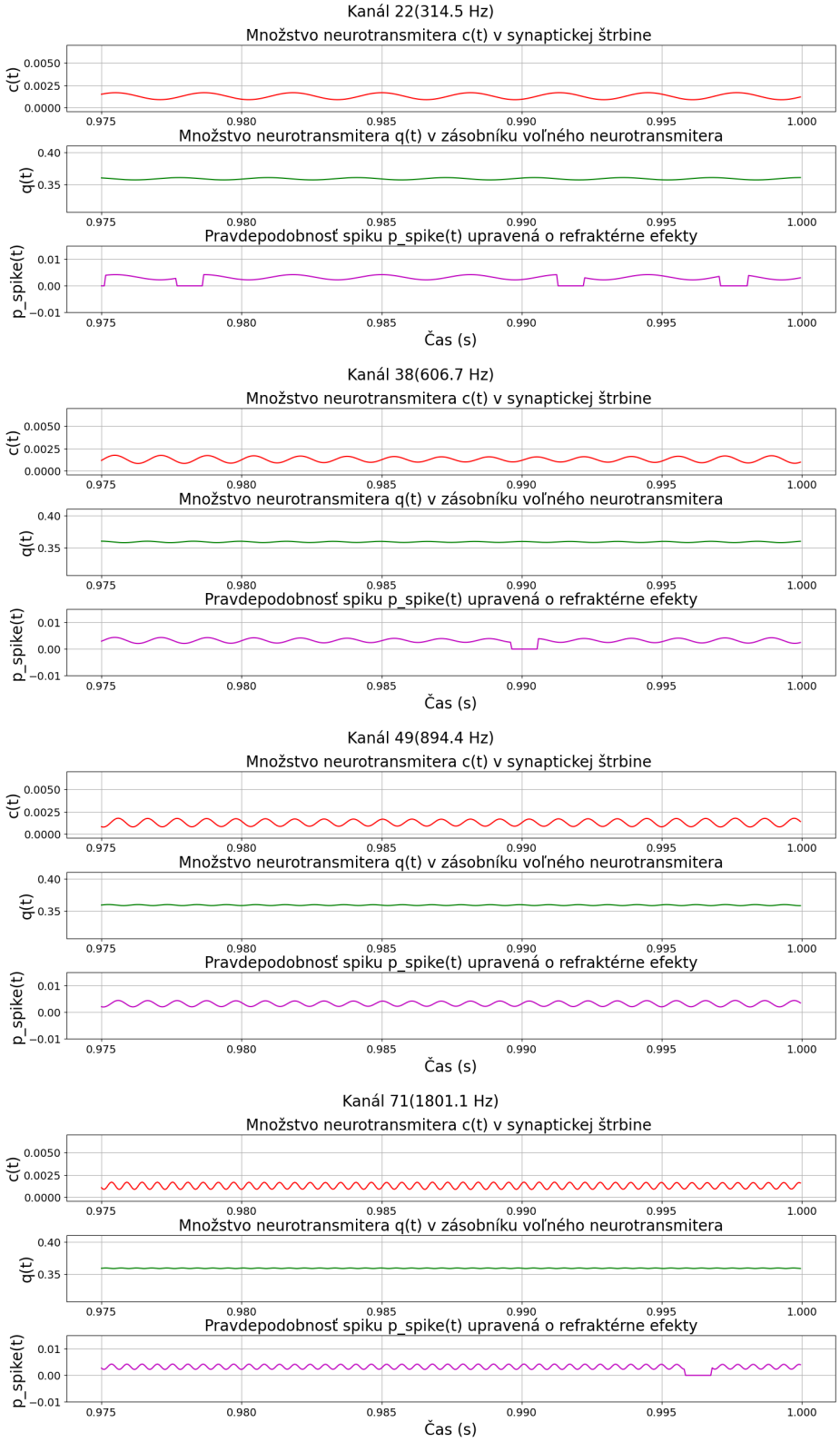
s rôznymi výškami.

Preto nás zaujímalo, ako bude na takto vygenerovaný zvukový signál reagovať Meddis–Hewittov model v implementácii realizovanej našim programom. Tak ako v prípade trúbky sme nahrávku zvukového signálu generovaného spievajúcou miskou realizovali programom Audacity. Spektrálna analýza realizovaná týmto programom ukázala, že najväčší prvok amplitúdového spektra tohto signálu sa nachádzal na frekvencii 315 Hz. Výstupy z jednotlivých fáz programu pri spracovaní zvuku spievajúcej misky sú zachytené na Obr. 26 – Obr. 29.

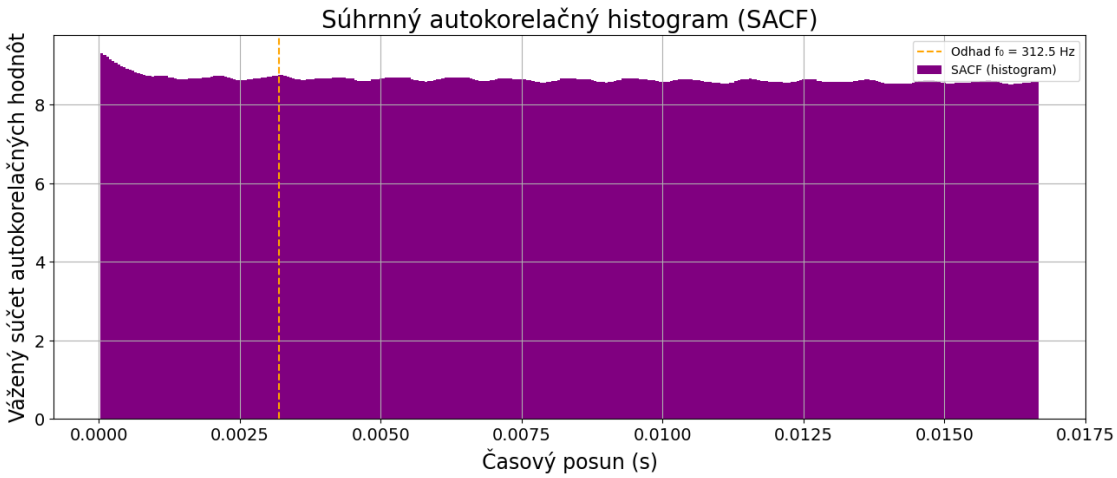
V porovnaní s výsledkami pre periodický tón (napr. tón trúbky) je vidieť podstatne nižšiu pravdepodobnosť spikov. Maximálna hodnota pravdepodobnosti spiku dosiahla v tomto prípade $p_spike \approx 0.005$. Výsledná rýchlosť generovania spikov bola teda:

$$\text{Firing rate} = \text{spike} \cdot \text{vzorkovacia frekvencia} = 0.005 \cdot 20,000 = 100 \text{ spike/s}. \tag{26}$$

V úvode tohto odseku sme uviedli, že zvukový signál spievajúcej misky obsahuje viacero frekvenčných zložiek, ktoré nie sú v jednoduchých harmonických pomeroch. Takýto spektrálny obsah vedie k spikom v rôznych filtrovaných kanáloch, ktorých časové rozmiestnenie nie je pravidelné a nevyvoláva synchronizované autokorelačné vzory. Výsledkom je, že spiky medzi kanálmi sa navzájom nezhodujú v typických časových intervaloch a SACF




**Obr. 28:** Dynamika premenných $q(t)$, $c(t)$ a $p_spike(t)$ charakterizujúcich mechnicko-nervovú transdukciu v štyroch vybraných kanáloch súboru gamatónových filtrov s centrálnymi frekvenciami 314.5 Hz, 606.7 Hz, 894.4 Hz a 1801.1 Hz pre zvukový signál spievajúcej misky (Zdroj: vlastné spracovanie)




**Obr. 29:** Súhrnný autokorelačný histogram (SACF) pre zvukový signál spievajúcej misky. Vrchol pri 312.5 Hz naznačuje periodickú zložku, hoci celkový signál je neperiodický. (Zdroj: vlastné spracovanie)

sa preto javí ako relatívne rovnomerný a bez výrazných maxím.

Ako možno vidieť na Obr. 29, použitý algoritnus však určil maximum pre frekvenciu $f_0 \approx 312.5$ Hz, ktorá podľa vyššie uvedenej spektrálnej analýzy vstupného signálu programom Audacity prakticky zodpovedá frekvencii s najvyššou hodnotou jeho amplitúdového spektra. Možno teda očakávať, že algoritmus dokáže identifikovať, aspoň v niektorých prípadoch, aj najvýraznejšie zložky neperiodického signálu.

# Záver

Cieľom tejto diplomovej práce bolo analyzovať fyziologicky inšpirované matematické prístupy k modelovaniu vnímania výšky tónu, pričom hlavným predmetom bola charakterizácia a klasifikácia existujúcich modelov a následne podrobná analýza a programová implementácia Meddis–Hewittovho časového modelu vnímania výšky tónu, ktorý zohral pri vývoji v tejto oblasti veľmi významnú úlohu.

V práci sme prepojili poznatky z matematiky, akustiky, digitálneho spracovania signálov, anatómie a neurofyziológie auditórneho systému človeka a ich aplikáciu na modelovanie procesu vnímania výšky tónu, ktoré tvorí základ vnímania hudby. Významným výsledkom práce je plne funkčná vlastná programová implementácia viacstupňového modelu v jazyku Python a jej otestovanie na synteticky generovaných a aj na prirodzených zvukových signáloch.

Práca má jasne definovanú a prehľadnú štruktúru pozostávajúcu zo 6 kapitol a prílohy.

V prvej kapitole sme poskytli úvod do vzťahu medzi hudbou a matematikou. Predstavili sme pojem zvukového javu a vysvetlili sme princíp jeho šírenia prostredníctvom troch základných fáz: zdroj zvukového signálu, prenosové médium a prijímateľ.

Druhá kapitola sa zaoberá fyzikálnou podstatou zvuku a jeho matematickému opisu. Venovali sme sa v nej prvej a druhej fáze zvukového javu, zdroju zvukového signálu a prenosovému médiu. Vysvetlili sme dôležitosť Fourierovej analýzy a autokorelácie vrámci ich použitia v modeloch vnímania výšky tónu. Záverečnú časť kapitoly sme venovali digitálnym filtrom, ktoré pri modelovaní sluchového vnímania zohrávajú dôležitú úlohu.

V tretej kapitole sme sa sústredili na vysvetlenie fyziologických základov vnímania hudby. Osobitne sme sa venovali periférnej aj centrálnej časti auditórneho systému, pričom ich poznanie je dôležité pri pochopení fungovania Meddis–Hewittovho modelu.

Štvrtá kapitola poskytuje prehľad existujúcich modelov vnímania výšky tónu. Porovnali sme v nej spektrálny prístup, založený na analýze frekvenčných zložiek, s časovým prístupom, ktorý využíva periodicitu signálu. Práve druhý prístup tvorí základ Meddis–Hewittovho modelu.

Piata kapitola je venovaná teoretickému popisu samotného Meddis–Hewittovho modelu, v ktorom je identifikácia výšky tónu založená na extrahovaní časových informácií z aktivít sluchového nervu.

V poslednej kapitole sme predstavili programovú implementáciu modelu v prostredí jazyka Python, prostredníctvom nášho programu MH. Postupne sme implementovali každú fázu modelu. Na popísanie efektov vonkajšieho a stredného ucha sme použili digitálny pásmový filter. Funkciu bazilárnej membrány sme modelovali pomocou súboru gamatónových filtrov. Simulovali sme zmeny množstva neurotransmitera v zásobníku, synaptickej štrbine a recyklačnom zásobníku. Pomocou upravenej pravdepodobnosti spiku o refraktérne efekty sme skonštruovali autokorelačné histogramy. Nakoniec sme spriemerovaním týchto autokorelačných histogramov dostali súhrnnú autokorelačnú funkci (SACF). Model sme otestovali na viacerých typoch zvukových signálov, ktorými boli: syntetický harmonický signál s tromi jednoduchými sínusovými zložkami, signál s chýbajúcou fundamentálnou frekvenciou, reálny tón trúbky a neharmonický signál tibetskej spievajúcej misky.

Prínosom tejto práce je najmä ucelená implementácia komplexného fyziologicky inšpirovaného Meddis–Hewittovho modelu v prostredí jazyka Python vo forme programu MH, sprevádzaná detailným vysvetlením jednotlivých krokov a vizualizáciou výstupov častí modelu. Za ďalšie prínosy pokladáme niektoré vlastné pohľady na problematiku modelovania vnímania výšky tónu prezentované v kapitole 4.

# Literatúra


[1] Benson D.J.: *Music: a Mathematical Offering*, Cambridge University Press, UK, 2006

[2] Glasberg B. R., Moore B. C. J.: *Derivation of auditory filter shapes from notched-noise data* , Hearing Research 47 (1990), 103–138, dostupné na internete: https://www.sciencedirect.com/science/article/abs/pii/037859559090170T

[3] Goldstein J. L.: *An optimum processor theory for the central formation of the pitch of complex tones*, Journal of the Acoustical Society of America 54(6) (1973), 1496–1516, dostupné na internete: https://pubs.aip.org/asa/jasa/article-abstract/54/6/1496/765165/An-optimum-processor-theory-for-the-central?redirectedFrom=fulltext

[4] Hartmann W. M.: *Principles of Musical Acoustics*, Springer Science+Business Media, New York, 2013

[5] Herzová E.: *Harmónia v hudbe a úloha matematiky v modeloch jej auditívneho vnímania človekom*, diplomová práca, FMFI UK, Bratislava, 2023

[6] Holdsworth J., et al.: *Implementing a GammaTone Filter Bank*, Research paper, Cambridge, 1988, dostupné na internete: https://www.pdn.cam.ac.uk/system/files/documents/SVOSAnnexC1988.pdf

[7] Licklider J.C.R.: *A duplex theory of pitch perception*, Experientia 7 (1951), 128–133

[8] Licklider J.C.R.: *Three auditory theories*, In Psychology: A Study of Science 1 (1959), 41–144, edited by Koch S., New York: McGraw-Hill

[9] Meddis R., Hewitt M. J.: *Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I. Pitch identification*, Journal of the Acoustical Society of America 89(6) (1991), 2866–2882, dostupné na internete: https://pubs.aip.org/asa/jasa/article-abstract/89/6/2866/629983/Virtual-pitch-and-phase-sensitivity-of-a-computer?redirectedFrom=fulltext

[10] Meddis R.: *Simulation of mechanical to neural transduction in the auditory receptor*, Journal of the Acoustical Society of America 79 (1986), 702–711, dostupné na internete: https://pubs.aip.org/asa/jasa/article-abstract/79/3/702/665520/Simulation-of-mechanical-to-neural-transduction-in?redirectedFrom=fulltext

[11] Meddis R., et al: *Implementation details of a computation model of the inner hair-cell auditory-nerve synapse*, Journal of the Acoustical Society of America 87 (1990), 1813–1816, dostupné na internete: https://pubs.aip.org/asa/jasa/article-abstract/87/4/1813/930809/Implementation-details-of-a-computation-model-of?redirectedFrom=fulltext

[12] Moore B. C. J., Glasberg B. R.: *Suggested formulae for calculating auditory-filter bandwidths and excitation patterns*, Journal of the Acoustical Society of America 74 (1983), 750–753, dostupné na internete: https://pubs.aip.org/asa/jasa/article-abstract/74/3/750/793872/Suggested-formulae-for-calculating-auditory-filter?redirectedFrom=fulltext

[13] Patterson R. D., Holdsworth J.: *A functional model of neural activity patterns and auditory images*, Advances in Speech, Hearing and Language Processing 3 (1996), 547–563, dostupné na internete: https://www.pdn.cam.ac.uk/system/files/documents/PH96.pdf

[14] Plack C. J., et al.: *Pitch: Neural Coding and Perception*, Springer Science+Business Media, New York, 2005

[15] Profant M., Kabátová Z.: *Audiológia*, Grada, Bratislava, 2012

[16] Roederer Juan G.: *The Physics and Psychophysics of Music*, Springer Science+Business Media, New York, 2008

[17] Shenoi B. A.: *Introduction to digital signal processing and filter design*, John Wiley & Sons, USA, 2006

[18] Schnupp J., et al.: *Auditory Neuroscience: Making sense of sound*, The MIT Press, Cambridge, 2010

[19] Smith III. J. O.: *Introduction to digital filters*, Center for Computer Research in Music and Acoustics, Stanford, 2007

[20] Smith W. S.: *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Pub., 1997

[21] Terhardt E.: *Pitch, consonance, and harmony*, Journal of the Acoustical Society of America 55(5) (1974), 1061–1069, dostupné na internete: https://pubs.aip.org/asa/jasa/article-abstract/55/5/1061/748589/Pitch-consonance-and-harmony?redirectedFrom=PDF

[22] Yost W. A.: *Pitch perception*, Attention, Perception, & Psychophysics 71 (2009), 1701–1715

# Príloha A

## Program MH

Program MH bol vytvorený v Pythone 3.13.2. Nasledujúce externé knižnice sú potrebné k spusteniu programu: numpy, matplotlib, scipy, librosa, tqdm. Program MH je dostupný na stiahnutie na odkaze: https://github.com/ZuzkaTom/DP.

```
print("Program MH: Implementacia Meddis-Hewittovho modelu")

import numpy as np
import matplotlib.pyplot as plt
import librosa
from matplotlib.widgets import Button
from tqdm import tqdm
from scipy.signal import filtfilt

# Nacitanie suboru v pripade trubky a tibetskej spievajucej misky
#test_signal, sr = librosa.load("trubka.wav", sr=20000)
#test_signal, sr = librosa.load("bowls.wav", sr=20000)

# Orezanie signalu na 1 sekundu
#test_signal = test_signal[:sr]

# Parametre signalu
sr = 20000  # vzorkovacia frekvencia (20 kHz)
duration = 1.0  # trvanie signalu v sekundach
vzorky = 500 # pocet vzoriek pre vykreslovanie
dt = 1 / sr  #  t : vzorkovaci krok v sekundach
t = np.arange(0, duration, dt)

# Generovanie umeleho signalu
freq = 200
test_signal = np.sin(2 * np.pi * freq * t) + np.sin(2 * np.pi * 2 * freq
    * t) + np.sin(2 * np.pi * 3 * freq * t)
# Generovanie signalu, pripad chybajucej fundamentalnej frekvencie
#test_signal = np.sin(2 * np.pi * 600 * t) + np.sin(2 * np.pi * 1500 * t
   )+np.sin(2 * np.pi * 2400 * t)
```

```
29
30
31  # Digitalny pasmovy filter s koeficientmi
32  def bandpass_filter(x):
33      b = [0.8878, 0, -0.8878]
34      a = [1, 0.2243, -0.7757]
35      return filtfilt(b, a, x)
36
37  # Aplikujeme digitalny pasmovy filter na testovaci signal
38  filtered_signal = bandpass_filter(test_signal)
39  print("Efekty vonkajsieho a stredneho ucha : bandpass filter")
40
41  # Funkcia na vypocet ERB
42  def erb_scale(frequency):
43      return 24.7 * (4.37e-3 * frequency + 1)
44
45  # Funkcia na prevod frekvencie (v Hz) do ERB-rate skaly
46  def hz_to_erb_rate(frequency):
47      return 21.4 * np.log10(4.37e-3 * frequency + 1)
48
49  # Inverzna funkcia: prevod z ERB-rate spat na frekvenciu v Hz
50  def erb_rate_to_hz(erb_rate):
51      return (10**(erb_rate / 21.4) - 1) / 4.37e-3
52
53  # Vypocet centralnych frekvencii pre filterbank
54  def erb_center_frequencies(start_freq=80, end_freq=8000, erb_step=0.25):
55      erb_start = hz_to_erb_rate(start_freq)
56      erb_end = hz_to_erb_rate(end_freq)
57      erb_points = np.arange(erb_start, erb_end + erb_step, erb_step)
58      return erb_rate_to_hz(erb_points)
59
60  # Implementacia gammatonoveho filterbanku pomocou IIR filtracie
61  def gammatone_filterbank_iir(signal, center_freqs, order=4):
62
63      filtered_signals = []
64
65      for f in tqdm(center_freqs, desc="Filtracia bazilarnou membranou :
    gammatone filterbank"):
```

```
        z = np.zeros_like(signal, dtype=np.complex128)

        for k in range(len(signal)):
            z[k] = signal[k] * np.exp(-1j * 2 * np.pi * f * t[k])

        alpha = 1 - np.exp(-2 * np.pi * erb_scale(f) * dt)
        w = np.zeros_like(z, dtype=np.complex128)

        for k in range(1, len(z)):
            w[k] = w[k-1] + alpha * (z[k-1] - w[k-1])

        for _ in range(order - 1):
            w2 = np.zeros_like(w, dtype=np.complex128)

            for k in range(1, len(w)):
                w2[k] = w2[k-1] + alpha * (w[k-1] - w2[k-1])
            w = w2

        y = np.real(w * np.exp(1j * 2 * np.pi * f * t))

        filtered_signals.append(y)

    return np.array(filtered_signals)

# Vypocet centralnych frekvencii filtrov
center_freqs = erb_center_frequencies(start_freq=80, end_freq=8000,
    erb_step=0.25)
print(f"Pocet filtrov: {len(center_freqs)}")

# Aplikacia IIR gammatonoveho filterbanku - kazdy filter simuluje jeden
    kanal bazilarnej membrany
gammatone_signals = gammatone_filterbank_iir(filtered_signal,
    center_freqs)

# Mechanicko-nervova transdukcia
# Parametre modelu
M = 1.0  # Maximalne mnozstvo volneho neurotransmitera
A = 5  # Konstanta pre permeabilitu
```

```
B = 300  # Konstanta pre permeabilitu
g = 2000  # Konstanta pre permeabilitu
y = 5.05  # Rychlost doplnania neurotransmitera
L = 2500  # Rychlost strat zo synaptickej strbiny
r = 6580  # Rychlost navratu neurotransmitera do recyklacneho zasobnika
x = 66.31  # Rychlost uvolnenia neurotransmitera z recyklacneho
    zasobnika
h = 50000  # Konstanta pre pravdepodobnost spiku

# Parametre pre vypocet amplitudy stimulu
dB = 80  # Decibelova uroven signalu
amp = 10 ** (dB / 20 - 1.35)  # Amplituda signalu
endsilence = 0.01  # Trvanie pociatocneho ticha v sekundach

#Vahova funkcia pre refrakterne efekty
def W_function(t):
    return np.where((t > 0) & (t < 0.001), 1.0, 0.0)

# Vypocet pociatocnych podmienok pre q, c, w
kt_spontaneous = (g * A) / (A + B)  # Permeabilita pri spontannej
    aktivite
c_spontaneous = (M * y * kt_spontaneous) / (L * kt_spontaneous + y * (L
    + r))  # Mnozstvo neurotransmitera v synaptickej strbine
q_spontaneous = (c_spontaneous * (L + r)) / kt_spontaneous  # Mnozstvo
    volneho neurotransmitera
w_spontaneous = (c_spontaneous * r) / x  # Mnozstvo neurotransmitera v
    recyklacnom zasobniku

# Pole na okladanie vysledkov
kt_values = []
q_values = []
cleft_values = []
p_spike_values = []

for sig_idx, sig in enumerate(tqdm(gammatone_signals, desc="Mechanicko-
    nervova transdukcia a refrakterne efekty: simulacia Meddis-Hewitt")):
    # Resetovanie hodnot pre kazdy kanal
    spike_history = np.zeros_like(t)
```

```
    q = q_spontaneous  # Pociatocne mnozstvo volneho neurotransmitera
    c = c_spontaneous  # Pociatocne mnozstvo neurotransmitera v synaptickej strbine
    w = w_spontaneous  # Pociatocne mnozstvo neurotransmitera v recyklacnom zasobniku

    # Pole na ukladanie vysledkov pre kazdy kanal
    kt_values_channel = []
    q_values_channel = []
    cleft_values_channel = []
    p_spike_values_channel = []

    # Pre kazdy casovy krok simulacie
    for i, t_sim in enumerate(t):
        # Vyberieme jednu vzorku zo signalu pre cas t_sim
        stimulus = sig[int(t_sim * sr)]

        # Vypocet amplitudy stimulu
        if t_sim > endsilence:
            stimulus = amp * stimulus
        else:
            stimulus = 0

        # Vypocet permeability membrany
        if stimulus + A > 0:
            kt = (g * dt * (stimulus + A)) / (stimulus + A + B)
        else:
            kt = 0

        # Vypocet zmien v mnozstvach neurotransmitera
        if M >= q:
            replenish = y * dt * (M - q)  # Doplnanie volneho neurotransmitera
        else:
            replenish = 0
        eject = kt * q  # Uvolnenie neurotransmitera do synaptickej strbiny
        loss = L * dt * c  # Strata neurotransmitera zo synaptickej
```

```
    strbiny
167         reuptake = r * dt * c  # Navrat neurotransmitera do recyklacneho
     priestoru
168         reprocess = x * dt * w  # Presun neurotransmitera z recyklacneho
     priestoru do zasobnika
169 
170         # Aktualizacia
171         q = q + replenish - eject + reprocess
172         c = c + eject - loss - reuptake
173         w = w + reuptake - reprocess
174 
175         # Vypocet pravdepodobnosti spiku
176         p_spike = h * c * dt
177 
178         # V po et vplyvu predchadzajucich spikov pomocou vahovej
    funkcie
179         ref_effect = np.sum(spike_history[:i] * W_function(t_sim - t[:i
    ]))
180         ref_effect = np.clip(ref_effect, 0, 1)
181 
182         #  prava  pravdepodobnosti spiku na zaklade refrakterneho efektu
183         p_spike_adjusted = p_spike * (1 - ref_effect)
184 
185         # Rozhodnutie o spiku na zaklade upravenej pravdepodobnosti
186         if np.random.random() < p_spike_adjusted:
187             spike_history[i] = 1
188 
189         # Ulozenie vysledkov do zoznamov pre aktualny kanal
190         kt_values_channel.append(kt)  # Ukladame hodnotu kt
191         q_values_channel.append(q)  # Ukladame hodnotu q
192         cleft_values_channel.append(c)  # Ukladame koncentraciu
    neurotransmitera v synaptickej strbine
193         p_spike_values_channel.append(p_spike_adjusted)  # Ulozena
    upravena pravdepodobnost spiku
194 
195     # Ulozenie vysledkov pre kazdy kanal (pridanie do globalnych
    zoznamov)
196     kt_values.append(kt_values_channel)
```

```
197     q_values.append(q_values_channel)
198     cleft_values.append(cleft_values_channel)
199     p_spike_values.append(p_spike_values_channel)
200 
201 # Vypocet autokorelacnych histogramov
202 # Casova konstanta pouzita vo vahovacej funkcii e^(-T/tau)
203 time_constant = 0.0025
204 
205 # Dlzka, pocas ktorej sa beru do uvahy predchadzajuce spiky
206 acf_window = 0.0075
207 
208 # Vytvorenie vsetkych posunov  t , pre ktore budeme pocitat
    autokorelaciu
209 delta_t_values = np.arange(0.00005, 0.01667, dt)
210 
211 # Pocet  t  hodnot = velkost histogramu pre kazdy kanal
212 num_bins = len(delta_t_values)
213 
214 # Zoznam, do ktoreho budeme ukladat histogram pre kazdy kanal
215 histograms = []
216 
217 # Hlavny cyklus cez kanaly
218 for k, p_t in enumerate(tqdm(p_spike_values, desc="Distribucia casovych
    intervalov : vypocet histogramov")):
219     histogram = np.zeros(num_bins)
220     p_t = np.array(p_t)
221 
222     # Cyklus cez vsetky oneskorenia  t
223     for i, delta in enumerate(tqdm(delta_t_values, desc=f" t  pre kanal
    {k+1}", leave=False)):
224         # Vyjadrime  t  ako pocet vzoriek
225         delay_samples = int(delta / dt)
226         # Urcime, kolko vzoriek T mame v intervale 0-7.5 ms
227         max_T = int(acf_window / dt)
228 
229         # Cyklus cez vsetky casy T
230         for T in range(1, max_T):
231             # Berieme len tie t, pri ktorych vieme bezpecne siahnut do
```

```
    minulosti az o T+deltat, teda:
232             t_idx = np.arange(delay_samples + T, len(p_t))
233 
234             # Vypocitame indexy pre p(t-T)
235             t_minus_T = t_idx - T
236 
237             # Vypocitame indexy pre p(t-T-deltat)
238             t_minus_T_minus_delta = t_idx - T - delay_samples
239 
240             # Zabezpecime, ze indexy neprecitaju data mimo pola
241             valid = (t_minus_T >= 0) & (t_minus_T_minus_delta >= 0)
242 
243             if np.any(valid):  # Pokracujeme len vtedy, ked mame aspon
    jeden platny index
244                 # Extrahujeme hodnoty pravdepodobnosti spiku v case t-T
    a t-T-deltat
245                 pit_T = p_t[t_minus_T[valid]]               # p(t-T)
246                 pit_T_dt = p_t[t_minus_T_minus_delta[valid]]  # p(t-T-
    deltat)
247 
248                 # Vypocitame vahu podla vzdialenosti T - Lickliderov
    exponencialny utlm
249                 weights = np.exp(-T * dt / time_constant)  # e^(-T/tau)
250 
251                 # Vypocitame prispevok do histogramu pre dany deltat:
252                 histogram[i] += np.sum(pit_T * pit_T_dt * weights)
253 
254     # Po dokonceni cyklu cez deltat ulozime histogram pre aktualny kanal
255     histograms.append(histogram)
256 
257 # Vypocet SACF ako priemer histogramov
258 time_intervals = delta_t_values
259 sacf = np.mean(histograms, axis=0)
260 print("Scitanie napriec kanalmi : SACF")
261 
262 # Casove hranice pre vypocet fundamentalnej frekvencie
263 min_lag = 0.001       # 1 ms = 1000 Hz
264 max_lag = 0.0125      # 12.5 ms = 80 Hz
```

```
265
266 # Vyfiltrujeme hodnoty SACF v danom intervale oneskoreni
267 mask = (delta_t_values >= min_lag) & (delta_t_values <= max_lag)
268 sacf_in_range = sacf[mask]
269 delta_in_range = delta_t_values[mask]
270
271 # Najdeme index najvyssieho vrcholu v tomto intervale
272 peak_idx = np.argmax(sacf_in_range)
273 dominant_delta = delta_in_range[peak_idx]
274 f0_estimated = 1 / dominant_delta
275
276 print(f"Extrakcia vysky tonu : odhadnuta fundamentalna frekvencia ( f  )
    : {f0_estimated:.2f} Hz")
277
278 # Vykreslovanie
279 # Zoznam grafov, ktore chceme prepinat
280 graphs = [
281     (test_signal, "Vstupny signal"),
282     (filtered_signal, "Filtrovany signal"),
283     (gammatone_signals, "Vystup gamatonovej banky filtrov"),
284     (cleft_values, "Mnozstvo neurotransmitera c(t) v synaptickej strbine
    "),
285     (q_values, "Mnozstvo neurotransmitera q(t) v zasobniku volneho
    neurotransmitera"),
286     (p_spike_values, "Pravdepodobnost spiku p_spike(t)"),
287     (histograms, "Autokorelacne histogramy"),
288     (sacf, "Suhrnny autokorelacny histogram (SACF)")
289 ]
290
291 # Index aktualneho kanala pre histogramy
292 current_histogram_channel = 0
293 # Index aktualne zobrazovaneho grafu
294 current_graph = 0
295
296 # Vytvorenie hlavneho grafu
297 fig, ax = plt.subplots(figsize=(10, 4))
298 # Posunutie plochy grafu vyssie, aby bolo miesto pre tlacidla
299 plt.subplots_adjust(bottom=0.2)
```

```
300
301 # Funkcia na aktualizaciu grafu
302 def update_graph(idx):
303     # Budeme pracovat s globalnou premennou current_graph, a nie
    vytvarat novu lokalnu premennu
304     global current_graph
305     # Zabezpeci cyklicke prepinanie grafov (% modulo)
306     current_graph = idx % len(graphs)
307     ax.clear()
308     ax.set_title(graphs[current_graph][1])
309     ax.set_xlabel("cas (s)")
310     ax.set_ylabel("Amplituda")
311
312     # Ak zobrazujeme vystup gammatone bankz filtrov
313     if current_graph == 2:
314         # Vykreslujeme len kazdy druhy kanal kvoli prehladnosti
315         selected_indices = np.arange(0, len(center_freqs), 2)
316         # Vertikalny rozostup
317         spacing = 0.1
318         # Amplitudove skalovanie
319         scale_factor = 2.0  # 5 pre trubku, 20 pre tibetske misky
320         # Rovnomerne vizualne rozostupy
321         offsets = np.arange(len(selected_indices)) * spacing
322         min_fill = offsets[0] - spacing
323
324         for idx, i in enumerate(selected_indices):
325             sig = graphs[current_graph][0][i]
326             scaled_sig = sig * scale_factor
327             y_vals = scaled_sig[-vzorky:] + offsets[idx]
328             ax.fill_between(t[-vzorky:], y_vals, min_fill, linewidth
    =0.8, edgecolor='black', zorder=-i)
329
330         # Nastavenie osi
331         ax.set_yticks(offsets)
332         yticks_labels = [f"{center_freqs[i]:.0f} Hz" if idx % 3 == 0
    else "" for idx, i in enumerate(selected_indices)]
333         ax.set_yticklabels(yticks_labels, fontsize=14)
334         ax.set_ylabel("Centralne frekvencie (Hz)")
```

```
335         ax.set_xlabel("cas (s)")
336         ax.set_title("Vystupy suboru gamatonovych filtrov")
337 
338     # c(t) - mnozstvo neurotransmitera v synaptickej strbine
339     elif current_graph in [3]:
340         data = graphs[current_graph][0]
341         title = graphs[current_graph][1]
342 
343         # Najde extremov napriec kanalmi
344         cleft_min = min(np.min(cleft[-vzorky:]) for cleft in
    cleft_values)
345         cleft_max = max(np.max(cleft[-vzorky:]) for cleft in
    cleft_values)
346 
347         values = data[current_histogram_channel][-vzorky:]
348         ax.plot(t[-vzorky:], values)
349 
350         ax.set_title(f"{title}     kanal {current_histogram_channel + 1}
     ({center_freqs[current_histogram_channel]:.1f} Hz)")
351         ax.set_xlabel("cas (s)")
352         ax.set_ylabel("c(t)")
353         # Nastavime mierku pre kazdy kanal
354         ax.set_ylim(cleft_min-0.001, cleft_max+0.005)
355         ax.grid(True)
356 
357     # q(t) - mnozstvo neurotransmitera v zasobniku volneho
    neurotransmitera
358     elif current_graph in [4]:
359         data = graphs[current_graph][0]
360         title = graphs[current_graph][1]
361 
362         # Najde extremov napriec kanalmi
363         q_min = min(np.min(q[-vzorky:]) for q in q_values)
364         q_max = max(np.max(q[-vzorky:]) for q in q_values)
365 
366         values = data[current_histogram_channel][-vzorky:]
367         ax.plot(t[-vzorky:], values)
368 
```

```
369         ax.set_title(f"{title}     kanal {current_histogram_channel + 1}
    ({center_freqs[current_histogram_channel]:.1f} Hz)")
370         ax.set_xlabel("cas (s)")
371         ax.set_ylabel("q(t)")
372         # Nastavime mierku pre kazdy kanal
373         ax.set_ylim(q_min-0.001, q_max+0.005)
374         ax.grid(True)
375
376     # p_spike(t)
377     elif current_graph in [5]:
378         data = graphs[current_graph][0]
379         title = graphs[current_graph][1]
380
381         # Najde extremov napriec kanalmi
382         p_min = min(np.min(p[-vzorky:]) for p in p_spike_values)
383         p_max = max(np.max(p[-vzorky:]) for p in p_spike_values)
384
385         values = data[current_histogram_channel][-vzorky:]
386         ax.plot(t[-vzorky:], values)
387
388         ax.set_title(f"{title}     kanal {current_histogram_channel + 1}
    ({center_freqs[current_histogram_channel]:.1f} Hz)")
389         ax.set_xlabel("cas (s)")
390         ax.set_ylabel("p_spike(t)")
391         # Fixneme mierku pre kazdy kanal
392         ax.set_ylim(p_min-0.001, p_max+0.005)
393         ax.grid(True)
394
395     # Autokorelacny histogram
396     elif current_graph == 6:
397         # Vypocita najvacsiu hodnotu zo vsetkych histogramov
398         hist_max = max(np.max(h) for h in histograms)
399
400         ax.bar(delta_t_values, histograms[current_histogram_channel],
    width=dt, color='teal', alpha=0.7)
401         ax.set_title(f"Autokorelacny histogram     kanal {
    current_histogram_channel + 1} ({center_freqs[
    current_histogram_channel]:.1f} Hz)")
```

```
402             ax.set_xlabel("casovy posun (s)")
403             ax.set_ylabel("Vazeny sucet korelacnych hodnot")
404             ax.set_ylim(0, hist_max * 1.05)  # +5 % rezerva
405             ax.grid(True)
406 
407         # Ak zobrazujeme suhrnny autokorelacny histogram (SACF)
408         elif current_graph == 7:
409             ax.bar(delta_t_values, graphs[current_graph][0], width=dt, color
    ='purple', label="SACF (histogram)")
410             ax.axvline(x=dominant_delta, color='orange', linestyle='--',
    label=f"Odhad  f    = {f0_estimated:.1f} Hz")
411             ax.set_xlabel("casovy posun (s)")
412             ax.set_ylabel("Vazeny sucet korelacnych hodnot")
413             ax.set_title("Suhrnny autokorelacny histogram (SACF) s
    oznacenymi vrcholmi")
414             ax.legend()
415             ax.grid(True)
416 
417         # Pre povodny a filtrovany signal
418         else:
419             ax.plot(t[-vzorky:], graphs[current_graph][0][-vzorky:])
420 
421         # Obnovenie zobrazenia grafu
422         fig.canvas.draw()
423 
424 # Funkcia, ktora meni index aktualne zobrazovaneho kanala histogramu
425 def change_hist_channel(step):
426         global current_histogram_channel
427         current_histogram_channel = (current_histogram_channel + step) % len
    (histograms)
428         # Znovu vykresli graf s novym kanalom histogramu
429         update_graph(current_graph)
430 
431 # Prve zobrazenie grafu
432 update_graph(current_graph)
433 
434 # Tlacidlo "Dalsi"
435 ax_next = plt.axes([0.8, 0.05, 0.1, 0.075])
```

```
btn_next = Button(ax_next, 'Dalsi')
# Funkcia na prepnutie na dalsi graf
btn_next.on_clicked(lambda event: update_graph(current_graph + 1))

# Tlacidlo "Predchadzajuci"
ax_prev = plt.axes([0.65, 0.05, 0.1, 0.075])
btn_prev = Button(ax_prev, 'Predch.')
# Funkcia na prepnutie na predchadzajuci graf
btn_prev.on_clicked(lambda event: update_graph(current_graph - 1))

# Tlacidlo "Dalsi kanal pre histogram"
ax_next_hist = plt.axes([0.45, 0.05, 0.1, 0.075])
btn_next_hist = Button(ax_next_hist, 'Kanal +')
btn_next_hist.on_clicked(lambda event: change_hist_channel(1))

# Tlacidlo "Predchadzajuci kanal pre histogram"
ax_prev_hist = plt.axes([0.3, 0.05, 0.1, 0.075])
btn_prev_hist = Button(ax_prev_hist, 'Kanal -')
btn_prev_hist.on_clicked(lambda event: change_hist_channel(-1))

# Zobrazenie grafov
plt.show()
```