

UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Luka Štepančič

# Optimizacija parametrov v modelu sintetizacije zvoka

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM RAČUNALNIŠTVO IN  
INFORMATIKA

MENTOR: doc. dr. Mira Trebar

Ljubljana, 2013



Rezultati diplomskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

*Besedilo je oblikovano z urejevalnikom besedil L<sup>A</sup>T<sub>E</sub>X.*





Št. naloge: 01919/2013

Datum: 03.04.2013

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **LUKA STEPANČIČ**

Naslov: **OPTIMIZACIJA PARAMETROV V MODELU SINTETIZACIJE ZVOKA  
OPTIMIZATION OF SOUND SYNTHESIS MODEL PARAMETERS**

Vrsta naloge: Diplomsko delo univerzitetnega študija

Tematika naloge:

V diplomski nalogi naj kandidat predstavi aproksimativen fizikalni model sintetizacije zvoka akustičnih glasbil, kjer je informacija podana v obliki zvočnih vzorcev. Z minimalnim številom parametrov jo je potrebno preslikati tako, da ohranimo perceptivne lastnosti čim bolj podobne sistemu, ki signale generira. V postopku kalibracije je potrebno poiskati takšne parametre modela DSP, ki minimizirajo perceptivno razdaljo med lastnostmi ciljnega zvoka in lastnostmi iskanega zvoka. Metrika oz. razdalja značilik določa obliko kriterijske funkcije. Za iskanje takega nabora parametrov, ki minimizira kriterijsko funkcijo naj bo uporabljena večkriterijska optimizacijska metoda z diferencialno evolucijo.

Mentor:

doc. dr. Mira Trebar

Dekan:

prof. dr. Nikolaj Zimic





## IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Luka Stepančič, z vpisno številko **63030324**, sem avtor diplomskega dela z naslovom:

*Optimizacija parametrov v modelu sintetizacije zvoka*

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom doc. dr. Mire Trebar,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, dne 5. julij 2013

Podpis avtorja:



Zahvaljujem se mentorici doc.dr. Miri Trebar za pomoč in praktične nasvete, družini za vso podporo in vsem ostalim, ki so me spodbujali v času študija - in ki (še) niso obupali nad mano.



# Kazalo

Povzetek

Abstract

<b>1</b>	<b>Uvod</b>	<b>1</b>
<b>2</b>	<b>Osnove analize in sinteze zvočnih signalov</b>	<b>7</b>
2.1	Psihoakustične lastnosti zvoka . . . . .	7
2.1.1	Spektralno maskiranje . . . . .	8
2.1.2	Časovno maskiranje . . . . .	9
2.1.3	Zaznavanje osnovne frekvence tona . . . . .	9
2.2	Zvočne značilke . . . . .	10
2.3	Dekompozicija signalov . . . . .	12
2.3.1	STFT in spektrogram . . . . .	12
2.3.2	Dekompozicija signalov v funkcionalne sklope . . . . .	13
2.3.3	Inverzno modeliranje . . . . .	14
2.3.4	Sinusoidno modeliranje . . . . .	15
2.3.5	Linearno prediktivno kodiranje . . . . .	16
2.4	Tehnike sintetizacije zvoka . . . . .	18
2.4.1	Aditivna sinteza . . . . .	21
2.4.2	Subtraktivna sinteza . . . . .	22
2.4.3	Frekvenčna modulacija . . . . .	22
2.4.4	Valovne tabele . . . . .	23
2.4.5	Karplus-Strong algoritem . . . . .	23
2.4.6	Digitalno valovno vodilo ( <i>Digital Waveguide</i> ) . . . . .	24

2.4.7	Primerjava tehnik . . . . .	25
<b>3</b>	<b>Modeliranje z digitalnimi valovnimi vodili</b>	<b>27</b>
3.1	Modeliranje resonatorjev . . . . .	28
3.1.1	Dušilni filter . . . . .	30
3.1.2	Filter za fino uglasovanje ali ' <i>fine-tuning</i> ' . . . . .	30
3.1.3	Disperzijski filter in disharmoničnost . . . . .	31
3.1.4	Priklopni oz. ( <i>angl. coupling</i> ) filter . . . . .	33
3.2	Telesni filter . . . . .	34
3.3	Modeliranje vzbujanja . . . . .	35
3.3.1	Vzbujanje z vzorčenjem . . . . .	36
3.3.2	Subtraktivni model . . . . .	36
3.4	Analično določanje vrednosti parametrov . . . . .	38
3.4.1	Določitev koeficientov disperzije . . . . .	38
3.4.2	Določitev osnovne periode . . . . .	40
3.4.3	Določitev parametrov dušenja . . . . .	40
3.4.4	Določitev parametrov vzbujanja . . . . .	41
3.4.5	Določitev preostalih vrednosti parametrov . . . . .	41
3.5	Predstavitev rezultatov . . . . .	42
<b>4</b>	<b>Kalibracija modela</b>	<b>45</b>
4.1	Kriterijske funkcije . . . . .	46
4.2	Optimizacija . . . . .	49
4.2.1	Pareto optimalnost . . . . .	50
4.2.2	Učinkovitost iskanja . . . . .	51
4.2.3	Diferencialna evolucija . . . . .	52
4.3	Določitev območja iskanja . . . . .	54
4.4	Rezultati iskanja . . . . .	56
4.5	Možne izboljšave algoritma optimizacije . . . . .	58
<b>5</b>	<b>Zaključek</b>	<b>61</b>

# Povzetek

V diplomskem delu je predstavljen prototip modela sintetizacije zvoka, kjer je bil glavni cilj modeliranje in realistična reprodukcija zvoka akustičnih glasbil. Uporabljena je bila metoda fizikalnega modeliranja z digitalnimi valovnimi vodili na primeru klavirja. Opisani so postopek načrtovanja modela, analitični način določanja optimalnih parametrov in avtomatska kalibracija le-teh. V ta namen je bil predstavljen postopek kalibracije modelov sintetizacije zvoka z večkriterijsko minimizacijo perceptivne razdalje med izvornim zvokom in sintetiziranim zvokom. Perceptivna razdalja je določena kot skupek štirih značilnih in metrik, ki dobro opisujejo človeško zaznavanje podobnosti zvokov, obenem pa so hitro izračunljive. Za avtomatsko kalibracijo je bila uporabljena metoda večkriterijske optimizacije z diferencialno evolucijo. Na koncu so predstavljene tudi možnosti razširitve oz. izboljšav postopka kalibracije.

## *KLJUČNE BESEDE:*

evolucijski algoritmi, večkriterijska optimizacija, fizikalno modeliranje, sintetizacija zvoka, digitalno procesiranje signalov



# Abstract

In this document we present a sound synthesis model prototype where a realistic sound reproduction of acoustic musical instruments is the final goal. A physically informed digital waveguide synthesis method was chosen for modelling the instrument properties. The method was applied for the case of the piano. A procedure was presented for design, analytical derivation and automatic calibration of the synthesis model parameters. A procedure for calibration with multiobjective minimization of perceptual differences between the sources and synthesized sounds was proposed. Four perceptually based sound feature distances with fast computation capabilities were chosen for optimization. For parameter search, a differential evolution scheme for multiobjective optimization was used. Finally, extension possibilities of the calibration method and future directives are shown.

## *KEYWORDS:*

evolutionary algorithms, multiobjective optimization, physical modelling, audio synthesis, digital signal processing



### UPORABLJENE KRATICE:

ADSR	-	<i>Attack-Decay-Sustain-Release envelope</i>
AGC	-	<i>Adaptive gain control</i> , avtomatska kontrola jakosti
AM	-	<i>Amplitude modulation</i> , amplitudna modulacija
CWS	-	<i>Commutated waveguide synthesis</i>
DE	-	Diferencialna evolucija
DFT	-	Diskretna Fourierjeva transformacija
DSD	-	<i>Deterministic-stochastic decomposition</i> , deterministično-stohastična dekompozicija
DSP	-	<i>Digital signal processing</i> , digitalno procesiranje signalov
DWG	-	<i>Digital waveguide synthesis</i> , sinteza z digitalnimi valovnimi vodili
EKS	-	<i>Extended Karplus-Strong algorithm</i> , razširjeni Karplus-Strong algoritem
ERB	-	<i>Equivalent rectangular bandwidth</i> , ekvivalentna pravokotna pasovna širina
ES	-	Evolucijske strategije
FFT	-	<i>Fast Fourier transform</i> , hitra Fourierjeva transformacija
FIR	-	<i>Finite impulse response</i> , končni odziv na enotin impulz
FM	-	<i>Frequency modulation</i> , frekvenčna modulacija
GA	-	Genetski algoritem
IIR	-	<i>Infinite impulse response</i> , neskončni odziv na enotin impulz
KS	-	Karplus-Strong algoritem
LFO	-	<i>Low frequency oscillator</i> , nizko-frekvenčni oscilator
LPC	-	<i>Linear predictive coding</i> , linearno prediktivno kodiranje
LP	-	linearna predikcija
LTI	-	<i>Linear-time invariant</i> , linearno časovno invarianten
SMS	-	<i>Sinusoidal modelling synthesis</i> , sinteza s sinusoidnim modeliranjem
SNR	-	<i>Signal to noise ratio</i> , razmerje signal-šum
STFT	-	<i>Short-time Fourier transform</i> , kratko-časovna Fourierjeva transformacija
WTS	-	<i>Wavetable synthesis</i> , sinteza z zvočnimi tabelami



# Poglavje 1

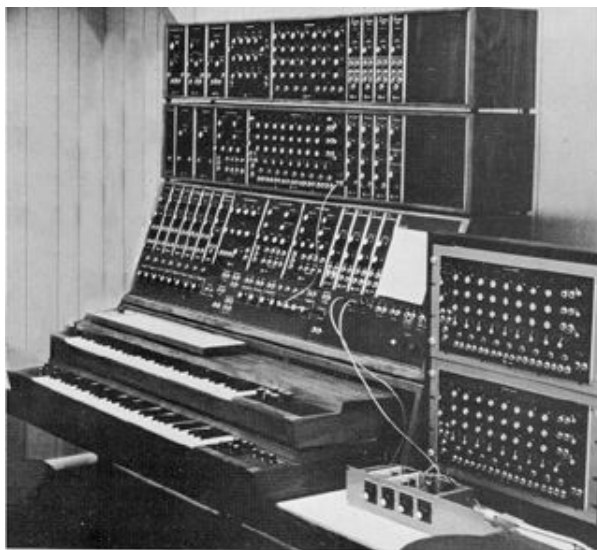
## Uvod

Že skoraj stoletje je minilo, odkar smo ljudje začeli ustvarjati umetna elektronska glasbila, ki poskušajo posnemati ali obogatiti zvoke akustičnih glasbil. Začelo se je z elektronskimi in elektro-akustičnimi napravami, sposobnimi generirati zvoke, ki so v marsičem spominjali na naravna glasbila. Kasneje so se začeli uveljavljati (analogni) elektronski sintetizatorji. Ti so dobili ime 'sintetizatorji' po težnji poustvarjanja zvoka akustičnih glasbil - sintetičnega zvoka. S pojavom digitalnih vezij in mikrokrmilniških sistemov, so analogne sintetizatorje kmalu nadomestili digitalni elektronski sistemi sintetizacije, kasneje pa osebni računalniki, saj so le-ti postali čedalje bolj dostopni.

Izdelovalci elektronskih glasbil si prizadevajo izpopolniti elektronske sintetizatorje zvoka do te mere, da bi bila razlika med naravnimi in sintetiziranimi neopazna. V tem tudi uspevajo, saj so dandanes laikom razlike med akustičnimi in sintetičnimi zvoki praktično neopazne. Kljub temu, se ljudje odločajo za nakup akustičnih glasbil zaradi ekspresivnosti, ki jih le-ta ponujajo.

Akustična glasbila (z izjemo družine glasbil s tipkami) so sposobna ustvarjati raznolik zvok že z načinom dovajanja energije v glasbilo, npr. način pihanja v jeziček saksofona ali pozicija ubiranja pri kitari. Vendar so proizvajalci in izdelovalci sintetizatorjev preveliko težnjo posvetili popolni reprodukciji stacionarnega zvoka, pozabili pa so na ekspresivnost. Razlog za to je bila velika težnja po diskretizaciji prostora parametrov upravljanja, bodisi zaradi popularnosti inštrumentov s tipkami ali zaradi tehničnih razlogov. Tako analogni kot digitalni sintetizatorji so imeli možnost dinamičnega upravljanja zveznih parametrov s pomočjo vrtlji-

vih ali drsnih potenciometrov. Ti dinamično spremenljivi parametri so povečali ekspresivnost, vendar so se relacije med parametri in lastnostmi zvoka precej razlikovale od akustičnih. Poleg tega so bile zveze med lastnostmi in parametri netrivialne, kar je otežilo umerjanje, predvsem pri kompleksnejših analognih elektronskih sintetizatorjih. Nekatere zgodnje različice sintetizatorjev so imele celo preveč le-teh, npr. znameniti *Moog* (slika 1.1). Najbolj obetaven pristop,



Slika 1.1: Kompleksni *Moog* sintetizator iz leta 1970

ki bi lahko združil tako možnost popolne zvočne reprodukcije, kot tudi ekspresivnost izvajanja, je zagotovo fizikalno modeliranje. Čeprav so fizikalni modeli akustičnih glasbil obstajali že nekaj časa v obliki diferencialnih enačb, niso nikoli zaživel v svetu sintetizacije in zaradi računske neučinkovitosti niso bili sposobni računanja in simulacije v realnem času.

Pravi prelom v svetu sintetizacije je nastal v 80-letih prejšnjega stoletja, ko sta leta 1983 K. Karplus and A. Strong razvila enostavno metodo za simulacije zvoka strun in tolkal, t.i. Karplus-Strong algoritem (KS). J. O. Smith je s sodelavci ta algoritem razširil in posplošil v tehnologijo, ki jo je poimenoval *digitalna valovna vodila - Digital Waveguide Synthesis (DWG)* in jo leta 1983 tudi patentiral [23].

Japonska družba Yamaha je le nekaj let kasneje tehnologijo licencirala, in leta 1994 je nastal znameniti *Yamaha VL1 'Virtual Acoustic' synthesizer*. VL1 je zaznamoval vstop fizikalnega modeliranja v domeno komercialnih sintetizatorjev

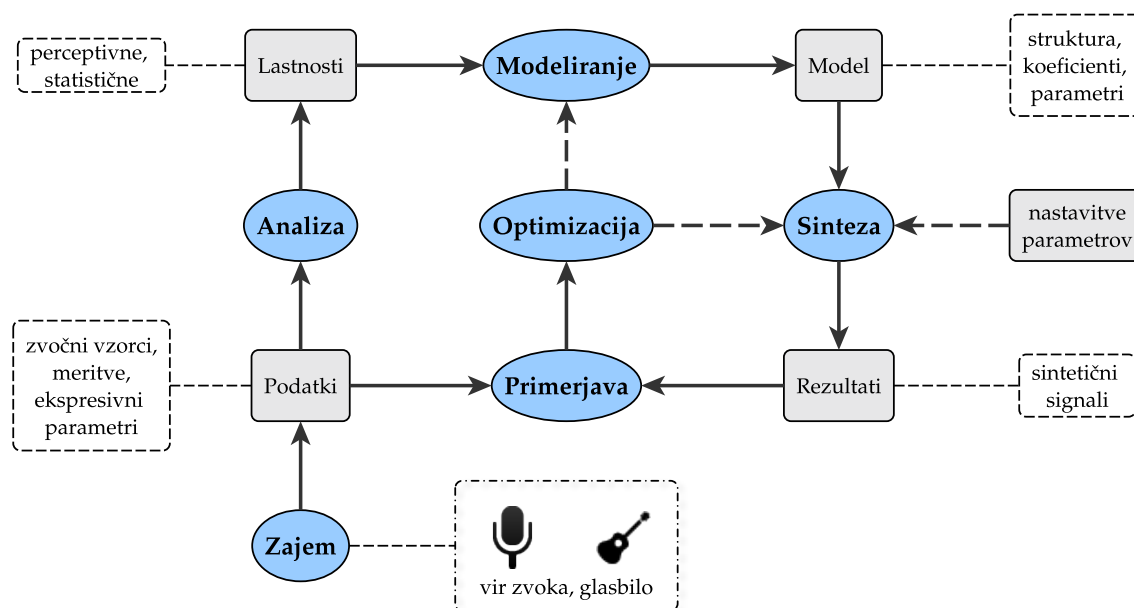
zvoka.

Fizikalno modeliranje zagotovo predstavlja pristop, ki bo postal *'de-facto'* tehnologija sintetizacije akustičnih glasbil. Danes so rezultati simulacij z DWG bržkone odlični. Tudi upravljanje takih modelov je možno na podoben način kot naravna akustična glasbila.

Pri modeliranju glasbil uporabljamo različne pristope in tehnike. Od načrtovalca je odvisno kakšen pristop bo uporabil. Če modeliramo glasbilo na podlagi vzorcev, lahko postopek posplošimo oz. strnemo v tri poglavitne faze:

1. Analiza in dekompozicija signalov.
2. Določitev primerne modela.
3. Parametrizacija in kalibracija modela.

Slika 1.2 prikazuje splošen postopek modeliranja glasbil. V osnovi je modeliranje oblikovanje abstraktnega sistema, ki ponazarja delovanje realnega sistema oz. naravnega procesa. Modeliranje je od ena ključnih sposobnosti človeka, da na



Slika 1.2: Razširjeni cikel sintetizacije (z optimizacijo)

čim bolj preprost način opiše svet, ki ga obkroža. Običajno skušamo ljudje analizirati nek proces realnega sveta, ki mu pravimo *okolje*. Analiza takega procesa se

izvaja z meritvami. V našem primeru so meritve posneti zvočni signali glasbila. Te meritve skušamo predstaviti z razgradnjo na čim bolj primitivne komponente in nam razumljive lastnosti. Procesu razgradnje pravimo tudi *analiza oz. dekompozicija*. Nato skušamo odkriti zakonitosti okolja, kar običajno delamo tako, da odkrivamo odvisnosti med meritvami oz. med komponentami, ki jih sestavljajo. Te zakonitosti zapišemo v obliki enačb in enostavnih procesov, ki mu pravimo *model*. Tako postavljen model poskušamo oceniti tako, da rezultate *simulacije* primerjamo z meritvami. Ker imamo opravka z modelom sintetizacije zvoka, je smiselno, da primerjamo *psihoakustične oz. perceptivne* lastnosti med seboj. Če se le da, razliko v percepciji opišemo z numerično izračunljivo razdaljo (*metriko*) med lastnostmi (zvočnimi značilkami) modela in realnega okolja. Perceptivna razdalja nam predstavlja napako, ki jo skušamo minimizirati tako, da prilagajamo parametre oz. koeficiente modela. Temu postopku pravimo *kalibracija*. Če ga lahko avtomatiziramo, imamo opravka z *optimizacijo*.

Do popolnega učinka sintetizacije pridemo s popolno zvočno reprodukcijo, ki vključuje tudi ekspresivnost izvajalca. Če povzamemo, *želja glasbenika-izvajalca oz. načrtovalca sintetizatorja je pustvariti simulacijsko okolje podobno naravnemu glasbilu, tako po akustičnih lastnosti, kot načinu upravljanja*. To pomeni, da se morajo:

- akustične lastnosti realnega glasbila in sintetizatorja ujemati in
- ukazi upravljanja preslikati v parametre sintetizatorja.

Fizikalno modeliranje to omogoča, vendar ima tudi določene pomanjkljivosti:

- zahteva dobro poznavanje okolja, ki ga skušamo opisati,
- detaljna implementacija glasbila je še vedno računsko zahtevna,
- modelirajo se tudi lastnosti zvoka, ki jih človek ne zaznava in
- pogosto se določene komponente modelirajo z drugimi pristopi sintetizacije, ki ne sodijo v fizikalno modeliranje.

Opisane točke predstavljajo izhodišče za razmišljanje v smeri hibridnega modeliranja. Kar pomeni, da se določene dele fizikalnega modela aproksimira zaradi

lažjega računanja, po možnosti z upoštevanjem omejitev človekovega sluha. Določene komponente imajo podobne lastnosti pri številnih glasbilih, zato se lahko ustvari bolj generičen model, ki opisuje bistveno več glasbil z relativno nizkim povešanjem kompleksnosti. Tak model lahko prilagodimo (*kalibriramo*) glede na zvočne vzorce oz. lastnosti, ki jih želimo pokriti, ponovno z upoštevanjem slušnih omejitev. Avtomatski postopek kalibracije s perceptivno optimizacijo modela predstavlja izhodišče za diplomsko delo, pri čemer se opiramo na naslednje trditve:

1. Vzorci, ki prihajajo iz istega vira (glasbila) imajo skupne lastnosti, ki jih lahko opišemo s strukturnim modelom.
2. Na podlagi zadostne količine vzorcev, za katere velja zgornja trditev, lahko glede na psihoakustične lastnosti postavimo nizkodimenzionalni model, ki predstavlja abstrakcijo realnega vira zvoka in ima podobne psihoakustične lastnosti.
3. Na podlagi zgornjih trditev, je možno z ustrezno preslikavo nadzornih lastnosti napovedati oz. interpolirati parametre modela in tako rekonstruirati manjkajoče vzorce.

V drugem poglavju so opisane teoretične osnove analize in sinteze zvoka. Začne se z obravnavo človekovega zaznavanja zvoka na nekaterih izsledkih nevrokognitivnih znanosti na področju sluha. Nato sledi opis nekaterih klasičnih prijemov za analizo signalov s pristopom spektralnega modeliranja. Za tem so opisani osnovni koncepti sintetizacije in predstavitev nekaterih uveljavljenih (klasičnih) tehnik sintetizacije. V tretjem poglavju bo predstavljena izgradnja modela, ki temelji na fizikalnem modeliranju glasbil z uveljavljeno tehnologijo digitalnih valovnih vodil. Modeliranje bo osredotočeno na fizikalni model klavirja. V četrtem poglavju bo opisan pristop za avtomatično kalibracijo modela na vhodne vzorce z uporabo optimizacije z diferencialno evolucijo. Prikazani bodo rezultati izvajanja in ovrednotenje uporabljenih metod celotnega postopka.



## Poglavje 2

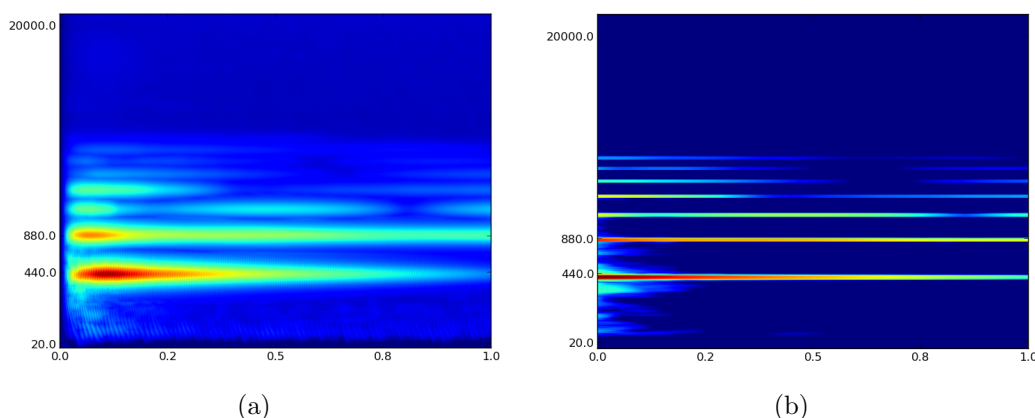
# Osnove analize in sinteze zvočnih signalov

Razumevanje lastnosti zvočnih signalov je predpogoj za nadaljno obdelavo. To poglavje opisuje človeško razumevanje zvoka in nekatere temeljne pristope k analizi in sintezi zvoka, ki bodo uporabljeni v nadaljevanju.

### 2.1 Psihoakustične lastnosti zvoka

Področje sluha ni domena tega dela, vendar je poznavanje lastnosti človeškega sluha ključnega pomena za analizo zvočnih signalov in utemeljitev uporabe psihoakustičnih lastnosti oz. značilk in perceptivne podobnosti pri optimizaciji v poglavju 4.

S fizikalnega stališča je zvočni oz. akustični signal valovanje zračnega pritiska v prostoru. Uho to valovanje pretvori v živčne impulze, ki se prenašajo po fizičnem komunikacijskem kanalu vse do možganov. Tam se izvaja nadaljna obdelava (filtriranje) in razni kognitivni procesi zaznavanja (asociacije). Signale v kognitivnem procesu možganov zaenkrat ne znamo oz. ne moremo izmeriti. Lahko pa izmerimo in modeliramo obliko signalov pred vstopom v možgane. Tako so znanstveniki dobili aproksimiran model odziva celotnega ušesa, ki se dandanes pogosto modelira s t.i. *Gammatone* in *Gammachirp* valčnimi funkcijami [7]. Časovno-frekvenčni predstavitvi signalov na koncu notranjega ušesa pravimo kroleogram, gammatone aproksimacija je prikazana na sliki 2.1(a), za primerjavo je



Slika 2.1: Kohleogram (a) in spektrogram v logaritemskem merilu (b) za klavirski ton ‘A4’ (440 Hz) s trajanjem ene sekunde

podan še spektrogram z ERB<sup>1</sup> transformacijo frekvenčne lestvice (slika 2.1(b)).

S systemskega stališča lahko torej obravnavamo človekov slušni organ kot kombinacijo akustičnega ojačevalca in specialnega frekvenčnega analizatorja, kar je že v 19. stoletju ugotovil nemški fizik Hermann von Helmholtz [15]. Ampak človekov sluh nikakor ni optimalen. Te pomanjkljivosti se izražajo kot zanimivi pojavi, ki nam dajejo natančnejšo sliko o delovanju človeškega zaznavanja zvoka in bodo opisani v nadaljevanju. Kljub temu, se lastnosti neoptimalnega zaznavanja zvoka lahko izkažejo koristne pri sintetizaciji oz. pri optimizaciji modela zvočnega vira, saj nam ni potrebno do popolnosti modelirati vseh lastnosti zvoka. Osredotočimo se samo na perceptivne.

### 2.1.1 Spektralno maskiranje

Če natančno poslušamo zvok dveh sinusoidnih oscilatorjev, kjer lahko vsakemu spreminjamo frekvenco, bomo ugotovili, da z bližanjem frekvenc, izgubljammo sposobnost razločevanja tonov. V neki točki ne bomo več zaznali dveh ločenih tonov, ampak samo enega z vmesno frekvenco in počasno modulacijo amplitude, ki se s približevanjem frekvenc še bolj upočasnjuje. Ta efekt se pojavi pri preslikavi me-

<sup>1</sup>ERB (Equivalent Rectangular Bandwidth) je empirično določena mera, ki aproksimira pasovne širine filtrov v človeškem slušnem organu in se uporablja za določitev srednje-prepustnih filtrov, ki aproksimirajo modeliranje sluha.

hanskega valovanja v elektrokemične impulze v notranjem ušesu in mu pravimo *spektralno maskiranje*.

Komunikacijski kanal med notranjim ušesom in centralnim živčnim sistemom prenaša spektralno informacijo v obliki frekvenčnih komponent in jo lahko opazimo že pri analizi kohleograma. Skupina senzornih celic, priklopljenih na eno živčno vlakno, se obnaša kot srednje prepustni filter, ki spusti ozek pas okoli frekvence na kateri je uglašena. Če sta frekvenci toliko blizu, da se preslikata v isti kanal, tvorita enotno električno stimulacijo v sistemu nevronov in tonov teoretično ne moremo več ločiti. Lahko pa zaznavamo amplitudno nihanje, ki se pojavi zaradi interference v filtriranem kanalu.

### 2.1.2 Časovno maskiranje

Če med poslušanjem glasbe zaslišimo nek glasen zvok (npr. pok), lahko z natančnim opazovanjem ugotovimo, da bomo takoj za tem težje razločevali oz. zaznali zvoke. Še bolj zanimivo je dejstvo, da tudi nekaj milisekund pred ‘pokom’ ne bomo slišali ničesar. Temu pojavu pravimo časovno maskiranje, traja pa približno 20 ms pred glasnim zvokom in 100 ms po njem. Ta pojav nas sicer ne zanima toliko v kolikor so zvoki razmeroma stacionarni. Pride pa v poštev pri dinamičnih zvočnih učinkih, ki so značilni za tolkala.

### 2.1.3 Zaznavanje osnovne frekvence tona

Človek natančneje zaznava osnovno frekvenco  $f_0$ , če so v zvoku prisotni tudi večkratniki osnovne frekvence ( $2 \times f_0, 3 \times f_0, 4 \times f_0, \dots$ ). Še več, izkaže se, da tudi brez prisotnosti osnovne frekvenčne komponente, lahko zaznamo osnovno frekvenco, če je le zvok sestavljen iz večkratnikov  $f_0$ . Navedeno značilnost lahko upoštevamo pri načrtovanju detektorja osnovne frekvence. Ta (človeška) lastnost zaznavanja je zelo podobna avtokorelacijski metodi detekcije  $f_0$ , ki bo opisana v nadaljevanju.

## 2.2 Zvočne značilke

Kohleogram (slika 2.1(a)) dobro aproksimira systemske lastnosti človeškega sluha, ampak njegov izračun je računsko potratna operacija, poleg tega ne zadovolji optimalnosti v kontekstu višjega (kognitivnega) nivoja slušnega zaznavanja.

Kognitivni nivo je zaenkrat prezapleten, da bi ga lahko podali v obliki enačb. Približek temu bi lahko bil model umetne nevrnske mreže, določene na osnovi učenja empirično ovrednotenih lastnosti zvoka. Tako učenje mreže je seveda zamudno, vendar bi dobili vsaj grobo aproksimacijo kognitivnega nivoja slušnega sistema posameznika. Tovrstne metode niso novost, uporabljajo se npr. pri klasifikaciji zvokov.

Podoben problem, kot pri določitvi perceptivnih lastnosti zvoka, imamo tudi pri primerljivosti dveh zvokov. Drugače povedano, ne moremo enolično in kvantitativno določiti koliko sta si zvoka medseboj podobna oz. kolikšna je njuna **perceptivna razdalja**. Mero perceptivne podobnosti potrebujemo tako pri ocenjevanju učinkovitosti modelov sintetizacije, kot tudi pri optimizaciji modela (poglavje 4). Zelo verjetno je metrika značilke, ki najbolje opisuje lastnosti zvoka, tudi najboljša mera podobnosti oz. perceptivne razdalje zvokov.

Če želimo dobro opisati zvok s perceptivnega stališča, moramo izbrati skupino značilk (atributov), ki bo čim natančneje opredelila človekovo zaznavanje zvoka. Zaželeno bi bilo, da skupina izbranih značilk ustreza naslednjim zahtevam:

- predstavitev v časovnem in frekvenčnem prostoru,
- odpornost na krajše (nezaznavne) časovne zamike,
- psihoakustična opredelitev,
- nizka računaska zahtevnost <sup>2</sup>.

Obstaja cela skupina nizko-dimenzionalnih značilk zvoka, ki se uporabljajo za določanje njegovih psihoakustičnih lastnosti. MPEG-7 standard določa bogato skupino značilk namenjenih predvsem razvrščanju in klasifikaciji zvoka [13]. Spo-

---

<sup>2</sup>Hiter izračun je zaželen predvsem zaradi postopka optimizacije pri avtomatski kalibraciji parametrov (več v poglavju 4).

daj so navedene perceptivno utemeljene značilke, ki skupaj ustrezajo zgornjim zahtevam.

- Amplitudna ovojnica nam določa amplitudni potek signala v času. Dobimo jo tako, da izračunamo povprečno jakost signala v časovnem okvirju z dolžino  $N$ ,

$$A(k) = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} |x(n + kN)|}. \quad (2.1)$$

- Frekvenčni spekter nam podaja informacijo o jakosti signala na posamezni frekvenci. Izračunamo ga kot absolutno vrednost kvadrata DFT transformacije signala,

$$P_x(\omega) = |\text{DFT}_N(x)|^2 = \left| \sum_{n=0}^{N-1} x(n) \cdot e^{-i2\pi kn/N} \right|^2. \quad (2.2)$$

- Težišče spektra oz. *spektralna centroida* je utežena srednja vrednost frekvenčnega spektra,

$$S_c = \frac{\sum_{n=0}^{N-1} P_x(\omega_k) w(\omega_k)}{\sum_{n=0}^{N-1} P_x(\omega_k)}, \quad (2.3)$$

kjer je  $w(n)$  monotono naraščajoča perceptivno motivirana obtežitvena funkcija vsake komponente diskretnega frekvenčnega spektra, pogosto kar linearna premica  $w(\omega_k) = \omega_k$ .

- Spektralna ovojnica se razlikuje od jakostnega frekvenčnega spektra predvsem po resoluciji v frekvencah, kjer se upoštevajo samo dominantne spektralne komponente. Lahko jo izračunamo z nizkoprepustnim filtriranjem *kepstra*<sup>3</sup> ali linearnim prediktivnim kodiranjem (LPC).
- Avtokorelacijski vektor predstavlja podobnost signala samemu sebi v vsaki točki v času. V diskretnem prostoru ga izračunamo s križno korelacijo signala s samim seboj,

$$R_x(j) = \sum_n x_n \bar{x}_{n-j}. \quad (2.4)$$

Lahko pa ga izračunamo kot DFT inverz jakostnega spektra, kar je običajno računsko ugodneje (pri dolgih signalih).

---

<sup>3</sup>Kepster je DFT inverz logaritma frekvenčnega spektra. Pogosto se uporablja pri analizi in razpoznavanju govora.

- Osnovna frekvenca tona ( $f_0$ ) je zelo pomembna v kontekstu glasbe. Ocenimo jo na podlagi pozicije maksimalnega vrednosti avtokorelacije (enačba 2.5), lahko pa tudi z razmikom med harmonskimi komponentami (pri izrazito harmonskih zvokih).

$$\hat{f}_0 = f_s / \arg \max_i \{R_x(i)\} \quad (2.5)$$

## 2.3 Dekompozicija signalov

Iz uvodnih trditev, predpostavljamo, da vsak vzorčen zvočni signal istega izvora vsebuje informacijo, ki je skupna vsem vzorcem in informacijo v obliki posameznih komponent. Dekompozicija signalov nam omogoča boljšo interpretacijo posameznih in skupnih lastnosti ter odkrivanje in interpretacijo relacij med vzorci. S pravilnim pristopom, lahko določimo tudi lastnosti manjkajočih zvočnih signalov, v kolikor imamo na voljo dovolj učnih signalov. Na ta način lahko oblikujemo tudi pravi model vira, ki ga obravnavamo. V nadaljevanju bodo obravnavane splošne metode za časovno-frekvenčno predstavitev signalov, tako kot metode za dekompozicijo signalov v funkcionalne sklope.

### 2.3.1 STFT in spektrogram

Zvočni vzorci imajo določen potek signala v času, zato lahko vzorec na nek način razgradimo tudi na frekvenčne komponente v času. Tu si lahko pomagamo s frekvenčno predstavitvijo signala v časovno krajših odsekih oz. okvirjih in jim pravimo **časovno-frekvenčne predstavitve** signalov [21].

Kratko-časovna Fourierjeva transformacija (*Short-time Fourier Transform - STFT*) je najpogostejša implementacija časovno-frekvenčne predstavitve signalov. Običajno jo računamo z hitro Fourierjevo transformacijo, zato deduje tudi njene težave, t.j. problem nedoločenosti bodisi v frekvencah ali v času. Po drugi strani, zaradi uporabe algoritma hitre Fourierjeve transformacije (FFT) jo odlikuje hitrost izračuna tako predstavitve, kot tudi njenega inverza. STFT s korakom  $K$  in širino okna  $N$  v diskretnem času  $n$  opišemo z enačbo:

$$X[m, \omega] = \sum_{n=-\infty}^{\infty} x(n)w(n - mK)e^{-j\omega n}, \quad (2.6)$$

kjer je  $m$  časovni odsek,  $w$  pa frekvenca v  $rad/s$ . Tudi inverz predstavitve STFT je izračunljiv, saj predstavitev hrani vso potrebno informacijo za rekonstrukcijo signala. V ta namen se uporablja metoda *Overlap-Add (OLA)*,

$$x(n) = \sum_m \text{DFT}^{-1} X[m, \omega]. \quad (2.7)$$

STFT inverz je ekvivalenten originalnemu signalu samo, če za uporabljeno oken-sko funkcijo velja:

$$\sum_m w(n - mK) = 1. \quad (2.8)$$

**Spektrogram** je časovno-frekvenčna predstavitev signala s časovnim pote-  
kom jakostnega spektra, ki ga običajno računamo s kvadratom absolutne vredno-  
sti kratko-časovne Fourierjeve transformacije,

$$S_x[m, \omega] = |X[m, \omega]|^2. \quad (2.9)$$

Velja poudariti, da iz spektrograma ne moremo rekonstruirati originalnega signala, saj je informacija o fazah signala izgubljena.

### 2.3.2 Dekompozicija signalov v funkcionalne sklope

Vsak zvočni signal lahko obravnavamo kot sestav različnih komponent glede na obnašanje sistema v časovnem ali frekvenčnem prostoru. Signal skušamo razčleniti na primitivne komponente, kar nam omogoča enostavnejše razumeva-nje in identifikacijo narave signala. Glede na operacije v času lahko razčlenimo signal na:

- **aditivne** komponente (sinusoide, šum, impulzi, rampe),
- časovno **multiplikativne** komponente (amplitudne ovojnice, okenske funk-cije),
- **filtri oz. konvolucijske** komponente.

Aditivne komponente predstavljajo linearno ločljive elemente tako v časovnem, kot v frekvenčnem prostoru. Klasični primer aditivne dekompozicije je Fou-rierjeva analiza, kjer vsaka točka predstavlja sinusoidno komponento signala z

določeno fazo in frekvenco. Skupni seštevek teh sinusoid predstavlja periodično obliko originalnega signala.

Časovno multiplikativne komponente so signali, ki nastopajo kot produkt vzorcev v časovni domeni. To so običajno signali umetne narave, amplitudne ovojnice in okenske funkcije, ki jih pogosto uporabljamo za predprocesiranje signala. Množenju v časovnem prostoru pravimo tudi amplitudna modulacija. V svetu sintetizacije pogosto srečamo nekatere primitivne multiplikativne komponente, npr. ADSR amplitudne ovojnice, LFO amplitudni modulatorji in RING modulatorji.

Konvolucijske komponente oz. filtri so plod operacije konvolucije dveh signalov oz. filtrov. Konvolucija v časovnem prostoru se preslika v množenje v frekvenčnem prostoru, in obratno. Pri analizi 'naravnih' signalov imamo opravka predvsem z aditivno in konvolucijsko dekompozicijo.

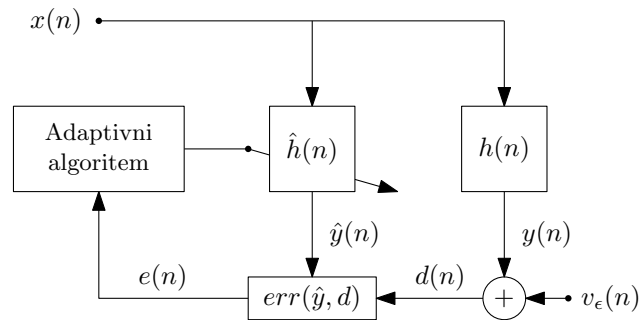
### 2.3.3 Inveržno modeliranje

Ker imamo opravka z dekompozicijo signala, nas zanima, kakšen je bil njegov izvor (vzbujanje) preden je vstopil v sistem s prenosno funkcijo  $h(n)$ , če poznamo izmerjeni signal  $y(n)$  (vzorčen zvok). Za določitev izvora potrebujemo inverz funkcije  $h(n)$ , inverzni operaciji pa pravimo **dekonvolucija**. Tu iščemo inverzni sistem  $h^{-1}(t)$ , za katerega velja  $x(t) = h^{-1}(t) * y(t)$ .

Dekonvolucija v frekvenčnem prostoru predstavlja deljenje frekvenčne predstavitev signalov in je (matematično) nedefinirana v vsaki točki, kjer je vrednost amplitude v imenovalcu ničelna. Realni signali sicer nimajo ničelnih frekvenčnih komponent, vendar se v izmerjenemu signalu ob zajemu običajno pojavlja šum, bodisi zaradi neidealne merilne opreme (mikrofonov), bodisi zaradi kvantizacijske napake na AD pretvorniku. To postopek dekonvolucije nekoliko oteži, kajti sedaj nastopa v sistemu tudi aditivni šum:

$$X = H^{-1}Y = H^{-1}(Y_S + \epsilon) \quad (2.10)$$

V tem primeru lahko dobimo le oceno inverznega filtra ( $\hat{H}^{-1}$ ). Pa še ta je močno zašumljena v predelih spektra, kjer je razmerje magnitud med signalom  $Y_S$  in šumom  $\epsilon$  (SNR) nizko. Če poznamo obliko močnostnega spektra aditivnega

Slika 2.2: Adaptivni filter za identifikacijo sistema  $h(n)$ 

šuma, si lahko pomagamo z Wienerjevo metodo dekonvolucije, ki zaduši šumne komponente [28],[5]. V kolikor narave šuma ne poznamo, lahko problem iskanja inverznega filtra prevedemo na problem **identifikacije sistema** (slika 2.2). V ta namen za dekonvolucijo uporabimo katerega izmed adaptivnih filtrov za izračun inverza sistema, npr. algoritmi NLMS, RLMS, MDF [5].

Seveda, če nam niti originalni signal oz. vir  $x(n)$  ni znan, je nemogoče določiti filter  $h^{-1}(n)$ . Če pa imamo vsaj oceno signala  $x(n)$  ali poznamo nekatere njegove statistične lastnosti, lahko ta sistem še vedno nekoliko rekonstruiramo. V tem primeru imamo opravka s t.i. slepo dekonvolucijo. Tu si lahko pomagamo samo z nekaterimi statističnimi ocenami signalov ali pa skušamo določiti skupne lastnosti signalov s spektralnimi metodami dekompozicije.

### 2.3.4 Sinusoidno modeliranje

Vsak zvok lahko zapišemo s kombinacijo sinusoidnih komponent, kar že vemo iz Fourierjeve analize. Posebnost sinusoidnega modeliranja je, da skušamo samo glavne spektralne komponente opisati z amplitudnimi in frekvenčnimi ovojnici [21]. Najpreprostejša implementacija je izvedena s sledenjem spektralnih vrhov pri STFT predstavitvi. Ker so signali večinoma odsekoma stacionarni, lahko to informacijo preoblikujemo tako, da odsekoma lineariziramo časovni potek amplitud in frekvenc na posameznih spektralnih komponentah. Imamo torej opravka s kombinacijo amplitudne in frekvenčne modulacije, kar omogoča relativno enostavno implementacijo s paralelno povezanimi oscilatorji, ki jim spreminjamo frekvenco in amplitudo. Proces sintetizacije predstavlja najosnovnejšo

tehniko spektralnega modeliranja, t.j. aditivna sinteza. Rekonstrukcijo oz. sintezo signala dobimo po formuli:

$$y(t) = \sum_{i=1}^N A_i(t) \sin[\theta_i(t)], \quad (2.11)$$

kjer so  $A_i(t)$  in  $\theta_i(t)$  vrednosti ovojnice komponente  $i$  v času  $t$ , ki jih dobimo z linearno interpolacijo.

Če sinusoidno modeliranje razširimo tako, da vključimo ločeno modeliranje šumnih komponent, dobimo tako imenovano *deterministično-stohastično dekompozicijo* - *DSD* [19]. Princip deluje tako, da sestavimo sinusoidni model signala s sledenjem STFT spektralnim vrhovom. Tega odštejemo od originalnega signala in tako dobimo ostanek (residualni signal), za katerega predpostavljamo, da je stohastične narave in ga lahko modeliramo s filtriranjem belega šuma (s spektralno ovojnico) - to je tudi osnovni pristop subtraktivne sinteze. Pri tem velja poudariti, da se običajno tako sinusoidni model, kot model šuma v času lahko spreminjata.

### 2.3.5 Linearno prediktivno kodiranje

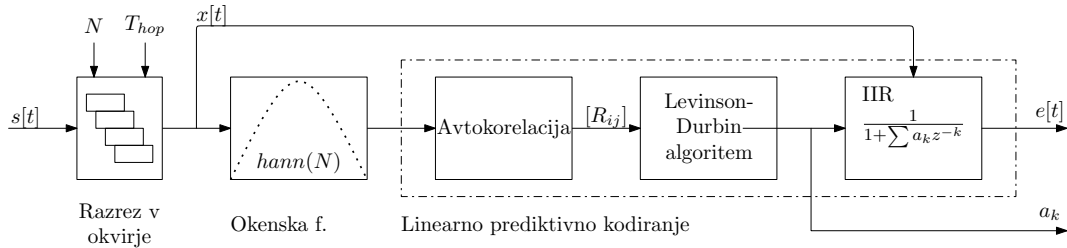
Linearno prediktivno kodiranje (LPC) je postopek za kodiranje, analizo in sintezo digitalnih signalov [12]. Odlikuje ga predvsem enostavnost in nizka računska zahtevnost, zato je takorekoč temelj vseh današnjih algoritmov za kodiranje zvoka (CELP, FLAC, ...). Izkaže se še posebej pri kodiranju govora, pogosto pa se uporablja tudi za oceno spektralne ovojnice signala kot alternativa glajenega *kepstra*. Postopek LPC analize je prikazan na sliki 2.3.

Cilj LPC metode je določiti koeficiente in napako linearne predikcije vzorca  $y(n)$  na podlagi predhodnih vzorcev,

$$y(n) = -a_1y(n-1) - a_2y(n-2) - \dots - a_My(n-M) + e(n), \quad (2.12)$$

kjer je  $M$  red linearne predikcije. Pri t.i. avtokorelacijski LPC metodi običajno skušamo minimizirati srednjo kvadratno napako med signalom in napovedjo,  $E[y(n) - \hat{y}_n]$ . S tem dobimo sistem enačb

$$\sum_{k=1}^M a_k R_y(j-k) = -R_y(j), \quad (2.13)$$



Slika 2.3: Analiza z LPC

kjer označujemo z  $R_y(j)$  avtokorelacijo signala  $y$  pri odmiku  $j$ ,

$$R_y(j) = \sum_n y(n) \bar{y}(n - j). \quad (2.14)$$

Ta sistem enačb<sup>4</sup> lahko zapišemo v matrični obliki, ki ima lastnosti simetrične Toeplitz-ove matrike dimenzij  $M \times M$ . Rešimo ga lahko s Gaussovo metodo, vendar se zaradi računske učinkovitosti pogosto uporablja aproksimacija z t.i. Levinson-Durbin algoritmom [12].

Vrednosti  $a_k$  tvorijo koeficiente realnega polinoma, ki jih lahko opišemo z racionalnim filtrom  $H_a$ , ki je sestavljen iz samih polov. Ostanek napovedi ali *residuuum* lahko izračunamo z inverznim filtriranjem, to je s konvolucijo signala in dobljenega filtra,  $Y_k H_a^{-1}$ , v časovni domeni:

$$e(n) = \sum_k a_k y(n - k) \quad (2.15)$$

Tako dobljeni ostanek predstavlja signal, ki ima oslABLJENE glavne spektralne komponente in posledično močno ojačane šumne komponente. Konvolucija originalnega signala z LP koeficienti skuša, torej, *izravnati spekter* signala. Ta ostanek se pri LPC običajno lahko predstavi z bistveno manj informacije s postopkom, ki mu pravimo vektorska kvantizacija (VQ). Zato je taka (kompresirana) predstavitev signala primerna za pošiljanje po komunikacijskih kanalih.

LPC sinteza je obratni postopek LPC analize in deluje na podobnem principu kot vokalni trakt pri govoru. Ta sistem lahko obravnavamo kot *izvor-filter* model, kjer formantni filter predstavlja LPC koeficienti  $H_a$ , izvor pa model vzbujanja s kombinacijo impulzov in šuma. Ta način zelo učinkovito modelira vokalni trakt, zato se je, pogosto uporabljal za sintezo umetnega govora.

<sup>4</sup>v literaturi jih srečamo pod imenom *Yule-Walker enačbe*

## 2.4 Tehnike sintetizacije zvoka

V splošnem je sintetizacija proces, kjer želimo simulirati skupino zvočnih signalov, ki so rezultat nekega zvočnega vira ali družine zvočnih virov. Želimo torej informacijo v obliki zvočnih vzorcev preslikati tako, da uporabimo čim manjše število parametrov, obenem pa ohranimo čim več informacije o obliki sistema, ki signale generira. Princip je zelo podoben kompresiji zvoka, vendar je tu eden primarnih ciljev ohraniti nadzor uporabnika nad tistimi parametri, ki vplivajo na nekatere perceptivne lastnosti signala, kot so npr. ton, jakost, svetlost, "barva", trajanje, vibracija, ipd.

Sedaj se pojavi vprašanje, kakšen delež informacij o ciljnem vzorcu naj hrani model v obliki tabel oz. uporabniku skritih parametrov? Odvisno je od cilja sintetizacije, torej, ali želimo popolnoma reproducirati originalne vzorce ali raje ohraniti fleksibilnost in omogočiti uporabniku ohraniti večji nadzor nad perceptivnimi lastnostmi. Glede na fleksibilnost oz. popolnost reprodukcije zvoka lahko uporabimo več različnih konceptov sintetizacije zvoka, ki bodo opisani v nadaljevanju [21],[20].

1. **Vzorčenje** je najpreprostejša tehnika sintetizacije, kjer posamezne zvoke glasbila posnamemo v ustreznem okolju, t.j. brez odmevov. Nato zvočne vzorce primerno obdelamo, da se prilagodijo vhodnim parametrom in predvajamo na različnih frekvencah in jakostih.

Kljub temu, vzorčenje omogoča samo omejen realizem. Vzorčenje je namreč optimalno samo v točkah posnetih vzorcev. Zato se pogosto uporabi interpolacija in predprocesiranje sosednjih vzorcev. In kar je bolj pomembno, z večanjem števila parametrov, število potrebnih zvočnih vzorcev za sintezo eksponentno narašča. Zato je pri večini glasbil nemogoče upoštevati vse okoliščine sistema, ki jih lahko izvajalec oz. glasbenik izrazi in jih mora nek sintetizator pokrivati. Zato običajno izpustimo nekatere ekspresivne lastnosti, kot so dinamično vzbujanje, medsebojna resonanca, ipd.

Nekateri proizvajalci so šli v skrajnost do te mere, da vzorčijo vsak ton glasbila posebej in z različnimi parametri. Seveda je tak pristop zelo požrešen, vsaj s stališča porabe pomnilnika. Ti vzorci lahko porabijo tudi več sto me-

gabajtov pomnilniškega prostora. Na primer, 16-bitno vzorčenje klavirja pri frekvenci vzorčenja  $48\text{ kHz}$ , ki upošteva samo najosnovnejše parametre: frekvenca tona (88 tonov) in jakost (samo 16 nivojev) terja 1.35GB pomnilnika, pri posnetkih, ki trajajo vsaj 10s.

$$S_{mem} = 88 \times 16 \times 48000\text{Hz} \times 2\text{Byte} \times 10\text{sec} = 1.35\text{GByte} \quad (2.16)$$

Druga slabost vzorčenja s stališča sintetizacije je nadzor nad perceptivnimi lastnostmi zvoka, ki jih sicer lahko do neke mere aproksimiramo z naknadnim procesiranjem signala z različnimi filtri.

2. **Strukturno vzorčenje** se razlikuje od navadnega vzorčenja po tem, da upošteva tudi nekatere koncepte fizikalnega ali abstraktnega modeliranja. Gre za neke vrste kombinacijo vzorčenja in modeliranja zvočnega vira. Namesto vzorčenja zvočnih signalov, vzorčimo samo nekatere osnovne komponente zvoka, kot so npr. odzivi na enotin impulz vzbujanja, ohišja, ipd. Prednost je ta, da lahko sistem opišemo z bistveno manj informacije kot pri vzorčenju, saj so nekateri karakteristični signali, npr. odzivi na enotin impulz, invariantni glede na nadzorne parametre [20].

Strukturno vzorčenje je kompromis, ki omogoča večjo fleksibilnost od navadne tehnike vzorčenja, obenem pa ohranja želeno kvaliteto signala. Primer take sintetizacije je komutirana sintetizacija z valovnimi vodili, kjer sta trup inštrumenta in vzbujevalni vir združena v enotnem (vzorčenem) signalu, medtem, ko so elementi resonatorjev posebej modelirani.

3. **Abstraktno modeliranje** je koncept sintetizacije, kjer je informacija o vzorcih modelirana v celoti parametrično v obliki abstraktnega strukturnega modela z minimalno količino informacij v obliki koeficientov. Pomemben cilj je minimizacija modela do take mere, da čedalje bolj pokrije ekspresivne lastnosti nekega glasbila. Pri tem želimo imeti čim nižje število parametrov, obenem pa želimo ohraniti fleksibilnost. Popolna perceptivna reprodukcija signala je tu sekundarnega pomena. Običajno se pri tem pristopu uporabljajo generični modeli sintetizatorjev, ki imajo nizko število parametrov in pokrivajo široko skupino zvokov oz. celotno družino glasbil.

V to skupino spadajo nekatere primitivne tehnike sintetizacije, kot npr. frekvenčna modulacija, amplitudna modulacija, fazna modulacija, fazna distorzija in seveda kakršna koli kombinacija le-teh.

4. **Fizikalno modeliranje** modelira informacijo o vzorcih v celoti parametrično v obliki aproksimativnega fizikalnega modela, vendar z minimalno količino informacij v obliki koeficientov. Fizikalno modeliranje se osredotoča na pravilno modeliranje fizikalne strukture glasbila. Popolna perceptivna reprodukcija signala je tudi tu sekundarnega pomena. Kljub temu je kvaliteta zvokov odlična. Glavno orodje fizikalnega modeliranja je običajno numerično reševanje sistemov *diferencialnih* enačb. Tipične tehnike sintetizacije pa so *modalna sinteza* in *digitalna valovna vodila* [22],[20].

5. Pri **spektralnem modeliranju** skušamo ugotoviti lastnosti inštrumenta samo na podlagi nekaterih lastnosti zvočnih vzorcev. Imamo torej opravka s pomanjkljivo ali neobstoječo informacijo o viru zvočnega signala (npr. glasbilu). Cilj spektralnega modeliranja je čim boljša (perceptivna) reprodukcija zvočnih vzorcev na račun nižje fleksibilnosti parametrizacije. Tu običajno analitično določimo parametre za vsak ton, ki ga želimo reproducirati. Število parametrov sintetizatorja je običajno zelo visoko, saj vključuje zapis nekaterih lastnosti v obliki vzorčenih tabel in koeficientov v časovnem in/ali frekvenčnem prostoru. Ker je model, ki generira zvočni signal, nepoznan, je spektralno modeliranje, s stališča sinteze zvoka, splošnejše od običajnega modeliranja oz. identifikacije sistema, saj moramo v tem primeru uporabljati generičen model sintetizacije.

Med tehnike spektralnega modeliranja štejemo analizo in sintezo s sinusoidnim modeliranjem (SMS), stohastično-deterministično dekompozicijo (DSD), linearnim prediktivnim kodiranjem (LPC) ter analizo in sintezo v frekvenčni domeni s hitro Fourierjevo transformacijo [21].

6. **Križna sintetizacija** je podzvrst spektralnega modeliranja, kjer je cilj križanje nekaterih perceptivnih lastnosti dveh različnih zvočnih virov. Pogost primer je kombinacija vzbujanja enega glasbila s spektralno ovojnico drugega. V ta namen se običajno uporabljajo tehnike spektralnega modeli-

ranja, npr. sintetizacija z linearnim prediktivnim kodiranjem (LPC), fazni vokoder<sup>5</sup> ali sinusoidno modeliranje. Lahko pa tudi križamo parametre abstraktnega ali fizikalnega modela, v kolikor imamo *enoten* model, ki je sposoben dobre reprodukcije zvoka obeh glasbil. Pri tem dobimo nov zvok, ki lahko, ne pa nujno, predstavlja kombinacijo perceptivnih lastnosti prvega in drugega glasbila.

Trenutno najbolj razširjena tehnika sintetizacije naravnih (akustičnih) glasbil je vzorčenje signalov. Vendar obstajajo določena glasbila, kjer se tehnika vzorčenja ne obnese najbolje, predvsem zaradi velike ekspresivnosti oz. raznolikosti glasbila. Lep primer so pihala, trobila in godala. Za te družine glasbil se dandanes vse bolj uveljavljajo tehnologije s katerimi lahko poljubno natančno simuliramo skorajda vsako glasbilo. Glavna predstavnika sta zagotovo fizikalno modeliranje in spektralno modeliranje [22],[20],[21].

V nadaljevanju so našteje nekatere uveljavljene tehnike sintetizacije zvoka, ki so se razvijale v 20. stoletju v obliki hibridnih elektroakustičnih sistemov in analognih elektronskih vezij. Danes se skoraj izključno pojavljajo v obliki računalniških programov ali digitalnih sistemov.

### 2.4.1 Aditivna sinteza

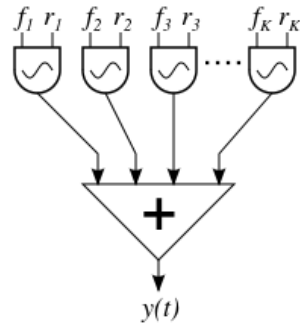
Aditivna sinteza je najpogosteje uporabljena tehnika v klasičnih sintetizatorjih prve generacije, predvsem v emulatorjih orgel, npr. *Hammond*. Njena poglobitna prednost je možnost določitve vseh harmonskih frekvenčnih komponent v zvoku, t.j. spektralne ovojnice. Tipična implementacija se izvaja s paralelnim blokom neodvisnih oscilatorjev z določeno amplitudo in frekvenco, kot kaže slika 2.4 in je opisana z enačbo:

$$y(t) = \sum A_k \cos(2\pi f_k t + \phi_k). \quad (2.17)$$

Vendar z naraščanjem frekvenčnih komponent raste tudi zahteva po procesni moči. Sploh, če je za oscilatorji uporabljen še kak način modulacije amplitude, npr. nizko-frekvenčna oscilacija (LFO) ali ADSR ovojnica.

---

<sup>5</sup> Fazni vokoder (*phase vocoder*) je sistem, ki z izkoriščanjem fazne informacije lahko spreminja spekter in trajanje odseka zvočnega signala [21].



Slika 2.4: Blok shema aditivne sinteze

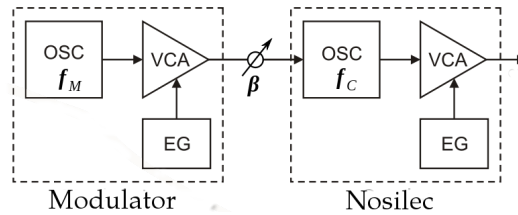
### 2.4.2 Subtraktivna sinteza

Je priljubljena tehnika, ki deluje preprosto tako, da vzamemo vzorec oz. signal z bogato harmonsko vsebino (npr. beli šum ali plaz impulzov) in ga filtriramo z resonančnimi filtri, da dosežemo želeni ton. Tako dosežemo harmonsko bogat zvočni vzorec z relativno zelo malo parametri, običajno okrog 10 do 20 parametrov. Ti so v glavnem koeficienti srednjega oz. nizkoprepustnega filtra, amplitudne (ADSR) ovojnice izvirnega signala (šuma). Pogosto so prisotni tudi nizkofrekvenčni oscilatorji za amplitudno in frekvenčno modulacijo signala.

### 2.4.3 Frekvenčna modulacija

Najpogosteje uporabljena tehnika v današnjih cenениh oz. klasičnih sintetizatorjih druge generacije, je zagotovo frekvenčna modulacija (*FM*). Njena poglobitna prednost je širok spektralni nabor harmonskih komponent pri minimalnem številu parametrov in praktično zanemarljivo porabo pomilnika in procesne moči. Poleg tega, v kombinaciji s filtriranjem, nizko frekvenčno oscilacijo (LFO) parametrov in različnimi načini moduliranja z ADSR ovojnicami lahko dosežemo emulacijo praktično katerega koli naravnega ali umetnega vira zvoka. Osnovno shemo FM modulacije prikazuje slika 2.5. Potek signala v časovni domeni je podan kot

$$y(t) = a(t) \sin(w_c t + I_m(t) \sin(w_m t)). \quad (2.18)$$



Slika 2.5: Blok shema frekvenčne modulacije

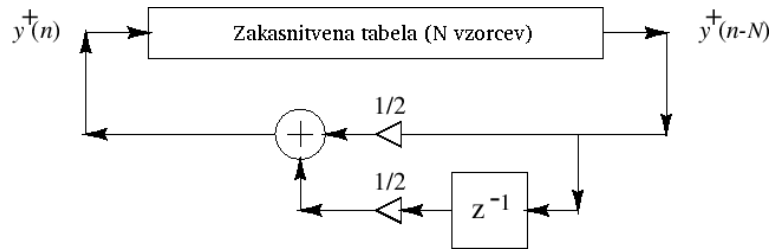
### 2.4.4 Valovne tabele

Naprednejša kombinacija tehnike vzorčenja, vsaj s stališča porabe pomnilnika, je sintetizacija z valovnimi tabelami *Wavetable* (*WT*). Tu imamo opravka z dinamično normaliziranim signalom, shranjenim v obliki tabele, ki jo predvajamo cirkularno in moduliramo na želeni osnovni frekvenci. Časovni potek običajno določimo z amplitudno ovojnico. Prednost te metode je predvsem v enostavnosti in široki paleti uporabnosti glede na nizko porabo pomnilnika in relativno nizko porabo procesorske moči. Zanimivo je dejstvo, da je to najpogostejši način implementacije sinusoidnega oscilatorja. Primer: če valovna tabela vsebuje  $N$  vzorcev in je frekvenca vzorčenja  $f_s$ , je za modulacijo zvoka na osnovni frekvenci  $f_0$ , potrebno preskakovati vzorce z zamikom  $f_s/(Nf_0)$  in uporabiti nek način interpolacije oz. nizkoprepustnega filtriranja. Jasno je, da bo kvaliteta signala precej odvisna od frekvence vzorčenja.

### 2.4.5 Karplus-Strong algoritem

Zanimiva metoda, ki je osnovana na dinamičnem obnavljanju oz. prepisovanju valovne tabele je znameniti **Karplus-Strong (KS)** algoritem [8]. Tabela je v osnovi krožni seznam v katerem se vzorci krožno pomikajo v isti smeri, kjer se na koncu seznama nekoliko preoblikujejo in nato 'vračajo' v začetek seznama (slika 2.6). Zvok oblikujemo tako, da v tabelo oz. zakasnitveno linijo vnesemo začetne vrednosti, običajno filtriran šum, lahko pa tudi poljuben vzorec. Na konec zanke vključimo filter, ki daje končno obliko zvoku v času in frekvencah. Modelu sintetizatorja, ki poleg filtra za dušenje, vključuje tudi vseprepustne filtre za 'uglaševanje' zakasnitvene linije, pravimo *razširjeni KS algoritem*.

Dinamično tabelo oz. seznam KS algoritma si lahko predstavljamo tudi kot



Slika 2.6: Karplus-Strong algoritem

zakasnitveno linijo s povratno zanko in ga lahko interpretiramo kot filter z neskončnim odzivom  $N$ -tega reda, kjer so vsi koeficienti, razen prvega in zadnjega, enaki nič. Taka razlaga predstavlja osnovo koncepta digitalnega valovnega vodila, ki bo opisan v nadaljevanju.

KS algoritem so včasih imenovali tudi *Digital*, saj je zelo učinkovito sintetiziral zvok kitare. Evolucija zvoka močno spominja na akustična glasbila, ker se začetna šumna spektralna vsebina počasi preoblikuje v urejeno harmonično strukturo. Kljub temu, da je KS algoritem osnovan na principu valovnih tabel, ki spadajo v družino vzorčenja, ga v novejši literaturi pogosteje uvrščajo med abstraktne metode sintetizacije [22].

### 2.4.6 Digitalno valovno vodilo (*Digital Waveguide*)

Digitalno valovno vodilo je učinkovit način za modeliranje prenosa valovanja po fizičnih medijih. Implementirano je s pomočjo zakasnitvenih linij in predstavlja čisto analogijo potovanja signala po fizičnem mediju, le da imamo tu opravka z diskretnimi, časovno kvantiziranimi signali. Sintetiziranje z digitalnimi valovnimi vodili se vse pogosteje uporablja predvsem zaradi računske nezahtevnosti, obenem pa so zaradi fizikalne analogije rezultati zelo obetavni za nekatere družine glasbil.

Ta tehnika se je razvila kot posplošitev EKS algoritma za poljuben fizikalni model glasbila. Kljub temu, se DWG in KS algoritem konceptualno razlikujeta. KS ima za osnovo valovno tabelo, kjer je zakasnitvena linija implementirana istočasno kot 'vir' vzbujanja in resonator. V digitalnih valovnih vodilih je zakasnitvena linija samo en izmed gradnikov (filtrov), ki običajno predstavlja le del digitalnega resonatorja.

Poseben primer kombinacije vzorčenja in modeliranja oz. strukturnega vzor-

čenja z valovnimi vodili predstavlja sintetizacija s **komutacijo** valovnih vodil - *Commutated Waveguide Synthesis (CWS)*. Izmerjen odziv na enotin impulz trupa glasbila ter vir vzbujanja sta tu združena v enoten signal, medtem, ko so resonatorski elementi modelirani ločeno na podlagi fizikalnih lastnosti ali pa na podlagi (izhodnih) zvočnih vzorcev sistema.

Ta metoda sintetizacije omogoča zadostno fleksibilnost, obenem pa ohranja skoraj popolni realizem predvsem pri glasbilih, ki se jih da opisati z linearnimi ter časovno invariantnimi sistemi. Poleg tega je računsko zelo nezahtevna metoda sintetizacije, zato je idealni kandidat za perceptivno realistično sintetizacijo množice glasbil.

### 2.4.7 Primerjava tehnik

Če izvzamemo vzorčenje in valovne tabele kot poseben primer vzorčenja, kjer je sintetizacija načeloma trivialna, nas zanima predvsem primerjava preostalih konceptov sintetizacije: abstraktne metode ter fizikalno in spektralno modeliranje. Tabela (2.1) nam podaja primerjavo teh skupin oz. pristopov sintetizacije.

Sama izbira pristopa k sintetizaciji je zgolj informativne narave, ampak kljub temu lahko povzamemo nekaj napotkov pri izbiri tehnik sintetizacije:

- Abstraktne metode so uporabne predvsem za generiranje umetnih zvokov in jih potrebujemo v aplikacijah, kjer računska zahtevnost lahko predstavlja problem.
- Spektralno modeliranje je uporabno predvsem za analizo, zapis in perceptivno kodiranje zvoka v nižjedimenzionalne sheme. Pa tudi za enostavno preoblikovanje zvočnih signalov (posnetkov), npr. modifikacijo osnovnega tona in drugih spektralnih lastnosti.
- Fizikalno modeliranje je uporabno tako za simulacijo obstoječih akustičnih glasbil, kot tudi za ustvarjanje novih modelov na osnovi akustičnih glasbil, ki jih sicer v realnem svetu ni mogoče realizirati. Poleg tega je primerno za aplikacije, kjer želimo imeti popoln nadzor tako nad fizikalnimi, kot izvajalnimi parametri.

Abstraktne metode (FM,AM,...)
<ul style="list-style-type: none"> <li>+ običajno zelo nizka računska zahtevnost</li> <li>+ nizko število parametrov</li> <li>+ možna <i>on-the-fly</i> modulacija parametrov</li> <li>– zahtevna perceptivna reprodukcija naravnih glasbil</li> <li>– specifična sintetizacija, primerna predvsem za umetne zvoke</li> <li>+ pester nabor umetnih zvokov</li> <li>– zahtevna in neintuitivna določitev parametrov</li> <li>+ enostavno upravljanje in nadzor v realnem času</li> </ul>
Spektralno modeliranje (SMS,DSD,LPC,...):
<ul style="list-style-type: none"> <li>– visoka računska zahtevnost</li> <li>– veliko parametrov (koeficientov)</li> <li>– zahtevna <i>on-the-fly</i> modulacija parametrov</li> <li>+ popolna (perceptivna) reprodukcija poljubnih zvokov</li> <li>+ generična sintetizacija</li> <li>+ poljuben nabor zvokov</li> <li>+ enostavna, analitična določitev koeficientov in parametrov</li> <li>– zahtevno oz. nemogoče upravljanje in nadzor v realnem času</li> </ul>
Fizikalno modeliranje (EKS,DWG,...):
<ul style="list-style-type: none"> <li>– računska zahtevnost odvisna od natančnosti modela</li> <li>+ nizko število parametrov</li> <li>+ možna <i>on-the-fly</i> modulacija parametrov</li> <li>+ poljubno dobra (perceptivna) reprodukcija</li> <li>– specifična sintetizacija, primerna samo za nekatere družine glasbil</li> <li>– ozek nabor zvokov specifičnega fizikalnega modela</li> <li>– zahtevna določitev parametrov in koeficientov</li> <li>+ enostavno upravljanje in nadzor in nadzor v realnem času</li> </ul>

Tabela 2.1: Primerjava tehnik sintetizacije

## Poglavje 3

# Modeliranje z digitalnimi valovnimi vodili

Implementaciji določene metode ali skupku metod sintetizacije pravimo **model sintetizatorja** oz. kar sintetizator. Model je tu mišljen kot abstraktna množica funkcij in nabora parametrov, katerih namen je modelirati akustične lastnosti glede na vhodne parametre in izračunati potek zvočnega signala. V tem poglavju skušamo modelirati tone z dekompozicijo na *izvor-filter* oz., gledano s fizikalnega stališča, na modeliranje *vzbujanje-resonator(ji)* z uporabo tehnike digitalnih valovnih vodil. Postopek bo osredotočen na modeliranje klavirskega tona, vendar bo tako dobljen (generičen) model primeren tudi za druge vrste podobnih glasbil.

Teoretična podlaga za digitalna valovna vodila je posplošena D'Alembertova rešitev  $y(t, x)$  valovne enačbe 3.1, ki predstavlja superpozicijo potovanja signala iz leve in desne strani [20].

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} \quad (3.1)$$

$$y(t, x) = y_r(t - x/c) + y_l(t + x/c) \quad (3.2)$$

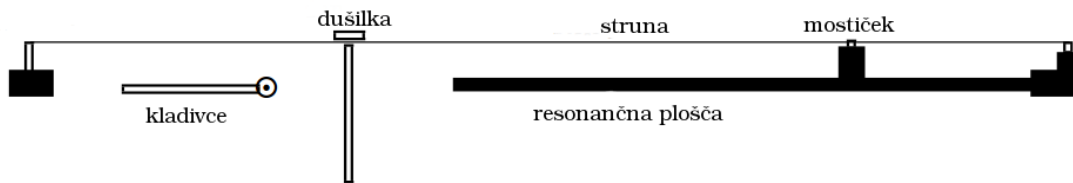
Pri modeliranju z digitalnimi valovnimi vodili je valovanje vzorčeno, zato v enačbo vpeljemo diskretni čas  $nT$ , kjer je  $T$  perioda vzorčenja in dobimo diskretno obliko D'Alembertove enačbe,

$$\begin{aligned} y(nT, mx) &= y_r(nT - mx/c) + y_l(nT + mx/c), & (x \triangleq cT) \\ &= y_r(nT - mT) + y_l(nT + mT) \end{aligned} \quad (3.3)$$

Tovrstno potovanje lahko učinkovito opišemo z zakasnitvenimi linijami, ki predstavljajo bistvo tehnike modeliranja z digitalnimi valovnimi vodili [23]. Ker valovi potujejo po istem mediju v obe smeri, moramo za pravilno modeliranje potujočega vala uporabiti vsaj dve zakasnitveni liniji. Če želimo modelirati hkrati horizontalna in vertikalna tranzverzalna ter longitudinalna valovanja, lahko uporabimo tudi več zakasnitvenih linij. Vsaka komponenta valovanja zahteva svojo linijo.

### 3.1 Modeliranje resonatorjev

Za implementacijo klavirske strune z digitalnimi valovnimi vodili je potrebno najprej določiti aproksimacijo modela strune, prikazane na sliki 3.1. Skratka, imamo točko, v kateri struno vzbudimo (s klavivcem). Signal se nato prenaša levo in desno po struni. Ko signal pride do tega terminala, se odbije in potuje nazaj. Med potjo se signal zaduši in popači - te pojave skušamo čim boljše modelirati s refleksijskim filtrom  $H_r$ . Nekaj energije signala se absorbira med potjo zaradi notranje torzije in upora, del signala pa se prenese na resonančno ploščo in od tod pride večina zvočnega valovanja do našega ušesa [1],[3].

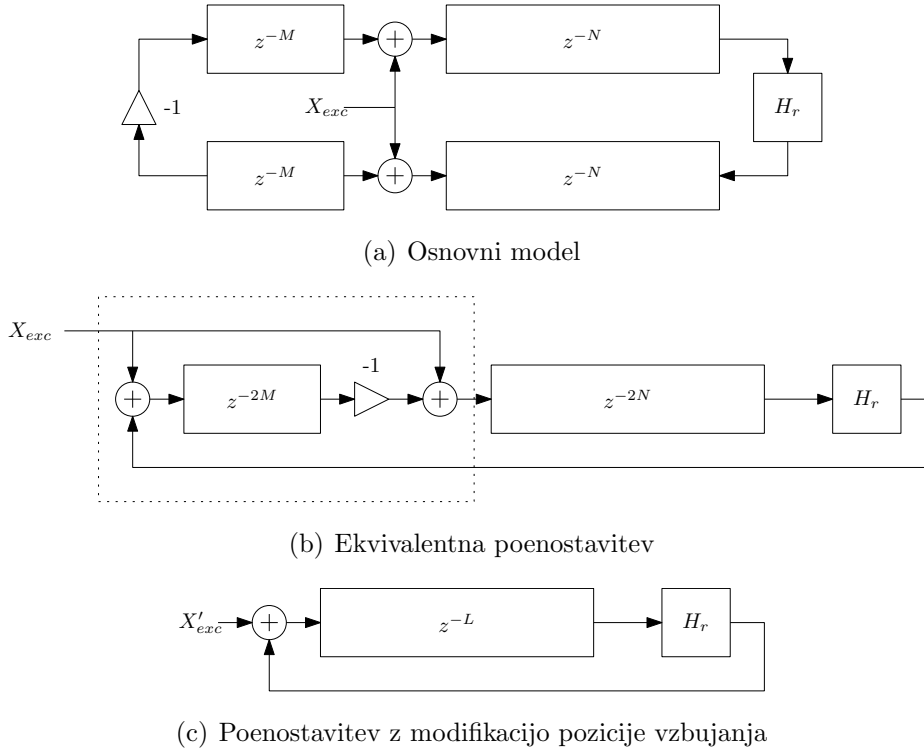


Slika 3.1: Prerez mehanizma klavirja

To abstrakcijo prikazuje slika 3.2(a). Slika 3.2(b) pa prikazuje njegovo skoraj enakovredno poenostavitev. Poenostavitev je mogoča, ker imamo opravka z linearno časovno invariantno (LTI) aproksimacijo sistema. Sledi formalni dokaz:

$$\begin{aligned} Y(z) &= z^{-N}(X_{exc}(z) + z^{-M}(-z^{-M}(X_{exc}(z) + z^{-N}(H_r Y(z)))))) \\ &= z^{-N}X_{exc}(z) - z^{(-2M-N)}X_{exc}(z) - z^{(-2M-2N)}H_r Y(z) \end{aligned} \quad (3.4)$$

$$H(z) = z^{-N} \frac{1 - z^{-2M}}{1 + H_r z^{(-2M-2N)}} \quad (3.5)$$



Slika 3.2: Model linijskega resonatorja z valovnim vodilom

Ker ne potrebujemo odvečne zakasnitve  $z^{-N}$ , jo lahko zapišemo kot

$$H(z) = \frac{1 - z^{-2M}}{1 + H_r z^{(-2M-2N)}}. \quad (3.6)$$

Funkcijo nato prevedemo v obliko

$$\begin{aligned} Y(z) &= X_{exc}(z) - z^{-2M} X_{exc}(z) - H_r(z) z^{(-2M-2N)} \\ &= X_{exc}(z) - z^{-2M} (X_{exc}(z) + H_r(z) z^{(-2N)}), \end{aligned} \quad (3.7)$$

ki je ravno funkcija sistema na sliki 3.2(b). Ta poenostavitev nam združi dve zakasnitveni liniji, kar delno olajša preračunavanje.

Naslednja poenostavitev, ki je prikazana na sliki 3.2(c), je osnovana na modifikaciji pozicije vzbujanja strune. Če ustrezno transformiramo vzbujevalni vir  $X_{exc}$ , tako, da aditivno komponento zakasnitvene linije ( $z^{-2M}$ ) vključimo v vzbujanje, dosežemo ekvivalentni odziv modela na sliki 3.2(b), ampak tudi nekoliko nižjo računsko zahtevnost resonatorja. Sedaj linijski resonator bistveno bolj spominja na rekurenčno tabelo pri KS algoritmu. V resnici je ta model strune hkrati osnovni model EKS algoritma.

Ton oz. osnovno frekvenco nihanja določa celotna dolžina resonatorja. Najtežjo nalogo nam predstavlja načrtovanje *refleksijskega* filtra  $H_r$ , ki ga opišemo kot kaskadni sistem treh filtrov, dušilnega ( $H_a$ ), disperzijskega ( $H_d$ ) in uglasovalnega filtra ( $H_t$ ).

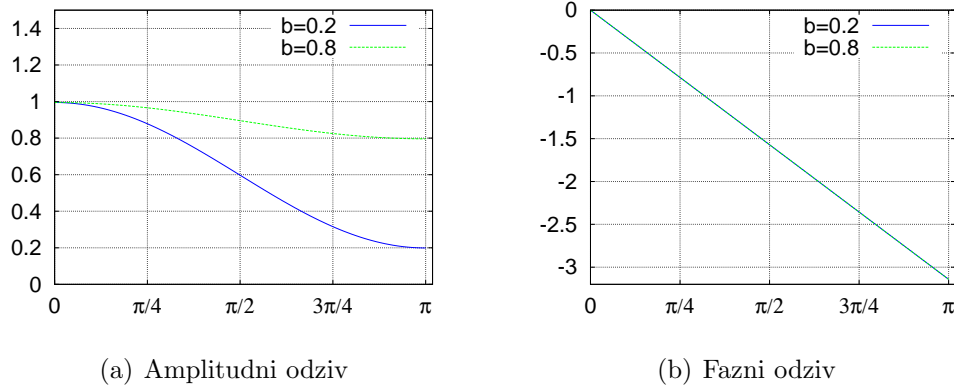
$$H_r = -H_a H_d H_t. \quad (3.8)$$

### 3.1.1 Dušilni filter

Zven strune s časom pojenja, zato potrebujemo filter, ki opisuje dušenje signala znotraj resonatorske zanke. Podobno kot pri Karplus-Strong algoritmu si tu pomagamo s FIR filtrom, ker ima linearno fazo in zato ne vnaša frekvenčno odvisnih zakasnitev v sistem. V osnovni implementaciji sem uporabil FIR filter 2. reda s prevajalno funkcijo

$$H_a(z) = \frac{g}{4} \left( (1-b) + (2+2b)z^{-1} + (1-b)z^{-2} \right), \quad (3.9)$$

kjer koeficient  $b$  predstavlja svetlost zvoka (*angl. brightness*) in obsega vrednosti med 0 in 1. Koeficient  $g$  pa določa dušenje in je običajno med 0.9 in 1.0. Slika 3.3 prikazuje frekvenčni odziv za primer, ko imamo  $b = 0.2$  oz.  $b = 0.8$ .



Slika 3.3: Dušilni filter za  $b = 0.2$  oz.  $b = 0.8$

### 3.1.2 Filter za fino uglasovanje ali ‘*fine-tuning*’

Pri linijskem resonatorju z valovnim vodilom se pojavi problem natančne uglasitve. Dolžina zakasnitvene linije je lahko le celoštevilski, kar pomeni, da lahko

struno ‘uglasimo’ le na določene frekvence. Ena rešitev tega problema bi bila povečanje frekvence vzorčenja, katera pa zahteva več procesorske moči in še vedno ni najbolj natančna. Visoko natančnost uglasitve potrebujemo predvsem pri tonih, ki so sestavljeni iz dveh ali treh strun, kajti strune istega tona so razglašene za približno 0.1% srednje frekvence. Sedaj se pojavi vprašanje, kako zakasniti linijo resonatorja za racionalno število vzorcev? Eno rešitev za to težavo nam ponuja linearna interpolacija vzorcev, ki vnaša dodatno dušenje v sistem. Drugo pa družina t.i. **Thiranovih** vseprepustnih filtrov [27].

Družina Thiranovih filtrov ima posebno lastnost enake skupne zakasnitve pri nizkih frekvencah. Ker gre za vseprepustni filter, je amplitudni odziv povsod enak 1. Prevaljalna funkcija Thiranovega filtra  $N$ -tega je določena kot

$$H_t(z) = \frac{z^{-N}A(z^{-1})}{A(z)} = \frac{a_N + a_{N-1}z^{-1} + \dots + a_1z^{-(N-1)} + z^{-N}}{1 + a_1z^{-1} + \dots + a_{N-1}z^{-(N-1)} + a_Nz^{-N}}, \quad (3.10)$$

koeficiente  $a_k$  filtra pa določimo z

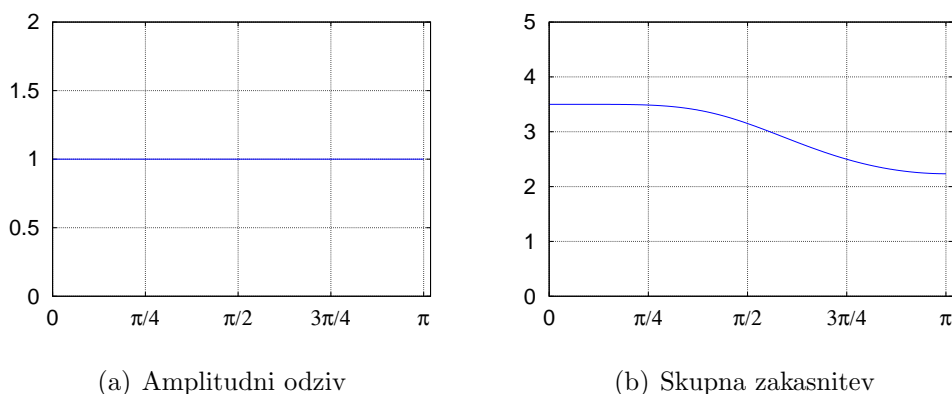
$$a_k = (-1)^k \binom{N}{k} \prod_{n=0}^N \frac{\Delta - N + n}{\Delta - N + k + n}, \quad k = 0, 1, 2, \dots, N. \quad (3.11)$$

Vendarle ima filter tudi pomanjkljivosti. Pri višjih frekvencah skupna zakasnitev pada, vendar to ne predstavlja težave, saj nas v glavnem zanima natančna uglasitev blizu osnovne frekvence tona. Visoki harmonski večkratniki (nad 6 kHz) so običajno zadušeni in nas tu ne zanimajo. Druga pomanjkljivost je minimalna zakasnitev filtra, ki je enaka redu filtra ( $N$ ). To moramo seveda upoštevati in ustrezno skrajšati zakasnitveno linijo resonatorja. Thiranov filter 3. reda je dovolj natančen za fino uglasjevanje strun.

Slika 3.4 prikazuje lastnosti Thiranovega filtra 3. reda, ki ima določeno zakasnitev 0.5 vzorca, kar skupno znaša 3.5 vzorca. Razvidno je, da ima filter takšno zakasnitev vsaj do frekvence  $\pi/4$ , kar ustreza frekvenci 6 kHz pri frekvenci vzorčenja 48000 kHz. Nato zakasnitev pada in se ustali pri vrednosti 2.2.

### 3.1.3 Disperzijski filter in disharmoničnost

Zaradi fizikalnih lastnosti neidealne strune, višje frekvence potujejo hitreje od nižjih, kar je opazno zlasti pri višjih harmonskih frekvencah. Te niso pravi

Slika 3.4: Thiranov filter 3. reda,  $\Delta = 3.5$ 

večkratniki osnovne frekvence tona. Višje harmonske frekvence dobimo z enačbo

$$f_k = k f_0 \sqrt{1 + Bk^2}, \quad (3.12)$$

kjer je  $f_0$  osnovna frekvenca tona,  $B$  koeficient disharmoničnosti in  $k$  večkratnik frekvence. Koeficient disharmoničnosti določa, koliko frekvenca višje harmonske komponente odstopa od čistega večkratnika, in je povsem odvisna od fizikalnih lastnosti strune, obsega pa vrednosti od 0.0001 do 0.001.

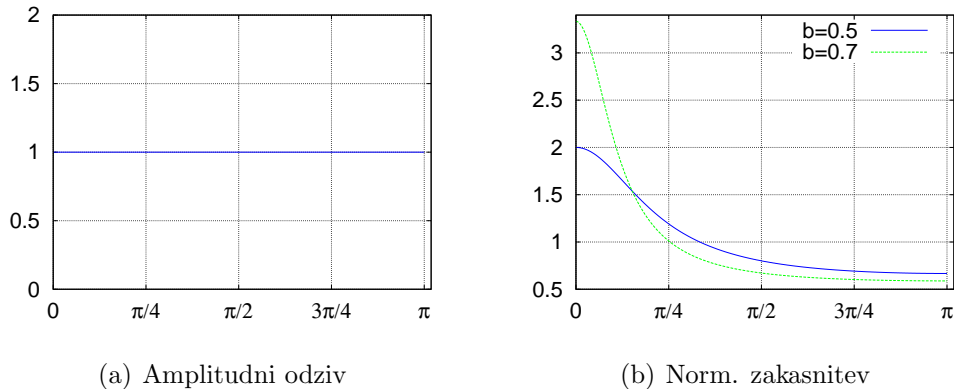
S stališča celotnega sistema, predstavlja načrt disperzijskega filtra določitev frekvenčno neenakomerne skupne zakasnitve resonatorske linije. Hitro ugotovimo, da imamo opravka s vseprepustnim filtrom, kjer želimo doseči določeno fazno popačenje sorazmerno z disperzijo na posamezni frekvenci.

V literaturi je predlagana rešitev iz več kaskadno povezanimi vseprepustnimi filtri 1. reda [9]. Kot bomo videli, en sam filter 1. reda že dovolj dobro aproksimira disperzijske lastnosti, vsaj s stališča percepcije zvoka. Za natančnejše prilaganje disperzijski krivulji, opisani v enačbi 3.12, lahko vedno uporabimo vseprepustne filtre višjega reda ali kaskadno vezavo enostavnih filtrov. Predlagani filter ima prevajalno funkcijo

$$H_d(z) = \frac{k_B + z^{-1}}{1 + k_B z^{-1}}, \quad (3.13)$$

kjer je  $k_B$  je koeficient, s katerim naravnamo disperzijo. Slika 3.5(a) prikazuje amplitudni odziv. Na sliki 3.5(b) so prikazane normalizirane<sup>1</sup> zakasnitve za različne

<sup>1</sup> zakasnitev filtra, zamaknjena za vrednost maksimalne zakasnitve

Slika 3.5: Disperzijski filter za  $k_d = -0.4$  oz.  $k_d = -0.5$ 

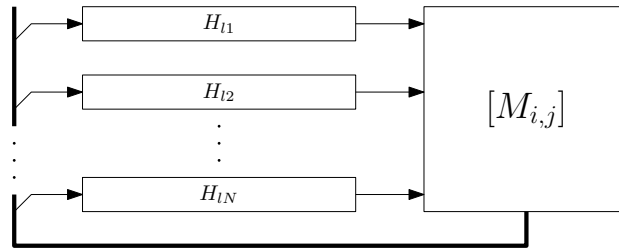
vrednosti koeficientov  $k_B$  in zelena zakasnitev za klavirski ton s frekvenco  $440\text{Hz}$  (črtkano), ki jo izračunamo na podlagi enačbe 3.12.

### 3.1.4 Priklopni oz. (*angl. coupling*) filter

Če imamo ton sestavljen iz več linijskih resonatorjev, potrebujemo tudi priklopni filter. Ta poskrbi za interakcijo med resonatorji in ne prispeva bistveno k zvoku inštrumenta. Kljub temu ga je potrebno upoštevati, predvsem, če želimo modelirati naprednejše lastnosti glasbila, kot so medsebojna resonanca strun in prehajanje energije signala iz horizontalne v vertikalno polarizacijo, ki ustvarajajo očitne perceptivne učinke, kot je npr. večstopenjsko upadanje [20]. Na primer, pri pritisku *sostenuto* pedala na klavirju je preklopni filter še kako pomemben, saj so vse strune aktivne<sup>2</sup> in resonirajo med seboj. Rezultirajoči zvočni učinek je podoben bogatemu odmevu in je očiten izvajalcu, kakor tudi oddaljenemu poslušalcu z izostrenim sluhom.

Pri klavirju se energija valovanja strune prenaša po ‘mostičku’ do kovinske resonančne plošče. Plošča nato prenaša energijo do vseh strun, ki niso zadušene. Če pritisnemo na *sostenuto* pedal, se na vzbujanje odzovejo vse strune, nekatere več, nekatere manj, odvisno od ujemanja harmonskih komponent v amplitudnem odzivu strun. Ta filter se lahko implementira v obliki utežene rekombinacije izhodov posameznih linijskih resonatorjev v vhod ostalih, prek množenja s simetrično ma-

<sup>2</sup>*Sostenuto* pedal odklopi mehanizem dušenja strun za vse tone, zato so strune aktivne in se odzovejo zunanjemu vzbujanju



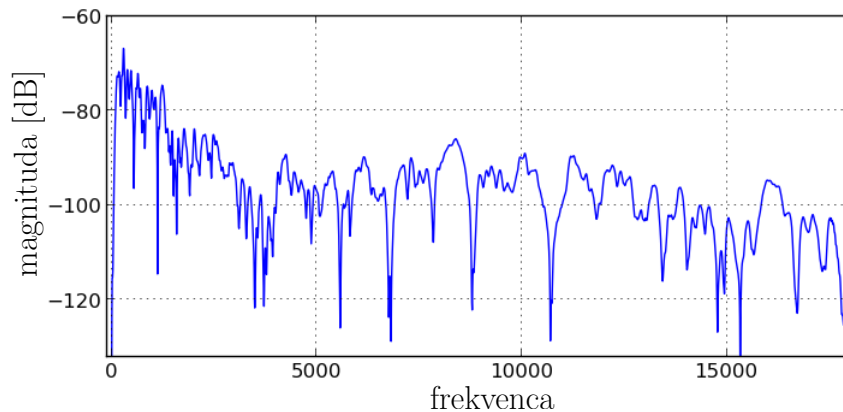
Slika 3.6: Priklopni filter

triko, kjer elementi matrike  $m_{ij}$  predstavljajo prispevek  $i$ -tega resonatorja  $j$ -temu (slika 3.6). Če imamo sistem z nizko stopnjo linijskih resonatorjev, je ta način modeliranja primeren za implementacijo sistema v realnem času. Na primer, pri 6-strunski kitari zadostuje 12 do 18 linijskih resonatorjev. Zato je tovrsten pristop ugoden za modeliranje medsebojne resonance. Pri klavirju se število strun giblje tudi do 200, saj ima vsak ton lahko več strun. Če upoštevamo tudi longitudinalne valove imamo torej 600 resonatorjev in matriko dimenzije  $600 \times 600$ . Ker je ta pristop neugoden za računanje, se v tem primeru medsebojno resonanco modelira z drugačnimi metodami [2],[24].

## 3.2 Telesni filter

Vsako naravno glasbilo ima preklapljene tonske resonatorje na neko telo (ohišje), ki običajno tudi samo resonira tako, kot vsi ostali primitivni gradniki glasbila. S stališča modeliranja resonančnega filtra je glavna razlika telesa, v primerjavi z zgoraj omenjenimi resonatorji, njegova (več)dimenzionalnost. Pri klavirju je to lesena ali kovinska plošča, ki poskrbi za prenos energije valovanja v mediju v zračno valovanje, zato ji pravimo tudi resonančna plošča. Pri akustični kitari je to votel trup oz. telo kitare.

Če ima telesni resonator izrazito linearno karakteristiko, ga lahko dovolj natančno aproksimiramo z LTI sistemom, ki ga bomo označevali kot *telesni filter*  $H_B$ . Ponavadi ga dobimo iz meritev odziva telesa oz. ohišja na enotin impulz, ki ga z ustrežno numerično aproksimacijo predstavimo kot FIR filter višjega reda. Slika 3.7 prikazuje izmerjen frekvenčni odziv na enotin impulz klavirske resonančne plošče.



Slika 3.7: Frekvenčni odziv klavirske resonančne plošče

Žal, na primeru klavirja ne moremo modelirati resonančno ploščo kot en sam filter, saj so meritve pokazale, da se frekvenčni odziv na enotin impulz razlikuje glede na lokacijo priklopa strune. Te razlike niso zanemarljive. Model telesnega filtra moramo določiti za skupine tonov ali celo za vsak klavirski ton posebej. Za natančno obliko frekvenčnega odziva telesnega filtra imamo opravka s filtri reda okoli 1000 in več. Ker je konvolucija v času razmeroma potratna operacija, lahko telesni filter združimo z vzbujevalno komponento, v kolikor vzbujevalniku sledijo samo komponente brez nelinearnih elementov. To je konkreten primer sintetizacije s komutacijo valovnih vodil [6].

### 3.3 Modeliranje vzbujanja

Eden ključnih vzrokov za razliko v zvoku kitare in klavirja je zagotovo način vzbujanja. Vemo, da se signal prenaša naprej in nazaj po valovnem vodu, vendar moramo signal nekje ustvariti. Vzbujevalnik je tisti element, ki dovaja sistemu energijo, zato je smiselno posvetiti posebno pozornost le-temu.

Pri večini glasbil je vzbujanje nek nelinearen sistem, ki je povratno povezan z resonatorjem. Ta aspekt je v marsikaterem pogledu problematičen s stališča načrtovalca sistema. Pri klavirju, na primer, bi morali upoštevati celoten čas vzbujanja strun in na podlagi meritev opisati stabilen sistem z diferenčnimi enačbami. Vendar je posledica natančnega načrtovanja takega sistema višja računaska zahtevnost. Poleg tega je končni izplen nizek, saj je npr. pri klavirju

trajanje vzbujanja razmeroma kratko in daje potujočemu valu le začetno obliko, ki določi katere frekvence bodo ojačane in katere ne, kar seveda lahko modeliramo tudi na druge načine.

Drugačen pristop k modeliranju vzbujanja je uporaba subtraktivne sinteze (poglavje 2.4.2), ki nikakor ne spada v fizikalno modeliranje. Ampak ima kljub temu veliko prednost, saj lahko s to metodo enostavno modeliramo skoraj poljubni način vzbujanja, tudi za glasbila, ki imajo dinamično vzbujanje, npr. pihala, trobila in godala. Vse kar potrebujemo je izvor šuma ali serije impulzov, spektralno ovojnico nizkega reda in amplitudno ovojnico.

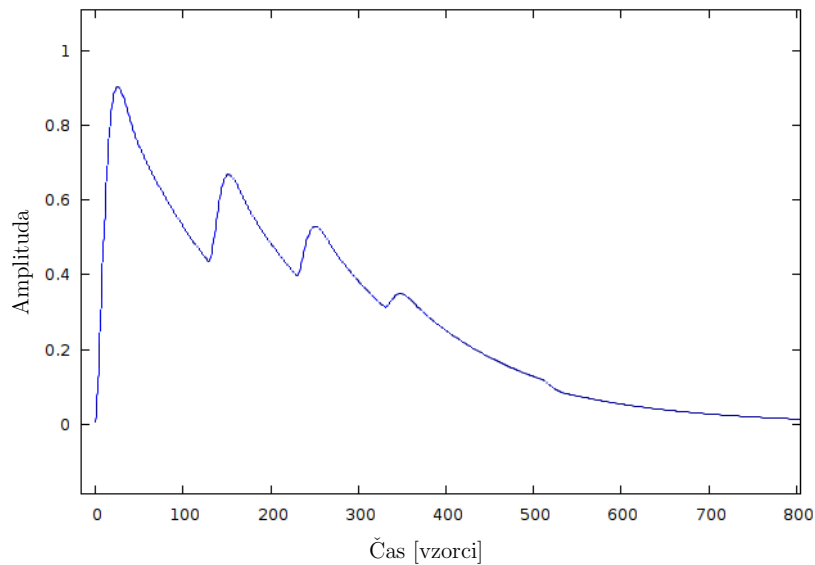
### 3.3.1 Vzbujanje z vzorčenjem

Če za implementacijo vzbujevalnika uporabimo kar vzorce konvolucije odziva na enotin impulz vira energije (npr. klavir) in telesnega filtra, imamo opravka s strukturnim vzorčenjem. Strukturne vzorce (konvolucijo odzivov) lahko izmerimo in izračunamo, lahko pa analitično določimo s kakšnim od postopkov dekompozicije opisanih v prejšnjem poglavju. Ta pristop sicer ne spada v fizikalno modeliranje, vendar je rezultat tega postopka izjemno kvalitetna perceptivna reprodukcija zvoka, kar lahko opazimo glede na priljubljenost tovrstne tehnike sintetizacije v primeru *komutiranega klavirja* [6].

### 3.3.2 Subtraktivni model

Kot osnovo implementacije je uporabljen model, ki je bil zasnovan [6] s filtriranjem (serije) impulzov - ta je določen v glavnem iz izmerjenih podatkov, časovni odziv je prikazan na sliki 3.8. Ta model je sicer uporaben, vendar je moč zaznati občutno pomanjkanje 'stohastičnih' oz. šumnih komponent v zvenu vzbujene strune. S stališča sinteze zvoka, je ta pristop ekvivalenten subtraktivni sintezi, zato smo poskusili uporabiti bolj splošen subtraktivni model, z nekoliko več možnostmi parametrizacije. Tako smo vključil tri osnovne gradnike subtraktivnega vzbujanja:

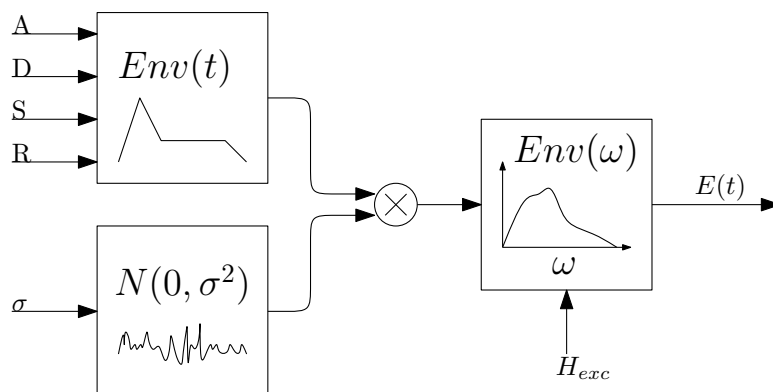
- naključni generator Gaussovega šuma,
- ovojnico za moduliranje amplitude ( $a(t) = e^{-\lambda t}$ ) in



Slika 3.8: Amplituda kladivca pri osnovnem modelu

- spektralno ovojnico oz. srednjeprepustni filter.

Model filtriranega šuma (slika 3.9) ni nobena novost v modeliranju vzbujanj digitalnih valovnih vodil. Izrazito se uporablja npr. pri pihalih. Poleg tega je tudi KS algoritem, kot predhodnik DWG, imel prisoten filtriran šum za oblikovanje začetne tabele. Z uporabo ADSR ovojnice s samo štirimi parametri (A) čas vzpona, (D) čas upada, (S) nivo mirovanja, (R) čas zvena, lahko modeliramo amplitudni potek bogate množice glasbil s strunami.



Slika 3.9: Subtraktivni model vzbujanja

### 3.4 Analitično določanje vrednosti parametrov

Kalibracijo z optimizacijo kriterijskih funkcij, ki bo opisana v naslednjem poglavju, lahko bistveno izboljšamo, če za začetne vrednosti uporabimo take parametre, ki že v osnovi minimizirajo perceptivno razdaljo. Začetne vrednosti lahko vedno empirično določimo, kar je sicer zamudno opravilo, še posebej, če je veliko možnih kombinacij parametrov in veliko število zvočnih signalov, ki jih želimo reproducirati. Zato, če je le mogoče, parametre določimo analitično. Nekatere lahko določimo direktno iz spektralnih lastnosti, npr. koeficiente filtrov za dušenje, disperzijo, uglaševanje. Druge lahko določimo na podlagi ustrezne dekompozicije signala, npr. parametre vzbujevalnih elementov. Vse ostale parametre moramo določiti empirično.

#### 3.4.1 Določitev koeficientov disperzije

Koeficient disharmoničnosti  $B$  lahko izračunamo iz meritev spektralnih vrhov, ki ustrezajo večkratnikom osnovne frekvence  $f_0$ . Enačbo za disharmoničnost strune (3.12) preoblikujemo tako, da dobimo inverz:

$$\hat{B}_k = \frac{\frac{f_k}{kf_0} - 1}{k^2}. \quad (3.14)$$

Iz tega lahko izračunamo zadostno oceno za koeficient  $B$ , če vzamemo povprečje izmerjenih koeficientov na posamezni komponenti,

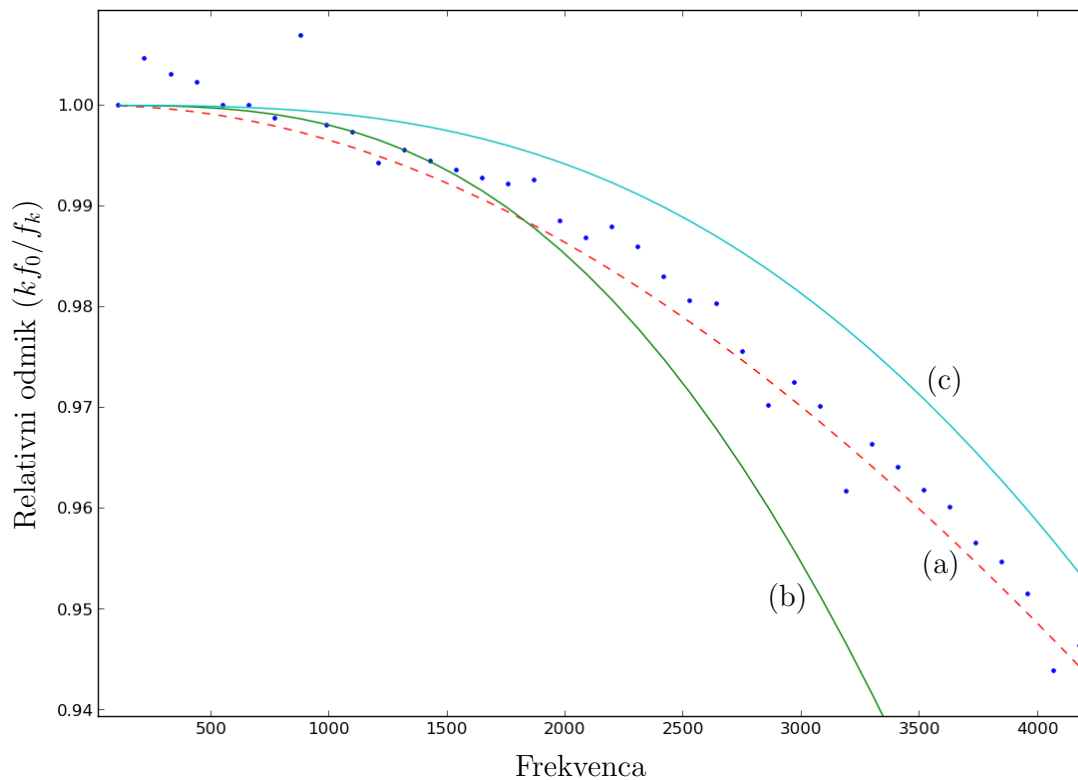
$$B = \frac{1}{N-1} \sum_{k=1}^N \hat{B}_k. \quad (3.15)$$

1.	$k_1, k_2, k_3 = (-0.002658, -0.014811, -2.9018)$
	$C_1, C_2 = (0.071089, 2.1074)$
2.	$K_d(B) = \exp(k_1 \ln(B)^2 + k_2 \ln(B) + k_3)$
3.	$C_d = \exp(C_1 \ln(Bc) + C_2)$
4.	$I_{\text{key}}(f_0) = \frac{\ln(f_0 \sqrt[12]{2}) - \ln(27.5)}{\ln(\sqrt[12]{2})}$
5.	$D = \exp(C_d - I_{\text{key}}(f_0) K_d)$
6.	$k_B = -(1 - D)/(1 + D)$

Tabela 3.1: Postopek za izračun disperzijskega filtra [16].

Sedaj, ko imamo določen koeficient inharmoničnosti, lahko določimo koeficiente disperzijskega filtra [16]. Postopek je podan v tabeli 3.1, kjer je  $B$  koeficient disperzije.

Graf na sliki 3.10 prikazuje točke in krivulje odstopanja harmonskih komponent od čistega večkratnika  $kf_0/f_k$  na zvočnem posnetku klavirja z osnovno frekvenco 110 Hz, kjer je izmerjen koeficient disharmoničnosti  $B$  enak  $8.3 \times 10^{-5}$ . Po postopku 3.1 dobimo koeficient  $k_B = -0.6372$  za vseprepustni filter 1. reda s prevajalno funkcijo iz enačbe 3.13. Ocenjena krivulja disharmoničnosti (enačba 3.12) sledi izmerjenim točkam, kar ustreza fizikalnemu modelu. Frekvenčno odstopanje analitično izračunanega vseprepustnega filtra pa nekoliko odstopa od želene vrednosti pri višjih frekvencah. To odstopanje je dopustno, vsaj s perceptivnega stališča, ker je resolucija človeškega sluha pri višjih frekvencah precej slabša od resolucije pri nizkih frekvencah.



Slika 3.10: Relativno odstopanje harmonskih komponent; točke izmerjenih odstopanj, ocenjena krivulja disharmoničnosti (a), umerjeni filter  $k_B = 0.6372$  (b) in filter s  $k_B = 0.5$  za primerjavo (c).

Vseprepustni filter za disperzijo ima pri osnovni frekvenci  $f_0$  zakasnitev

$$\begin{aligned}\tau_B(f_0) &= -\frac{d}{d\omega} \angle H_{ap}(e^{j\omega}) \\ &= \frac{1}{\omega T} \left( \tan^{-1} \frac{\sin(\omega_0 T)}{k_B + \cos(\omega_0 T)} - \tan^{-1} \frac{\sin(\omega_0 T)}{(1/k_B) + \cos(\omega_0 T)} \right),\end{aligned}\quad (3.16)$$

kjer je  $\omega_0 T = 2\pi f_0 / f_s$ .

### 3.4.2 Določitev osnovne periode

Osnovna frekvenca tona določa periodo valovanja  $\tau_0 = f_s / f_0$ , in posledično tudi dolžino resonatorja oz. zakasnitvene linije. Vsak filter, ki nima linearne faze, prispeva k zakasnitvam znotraj resonatorske zanke. To zakasnitev moramo kompenzirati z ustreznim skrajšanjem zakasnitvene linije. Dušilni filter (3.9) ima linearno fazo, zato ne prispeva k faznim zakasnitvam. Brez upoštevanja ostalih elementov, ki v sistem vnašajo dodatne zakasnitve, izračunamo dolžino po formuli

$$L_0 = \lfloor \tau_0 \rfloor = \lfloor f_s / f_0 \rfloor. \quad (3.17)$$

Periodo oz. dolžino resonatorja dobimo tako, da od teoretične dolžine odštejemo skupne zakasnitve ostalih filtrov v sistemu,

$$\tau'_{loop} = \lfloor f_s / f_0 - \tau_B(f_0) - \tau_t(f_0) \rfloor. \quad (3.18)$$

### 3.4.3 Določitev parametrov dušenja

Dušenje znotraj zanke resonatorja (3.9) oz. koeficienta  $g$  in  $b$  lahko določimo na podlagi upadanja amplitudnega odziva v času [20]. Postopek je sledeč:

1. Izračunamo STFT s Hammingovo okensko funkcijo.
2. Določimo amplitudni potek spektralnih vrhov (s sinusoidnim modeliranjem).
3. Za vsako sinusoidno komponento določimo eksponentno funkcijo prilaganja. V logaritemskem merilu se ta operacija prevede na linearno interpolacijo.
4. Iz naklonov interpolacijskih premic (posameznih sinusoidnih komponent) določimo celoten amplitudni odziv filtra za dušenje.

5. Z uporabo Prony-eve metode ali Steiglitz-McBride algoritma določimo IIR filter N-tega reda, ki aproksimira amplitudni odziv.

Ker dušenje znotraj naše linije ustvarja FIR filter 2.reda s prevajalno funkcijo 3.9, lahko uporabimo Remezov algoritem za načrtovanje FIR filtrov. Pri tem nižje frekvence obtežimo nekoliko več zaradi lastnosti sluha. Parametra dušenja  $g$  in  $b$  dobimo direktno iz koeficientov dobljenega simetričnega filtra, kjer so  $\mu_i$  koeficienti FIR filtra s prevajalno funkcijo

$$\begin{aligned} H_a(z) &= \mu_0 + \mu_1 z^{-1} + \mu_2 z^{-2} \\ &= \frac{g}{4} ((1-b) + (2+2b)z^{-1} + (1-b)z^{-2}) \end{aligned} \quad (3.19)$$

$$g = \mu_1 + 2\mu_0 \quad (3.20)$$

$$b = \frac{\mu_1 - 2\mu_0}{\mu_1 + 2\mu_0}. \quad (3.21)$$

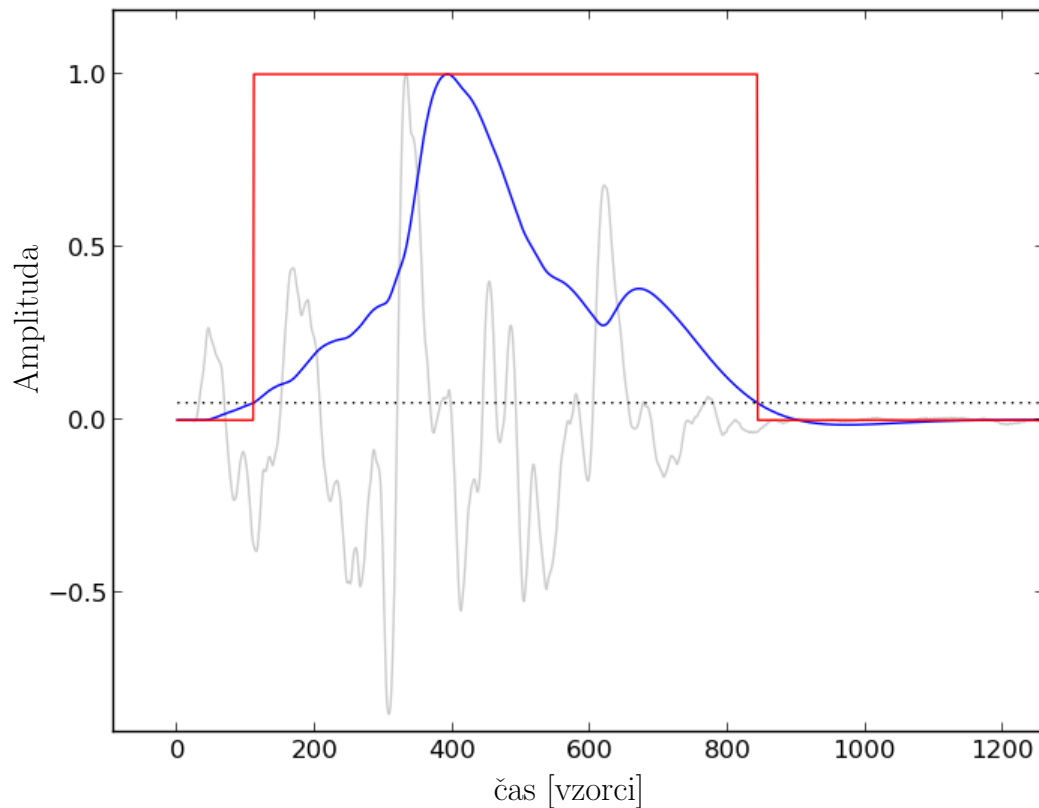
### 3.4.4 Določitev parametrov vzbujanja

Vzbujanje modela opisanega v poglavju 3 temelji na stohastičnem procesu, zato je smiselno uporabiti tehnike spektralnega modeliranja, kot so deterministično-stohastična dekompozicija ali linearna predikcija za izločanje aditivnih oz. konvolucijskih šumnih komponent.

Slika 3.11 prikazuje izluščen signal klavirca pri glasnejšem tonu s frekvenco 110 Hz. Izračunana vrednost za časa padca vzbujanja je 0.018 ms. Meja aktivnega vzbujanja je (grobno) določena na 5% vrednosti jakostne amplitude. Na podlagi točke v času, ko amplituda doseže to mejno vrednost, lahko določimo čas upadanja vzbujanja.

### 3.4.5 Določitev preostalih vrednosti parametrov

Preostale parametre določimo na podlagi empirične ocene pri posameznih zvočnih signalih. S tem določimo tudi območje iskanja, v kolikor želimo uporabiti avtomatsko kalibracijo.



Slika 3.11: Šumne komponente vzbujanja, amplitudna ovojnica (modro), ocena trajanja aktivnega vzbujanja (rdeče)

### 3.5 Predstavitev rezultatov

V tem poglavju je bil predstavljen primer modeliranja z digitalnimi valovnimi vodili, na primeru klavirskega tona. Poglavitna razlika med tem načinom modeliranja in ostalimi tehnikami sintetizacije je v:

1. razumljivosti in analogije s fizikalnim modelom in
2. sposobnosti kvalitetne reprodukcije zvoka (v perceptivnem smislu).

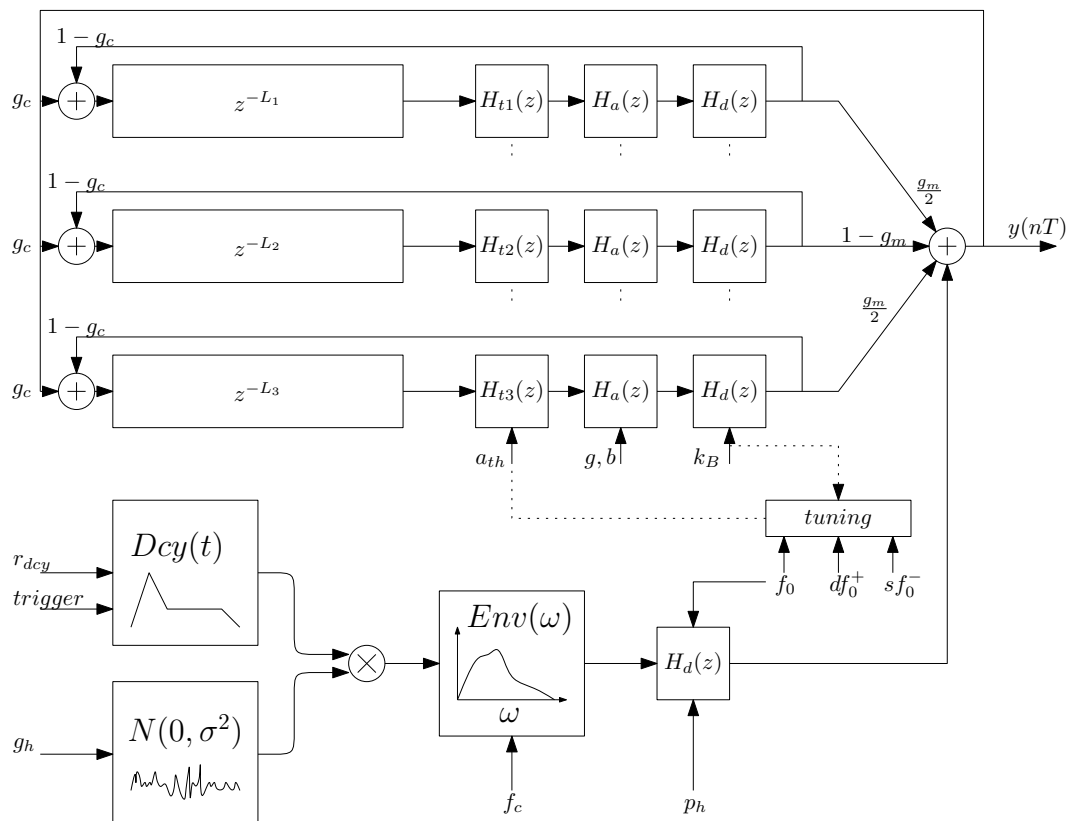
Na sliki 3.12 je prikazan celoten DWG model sintetizacije, ki je prilagojen simulaciji klavirske strune. Sintetizator ima 12 nastavljivih parametrov, ki so prikazani v tabeli 3.5.

---

$b$	-	koeficient svetlosti
$k_B$	-	koeficient disperzije
$df_0^+$	-	faktor pozitivne razglasitve pomožne strune
$df_0^-$	-	faktor negativne razglasitve pomožne strune
$f_0$	-	osnovna frekvenca
$g$	-	koeficient dušenja
$D_h$	-	čas upadanja vzbujanja
$g_h$	-	jakost vzbujanja
$f_c$	-	frekvenca nizko prepustnega filtra za vzbujanje
$p_h$	-	filter za pozicijo točke vzbujanja
$g_c$	-	jakost povratne vezave
$g_m$	-	razmerje jakosti mešanja signalov strun

---

Tabela 3.2: Parametri sintetizatorja



Slika 3.12: Shema celotnega sistema

V uvodu je bilo že predstavljeno, da je predpogoj realizacija sintetizatorja sposobnega izvajanja v realnem času. Tabela 3.3 prikazuje rezultate izvajanja pri različnem številu aktivnih strun. Merili smo št. ciklov procesorja za bločno izvajanje DSP zanke pri velikosti medpomnilnika velikosti 2048 vzorcev. Groba ocena časovne zahtevnosti je pribl. 85 ciklov + 60 ciklov/struno. Sintetizator

Št.strun	ciklov	ciklov/vzorec
1	174440	85
3	534376	260
5	877688	428
10	1067584	521
50	6671272	3257

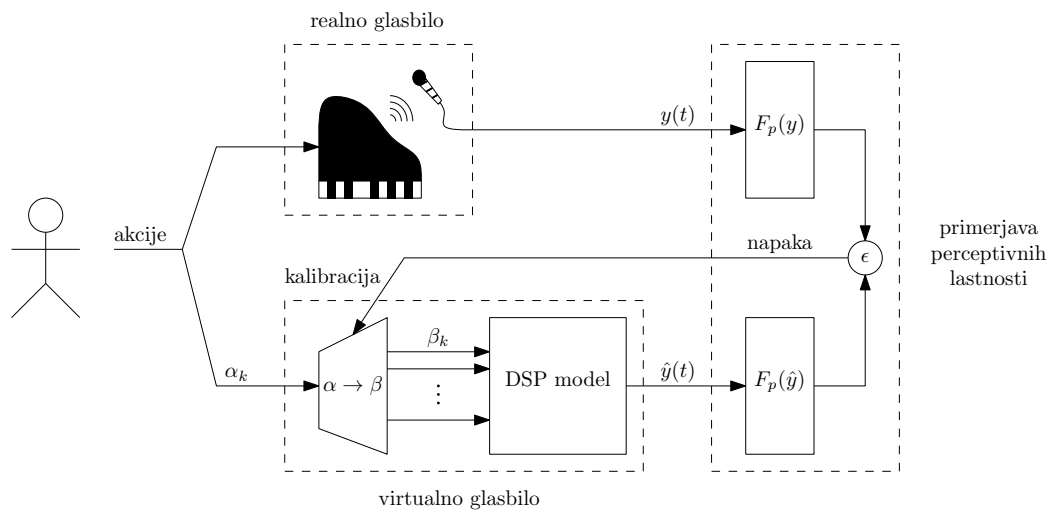
Tabela 3.3: Računska zahtevnost izvajanja v realnem času

ima praktično zanemarljivo porabo procesorskega časa in pomnilnika, saj pri hkratnem izvajanju 50 strun obremeni manj kot 5% enojedernega procesorja (s taktom 2 GHz). Tudi poraba pomnilnika je zanemarljiva.

# Poglavje 4

## Kalibracija modela

Sedaj, ko imamo postavljen model, definirane nadzorne parametre modela ter določene vrednosti parametrov se pojavi vprašanje, ali lahko poiščemo boljše približke parametrov, da bo sintetični zvok čim bolj podoben originalnim vzorcem. Iščemo torej take parametre, ki minimizirajo perceptivno razdaljo med lastnostmi ciljnega zvoka in lastnostmi iskanega zvoka. Metrika oz. razdalja značilk določa obliko naše kriterijske funkcije. Za iskanje takega nabora parametrov, ki minimizira kriterijsko funkcijo, se uporabljajo algoritmi za optimizacijo. Celoten



Slika 4.1: Kalibracija parametrov

postopek avtomatske kalibracije prikazuje slika 4.1:

1. Izvajalec zaigra skupino tonov, ki jih želimo reproducirati. Vsako izva-

jalčevo akcijo  $\alpha_k$  zabeležimo (shranimo), tako kot ustrezni zvočni signal  $y(t)$  iz realnega glasbila.

2. Iz zvočnih signalov izluščimo perceptivne lastnosti in jih primerjamo med seboj. Na podlagi napake oz. razdalje med perceptivnimi lastnostmi optimiziramo parametre modela  $\beta_k$  za trenutno akcijo  $\alpha_k$ .
3. Ko je perceptivna napaka manjša od zelene vrednosti <sup>1</sup> oz. po maksimalnem številu iteracij, postopek kalibracije zaključimo.
4. Za manjkajoče akcije, t.j. tone, kjer nimamo pripadajočega zvočnega signala, parametre ( $\beta_k$ ) intepoliramo in dobimo preslikavo  $\alpha \rightarrow \beta$ .

Prvi korak v kalibraciji parametrov je ustrezno predprocesiranje vzorcev. Glede na to, da vzorce primerjamo po frekvencah in času, jih moramo najprej časovno poravnati. To lahko naredimo tako, da zaznamo začetek vzorca in ga ustrezno premaknemo nazaj na časovno os. Začetek lahko določimo na enak način kot smo določili čas upadanja v prejšnjem poglavju. Le da tu vzamemo čas, kjer amplituda zraste nad 5% vrednosti (slika 3.11). Tako bosta ciljni zvočni signal in sintetični signal sinhronizirana. Tem bolj, ko bodo vzorci sinhronizirani, boljši bo rezultat optimizacije.

## 4.1 Kriterijske funkcije

Izbira najprimernejših kriterijskih funkcij je subjektivne narave. To pomeni, da moramo empirično določiti, katera funkcija najboljše aproksimira perceptivno podobnost zvočnih signalov. Izhodišče za izbiro funkcije je razumevanje delovanja človeškega slušnega sistema na fizikalnem in nekoliko tudi na kognitivnem nivoju. Za določanje podobnosti signala si lahko pomagamo z nekaterimi značilkami signala, ki z minimalno količino informacije dobro opišejo perceptivne lastnosti zvoka in so razmeroma enostavni za računanje. Metrika značilk (perceptivna razdalja) je običajno določena kot evklidska razdalja oz.  $L_2$ -norma perceptivnih

---

<sup>1</sup>Vrednost perceptivne napake empirično določimo, na ta način, da poiščemo mejo nezaznavnosti razlike v zvokih

značilka,

$$d(\mathbf{X}_a, \mathbf{Y}_a) = \sqrt{\left(\sum_{i=1}^n |X_a(i) - Y_a(i)|^2\right)}, \quad (4.1)$$

kjer sta  $\mathbf{X}_a$  in  $\mathbf{Y}_a$  vektorja značilke  $a$  za sintetični signal  $x$  oz. ciljni signal  $y$ . Seveda se lahko uporabi tudi druge metrike, npr. Manhattan normo ( $L_1$ ) ali normo Čebiševa ( $L_\infty$ ). Pomembno je, da metrike izpolnjujejo naslednje zahteve:

$$d(X, Y) \geq 0 \text{ (nenegativnost)} \quad (4.2)$$

$$d(X, Y) = 0, \text{ če velja } X = Y \text{ (identiteta)} \quad (4.3)$$

$$d(X, Y) = d(Y, X) \text{ (simetričnost)} \quad (4.4)$$

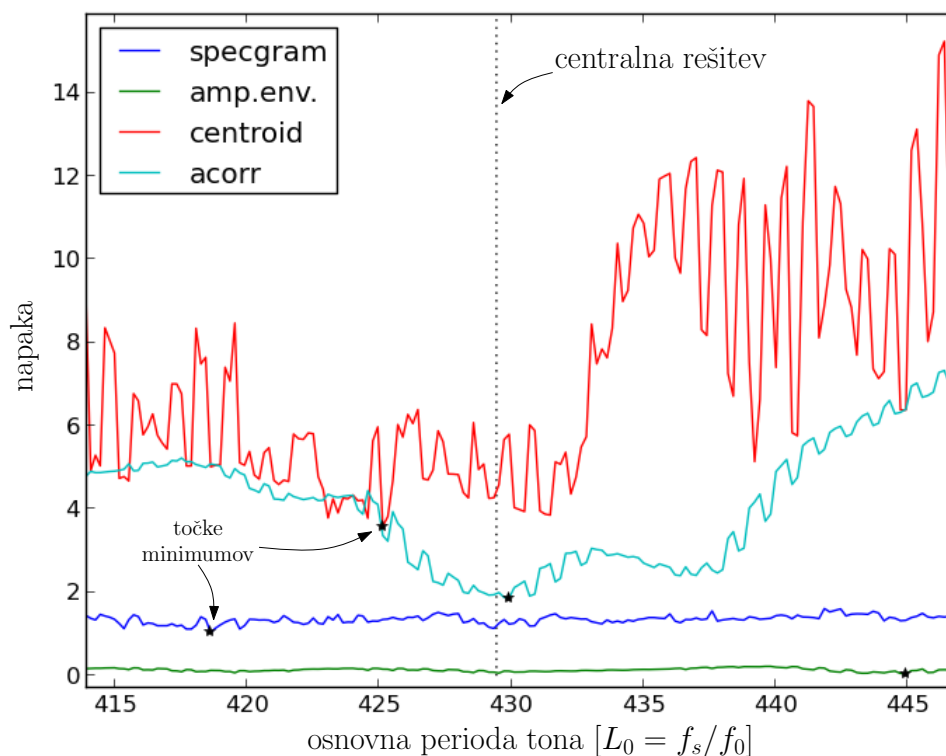
Zaželeno je tudi, da se izbrane metrike normalizira oz. ustrezno obteži. Tako se proces iskanja usmerja enakovredno glede na izbrane funkcije. Naslednje empirično izbrane razdalje značilk so bile uporabljene za optimizacijo in precej dobro opišejo časovno-frekvenčno obliko zvoka:

- Absolutna napaka logaritma spektrogramov
- Absolutna napaka amplitudnih ovojníc
- Absolutna napaka centroid
- Kvadratna napaka avtokorelacij<sup>2</sup>

Vse zgoraj navedene metrike lahko izračunamo neposredno iz spektrograma, kar bistveno skrajša čas izvajanja optimizacije. Sliki 4.2 in 4.3 prikazujeta primer odvisnosti vrednosti kriterijskih funkcij glede na parametra dolžine resonatorja (frekvenca tona) oz. disperzijskega koeficienta, kjer so ostali parametri empirično določeni. Minimumi posameznih funkcij so označeni z zvezdico. Najboljša vrednost parametrov<sup>3</sup> je označena s črtkano. Opazimo lahko, da imajo kriterijske funkcije veliko minimumov in ne nujno na istih mestih. Predvsem pri zvezni spremembi osnovne periode lahko opazimo neidealni učinek filtra za fino uglaševanje.

<sup>2</sup>Napaka avtokorelacij je vključena namesto jakostnega spektra, saj pri nizkih tonih nekoliko bolje določi osnovno frekvenco od spektra.

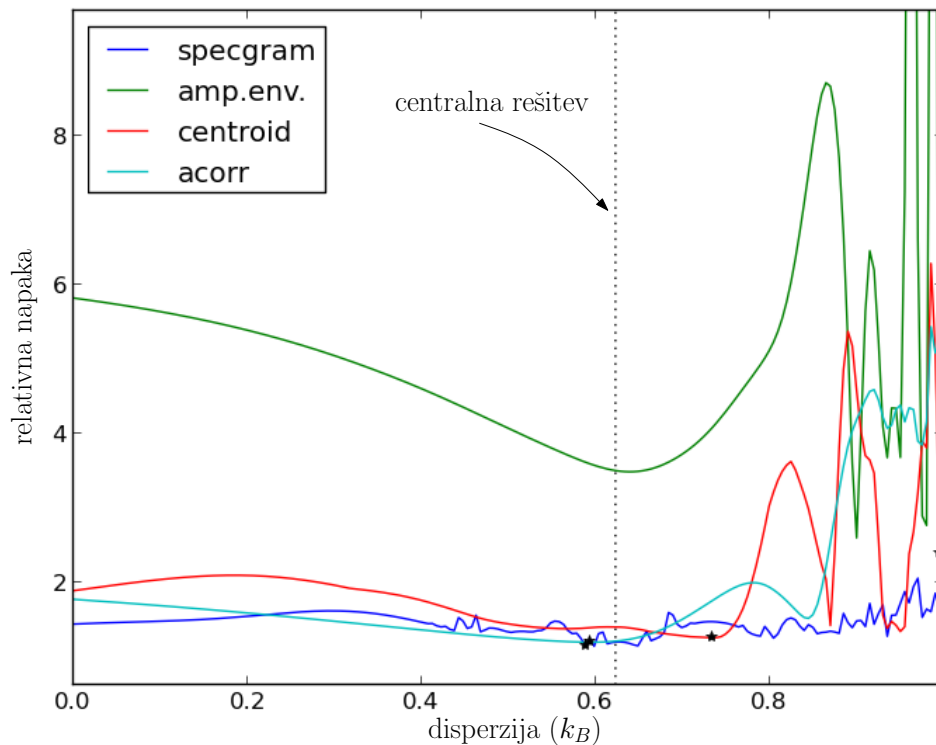
<sup>3</sup>Centralna vrednost enakomerno uteženih kriterijskih funkcij.



Slika 4.2: Odvisnost izbranih kriterijskih funkcij od dolžine resonatorja oz. osnovne periode tona

Obtežitev metrik bi lahko bil naslednji korak pri enokriterijski optimizaciji, ki bi omogočil pridobiti enotno kriterijsko funkcijo. Ker bomo imeli opravka z večkriterijsko optimizacijo, lahko ta korak (zaenkrat) izpustimo.

Sedaj, ko imamo določene metrike, je potrebno določiti tudi območje iskanja parametrov, kjer pričakujemo optimalno podobnost. Na ta način bo postopek optimizacije hitreje konvergiral k najboljši rešitvi. Lahko pa sprožimo iskanje po neomejenem prostoru parametrov. Vendar pri algoritmih optimizacije to predstavlja težavo, saj lahko ti obtičijo v lokalnem minimumu kriterijske funkcije. V našem primeru je prostor parametrov visoko dimenzionalen, npr. DWG model z dvema resonatorjema ima običajno več kot 10 parametrov. Poleg tega je prostor kriterijske funkcije razmeroma kaotičen, ker primerjamo lastnosti realnega in sintetičnega signala. Zato je jasno, da nam klasične (gradientne) metode optimizacije ne bodo v korist. Ker želimo optimizirati prostor parametrov brez nikakršnega predznanja o lastnostih okolja (sintetizatorja), imamo opravka s t.i.

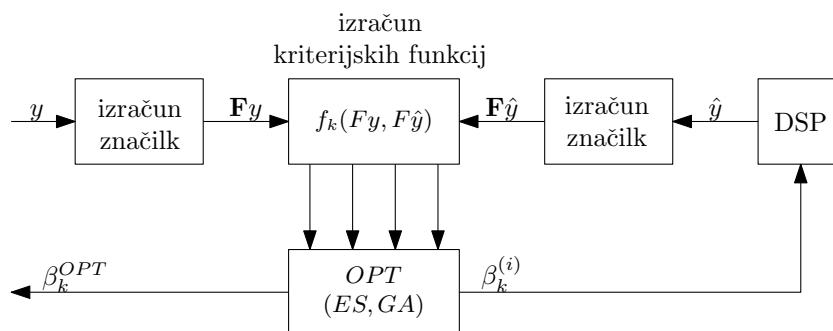


Slika 4.3: Odvisnost izbranih kriterijskih funkcij od disperzijskega koeficienta (relativna napaka)

optimizacijo ‘črne škatle’ (*angl. black-box optimization*). Tu so veliko bolj primerne *populacijske* metode, ki ne zahtevajo nobenih predpostavk o problemu, ki ga rešujejo. Med te štejemo npr. algoritem optimizacije s kolonijo mravlj, optimizacijo z rojem delcev ter evolucijske algoritme za optimizacijo.

## 4.2 Optimizacija

V zadnjih petdesetih letih je evolucijsko računanje doživelo popoln razmah zaradi večanja zmogljivosti računalniških sistemov in vsestranske uporabnosti na različnih področjih znanosti in ekonomije. Iz literature je razvidno, da so se predvsem genetski algoritmi izkazali za izredno praktične pri reševanju problemov, kot je optimizacija parametrov sintetizacije [17],[10],[11]. Pri genetskih algoritmih je prostor parametrov potrebno diskretizirati. To pomeni, da je zaželeno vnaprej vedeti vsaj približno kakšna je varianca in pričakovana vrednost parametrov, ki



Slika 4.4: Večkriterijska optimizacija parametrov

bo peljala k smiselnim rešitvam. Izpeljankam genetskih algoritmov za zvezni prostor parametrov pravimo *evolucijske strategije*. Če želimo doseči bistveno boljši učinek optimizacije s perceptivnega stališča uporabimo **večkriterijsko** optimizacijo, kjer v proces iskanja vstopa več kriterijskih funkcij, ki se ločeno optimizirajo (slika 4.4). Rezultat iskanja pa je skupina t.i. optimalnih točk. Zitzler (2000) podaja kvantitativno študijo nekaterih evolucijskih algoritmov za večkriterijsko optimizacijo in navaja primerne metrike za ocenjevanje učinkovitosti glede na *Pareto* optimalnost [29].

### 4.2.1 Pareto optimalnost

Ker uporabljamo večkriterijsko optimizacijo, vemo, da bodo posamezne napake kriterijev gotovo enake 0 samo v primeru, če bo iskani signal identičen ciljnemu. To je seveda nemogoče doseči, ker je ta model v osnovi groba aproksimacija realnega okolja (glasbila). Pri enokriterijski optimizaciji bo optimizacija posamezne kriterijske funkcije v idealnih pogojih dosegla samo eno (globalno) optimalno vrednost in od tu dalje običajno ustavimo iskanje. Ampak, ker imamo opravka z večkriterijsko optimizacijo, bodo v nekem trenutku kriteriji med seboj delovali destruktivno. To pomeni, da od nekega trenutka dalje ne bomo mogli optimizirati napake na nekem kriteriju, brez da bi poslabšali katerega izmed ostalih kriterijev. Takim rešitvam pravimo *Pareto* optimalne rešitve in nakazujejo, da smo dosegli (Pareto) optimalno območje skupine kriterijskih funkcij in lahko ustavimo algoritem iskanja oz. optimizacije.

V večkriterijski optimizaciji je vsaka rešitev oz. točka v prostoru parametrov,

ki minimizira določeno kriterijsko funkcijo, optimalna v *Pareto* smislu. Zato rešitve primerjamo med seboj z metriko, ki ji pravimo *dominanca*. Tako je pri dveh rešitvah boljša oz. *dominantna* tista, pri kateri velja, da je vrednost vsaj ene kriterijske funkcije manjša in nobena vrednost kriterijskih funkcij večja (enačba 4.5). Ekvivalentno velja, da je slabša rešitev, tista, pri kateri je katerakoli vrednost kriterijskih funkcij večja.

$$\exists k, f_k(\beta_a) < f_k(\beta_b) \wedge f_i(\beta_a) \leq f_i(\beta_b), \forall i \quad (4.5)$$

### 4.2.2 Učinkovitost iskanja

Uspešnost algoritma za optimizacijo modela moramo nekako oceniti. Učinkovitost iskanja v kontekstu optimizacije označujemo kot lastnost algoritma, da pripelje iskanje parametrov v globalni optimum iskane funkcije s čim manjšim številom evalvacij funkcije oz. iteracij algoritma. Dobra metoda za določitev učinkovitosti takega algoritma, kot kriterijskih funkcij je seveda test na takem signalu, ki ga lahko ustvari sintetizacijski model. Pričakujemo lahko, da bo optimizacija pripeljala iskanje do takih parametrov, ki stoddostno reproducirajo ciljni signal. Gre za t.i. *self-test* optimizacije modela. Pri tem lahko učinkovitost ocenimo z naslednjimi ugotovitvami:

1. kolikšno je število iteracij do dosežene zelene vrednosti oz. kolikšna je vrednost kriterijske funkcije pri fiksnem številu iteracij pri **različnih nastavitvah iskanja**
2. kolikokrat je **iskanje pripeljalo do istih rešitev** oziroma, kolikšna je varianca parametrov po večkratnem izvajanju iskanja z istimi nastavitvami - s tem lahko ocenimo verjetnost konvergence v **globalni optimum** na območju iskanja
3. ali pripelje iskanje testnega signala do identičnih testnih parametrov?
4. kolikšna je (evklidska) razdalja od optimalnih parametrov do ciljnih (testnih) parametrov sintetičnega signala

### 4.2.3 Diferencialna evolucija

Diferencialna evolucija (DE) je optimizacijski algoritem, ki ima določene lastnosti analogne naravni evoluciji [25]. Za razliko od ostalih evolucionskih algoritmov, tu ni očitne mutacije in selekcija je običajno lahko popolnoma naključna. Je pa prisotna rekombinacija (analogija križanja). Ker gre v osnovi za optimizacijo ‘črne škatle’ z zveznimi parametri, je algoritem zagotovo primeren za iskanje parametrov sintetizatorja.

Osnovni algoritem DE sicer ni podpiral večkriterijske optimizacije, ampak je zelo enostavno razširljiv za večkriterijske probleme. Tabela 4.1 prikazuje enostaven primer programske kode klasičnega DE algoritma razširjenega za večkriterijsko optimizacijo. Bistvo algoritma predstavlja iteracija opisana v naslednjih petih korakih:

1. Določimo začetno porazdelitev populacije.
2. Za vsak element populacije naključno izberemo tri pomožne kandidate.
3. Naključno določimo koliko parametrov se bo križalo.
4. Izvedemo diferencialno<sup>4</sup> križanje s kombinacijo izbranih pomožnih kandidatov in dobimo novega kandidata.
5. Če novi kandidat dominira obstoječega, ga zamenjamo.

Naprednejše izpeljanke večkriterijske diferencialne evolucije (*Crowding-DE*, *DE-MO/closest*) skušajo poleg optimizacije hraniti tudi več optimalnih točk z avtomatskim razvrščanjem točk v populaciji. To dosežejo z izločanjem najbližjih točk, t.j. tistih, ki imajo najnižjo medsebojno evklidsko razdaljo v prostoru parametrov [26],[18]. Temu pravimo združevanje v skupine oz. *angl. crowding*.

---

<sup>4</sup>diferencialno križanje je križanje izbranega kandidata  $X$  z diferencialnim pomikom pomožnega kandidata  $X_a$  v smeri  $F(X_b - X_c)$ , kjer je  $F$  faktor diferencialnega pomika.

---

```
# določimo parametre diferencialne evolucije
N = ... # velikost populacije
CR = ... # verjetnost križanja
F = ... # faktor diferencialnega pomika

# naključno porazdelimo populacijo z N elementi
P = randomize_population(distribution)
...
# zanka ene iteracije diferencialne evolucije
for i,X in enumerate(self.P):
    # za vsak kromosom X z indeksom i v populaciji P
    # izberemo 3 naključne kromosome iz populacije, ki so različni od X
    Xa,Xb,Xc = select(P,X)
    # določimo naključni binarni rekombinacijski vektor,
    # kjer je vsaj en gen rekombinacije aktiven
    r = (rand(D) > CR) + 0
    r[randint(D)] = 1
    # določimo nov kromosom z rekombinacijo (križanjem)
    # izbranega kromosoma X in diferencialnim pomikom
    # kromosoma Xa v smeri  $F*(Xb-Xc)$ 
    Y = (1-r)*X + r*(Xa + F*(Xb-Xc))
    # nato izračunamo vrednosti kriterijskih funkcij za Y
    # in X (te so v praksi shranjene v seznamu)
    Qy = fit_fn(Y)
    Qx = fit_fn(X)
    # če je v Qy vsaj ena krit. funkcija večja od Qx
    # zamenjamo izbrani kromosom (X) z novim (Y)
    if (Qy<Qx).any():
        P[i] = Y
```

---

Tabela 4.1: Zanka DE iteracije za večkriterijske funkcije (Python impl.)

### 4.3 Določitev območja iskanja

Pri evlucijskem iskanju optimalnih parametrov moramo določiti območje iskanja in porazdeliti začetno populacijo. Zato lahko uporabimo dva pristopa:

- **neomejeno** območje iskanja, kjer začetno porazdelitev določimo z Gaussovo porazdelitveno funkcijo  $N(\mu, \sigma^2)$  in
- **omejeno** območje iskanja, kjer za vsak analitično določen parameter določimo spodnjo in zgornjo mejo iskanja, začetno populacijo pa porazdelimo enakomerno ali z omejeno Gaussovo porazdelitvijo.

Pri neomejenem območju je center oz. povprečna vrednost enaka analitično izračunani vrednosti parametra ( $\mu = \hat{\beta}_i$ ). Standardni odklon pa določimo tako da bo večji del populacije znotraj analitično ali empirično določenih mej območja, npr.

$$\sigma_i = \max\{\text{abs}(\hat{\beta}_i - \hat{\beta}_i^{\min}), \text{abs}(\hat{\beta}_i - \hat{\beta}_i^{\max})\}.$$

Pri omejenem iskanju, začetno porazdelitev populacije določimo po enakomerni verjetnostni porazdelitvi  $U(\beta_i^{\min}, \beta_i^{\max})$ , ali po omejeni Gaussovi porazdelitvi (enačba 4.6), ki privilegira empirično oz. analitično izračunano začetno vrednost parametra  $\hat{\beta}_i$ ,

$$f(x; \mu, \sigma, a, b) = \frac{\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad (4.6)$$

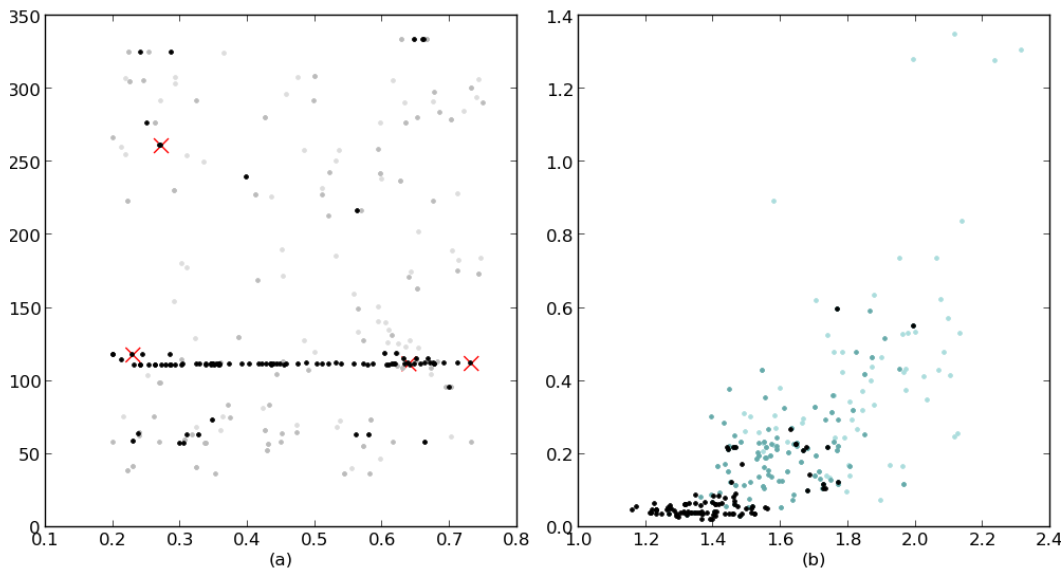
kjer so parametri verjetnostne porazdelitve  $\mu = \hat{\beta}_i$ ,  $a = \beta_i^{\min}$ ,  $b = \beta_i^{\max}$ , funkciji  $\phi(x)$  in  $\Phi(x)$  pa porazdelitvena funkcija in kumulativna porazdelitvena funkcija normalne porazdelitve,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (4.7)$$

$$\Phi(x) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x}{\sqrt{2}} \right) \right]. \quad (4.8)$$

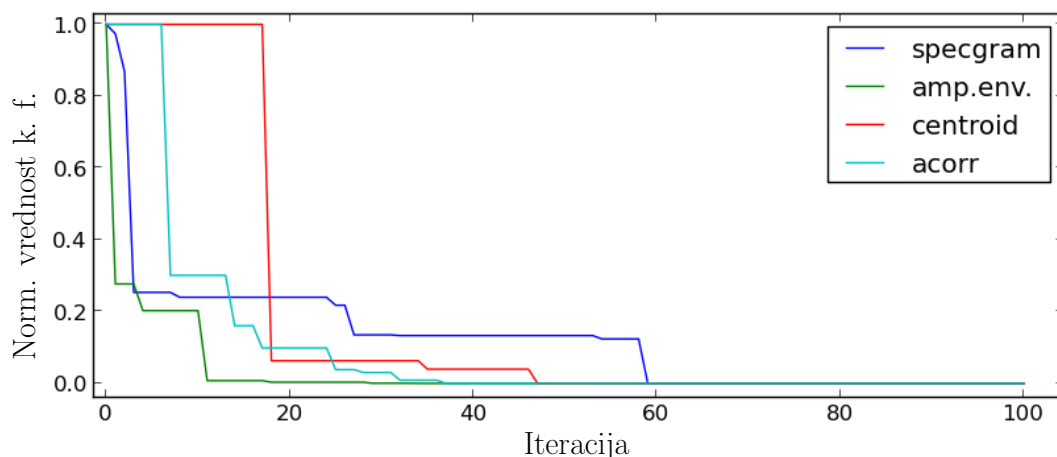
Originalni DE algoritem išče po neomejenem območju. To pomeni, da kljub temu, da smo začetno porazdelitev omejili, se sredi procesa iskanja lahko ustvarijo točke oz. parametri, ki so zunaj mej območja začetne populacije. Najpreprostejša

rešitev je, da le-te zaznamo in preprosto zavržemo. Druga možnost je, da točke zunanaj meja iskanja postavimo v najbližjo točko znotraj območja, tako, da vrednosti parametrov ( $\beta_i$ ) postavimo na meje območja. Ta rešitev lahko v določenih primerih sproži hitro konvergenco velikega dela populacije na rob območja parametra, ki sam po sebi lahko predstavlja lokalni minimum, zato se ji raje izogibamo.



Slika 4.5: Razvoj območja iskanja DE algoritma: od začetnega območja (svetlo), do območja po 50 iteracijah (temno) za prostor parametrov (a) in porazdelitev kriterijskih funkcij (b).

Slika 4.5 prikazuje dinamiko iskanja na enostavnem primeru iskanja optimalnih parametrov, kjer iščemo samo koeficient disperzije in frekvenco tona (dolžino resonatorja). Z rdečo so označene vrednosti najboljših kandidatov glede na posamezno kriterijsko funkcijo. Slika 4.6 pa prikazuje potek vrednosti (normaliziranih) kriterijskih funkcij za prvih 100 iteracij DE algoritma za isti primer. Po 60 iteracijah se vrednost kriterijskih funkcij praktično ne spreminja več. Iz slike 4.5 lahko opazimo, da že po 50 iteracijah iskanje parametrov konvergira v Pareto območje. Ostane nam torej več rešitev, ki nakazujejo zvezo med iskanima parametra glede na kriterijske funkcije. S stališča le-teh so vse **enako** dobre. Pa vendar, pri poslušanju simuliranega zvoka najboljših kandidatov posameznih kriterijskih funkcij opazimo, da je avtokorelacijska rešitev perceptivno najboljša. To je smiselno, ker smo optimizirali dva parametra, ki vplivata v glavnem na



Slika 4.6: Normalizirana vrednost kriterijskih funkcij v izvajanju 100 iteracij algoritma.

zaznavanje frekvence tona. Kot najboljšo rešitev lahko vzamemo tisto, ki ima najnižjo skupno perceptivno napako v populaciji. Dobra rešitev, ki ima te lastnosti je npr. točka, kjer je normalizirana vsota vseh kriterijskih funkcij najnižja oz. ‘centralna’ rešitev v populaciji,

$$i_c = \arg \min_i \left\{ \sum_k \frac{f_k(i) - \min(f_k(i))}{\max(f_k(i)) - \min(f_k(i))} \right\}. \quad (4.9)$$

V tabeli 4.2 so prikazane začetna ocena, začetno območje ter optimalne rešitve glede na posamezne kriterijske funkcije in ‘centralna’ rešitev po 100 iteracijah algoritma.

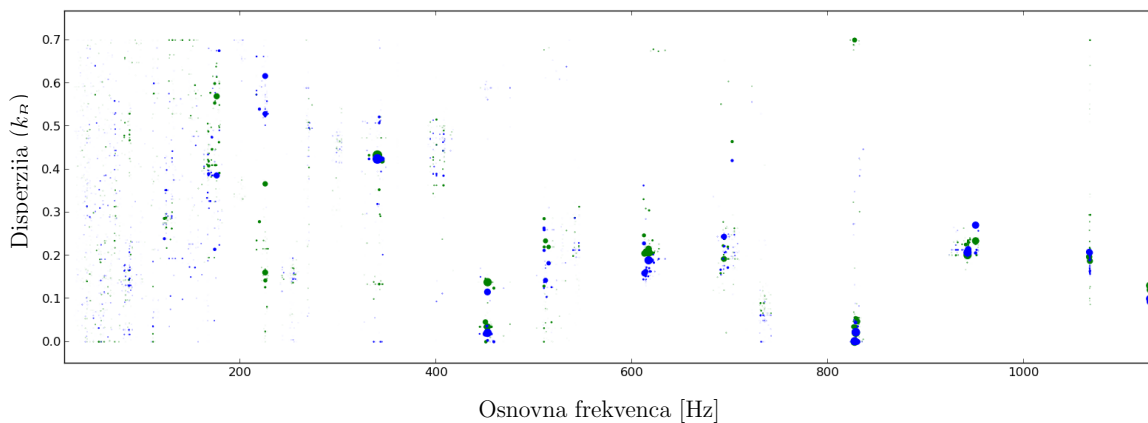
## 4.4 Rezultati iskanja

Rezultati iskanja so zelo odvisni od začetne porazdelitve in parametrov DE algoritma. Če imamo območje omejeno zelo blizu pričakovane rešitve, bo iskanje verjetno pripeljalo do (globalno) optimalne rešitve. Sicer bo zelo verjetno obtičalo v nekem lokalnem minimumu, zaradi narave kriterijskih funkcij. Zato je smiselno večkrat ponoviti iskanje, po možnosti z različnimi parametri iskanja. Slika 4.7 prikazuje rešitve pri večkratnem izvajanju evolucijskega algoritma pri različnih zvočnih posnetkih, kjer je algoritem skušal optimizirati kriterijske funkcije glede na vse parametre modela. Prikazana sta parametra disperzije  $k_B$  in frekvence  $f_0$ .

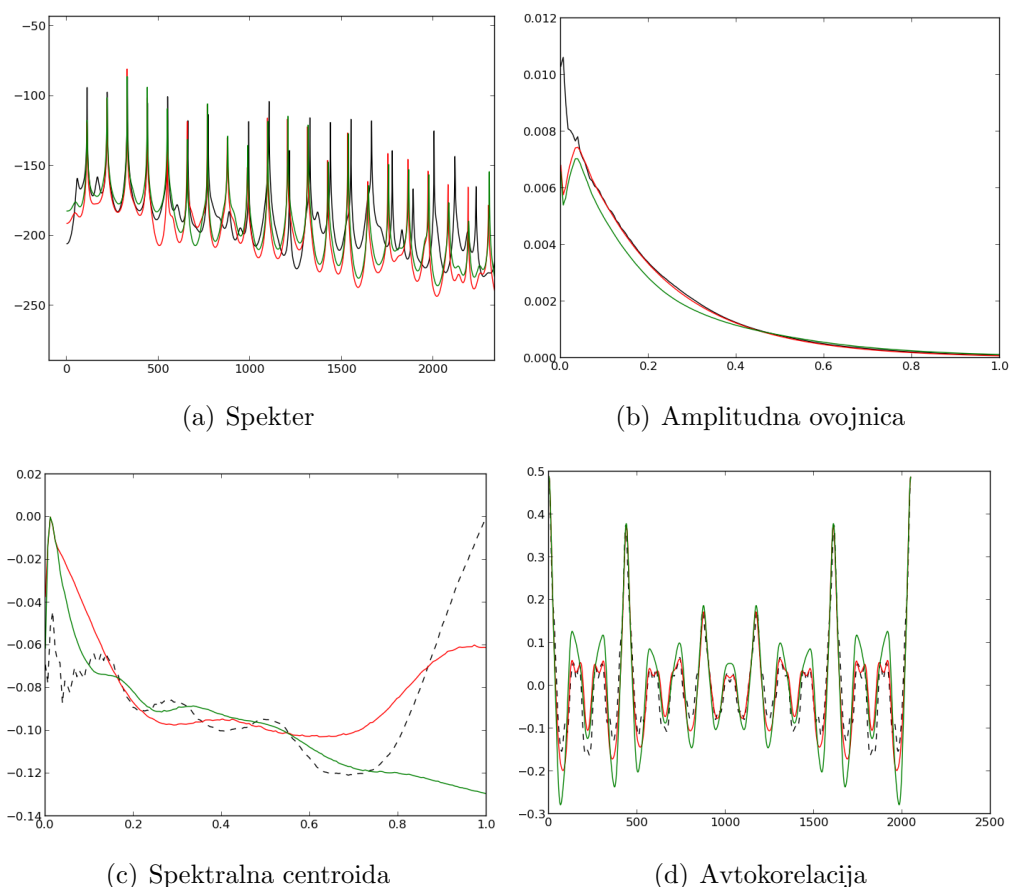
	$k_B$	$f'_0[\text{Hz}]$
začetna ocena	0.646	110.599
začetni min.	0.200	36.0
začetni max.	0.750	334.0
spektrogram	0.674	115.101
amp. ovojnica	0.270	117.967
centroida	0.277	261.212
avtokorelacija	0.626	111.471
centralna reš.	0.428	110.986

Tabela 4.2: Začetne in optimizirane vrednosti

Slika 4.8 pa prikazuje rezultate za ton s frekvenco 110 Hz pri najboljših kandidatih vsake kriterijske funkcije (rdeča) in pri skupnemu najboljšemu kandidatu (zelena). V splošnem lahko opazimo zelo dobro ujemanje na amplitudi in avtokorelaciji, nekoliko slabše spektralno ujemanje in zelo slabo ujemanje centroida. Enake ugotovitve lahko povzamemo v perceptivem smislu. To nakazuje da model ne more simulirati vseh perceptivnih lastnosti zvoka in je potreben razširitve.



Slika 4.7: Rezultati večkratnih poganjanj optimizacije različnih tonov (frekvenca in disperzija)



Slika 4.8: Rezultati za ton 110 Hz. Prikaz značilk pri najboljših kandidatih za posamezne kriterije (rdeče), za optimalno vrednost (zeleno) in ciljno vrednost (črtkano)

## 4.5 Možne izboljšave algoritma optimizacije

Ker imamo opravka s percepcijo zvoka, nas v praksi ne zanima globalno optimalen hiperprostor rešitev. Pomembneje je doseči območje rešitev, kjer so napake pri vsaki rešitvi perceptivno nezaznavne. V tem kontekstu lahko v DE algoritmu iskanje omejimo tako, da določimo ciljno oz. zeleno vrednost za vsako kriterijsko funkcijo. Pod to vrednostjo, algoritem ne bo skušal ohranjevati najnižjo vrednost kriterijske funkcije, ampak bo točko zamenjal, v kolikor vrednost kriterijske funkcije ne bo višja od ciljne vrednosti. Na ta način bo algoritem hitreje konvergiral proti idealnemu območju.

Slabost evolucijskih algoritmov je v visokem številu potrebnih evalvacij kri-

terijske funkcije, ki jo skušamo optimizirati. Če je izračun kriterijske funkcije razmeroma računsko zahteven, se čas optimizacije bistveno poveča, saj je čas izvajanja produkt števila evalvacij funkcije in časa posamezne evalvacije. Poleg tega je generiranje populacije delno naključno. To pomeni, da bo v vsaki iteraciji algoritma nastalo nekaj rešitev (parametrov), ki bodo zelo blizu nekaterim trenutnim ali predhodnim instancam, ki so se izkazale za slabe, vendar populacijsko-osnovani algoritem to informacijo v naslednji iteraciji pozabi. Neugodna posledica tega je, da je potrebna ponovna evalvacija računsko potratne funkcije, kljub temu, da pričakujemo neugoden izid.

Obstaja tudi izboljšava, ki odpravi oz. nekoliko omili zgoraj omenjene slabosti. Če lahko z veliko gotovostjo napovemo kakšna je vrednost kriterijske funkcije v novonastali točki, lahko napoved uporabimo za odločitev ali naj izvedemo simulacijo parametrov ali ne. Tudi če ta metoda ne napove optimalne točke, nam vsaj zmanjša število evalvacij kriterijskih funkcij in s tem bistveno skrajša čas iskanja. Problem je v tem, da za je za kaotične funkcije rezultat napovedi običajno slab.

Druga možnost je ta, da namesto napovedi kriterijskih funkcij, skušamo napovedati vrednost izbranih nizko-dimenzionalnih perzeptivnih značilnk. V tem primeru lahko natančneje napovemo ali je vrednost rešitve perzeptivno 'blizu' ali ne.



# Poglavje 5

## Zaključek

V pričujočem delu je bil predstavljen prototip modela sintetizatorja, kjer je bil glavni cilj modeliranje zvoka klavirja. Zasnovan je bil z metodo sintetizacije z digitalnimi valovnimi vodili, ki učinkovito simulira nekatere elemente akustike.

Prikazan je bil tako analitični način določanja optimalnih parametrov, kot avtomatizacija iskanja le-teh z minimizacijo perceptivne razdalje med zvoki. Perceptivna razdalja je določena kot skupek značilnk in metrik, ki čim natančneje opisujejo človeško zaznavanje zvoka, obenem pa so numerično hitro izračunljive.

Optimizacija enokriterijske funkcije lahko privede do enostranskih rešitev, kjer je določen kriterij močno optimiziran, ostali pa so zanemarjeni. V tem primeru je perceptivna razlika v zvoku lahko zelo očitna. V ta namen je bil predstavljen postopek kalibracije modelov sintetizacije zvoka z večkriterijsko minimizacijo perceptivne razdalje med izvornim zvokom in sintetiziranim zvokom. Metoda večkriterijske optimizacije z diferencialno evolucijo se je izkazala za hkrati enostavno in učinkovito orodje pri perceptivni optimizaciji. Kljub temu, ne moremo pričakovati (perceptivno) idealnih rezultatov, zaradi naslednjih razlogov:

1. Izbran model ni idealen in ne more opisovati vseh fizikalnih pojavov, ki tvorijo očitne zvočne učinke, katere človek lahko zaznava.
2. Izbrane značilke in metrike ne odražajo v celoti človekovega zaznavanja razlike med zvoki.
3. Tako prostor parametrov, kot prostor značilnk sta visokodimenzionalna. Zveza med parametri modela in značilkami je nelinearna. Prostor kriterijskih

funkcij je zato izredno kaotičen. To okolje je zahtevno za optimizacijo, saj v prostoru kriterijskih funkcij obstaja veliko (lokalnih) minimumov.

Na koncu prejšnjega poglavja so bili predlagani nekateri pristopi strojnega učenja, ki lahko doprinesejo k učinkovitosti optimizacije. Kljub temu, osnovno težavo predstavlja določitev idealnih ali vsaj boljših nizkodimenzionalnih značilk, ki opisujejo človeško zaznavanje tona oz. zvoka. Brez teh je vsakršna optimizacija nesmiselna, saj se iščejo kandidati, ki niso perceptivno podobni zvočnemu viru. Tega seveda ne želimo.

Večkriterijska optimizacija ima poleg omenjenih slabosti tudi dobro plat. Omogoča nam, da npr. določimo ali celo narišemo zeleni potek značilk in poiščemo najboljše kandidate sintetizacijskega okolja. Tako lahko raziskujemo lastnosti sintetizacijskih modelov oz. določimo nova umetna glasbila brez implementacije računsko zahtevnih algoritmov za modulacijo značilk. V ta namen so digitalna valovna vodila praktično idealna. Odpirajo nam možnosti raziskovanja in izdelave neobstoječih glasbil, ki lahko zvenijo naravno, a imajo drugačne lastnosti od obstoječih akustičnih glasbil.

# Literatura

- [1] F. Avanzini, B. Bank, G. Borin, G. De Poli, F. Fontana, and D. Rocchesso. Musical instrument modeling: the case of the piano. In *Proc. Int. Workshop on Current Research Directions in Comp. Mus.*, pages 124–133, Barcelona, 2001.
- [2] B. Bank. Accurate and efficient modeling of beating and two-stage decay for string instrument synthesis. In *Proc. MOSART Workshop on Curr. Res. Dir. in Computer Music, Barcelona, Spain*, 2001.
- [3] B. Bank, F. Avanzini, G. Borin, G. De Poli, F. Fontana, and D. Rocchesso. Physically informed signal processing methods for piano sound synthesis: A research overview. In *EURASIP Journal on Applied Signal Processing 2003:10*, c Hindawi Publishing Corporation, page 941–952, 2003.
- [4] Scott A. Van Duyne and Julius O. Smith III. Developments for the commuted piano, 1999.
- [5] S.S. Haykin. *Adaptive Filter Theory, 4/E*. Pearson Education, 2002.
- [6] Julius O. Smith III and Scott A. Van Duyne. Commuted piano synthesis, 2008.
- [7] Toshio Irino and Roy D. Patterson. A time-domain, level-dependent auditory filter: The gammachirp. *J. Acoust. Soc. Am. Volume 101, Issue 1*, pages 412–419, 1997.
- [8] K. Karplus and A. Strong. Digital synthesis of plucked string and drum timbres. *Computer Music Journal, Vol. 7, No. 2*, page 43–55, 1983.

- 
- [9] Jari Kleimola. Dispersion modulation using allpass filters. In *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Finland, 2008.
- [10] Yuyo Lai, Shyh kang Jeng, Der tzung Liu, and Yo chung Liu. Automated optimization of parameters for fm sound synthesis with genetic algorithms, 2006.
- [11] Matthieu Macret, Philippe Pasquier, and Tamara Smyth. Automatic calibration of modified fm synthesis to harmonic sounds using genetic algorithms. In *Proceedings of the 9th Sound and Music Computing Conference, Copenhagen, Denmark, p.387-394*, 2012.
- [12] J. Makhoul. Linear prediction: A tutorial review. In *Proceedings of the IEEE, vol. 63, pp. 561-580*, 1975.
- [13] B.S. Manjunath, Philippe Salembier, and Thomas Sikora. *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley & Sons, 2002.
- [14] B.C.J. Moore. *Hearing*. Handbook of Perception and Cognition, Second Edition. Elsevier Science, 1995.
- [15] R. Plomp. *Aspects of tone sensation: a psychophysical study*. Academic press series in cognition and perception /ed. by Edward Carterette, Morton P. Friedman. Academic Press, 1976.
- [16] Jukka Rauhala and Vesa Välimäki. Dispersion modeling in waveguide piano synthesis using tunable allpass filters. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 71–76, Montreal, Quebec, Canada, Sept. 18–20, 2006.
- [17] Janne Riionheimo and Vesa Välimäki. Parameter estimation of a plucked string synthesis model using a genetic algorithm with perceptual fitness calculation. *EURASIP J. Adv. Sig. Proc.*, 2003:791–805, 2003.
- [18] Tea Robič and Bogdan Filipič. Demo: Differential evolution for multi-objective optimization. In *In Proceedings of the 3rd International Conference on Evolutionary MultiCriterion Optimization (EMO 2005)*, pages 520–533. Springer. LNCS, 2005.

- 
- [19] X. Serra and J. O. Smith. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, vol. 14, no. 4, pp. 12-24, 1990.
- [20] J.O. Smith. *Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects*. W3K Publishing, 2010.
- [21] J.O. Smith and Stanford University. *Spectral Audio Signal Processing*. W3K Publishing, 2011.
- [22] Julius O. Smith. Viewpoints on the history of digital synthesis. In *Proceedings of the International Computer Music Conference*. International Computer Music Association, 1991.
- [23] Julius O. Smith. Physical modelling using digital waveguides. In *Computer Music Journal*, Vol 16, no. 4, pp. 74-91, 1992.
- [24] Zambon Stefano, Heidi-Maria Lehtonen, and Balázs Bank. Simulation of piano sustain-pedal effect by parallel second-order filters. In *Submitted to the DAFx-08 Int. Conference, Espoo*, 2008.
- [25] Rainer Storn and Kenneth Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *J. of Global Optimization*, 11(4):341–359, December 1997.
- [26] R. Thomsen. Multimodal optimization using crowding-based differential evolution. In *IEEE Congress on Evolutionary Computation*, 2004.
- [27] Vesa Välimäki. Simple design of fractional delay allpass filters. In *Proc. Euro. Conf. Circuit Theory Design. Vol. 1*, 2000.
- [28] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.
- [29] Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8:173–195, 2000.