

U N I V E R Z A   V   L J U B L J A N I

Fakulteta za elektrotehniko

Janez Žibert

**ČASOVNO–FREKVENČNE PREDSTAVITVE  
GOVORNIH SIGNALOV**

MAGISTRSKO DELO

Mentor: prof. dr. France Mihelič

Ljubljana, 2001



---

**Ključne besede:** časovno–frekvenčne predstavitve, časovno–frekvenčne porazdelitve, energijske porazdelitve, kvadratne časovno–frekvenčne predstavitve, linearne časovno–frekvenčne predstavitve, diskretne časovno–frekvenčne predstavitve, Cohenov razred, afine predstavitve, Wigner–Villejeva predstavitev, obdelava govornega signala, akustične značilke govora, koeficienti kepstra, modeliranje govora, razpoznavanje govora, razpoznavanje glasov, razpoznavanje glasovnih prehodov.



---

# Zahvala

Za nastanek magistrskega dela se moram najprej iskreno zahvaliti mentorju prof. dr. Francetu Miheliču. Pod njegovim vodstvom sem uspel pridobiti ustrezno znanje iz področja tehnologij govornega (slovenskega) jezika. Njegova pripravljenost za pomoč, izvrstno poznavanje področja in izkušnost so mi prihranile prenekatero uro, ki bi jo bil sicer prisiljen žrtvovati za iskanje odgovorov na vsa svoja vprašanja. Njegova raziskovalna pot pa pomeni zame tudi veliko vzpodbudo.

Hvala tudi moji družini, ki mi je s svojo ljubeznijo, delom in prijaznim domom omogočila možnost brezskrbnega študija in hkrati pomenila veliko gonilo v življenju.

Predvsem pa bi se rad zahvalil Mateji. Hvala za vso potrpežljivost, vsestransko pomoč in razumevanje. Hvala, ker si z mano, me vzpodbujaš in mi daješ energijo.

Rad bi se zahvalil tudi vsem sodelavcem Laboratorija za umetno zaznavanje, sisteme in kibernetiko: vodji prof. dr. Nikoli Pavešiču za prijaznost in pripravljenost za pomoč, Simonu, Jerneji, Ankici, Mariu, Ivotu, Tonetu, Boštjanu in Amirju. Hvala, ker ste me tako prijazno sprejeli v laboratorij in bili vedno pripravljeni za pogovor in pomoč. Še posebej bi rad izpostavil Simona, ki me je s svojimi idejami in nespornim znanjem naučil veliko novega, predvsem pa mi je odprl veliko novih in zanimivih raziskovalnih izzivov. Hvala tudi za čas, ki si ga porabil zame, in številne programe, ki sem jih lahko uporabljal za izvedbo svojih idej.

Nazadnje se zahvaljujem tudi vsem svojim profesorjem in učiteljem za pridobljeno znanje.



---

# Povzetek

V magistrskem delu se ukvarjamo s časovno–frekvenčnimi predstavitvami govornih signalov.

V prvem delu podajamo splošne principe in prijeme obravnave signalov z matematičnega stališča. V časovno–frekvenčni ravnini jih obravnavamo na dva načina: z linearnimi in kvadratnimi predstavitvami, kjer preučujemo porazdelitve energije signala v ravnini časa in frekvence. Zato takšnim predstavitvam pravimo tudi energijske porazdelitve. V tem delu vpeljemo nekaj osnovnih idej izvedbe in analiziramo lastnosti takšnih predstavitev. Posebna pozornost je namenjena predstavitvam Cohenovega razreda, manj pa drugi večji skupini - afinim predstavitvam.

V nadaljevanju se posvečamo diskretizaciji časovno–frekvenčnih predstavitev in problemom, ki pri tem nastanejo. Podajamo dve metodi konstrukcije diskretnih predstavitev: neposredne prevedbe zveznih verzij in izpeljave diskretizacije direktno iz vzorčenega signala s pomočjo operatorske teorije. Drugi način je predvsem pomemben za ugotavljanje lastnosti diskretnih časovno–frekvenčnih predstavitev.

Glavni poudarek magistrskega dela pa je namenjen časovno–frekvenčni obdelavi govornih signalov. Tu se posvečamo dvema konkretnima primeroma: prikazovanju slik časovno–frekvenčnih predstavitev in ustreznemu pridobivanju informacije iz teh slik za nadaljnje razpoznavanje govora. Predlagamo nekaj metod obdelave govornih signalov s temi predstavitvami, ki jih uspešno uporabimo za pridobivanje ustreznih značilk govornega signala za razpoznavanje govora. Tu kombiniramo različne izvedbe pridobivanja značilk iz govornega signala na podlagi različnih časovno–frekvenčnih predstavitev in primerjamo rezultate razpoznavanja na primeru govorne zbirke.





---

# Abstract

This master thesis focuses on time–frequency representations of speech signals.

First part overviews some theoretical grounds and basic concepts of time–frequency analysis of signals in general. Joint time–frequency signal representations characterize signals over time–frequency plane. Therefore they combine time domain and frequency domain analyses to yield a potentially more revealing picture of the temporal localization of signal’s spectral components. Time–frequency representations are divided into two main classes: linear (also known as atomic decompositions) and quadratic (energy distributions) representations. In this section we review some fundamental ideas and basic properties of each class of representations with emphasis to Cohen class of energy distributions.

The next part of our work gives discrete versions of time–frequency representations. Here we are studying two basic concepts of discretization of continuous time–frequency representations: uniform sampling of continuous counterparts and direct derivation of discrete representations from sampled signals using the concepts of operator theory. The second approach is important for studying the properties of discrete time - discrete frequency representations.

The main emphasis of this work has been to study and to find potentially good time–frequency representations for analysis and further processing of speech signals. Our main goal has been to find the best method and representation for deriving features from speech signals for automatic speech recognition. Here we present different methods and compare several time–frequency representations to achieve the best performance of speech recognition system based on hidden Markov models.



---

# Kazalo

<b>Zahvala</b>	<b>iii</b>
<b>Povzetek</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>1 Uvod</b>	<b>1</b>
1.1 Motivacije za delo . . . . .	2
1.2 Pregled vsebine dela . . . . .	3
<b>2 Časovno–frekvenčne predstavitve signalov</b>	<b>4</b>
2.1 Analiza signalov . . . . .	5
2.1.1 Časovna in frekvenčna predstavitev signala . . . . .	5
2.1.2 Princip nedoločenosti . . . . .	6
2.1.3 Preproste časovno–frekvenčne predstavitve . . . . .	7
2.2 Linearne časovno–frekvenčne predstavitve . . . . .	9
2.2.1 Kratkočasovna Fourierjeva transformacija . . . . .	10
2.2.2 Valčna transformacija . . . . .	13
2.2.3 Diskretne linearne predstavitve . . . . .	15
2.3 Kvadratne časovno–frekvenčne predstavitve . . . . .	18
2.3.1 Prehod iz linearnih h kvadratnim predstavitev . . . . .	19
2.3.2 Cohenov razred . . . . .	22
2.3.3 Afine predstavitve . . . . .	37
<b>3 Diskretne časovno–frekvenčne predstavitve</b>	<b>41</b>

---

3.1	Vzorčenje zveznih predstavitev . . . . .	42
3.1.1	Diskretna Wigener–Villejeva porazdelitev . . . . .	42
3.2	Vzorčenje s pomočjo operatorske teorije . . . . .	46
3.2.1	Karakteristične funkcije zveznih predstavitev . . . . .	46
3.2.2	Prevedba v diskreten prostor . . . . .	48
3.2.3	Lastnosti operatorjev . . . . .	49
3.2.4	Izpeljava diskretnih časovno–frekvenčnih predstavitev . . . . .	50
<b>4</b>	<b>Analiza in razpoznavanje govornega signala</b>	<b>55</b>
4.1	Časovno–frekvenčne slike govornega signala . . . . .	56
4.1.1	Slike časovno–frekvenčnih predstavitev . . . . .	56
4.1.2	Primerjava predstavitev . . . . .	58
4.1.3	Popravljenе predstavitve . . . . .	64
4.2	Razpoznavanje govora . . . . .	68
4.2.1	Razpoznavalnik govora . . . . .	68
4.2.2	Razpoznavanje govora z različnimi predstavitvami . . . . .	72
<b>5</b>	<b>Zaključek</b>	<b>93</b>
5.1	Sklepne ugotovitve . . . . .	94
5.2	Nadaljnje delo . . . . .	95
5.2.1	Renyijeva entropija . . . . .	96
5.2.2	Houghova transformacija . . . . .	96
5.2.3	Učenje jedra predstavitve . . . . .	97
	<b>Viri in literatura</b>	<b>99</b>
<b>A</b>	<b>Pregled časovno–frekvenčnih predstavitev</b>	<b>104</b>
A.1	Linearne časovno–frekvenčne predstavitve . . . . .	104
A.2	Kvadratne časovno–frekvenčne predstavitve . . . . .	104
A.3	Ostale časovno–frekvenčne predstavitve . . . . .	107
	<b>Slovar izrazov</b>	<b>108</b>

---

---

# Slike

2.1	Časovna predstavitev signala in njegov (amplitudni) spekter. Fourierjeva transformacija $\mathcal{F}$ predstavlja vez med časovno in frekvenčno predstavitev signala. . . . .	5
2.2	Povprečna frekvenca $f_m$ in standardna deviacija $B$ v primeru realnega signala (a) in dejanske iskane vrednosti (b). . . . .	6
2.3	Prikaz trenutne frekvence in skupinske zakasnitve signala iz zgornje slike. Odebeljena črta predstavlja oceno trenutne frekvence, tanjša pa skupinske zakasnitve cvrka. Tu gre za monokomponentni nestacionarni signal. . . . .	8
2.4	Prikaz trenutne frekvence in skupinske zakasnitve signala $x(t) = e^{j2\pi f_1 t} + e^{j2\pi f_2 t}$ , kjer je $\frac{f_1}{f_s} = 0.1$ in $\frac{f_2}{f_s} = 0.3$ . Tu gre za dvokomponentni stacionarni signal. Iz slike je razvidno, da obravnavani količini v tem primeru nezadostno opišeta potek signala. . . . .	9
2.5	Delovanje okna na signal $x(u)$ pri kratkočasovni Fourierjevi transformaciji. .	11
2.6	Različni elementarni funkciji (atoma) dobljeni iz osnovne funkcije (okna analize) s premiki po času in po frekvenci. Na sliki sta prikazani združeno. . . .	12
2.7	Dobra frekvenčna in slaba časovna ločljivost nestacionarnega signala s STFT z uporabo daljšega Hammingovega okna. . . . .	13
2.8	Dobra časovna in slabša frekvenčna ločljivost nestacionarnega signala s STFT z uporabo krajšega Hammingovega okna. . . . .	14
2.9	Saklirni faktor $a$ vpliva na dolžino in frekvenco valčkov, oblika pa se ohranja. Tu je primer dveh valčkov pridobljenih iz osnovnega Morletovega valčka $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{j\omega_0 t} e^{-\frac{t^2}{2}}$ . . . . .	15
2.10	Časovno–frekvenčna ločljivost CWT na primeru Diracovega impulza. . . . .	16
2.11	Diskretizacije časovno–frekvenčne ravnine: vzorčenje po času (Shannon) vzorčenje po frekvenci (Fourier), ekvidistantna razdelitev časovno–frekvenčne ravnine (Gabor) in neenakomerna razdelitev (valčna transformacija). . . . .	18
2.12	Spektrograma signala sestavljenega iz linearno moduliranih signalov. Na levi strani (a) imamo izrazit pojav interference, desno (b) pa ne. Obakrat smo uporabili Gausovo okno, dolžine 23 točk. . . . .	21

---

2.13	Primer skalograma z Morletovimi valčki, $\psi(t) = \frac{1}{\sqrt{2\pi}}e^{j\omega_0 t}e^{-\frac{t^2}{2}}$ , na signalu sestavljenem iz dveh sinusnih valovanj. . . . .	22
2.14	Slika WVD cvrka. . . . .	23
2.15	Pojav interference med dvema točkama v časovno–frekvenčni ravnini. Pri WVD se interferenčni členi manifestirajo v geometričnem središču med dvema točkama, pasovi pa so vzporedni daljici, ki ju povezuje [Auger-97a]. . . . .	26
2.16	Slika WVD in PWD štirih Gaussovih atomov lociranih na ogliščih pravokotnika v časovno–frekvenčni ravnini. (a) WVD porazdelitev poleg slik posameznih atomov imamo še 6 področji interference (dva v sredini se prekrivata). (b) PWVD porazdelitev: zaradi glajenja interferenca ni tako izrazita (predvsem po frekvenčni smeri), frekvenčna ločljivost pa je slabša (atomi so v smeri frekvence bolj razmazani, tudi interferenčni členi so odebeljeni). . . . .	27
2.17	Primerjava slik sintetičnega signala z WVD, PWVD in SPWVD. . . . .	30
2.18	Primerjava slik WVD in AF preprostih sintetiziranih signalov [Hlawatsch-92].	31
2.19	Sliki signala sestavljenega iz dveh cvrkov: (a) WVD predstavitev (b) predstavitev z AF funkcijo. . . . .	32
2.20	Slika prikazuje pojav interference pri Margenau–Hillovi predstavitvi. Gre za drugačno geometrijo interference kot pri WVD. Signal je sestavljen iz dveh Gaussovih atomov. . . . .	34
2.21	Pojav interference pri različnih položajih štirih Gaussovih atomov. Uporabili smo CW pri $\sigma = 1$ . . . . .	36
2.22	Slike jeder v AF ravnini (tu gre za množenje jedra z osnovno predstavitvijo): (a) CW, manjša $\sigma$ , (b) CW, večja $\sigma$ , (c) ZAM, manjša $\alpha$ , (d) ZAM, večja $\alpha$ . . . . .	37
3.1	Slika prikazuje pojav prekrivanja pri DTWVD. Levo je DTWVD s prekrivanjem, desno pa slika brez prekrivanja. . . . .	43
3.2	Podvojevanje frekvence vzorčenja signala z interpolacijo. . . . .	44
3.3	Slika dveh Gaussovih atomov z diskretno verzijo WVD: (a) realni signal, (b) analitičen signal tvorjen iz realnega signala. . . . .	45
4.1	Različne slike Wigner–Villejeve porazdelitve ene osnovne periode fonema /e/: (a) 3D slika WVD v linearni skali (ptičja perspektiva), (b) 3D slika WVD v linearni skali (stranski pogled), (c) 2D slika WVD v linearni skali. . . . .	57
4.2	Slike različnih predstavitev dveh osnovnih period fonema /e/: (a) spektrogram, dolgo okno, (b) spektrogram, kratko okno, (c) Wigner–Villejeva predstavitev, (d) psevdo Wigner–Villejeva predstavitev, (e) zglajena psevdo Wigner–Villejeva predstavitev, (f) Pageova predstavitev. . . . .	60

---

4.3	Slike različnih predstavitev dveh osnovnih period fonema /e/ (nadaljevanje slike 4.2): (g) Margenau–Hillova predstavitev (h) psevdo Margenau–Hillova predstavitev, (i) Born–Jordanova predstavitev (zglajena verzija), (j) Choi–Williamsova predstavitev (zglajena verzija), (k) Zhao–Atlas–Marksova predstavitev (zglajena verzija). . . . .	61
4.4	Slika skalograma dveh osnovnih period fonema /e/. . . . .	63
4.5	Primerjava spektrogramov: (a) idealen potek signala v časovno-frekvenčni ravnini, (b) navaden spektrogram, (c) realocirani spektrogram. . . . .	66
4.6	Slike različnih realociranih predstavitev dveh osnovnih period fonema /e/: (a) realocirani spektrogram, (b) realocirana zglajena psevdo Wigener–Villejeva predstavitev, (c) realocirana psevdo Margenau–Hillova predstavitev. . . . .	67
4.7	Shema razpoznavalnika govora, ki temelji na Bayesovem odločitvenem pravilu. . . . .	68
4.8	Levo desni PMM. . . . .	70
4.9	Toopologija PMM za razpoznavanje glasov. . . . .	71
4.10	Razporeditev trikotnih filtrov po melodični skali. . . . .	74
4.11	Spektrogram besede /vOUk/. . . . .	77
4.12	WVD predstaviti besede /vOUk/: (a) WVD1, (b) WVD3 z uporabo Hammingovega okna pri $N_w = 512$ in $N_f = 160$ . . . . .	81
4.13	Predstavitve besede /vOUk/: (a) porazdelitev energije s CKD, (b) slika prvega odvoda predstavitve CKD. . . . .	86
4.14	Tri različni pristopi modeliranja prehodov po posameznih frekvenčnih pasovih predstavitve CKD. . . . .	88
4.15	Rezultati razpoznavanja monofonov z različnimi predstavitvami. . . . .	89
4.16	Rezultati razpoznavanja s SPEC, CKD6 in PMH2 predstavitvami posameznih monofonov govorne zbirke K211d. . . . .	90
4.17	Primerjava porazdelitve napak razpoznavanja: (a) SPEC - CKD6, (b) SPEC - PMH2. . . . .	91
4.18	Rezultati razpoznavanja difonov s posameznimi predstavitvami. . . . .	92

---

# Tabele

2.1	Povezava med lastnostmi jedra večlične funkcije in lastnostmi (str. 24) tako dobljenih predstavitev. . . . .	33
4.1	Alofoni širšega fonetičnega prepisa. Zapis v MRPA [Wells-92] formatu. . . .	70
4.2	Rezultati razpoznavanja z MFCC značilkami z uporabo PMM s 3 stanji na model. . . . .	78
4.3	Rezultati razpoznavanja z WV predstavitevijo (3 stanja na model). . . . .	82
4.4	Rezultati razpoznavanja s PWV predstavitevijo (3 stanja na model). . . . .	83
4.5	Rezultati razpoznavanja s PMH predstavitevijo (3 stanja na model). . . . .	84
4.6	Rezultati razpoznavanja s CKD predstavitevijo (3 stanja na model). . . . .	87



---

# 1 Uvod

---

1.1 Motivacije za delo

1.2 Pregled vsebine dela

---

V uvodnem poglavju bomo podali glavne ideje in motive za nastanek magistrskega dela. Poskušali bomo utemeljiti smiselnost uporabe časovno–frekvenčnih predstavitev, ki jih bomo podrobneje preučevali v naslednjih poglavjih tega dela.

V drugem delu pa bomo opisali kratko vsebino in zastavljene naloge, ki smo jih obravnavali v pričujočem delu.

---

## 1.1 Motivacije za delo

---

Pri analizi in obdelavi signalov se v veliki meri ukvarjamo z lastnostmi, naravo in nastankom signalov. Signal lahko v matematičnem smislu razumemo kot funkcijo ene ali več spremenljivk.

Najbolj osnoven zapis signala je obravnava signala kot *funkcije časa*. Tu nas zanima časovno spreminjanje količine, ki jo opisuje signal. Ta zapis je človeku najbolj naraven. Včasih pa nam ta zapis ne pove dovolj o lastnostih signala, ki ga opazujemo. Zato imamo na voljo druge možnosti opazovanja, druge predstavitve. V matematičnem smislu lahko signal kot funkcijo časa razstavimo po elementarnih baznih funkcijah na neskončno načinov. Se pravi, imamo lahko neskončno predstavitev signala. Med njimi izberemo tisto, s katero predstavimo signal na nam najbolj razumljiv in/ali fizikalno razložljiv način, ki dovolj dobro opiše karakteristike signala.

Poleg predstavitve signala po času, poznamo še eno prav tako temeljno predstavitev, predstavitev po frekvenci. Matematična orodja za tak opis funkcij je prvi razvil Fourier (1807), ki je preučeval enačbo širjenja toplote. Dokončno pa sta njegovo delo povzdignila Bunsen in Kirchhoff šestdeset let kasneje, ko sta proučevala lastnosti in naravo svetlobe. Danes lahko uvedbo frekvenčne analize signalov, ki jo imenujemo tudi *spektralna analiza*, utemeljujemo na več načinov. Prvi je, da nam spekter pove nekaj o viru signala. Hkrati je to danes tudi povsem naraven opis signala, ki ga tako obravnavamo kot vsoto posameznih valovanj. Valovanje pa je fizikalno povsem naraven in skoraj povsod prisoten pojav. Preslikava signala iz časovnega prostora v frekvenčni se imenuje *Fourierjeva transformacija*.

Tretji pogled na signal je, da opišemo *porazdelitev gostote energije* signala po času in/ali frekvenci. Energija je še ena temeljna fizikalna količina. Pri signalu definiramo energijo kot kvadrat absolutne vrednosti signala v času ali kot kvadrat magnitude Fourierjeve transformacije signala po frekvenci. S porazdelitvijo energije lahko tako na tretji način predstavimo signal. Opisovanje energije samo po času ali po frekvenci pa ni vedno dovolj. Tako namreč ne moremo ugotoviti odvisnosti ene količine od druge. Iz spektra lahko odčitamo, katere frekvence so prisotne v signalu, ne vemo pa kdaj v času so nastopile. V primeru signalov, kjer se spekter spreminja s časom, nam takšna predstavitev zato nezadostno opiše lastnosti signala. V ta velik razred spadajo takšni signali, kjer na nastanek in spreminjanje določenih frekvenc (spektralnih komponent) v signalu vplivamo s spreminjanjem fizikalnih parametrov vira oddajanja signala.

Tipičen primer takšnih signalov predstavlja človeški govor. Temu smo se v magistrskem delu tudi najbolj posvetili. Vokalni trakt pri človeku lahko razumemo kot cev z odprtino na eni strani. S pihanjem (potiskanjem zraka iz pljuč) v cev povzročimo na odprtem koncu cevi valovanje zraka z določeno frekvenco. Če je dolžina cevi vedno enako dolga in se njena oblika ne spreminja, bomo na koncu cevi dobili vedno enak zvok (valovanje s konstantno periodo). S spreminjanjem oblike in dolžine cevi pa se spreminja tudi valovanje. To pa je približen model obnašanja govoril pri človeku pri produkciji govora.

Takšne signale najboljše opišemo s časovno–frekvenčnimi porazdelitvami energije ali bolj natančno z vezanimi porazdelitvami gostote energije hkrati po času in frekvenci. Sig-

nale tako predstavljamo z delčki energije v določenem trenutku in v določeni frekvenci. Takšnim predstavitev pravimo *časovno–frekvenčne predstavitve*, katerim smo se v magistrskem delu najbolj posvetili predvsem v zvezi z analizo in razpoznavanjem govornega signala.

## 1.2 Pregled vsebine dela

---

Magistrsko delo lahko razdelimo na tri večje sklope. V prvem delu se bomo posvetili teoriji časovno–frekvenčnih predstavitev, v drugem diskretnim predstavitev, tretji del pa bo obravnaval predstavitve govornih signalov.

V prvem delu bomo razvrstili časovno–frekvenčne predstavitve glede na njihove lastnosti na več skupin. Najprej bomo podali nekaj osnovnih časovno–frekvenčnih meritev energije v signalu in nato prešli na zahtevnejše metode. V nadaljevanju bomo tako obravnavali linearne in kvadratne predstavitve. Bolj se bomo posvetili drugi skupini predstavitev, ki jo nadalje razdelimo na predstavitve Cohenovega razreda in afine predstavitve. Obravnavali bomo številne matematične lastnosti takšnih preslikav in ugotavljali njihovo primernost za analizo različnih tipov signalov. Ustrezno bomo umestili znane in že precej uveljavljene predstavitve v ta koncept pregleda porazdelitev. Največ časa bomo posvetili obravnavi interferenčnih členov, ki so posledica konstrukcije in definicije kvadratnih predstavitev. Podali bomo nekaj metod odprave oziroma zmanjševanja vpliva teh členov. Nekaj manj pozornosti pa bomo namenili afinim predstavitev, ki tvorijo drugo pomembno skupino energijskih (kvadratnih) porazdelitev. Prvi del skupaj z dodatkom A, v katerem bomo podali celoten pregled predstavitev, moramo predvsem razumeti kot pregled teorije časovno–frekvenčnih predstavitev signalov.

V drugem delu bomo obravnavali diskretne verzije predstavitev. Diskretizacija zveznih predstavitev predstavlja resno oviro uporabe teorije iz prvega dela. Izkaže se namreč, da vzorčenje časovno–frekvenčnega prostora in s tem tudi predstavitev pokvari številne lastnosti teh porazdelitev, ki potem lahko ne ustrezajo danemu primeru uporabe. Pogledali si bomo direktno diskretizacijo predstavitev in probleme, ki pri tem nastanejo. V drugem delu pa bomo podali alternativno posredno pot iskanja diskretnih predstavitev z uporabo karakterističnih funkcij in znanja linearne algebre.

Tretji del magistrskega dela bo obravnaval časovno–frekvenčne predstavitve govornih signalov. Na primeru slik časovno–frekvenčnih predstavitev govornih signalov si bomo ogledali lastnosti posameznih predstavitev. Ukvarjali se bomo tudi z izboljšavo predstavitev za namen nadaljnje analize govornih signalov. V zadnjem delu pa se bomo posvetili razpoznavanju govora na podlagi uporabe različnih časovno–frekvenčnih predstavitev. Podali bomo postopke pridobivanja različnih značilk govornega signala in opisali probleme, ki smo jih morali rešiti za uspešno uporabo novih predstavitev.

V zaključku bomo podali sklepne ugotovitve in pomembne prispevke dela ter smernice za nadaljnje delo.

---

# 2 Časovno–frekvenčne predstavitev signalov

---

## 2.1 Analiza signalov

## 2.2 Linearne časovno–frekvenčne predstavitve

## 2.3 Kvadratne časovno–frekvenčne predstavitve

---

Signale v časovno–frekvenčni prostoru opisujemo s časovno–frekvenčnimi predstavitvami. S kombinacijo analize signala skupno po času in frekvenci tako dobimo boljšo sliko časovnih potekov posameznih spektralnih komponent signala, saj takšne predstavitve opisujejo, katere spektralne komponente so v določenem trenutku prisotne v signalu.

Časovno-frekvenčne predstavitve signalov uporabljamo predvsem za analizo in sintezo nestacionarnih naključnih signalov. S tridimenzionalnimi slikami časovno–frekvenčnih preslikav lahko analiziramo, kako se spektralne komponente signalov spreminjajo s časom. Obratne preslikave pa uporabljamo za sintezo signalov bodisi za časovno spremenljivo filtriranje, odstranjevanje šumov v signalih, časovno prileganje ipd. V našem primeru bomo uporabili časovno–frekvenčne preslikave za statistično obdelavo, razpoznavanje in razvrščanje (govornih) signalov.

V tem poglavju si bomo najprej pogledali različne predstavitve signalov in podali lastnosti takšnih transformacij. V splošnem lahko ločimo časovno–frekvenčne predstavitve na linearne in kvadratne ali pa na točkovne in energijske predstavitve. V prvem primeru gledamo predvsem na lastnosti takšnih preslikav, v drugem pa na način konstrukcije posameznih transformacij. Tako točkovne predstavitve merijo podobnost signala in posameznih delov časovno–frekvenčnega prostora opisanega z elementarnimi funkcijami, ki jih poimenujemo *atomi*, energijske predstavitve pa ocenjujejo *porazdelitve energije* signala v časovno–frekvenčni ravnini.

Pri linearnih predstavitvah bomo podali predvsem primerjavo med kratkočasovno Fourierjevo in valčno transformacijo. V naslednjem razdelku pa si bomo ogledali najbolj znane predstavnike kvadratnih (bilinearne) predstavitev, ki jih nadalje razdelimo na razred *Cohenovih porazdelitev* in na razred *afinih predstavitev*.

V tem poglavju se naslanjamo na dela [Hlawatsch-92], [Cohen-95], [Qian-96] in [Auger-97a]. Slike predstavitev smo naredili z Matlabovim orodjem za časovno–frekvenčne predstavitve [Auger-97].

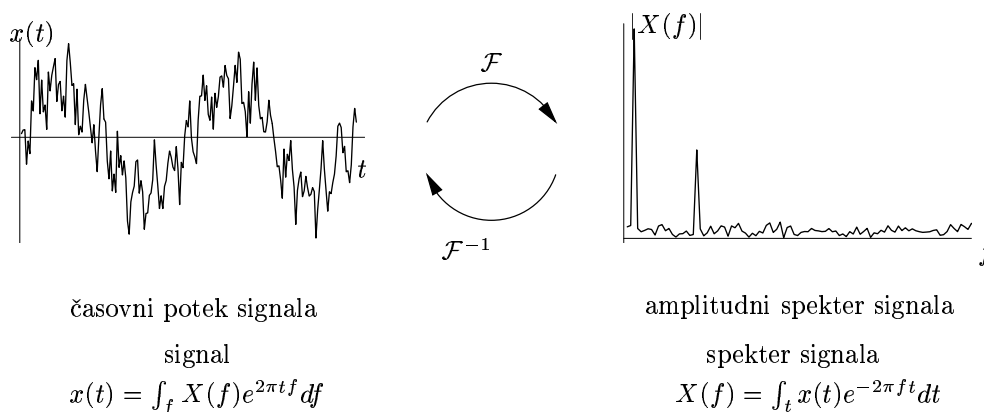
## 2.1 Analiza signalov

Z matematičnega stališča lahko dan signal predstavimo na več načinov. V naslednjih podpoglavjih si bomo pogledali osnovni predstavitvi signalov ločeno po času in frekvenci in podali nekaj osnovnih pojmov v zvezi z analizo in obdelavo signalov v časovno–frekvenčni ravnini.

### 2.1.1 Časovna in frekvenčna predstavitev signala

Najbolj naravno je, da obravnavamo signal kot (eno ali več dimenzionalno) funkcijo časa. S tako predstavitvijo signala pridobivamo informacijo o časovni spremembi dane (merjene) količine.

Druga povsem ekvivalentna (v smislu transformacije) predstavitev pa je prikaz signala v frekvenčnem prostoru. Tu lahko opazujemo, kako pogosto se dane količine v signalu spreminjajo. Vez med predstavitvami signalov v času in frekvenci predstavlja *Fourierjeva transformacija*, slika 2.1.



Slika 2.1: Časovna predstavitev signala in njegov (amplitudni) spekter. Fourierjeva transformacija  $\mathcal{F}$  predstavlja vez med časovno in frekvenčno predstavitvijo signala.

Frekvenčna predstavitev nam opiše zastopanost posameznih frekvenčnih komponent v signalu, ne podaja pa nam informacije, kdaj te frekvence zasledimo v signalu.

Tretja možnost predstavitve signala pa je seveda kombinacija časovne in frekvenčne predstavitve, kjer poskušamo opisati časovne poteke posameznih spektralnih (frekvenčnih) komponent v signalu.

Izbira predstavitve je seveda odvisna tudi od same analize in nadaljnje obdelave signala ter od signala samega. Tako nam lahko frekvenčna predstavitev zelo dobro opiše stacionarne signale, kjer se spekter signala s časom ne spreminja, slabo oziroma nepopolno pa nestacionarne.

V nadaljevanju bomo tako podali nekaj osnovnih idej in pojmov pri pridobivanju časovno–frekvenčne informacije iz signala.

### 2.1.2 Princip nedoločenosti

Prvi precej preprost način predstavitve signala v časovno–frekvenčni ravnini je, če obravnavamo količini  $|x(t)|^2$  in  $|X(f)|^2$  signala kot funkciji gostote verjetnosti. Pri tem moramo predpostaviti, da je energija signala končna:

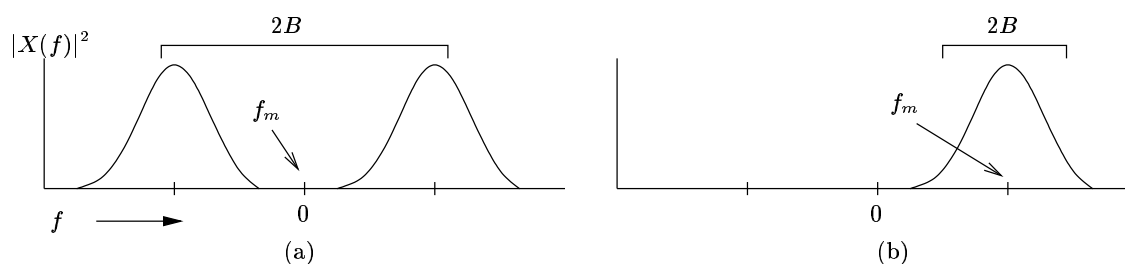
$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt < +\infty.$$

Tako lahko definiramo povprečno vrednost in standardno deviacijo obeh količin kot:

$$\begin{aligned} t_m &= \frac{1}{E_x} \int_{-\infty}^{\infty} t |x(t)|^2 dt && \text{povprečen čas,} \\ f_m &= \frac{1}{E_x} \int_{-\infty}^{\infty} f |X(f)|^2 df && \text{povprečna frekvenca,} \\ T^2 &= \frac{4\pi}{E_x} \int_{-\infty}^{\infty} (t - t_m)^2 |x(t)|^2 dt && \text{časovna širina,} \\ B^2 &= \frac{4\pi}{E_x} \int_{-\infty}^{\infty} (f - f_m)^2 |X(f)|^2 df && \text{frekvenčna širina.} \end{aligned}$$

S tem pa smo že predstavili signal v časovno–frekvenčni ravnini in sicer s povprečjem  $(t_m, f_m)$  in standardnimi odklonoma  $T$  po času in  $B$  po frekvenci.

Če malo bolje analiziramo količino  $|X(f)|^2$ , ugotovimo, da je v primeru realnega signala gostota spektralne energije  $|X(f)|^2$  simetrična funkcija frekvence  $f$ , saj velja  $X(-f) = X^*(f)$ . To pa pomeni, da imamo v tem primeru povprečno frekvenco  $f_m$  vedno 0, pa tudi standardna deviacija nam ne pove dovolj. Razmere v primeru realnega signala so prikazane na sliki 2.2(a), mi pa bi radi dobili “prave” vrednosti prikazane na sliki 2.2(b).



Slika 2.2: Povprečna frekvenca  $f_m$  in standardna deviacija  $B$  v primeru realnega signala (a) in dejanske iskane vrednosti (b).

To težavo lahko preprosto rešimo na dva načina. Najpreprosteje je, če v primeru realnih signalov računamo ti dve količini samo za pozitivni del. Tako dobimo  $B = \frac{2\pi}{E_x} \int_0^{\infty} (f - f_m)^2 |X(f)|^2 df$ . Druga možnost pa je uvedba analitičnega signala, ki ga bomo spoznali v nadaljevanju.

Produkt  $T \times B$  ima zanimivo lastnost. Iz definicije Fourierjeve transformacije ob upoštevanju Cauchy-Schwarzove neenakosti lahko pokažemo, da je

$$T \times B \geq 1. \quad (2.1)$$

To lastnost poznamo kot *Heisenberg-Gaborjevo neenakost* [Papo-77] ali kot *princip nedoločenosti*. Spodnja meja produkta (2.1) predstavlja omejitev pri opazovanju časovno–frekvenčnih karakteristik signala. Pove namreč, da ne moremo sočasno doseči visoke časovne in frekvenčne ločljivosti.

Spodnjo mejo produkta  $T \times B = 1$  dosežemo lahko le z Gaussovimi signali

$$x(t) = C \exp[-\alpha(t - t_m)^2 + j2\pi f_m(t - t_m)], \quad (2.2)$$

kjer sta  $C \in \mathbb{R}$  in  $\alpha \in \mathbb{R}_+$ .

### 2.1.3 Preproste časovno–frekvenčne predstavitve

Drugi preprosti način predstavitve nestacionarnih signalov sočasno v časovnem in frekvenčnem prostoru je ocenjevanje trenutne frekvence signala.

#### 2.1.3.1 Analitičen signal

Ob tem moramo najprej definirati pojem analitičnega signala. Eden izmed razlogov uvedbe analitičnega signala je primer, ki smo ga omenili že v prejšnjem poglavju (slika 2.2). Druge pa bomo spoznali v tem delu.

Analitični signal realnega signala  $x(t)$  je kompleksen signal  $x_a(t)$  podan kot

$$x_a(t) = x(t) + jHT(x(t)), \quad (2.3)$$

kjer je  $HT(x(t)) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(t')}{t-t'} dt'$  Hilbertova transformacija signala  $x$ . Signal  $x_a(t)$  je analitičen, ker ustreza Cauchy–Riemannovim pogojem o diferenciability.

Dá se pokazati [Cohen-95, str. 31], da je

$$E_x = \frac{1}{2}E_{x_a} \quad \text{in} \quad E_x = E_{HT(x)}. \quad (2.4)$$

To pa še ne pomeni, da je isto, če spremljamo porazdelitev energije realnega signala  $x(t)$  ali pa njegove analitične verzije  $x_a(t)$ .

Bistvena lastnost analitičnega signala je, da se znebimo problema simetričnosti. To lahko najbolj vidimo, če zapišemo pridružen analitični signal  $x_a(t)$  v frekvenčnem prostoru z  $X_a(f)$ , za katerega velja:

$$\begin{aligned} X_a(f) &= 0, & \text{če } f < 0, \\ X_a(f) &= X(0), & \text{če } f = 0, \\ X_a(f) &= 2X(f), & \text{če } f > 0. \end{aligned}$$

Tu se lepo vidi, da smo se znebili porazdelitve pri negativnih vrednostih frekvence  $f$ , hkrati pa smo tudi za dvakrat povečali energijo signala.

Zapis analitičnega signala v frekvenčnem prostoru predstavlja tudi način, kako iz danega signala tvorimo analitični.

### 2.1.3.2 Trenutna frekvenca in skupinska zakasnitev

Na analitičnih signalih lahko definiramo *trenutno amplitudo* in *trenutno frekvenco* signala kot

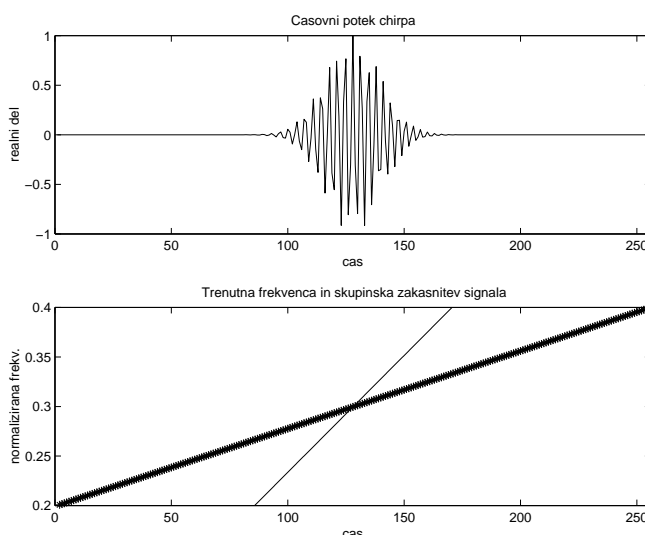
$$\begin{aligned} a_x(t) &= |x_a(t)| && \text{trenutna amplituda,} \\ f_x(t) &= \frac{1}{2\pi} \frac{d \arg x_a(t)}{dt} && \text{trenutna frekvenca.} \end{aligned} \quad (2.5)$$

$\arg x_a(t)$  pomeni fazo analitičnega signala  $x_a(t)$ , torej je trenutna frekvenca definirana kot odvod faze signala, kar si lahko razlagamo kot opisovanje spreminjanja frekvence po času.

Na podoben način v dualnem frekvenčnem prostoru lahko definiramo tudi *skupinsko zakasnitev* analitičnega signala kot

$$t_x(f) = -\frac{1}{2\pi} \frac{d \arg X_a(f)}{df}, \quad (2.6)$$

s katero ocenjujemo čas pojava posamezne frekvence v signalu.

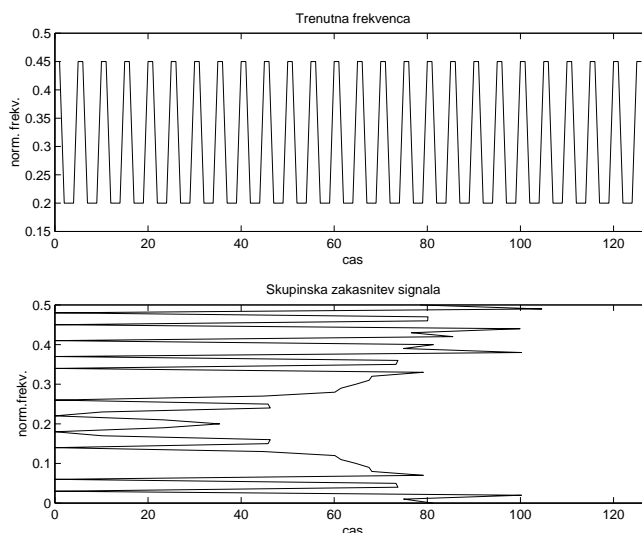


Slika 2.3: Prikaz trenutne frekvence in skupinske zakasnitve signala iz zgornje slike. Odebelenjena črta predstavlja oceno trenutne frekvence, tanjša pa skupinske zakasnitve cvrka. Tu gre za monokomponentni nestacionarni signal.

Načeloma sta ta dva pojma zelo sorodna, kar je lepo razvidno iz slike 2.3, kjer je z obema količinama prikazan Gaussovo amplitudno in linearno frekvenčno moduliran signal - *cvrk* (ang. *chirp*).

Slabost te metode pa je naslednja. Pri trenutni frekvenci implicitno privzamemo, da je v vsakem trenutku v signalu prisotna samo ena frekvenčna komponenta, katere spremembe merimo. Podobno je tudi z dualno količino - skupinsko zakasnitvijo signala. Tudi tu namreč privzamemo, da se dana frekvenca pojavi samo ob določenem trenutku v signalu. Na žalost pa je signalov, ki bi zadoščali tem predpostavkam v realnem svetu malo.





Slika 2.4: Prikaz trenutne frekvenca in skupinske zakasnitve signala  $x(t) = e^{j2\pi f_1 t} + e^{j2\pi f_2 t}$ , kjer je  $\frac{f_1}{f_s} = 0.1$  in  $\frac{f_2}{f_s} = 0.3$ . Tu gre za dvokomponentni stacionarni signal. Iz slike je razvidno, da obravnavani količini v tem primeru nezadostno opišeta potek signala.

Za primer si lahko pogledamo signal  $x(t) = e^{j2\pi f_1 t} + e^{j2\pi f_2 t}$ , kjer imamo v vsakem trenutku dve frekvenci. Tu s količinama trenutne frekvenca in skupinske zakasnitve v signalu ne moremo več pravilno analizirati signala, kot je razvidno iz slike 2.4.

Kot je razvidno iz primerov, nam ti dve količini ne dasta dovolj informacije v primeru večkomponentnih (frekvenčno) nestacionarnih signalov. To pa zato, ker je takšna predstavitev omejena samo na enodimenzionalne krivulje v časovno–frekvenčni ravnini (čas v odvisnosti od frekvenca in obratno). Zato si bomo v nadaljevanju ogledali predstavitve, kjer predstavimo signal z dvodimenzionalnimi ploskvami v časovno–frekvenčni ravnini.

## 2.2 Linearne časovno–frekvenčne predstavitve

Kot smo lahko videli iz prejšnjih primerov, nam osnovna analiza signalov ne pove dovolj v primeru nestacionarnih signalov, ki vsebujejo več kot eno samo frekvenčno komponento. Medtem ko sam časovni potek signala ne pove nič o frekvenčni sliki signala, nam spekter signala podaja samo informacijo o frekvenčni zastopanosti v signalu. Ravno tako nam ocene trenutne frekvenca in skupinske zakasnitve v signalu nezadostno opišejo časovno zastopanost frekvenčnih komponent in spreminjanje le-teh v odvisnosti od časa. Problem, ki je skupen vsem takšnim predstavitevam je, da poskušamo opisovati signal samo v odvisnosti od enega samega parametra bodisi od frekvenca ali časa.

V nadaljevanju si bomo ogledali predstavitve, kjer analiziramo signal v odvisnosti od časa in frekvenca skupaj. Se pravi, da bomo v časovno–frekvenčni ravnini opazovali dvodimenzionalne ploskve, ki ju definira par  $(t, f)$ . Zato takšnim predstavitevam pravimo časovno–frekvenčne predstavitve signalov. Formalno jih lahko zapišemo kot

preslikave:

$$\mathcal{T} : x(t) \rightarrow T_x(t, f).$$

V tem razdelku se bomo ukvarjali t.i. *linearnimi časovno–frekvenčnimi predstavitevami* (ang. *linear time–frequency representations*) ali *atomarnimi predstavitevami* (ang. *atomic decompositions*), kjer gre v bistvu za razvoj signala po temeljnih funkcijah (atomih) časovno–frekvenčnega prostora. Takšne transformacije signalov v vrsto po elementarnih funkcijah prostora so linearne, tako da za takšne predstavitve velja:

$$x(t) = c_1 x_1(t) + c_2 x_2(t) \Rightarrow T_x(t, f) = c_1 T_{x_1}(t, f) + c_2 T_{x_2}(t, f).$$

Najbolj znana in razširjena predstavnika atomarnih časovno–frekvenčnih predstavitev sta *kratkočasovna Fourierjeva transformacija* in *valčna transformacija*.

V nadaljevanju se bomo omejili samo na nekaj osnovnih definicij in lastnosti kratkočasovne Fourierjeve in valčne transformacije. Obravnavali ju bomo predvsem v luči linearnih predstavitev, poudarili njune razlike in pomanjkljivosti, ki nas bodo pripeljale k novim razredom časovno–frekvenčnih predstavitev.

## 2.2.1 Kratkočasovna Fourierjeva transformacija

### 2.2.1.1 Definicija

Kot smo že omenili Fourierjeva transformacija (FT) ne nosi informacije o času nastopa posameznih spektralnih komponent. Da bi dosegli časovno lokalizacijo FT, uporabimo okensko funkcijo  $h(u - t)$ , s katero pomnožimo signal, nad tem signalom pa potem izvedemo običajno FT. Če to počnemo v vsakem trenutku  $t$ , lahko definiramo *kratkočasovno Fourierjevo transformacijo STFT* (ang. *short–time Fourier transform*) kot

$$F_x(t, f; h) = \int_{-\infty}^{\infty} x(u) h^*(u - t) e^{-j2\pi f u} du. \quad (2.7)$$

Tu predstavlja funkcija  $h(t)$  kratkočasovno okno s središčem v točki  $t = 0$  in  $f = 0$  (glej sliko 2.5). Z množenjem signala  $x(u)$  s konjugirano okensko funkcijo  $h^*(u - t)$  dejansko izrežemo signal v okolici točke  $t$ , na katerem potem izračunamo spekter. Zato si lahko razlagamo *STFT* tudi kot izračun “lokalnega spektra” signala  $x(u)$  v okolici točke  $t$ .

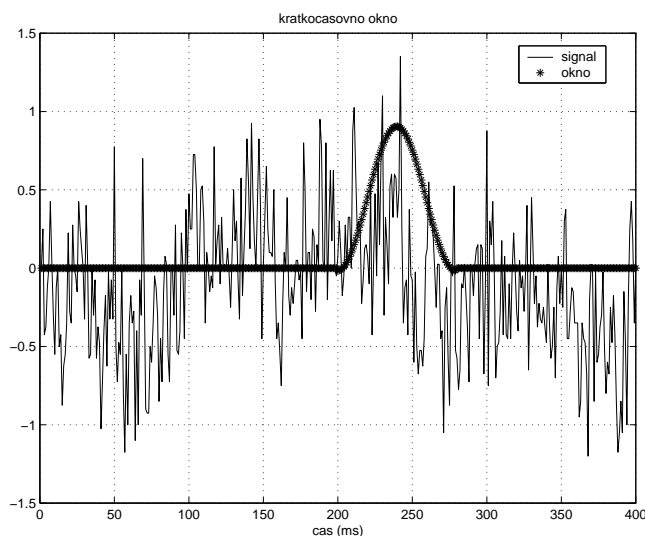
Ob predpostavki, da ima okenska funkcija končno energijo, lahko zapišemo signal  $x(t)$  kot

$$x(t) = \frac{1}{E_h} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_x(u, \xi; h) h(t - u) e^{j2\pi t \xi} du d\xi,$$

kjer je  $E_h = \int_{-\infty}^{+\infty} |h(t)|^2 dt$ . To pa pomeni, da smo signal razstavili po elementarnih funkcijah - atomih, oblike

$$h_{t,f} = h(u - t) \exp[j2\pi f u],$$

ki jih dobimo iz osnovne okenske funkcije  $h(t)$  s translacijo po času in po frekvenci (modulaciji). Na sliki 2.6 lahko vidimo, dve takšni bazni funkciji dobljeni iz Hammingovega okna.



Slika 2.5: Delovanje okna na signal  $x(u)$  pri kratkočasovni Fourierjevi transformaciji.

STFT pa lahko zapišemo tudi z *dualnimi* spektralnimi oblikami signala in okna:

$$F_x(t, f; h) = \int_{-\infty}^{\infty} X(\xi) H^*(\xi - f) e^{j2\pi(\xi - f)t} d\xi,$$

kjer sta  $X$  in  $H$  Fourierjevi transformaciji  $x$  in  $h$ . V tem primeru si lahko razlagamo STFT kot filtriranje signala  $x(u)$  z banko filtrov s prenosnimi funkcijami  $H^*(\xi - f)$ , [Rabiner-93, str. 84].

### 2.2.1.2 Osnovne lastnosti

- Poleg pogoja linearnosti ohranja *STFT* še pomike po frekvenci in času, velja namreč:

$$\begin{aligned} \tilde{x}(t) = x(t) e^{j2\pi f_0 t} &\Rightarrow F_{\tilde{x}}(t, f; h) = F_x(t, f - f_0; h), \\ \tilde{x}(t) = x(t - t_0) &\Rightarrow F_{\tilde{x}}(t, f; h) = F_x(t - t_0, f; h) e^{-j2\pi t_0 f}. \end{aligned}$$

- Signal  $x(t)$  lahko rekonstruiramo iz STFT tudi z uporabo drugih oken analize, saj velja:

$$x(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F_x(u, \xi; h) g(t - u) e^{j2\pi t \xi} du d\xi,$$

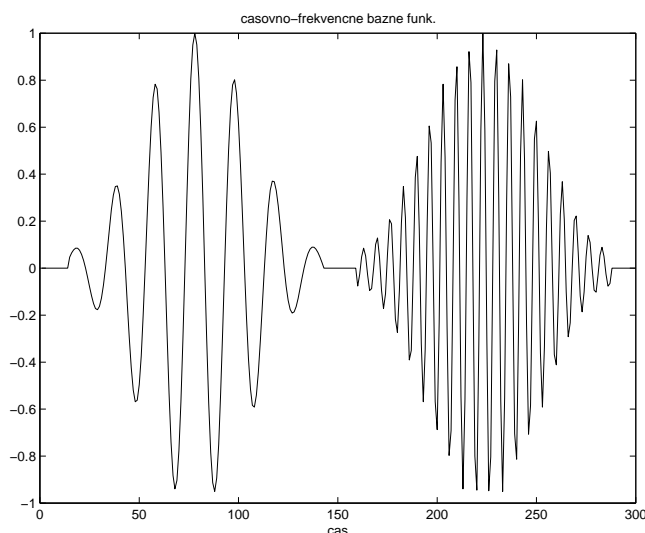
pri čemer izberemo  $g(u)$  takšen, da ustreza pogoju

$$\int_{-\infty}^{+\infty} g(t) h^*(t) dt = 1.$$

### 2.2.1.3 Ločljivost v časovno–frekvenčni ravnini

Časovno ločljivost STFT najbolj obravnavamo, če vzamemo za signal *Diracov impulz*:

$$x(t) = \delta(t - t_0) \Rightarrow F_x(t, f; h) = e^{-j2\pi t_0 f} h(t - t_0).$$



Slika 2.6: Različni elementarni funkciji (atoma) dobljeni iz osnovne funkcije (okna analize) s premiki po času in po frekvenci. Na sliki sta prikazani združeno.

Tako lahko ugotovimo, da je časovna ločljivost odvisna od trajanja osnovne okenske funkcije  $h$ . Podobno lahko analiziramo frekvenčno ločljivost, če vzamemo za signal kompleksni sinusoid (ali Diracov impulz v spektru):

$$x(t) = e^{j2\pi f_0 t} \Rightarrow F_x(t, f; h) = e^{-j2\pi t f_0} H(f - f_0).$$

Iz tega izračuna pa je razvidno, da je frekvenčna ločljivost STFT odvisna od pasovne (frekvenčne) širine okenske funkcije  $h$ . Tako smo dobili dva izključujoča si pogoja: za dobro časovno ločljivost bi namreč potrebovali kratko časovno okno  $h(t)$ , po drugi strani pa zahtevamo za boljšo frekvenčno ločljivost ozkopasovni filter, kar pomeni dolgo časovno okno  $h(t)$ . To pa nam onemogoča *princip nedoločenosti*.

Najbolj očiten primer najboljše časovne ločljivosti je, če si za okensko funkcijo izberemo Diracov impulz:

$$h(t) = \delta(t) \Rightarrow F_x(t, f; h) = x(t)e^{-j2\pi f t}.$$

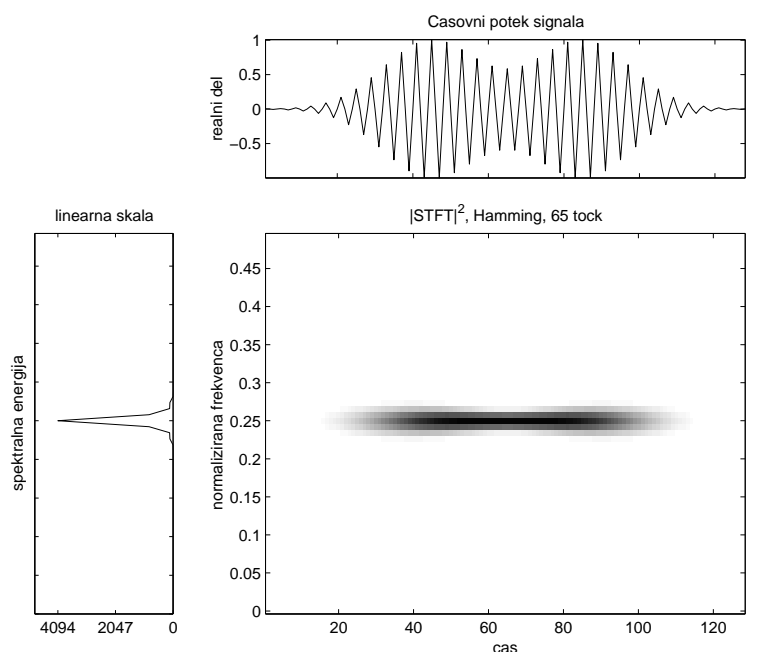
S tako STFT dosežemo najboljše časovno ločljivost, ne dobimo pa nobene informacije o spektru. Podobno velja tudi v primeru, ko si za okensko funkcijo izberemo konstanto:

$$h(t) = 1 \quad (H(f) = \delta(f)) \Rightarrow F_x(t, f; h) = X(f),$$

kjer spremenimo STFT v Fourierjevo transformacijo in tako ne dobimo nobene informacije o časovnem poteku signala.

Seveda je smiselno vzeti okna “nekje vmes” med obema skrajnostima. Izbira oken je odvisna od signalov, ki jih obdelujemo in analize same, [Harris-78].

Kot primer kako oblika in dolžina okna vplivata na časovno–frekvenčno ločljivost, si pogledjmo kombinacijo dveh Gaussovo amplitudno moduliranih signalov (Gaussovi atomi) skupne dolžine 128 odtipkov in enotne konstantne frekvence. V prvem primeru vzamemo za STFT Hammingovo okno dolžine 65 točk. Kot lahko vidimo iz slike 2.7



Slika 2.7: Dobra frekvenčna in slaba časovna ločljivost nestacionarnega signala s STFT z uporabo daljšega Hammingovega okna.

smo s širokim oknom dosegli dobro frekvenčno ločljivost, ne moremo pa v popolnosti ločiti komponent signala v času.

V drugem primeru, slika 2.8, pa smo vzeli krajše okno dolžine 17 točk in s tem poslabšali frekvenčno ločljivost ter izboljšali časovno, tako da že lahko ločimo čas nastopa posameznih komponent v signalu.

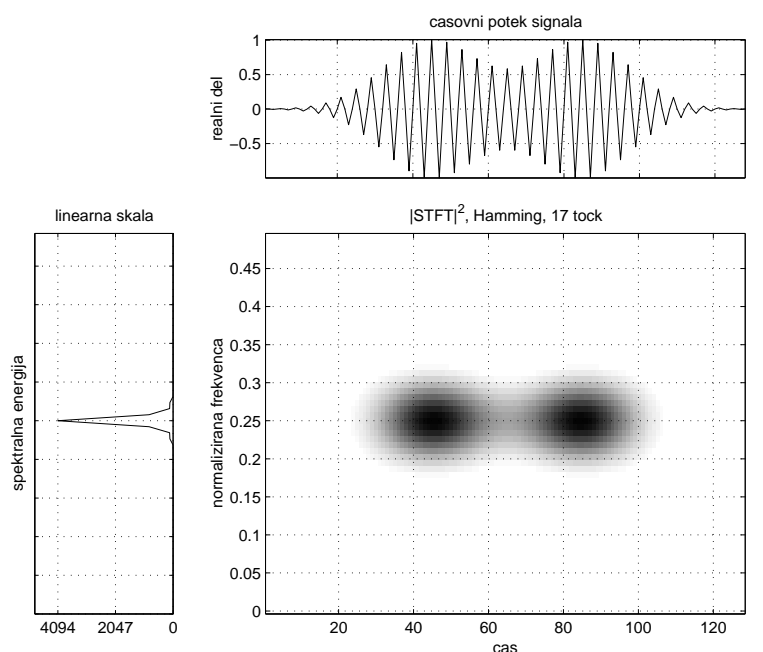
## 2.2.2 Valčna transformacija

Drugi primer linearnih časovno–frekvenčnih predstavitev je *valčna transformacija*, *WT* (ang. *wavelet transform*). Uporabo valčne transformacije pri analizi signalov je prvi predstavil Daubechies [Daubechies-90], [Daubechies-92]. Gre pa za to, da poskušamo visokofrekvenčne pojave v signalu opazovati z visoko časovno ločljivostjo in ob upoštevanju principa nedoločenosti z ustrezno nižjo frekvenčno ločljivostjo, nizkofrekvenčne komponente pa opazujemo z višjo frekvenčno in temu ustrezno nižjo časovno ločljivostjo.

### 2.2.2.1 Definicija

Tu se bomo omejili samo na *zvezne valčne transformacije* (*CWT*). Princip je podoben kot pri STFT, tudi tu namreč razstavimo signal  $x(t)$  v prostoru, ki ga razpenjajo valčne funkcije (*valčki*), ki jih dobimo s skaliranjem in translacijo *izhodiščne valčne funkcije* (ang. *mother wavelet*). Tako definiramo zvezno valčno transformacijo kot

$$T_x(t, a; \psi) = \int_{-\infty}^{+\infty} x(s) \psi_{t,a}^*(s) ds, \quad (2.8)$$



Slika 2.8: Dobra časovna in slabša frekvenčna ločljivost nestacionarnega signala s STFT z uporabo krajšega Hammingovega okna.

kjer je  $\psi_{t,a}(s) = |a|^{-1/2} \psi\left(\frac{s-t}{a}\right)$ . Parameter  $a$  predstavlja skalirni faktor, če je  $|a| > 1$  valček  $\psi$  raztegnemo, če je  $|a| < 1$  valček  $\psi$  skrčimo. Tako v primeru CWT zamenjamo translacije po frekvenci osnovnih elementarnih funkcij (STFT) s skaliranjem (krčenjem ali raztezanjem) po času. Formalno lahko interpretiramo CWT tudi kot časovno–frekvenčno predstavitev, če zapišemo  $a = f_0/f$ , kjer je  $f_0$  središčna frekvenca valčka, [Rioul-92].

Osnovna lastnost skalirnega faktorja  $a$  je, da se z njegovim spreminjanjem spreminja tako dolžina kot tudi frekvenčna širina valčka, oblika pa ostaja enaka (slika 2.9). Če to primerjamo s STFT, lahko ugotovimo, da se okenska funkcija pri STFT ohranja, samo premikamo jo po časovno–frekvenčni ravnini, medtem ko pri CWT uporabljamo kratko časovno okno pri visokih frekvencah in daljše pri nizkih, slika 2.10. Frekvenčna širina je proporcionalna frekvenci, velja namreč  $\frac{B}{f} = Q = \text{const}$ . S tem tako delno izboljšamo ločljivost (ni več enakomerna) navkljub principu nedoločenosti.

### 2.2.2.2 Lastnosti

- Poleg linearnosti ohranja valčna transformacija premike in raztege po času:

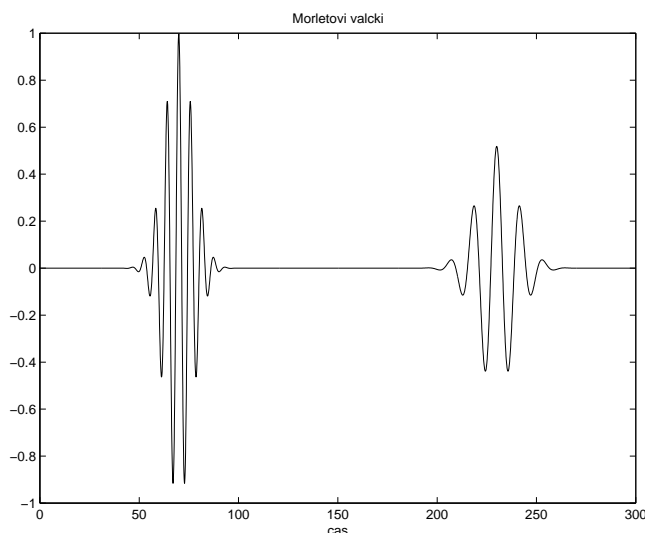
$$y(t) = \sqrt{|a_0|} x(a_0(t - t_0)) \quad \Rightarrow \quad T_y(t, a; \psi) = T_x(a_0^*(t - t_0), a/a_0; \psi).$$

Ne ohranja pa premike po frekvenci.

Takšnim preslikavam pravimo *afine preslikave*.

- Signal  $x$  lahko, podobno kot pri STFT, zapišemo z valčno transformacijo:

$$x(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} T_x(s, a; \Phi) \psi_{s,a}(t) ds \frac{da}{a^2},$$



Slika 2.9: Saklirni faktor  $a$  vpliva na dolžino in frekvenco valčkov, oblika pa se ohranja. Tu je primer dveh valčkov pridobljenih iz osnovnega Morletovega valčka  $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{j\omega_0 t} e^{-\frac{t^2}{2}}$ .

kjer valček  $\Phi$  ustreza pogoju

$$\int_{-\infty}^{+\infty} \Phi(f) \psi^*(f) \frac{df}{|f|} = 1.$$

- Tudi tu velja princip nedoločenosti, le da je tu ločljivost odvisna od frekvence: frekvenčna ločljivost (časovna ločljivost) se z višanjem frekvence slabša (boljša).

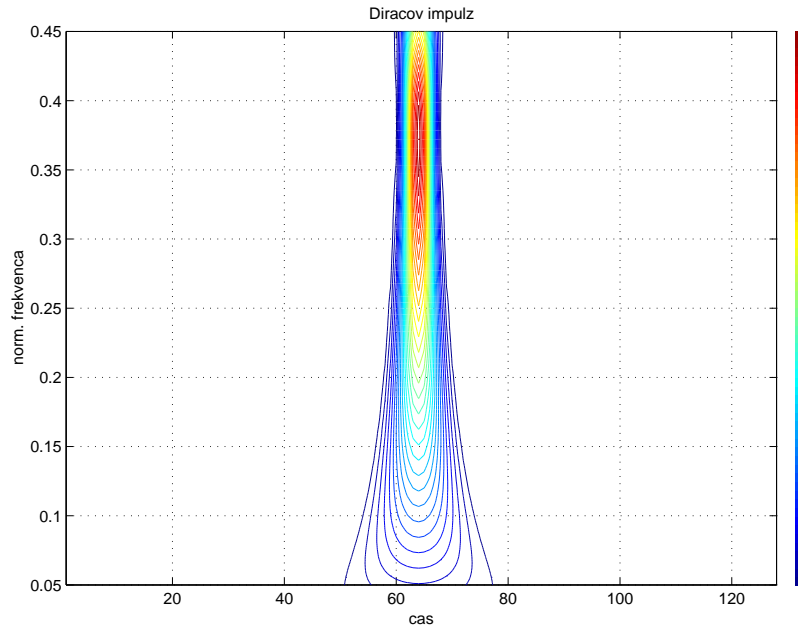
### 2.2.3 Diskretne linearne predstavitve

Pri analizi in obdelavi signalov si v večini primerov ne moremo privoščiti zvezno podanih signalov, ampak zgolj njihove vzorčene približke (diskretne verzije signalov). Zato je potrebno vso teorijo zveznih predstavitev tudi ustrezno prenesti v diskreten prostor, kar je v primeru nelinearnih predstavitev lahko zahtevna naloga.

#### 2.2.3.1 Diskretna kratkočasovna Fourierjeva transformacija

Pri zvezni STFT vsaka elementarna funkcija (atom) opisuje del časovno–frekvenčne ravnine s površino  $T \times B$  (dokaz v [Qian-96]). To pa pomeni, da se v zveznem primeru lahko deli, ki jih opisujejo posamezne bazne funkcije, med seboj prekrivajo. Takšna predstavitev pa je redundantna. Zato lahko STFT diskretiziramo, in sicer tako da izberemo ekvidistantno (pravokotno) mrežo ravnine (to je najbolj naravna izbira, saj pri STFT pokrivamo prostor z atomi, ki jih dobimo s translacijami po času in frekvenci):

$$F_x[n, m; h] = F_x(nt_0, mf_0; h) = \int_{-\infty}^{+\infty} x(u) h^*(u - nt_0) e^{-j2\pi mf_0 t_0} du,$$



Slika 2.10: Časovno–frekvenčna ločljivost CWT na primeru Diracovega impulza.

$m, n \in \mathbb{Z}$ , kjer je  $(nt_0, mf_0)$  središče premaknjene bazne funkcije  $h_{nt_0, mf_0}$ . Da bi minimizirali redundantnost diskretne predstavitve, morata  $t_0$  in  $f_0$  zadoščati pogoju

$$t_0 \times f_0 \leq 1.$$

V primeru, ko je  $t_0 \times f_0 > 1$  pokrijemo z atomi  $h_{nt_0, mf_0}$  premalo prostora (dobimo “luknje” v časovno–frekvenčnem prostoru), v primeru, ko je  $t_0 \times f_0 < 1$ , atomi  $h_{nt_0, mf_0}$  tvorijo ogrodje prostora (ne baze), v primeru  $t_0 \times f_0 = 1$ , pa dobimo ortonormirano bazo prostora časovno–frekvenčnih funkcij [Daubechies-92].

V primeru diskretne STFT rekonstruiramo signal po naslednji zvezi:

$$x(t) = \sum_n \sum_m F_x[n, m; h] g_{n, m}(t),$$

kjer je  $g_{n, m}(t) = g(t - nt_0) \exp[j2\pi mf_0 nt_0]$ . Podobno kot v zveznem primeru je tu zveza med različnima okenskima funkcijama izražena s pogojem

$$\frac{1}{f_0} \sum g\left(t + \frac{k}{f_0} - nt_0\right) h^*(t - nt_0) = \delta_k, \quad \forall t,$$

kjer je  $\delta_0 := 1$  in  $\delta_k := 0$ , za  $k \neq 0$ .

Če obdelujemo vzorčeni signal  $x[n]$ , dolžine  $N$ , s frekvenco vzorčenja  $\frac{1}{T}$ , potem izberemo  $t_0 = kT$ ,  $k \in \mathbb{N}$ . Tako dobimo znani zvezi

$$F_x[n, m; h] = \sum_k x[k] h^*[k - n] \exp[-j2\pi mk], \quad -\frac{N}{2} \leq m \leq \frac{N}{2} \quad (2.9)$$

$$x[k] = \sum_n \sum_m F_x[n, m; h] g[k - n] \exp[j2\pi mk], \quad (2.10)$$

ki ju učinkovito rešujemo z uporabo hitre Fourierjeve transformacije.



### 2.2.3.2 Gaborjeva transformacija

V splošnem pravimo izraža signala

$$x(t) = \sum_n \sum_m F_x[n, m; h] g_{n,m}(t),$$

kjer je  $g_{n,m}(t) = g(t - nt_0) \exp[j2\pi m f_0 n t_0]$ , *Gaborjeva transformacija* [Gabor-46]. Prvotno se je pri Gaborjevi transformaciji za elementarne funkcije  $g(t)$  izbiralo Gaussova okna (2.2), s katerimi lahko najboljše opišemo (v smislu ločljivosti) časovno–frekvenčni prostor. Gaborjevo transformacijo lahko posplošimo, če dovolimo uporabo poljubnih normiranih funkcij  $g(t)$ .

Funkcijam  $g_{n,m}$  pravimo *Gaborjevi nastavki* (ang. *logons*), koeficientom  $F_x[n, m; h]$  pa *Gaborjevi koeficienti*. Gaborjevi koeficienti tako merijo podobnost med signalom in Gaborjevimi funkcijami  $g_{n,m}$  v kvadratni okolici časovno–frekvenčne ravnine s središčem v točki  $(nt_0, m f_0)$ .

Tudi tu velja podobno kot pri diskretni STFT (ki je le primer Gaborjeve transformacije) potreben in ne zadosten pogoj  $t_0 f_0 \leq 1$  za pokritje prostora z Gaborjevimi funkcijami.

Ob dodatnem pogoju biortogonalnosti funkcij  $g$  in  $h$

$$\int_{-\infty}^{+\infty} g_{n,m}(t) h_{n',m'}^*(t) dt = \delta_{n-n'} \delta_{m-m'},$$

lahko uporabimo zvezi za izračun Gaborjevih koeficientov (2.9) in rekonstrukcijo signala iz Gaborjeve transformacije (2.10).

Natančnejši pregled teorije in uporabe Gaborjeve transformacije lahko najdemo v [Qian-96], [Qian-99], [Hlawatsch-92].

### 2.2.3.3 Diskretna valčna transformacija

Tu bomo podali samo idejo diskretizacije časovno–frekvenčne ravnine pri valčni transformaciji. Več o teoriji diskretizacije valčne transformacije si lahko preberete v [Daubechies-92], [Rioul-91], [Marušič-98].

Ker imamo pri valčni transformaciji neenakomerno časovno–frekvenčno ločljivost, tudi prostor diskretiziramo tako, in sicer

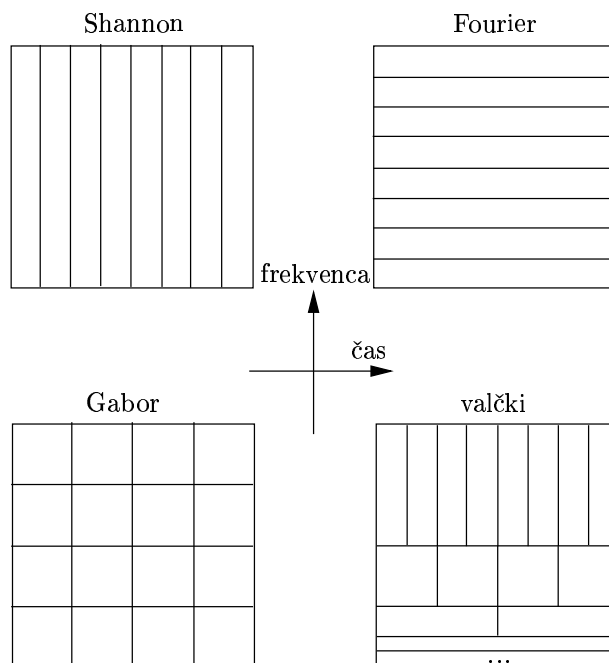
$$(t, a) = (nt_0 a_0^{-m}, a_0^{-m}), \quad t_0 > 0, a_0 > 0, \quad m, n \in \mathbb{Z}.$$

V tem primeru lahko definiramo *diskretno valčno transformacijo (DWT)* (ang. *discrete wavelet transform*):

$$T_x[n, m; \psi] = a_0^{m/2} \int_{-\infty}^{+\infty} x(u) \psi_{n,m}^*(u) du, \quad m, n \in \mathbb{Z},$$

kjer je  $\psi_{n,m}(u) = \psi(a_0^m u - nt_0)$ . Ob izbiri parametrov  $a_0 = 2$ ,  $t_0 = 1$  dobimo diskretizacijo časovno–frekvenčne ravnine v *diadični skali*, kot je prikazano na sliki 2.11.

V tem primeru dejansko računamo koeficiente po oktavah frekvenčnega prostora. Taka izbira diskretizacije pa tudi definira *ortonormalno bazo*  $\{\psi_{n,m}(u); m, n \in \mathbb{Z}\}$  prostora časovno–frekvenčnih funkcij [Daubechies-92].



Slika 2.11: Diskretizacije časovno–frekvenčne ravnine: vzorčenje po času (Shannon) vzorčenje po frekvenci (Fourier), ekvidistantna razdelitev časovno–frekvenčne ravnine (Gabor) in neenakomerna razdelitev (valčna transformacija).

Glavna slabost takšne diskretizacije je v tem, da DWT ni več neodvisna od premika po času.

## 2.3 Kvadratne časovno–frekvenčne predstavitve

V prejšnjem razdelku smo si ogledali linearne predstavitve signalov v časovno–frekvenčnem prostoru. Tem predstavitev pravimo tudi atomarne, ker v osnovi razstavljamo signal po elementarnih funkcijah (atomih), ki pokrivajo dele časovno–frekvenčnega prostora. Elementarne funkcije izbiramo tako, da te čimbolje pokrivajo ves prostor.

Drugi način opisa signala v časovno–frekvenčni ravnini je z opazovanjem porazdelitve *energij* signalov.

Energija signala  $x(t)$  je definirana kot

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(f)|^2 df. \quad (2.11)$$

Funkciji  $|x(t)|^2$  in  $|X(f)|^2$  lahko obravnavamo kot *porazdelitve* energije po času ali (v drugem primeru) po frekvenci. Na podoben način pa lahko definiramo porazdelitve

energije  $\rho(t, f)$  kot funkcije časa in frekvence:

$$E_x = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho(t, f) dt df. \quad (2.12)$$

Ker je energija signala v splošnem kvadratna funkcija signala, takšne predstavitve imenujemo *kvadratne predstavitve*. Nekako bolj naravno (v smislu izpeljave) pa take predstavitve imenujemo tudi *energijske porazdelitvene predstavitve* ali krajše *energijske porazdelitve*.

Iz zgornjih dveh definicij (2.11) in (2.12) je smiselno za porazdelitev  $\rho(t, f)$  zahtevati, da ustreza t.i. *robnima pogojem*:

$$\int_{-\infty}^{+\infty} \rho(t, f) dt = |X(f)|^2, \quad (2.13)$$

$$\int_{-\infty}^{+\infty} \rho(t, f) df = |x(t)|^2. \quad (2.14)$$

To pomeni, da z integriranjem  $\rho(t, f)$  po eni spremenljivki dobimo porazdelitev po drugi spremenljivki.

Več o energijskih porazdelitvah lahko najdemo v [Cohen-95], [Qian-96], [Hlawatsch-92].

### 2.3.1 Prehod iz linearnih h kvadratnim predstavitvam

V nadaljevanju bom predstavil dve zelo znani in razširjeni predstavitvi signalov, ki spadata k energijskim porazdelitvam in predstavljata osnovo za nastanek in razvoj energijskih predstavitev signalov v časovno–frekvenčni ravnini. To sta *spektrogram* in *skalogram*. Spektrogram je predstavnik večje skupine predstavitev, ki sodijo v *Cohenov razred* energijskih porazdelitev, skalogram pa je predstavnik t.i. *afinih predstavitev*, ki jih bomo natančneje opisali v naslednjih poglavjih.

#### 2.3.1.1 Spektrogram

*Spektrogram* opisuje spektralno energijo signala  $x(u)h^*(u-t)$ , ki smo ga predhodno omejili na območje okna, torej

$$S_x(t, f) = \left| \int_{-\infty}^{\infty} x(u)h^*(u-t)e^{-j2\pi fu} du \right|^2. \quad (2.15)$$

Ob predpostavki, ki smo jo upoštevali že pri STFT, da je energija okna 1, velja pogoj 2.12, torej:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S_x(t, f) dt df = E_x.$$

Spektrogram tako predstavlja realno nenegativno energijsko porazdelitev. S  $S_x(t, f)$  opazujemo energijo signala na območju časovno–frekvenčne ravnine s središčem v točki  $(t, f)$ .

## Lastnosti

- *ohranitev premikov*

Neposredno iz definicije spektrograma velja, da se vrednosti spektrograma za translacije po času in frekvenci ohranjajo:

$$\begin{aligned} y(t) = x(t - t_0) &\Rightarrow S_y(t, f) = S_x(t - t_0, f), \\ y(t) = x(t) \exp[j2\pi f_0 t] &\Rightarrow S_y(t, f) = S_x(t, f - f_0). \end{aligned}$$

Predstavitve, ki zadoščajo temu pogoju, uvrščamo v Cohenov razred, ki si ga bomo podrobneje ogledali v nadaljevanju.

- *časovno–frekvenčna ločljivost*

Časovno–frekvenčna ločljivost je enaka kot pri STFT. Ravno tako tudi za to predstavitev velja princip nedoločenosti, kar predstavlja glavno slabost te predstavitve.

- *pojavnost interference*

Spektrogram je kvadratna (bilinearna) predstavitev, torej tu ne velja lastnost linearnosti, da je vsota dveh signalov vsota dveh spektrogramov, ampak lastnost *kvadratne superpozicije*, in sicer:

$$y(t) = x_1(t) + x_2(t) \Rightarrow S_y(t, f) = S_{x_1}(t, f) + S_{x_2}(t, f) + 2\Re\{S_{x_1, x_2}(t, f)\},$$

kjer predstavlja  $S_{x_1, x_2}(t, f) = F_{x_1}(t, f)F_{x_2}^*(t, f)$  križni spektrogram signalov  $x_1(t)$  in  $x_2(t)$ . Križni členi predstavljajo pojav interference na slikah kvadratnih predstavitev. Hlawatsch [Hlawatsch-91] je pokazal, da so interferenčni členi izraziti, če se spektrograma  $S_{x_1}(t, f)$  in  $S_{x_2}(t, f)$  prekrivata. To pa pomeni, da je prispevek križnih členov spektrograma  $S_{x_1, x_2}(t, f)$  manjši, če so komponente  $x_1(t)$  in  $x_2(t)$  signala čimbolj narazen, kar je posledica slabe ločljivosti spektrograma. V tem primeru predstavlja torej slabša ločljivost prednost predvsem v primerjavi z ostalimi kvadratnimi predstavitvami, ki jih bomo še spoznali.

## Primer

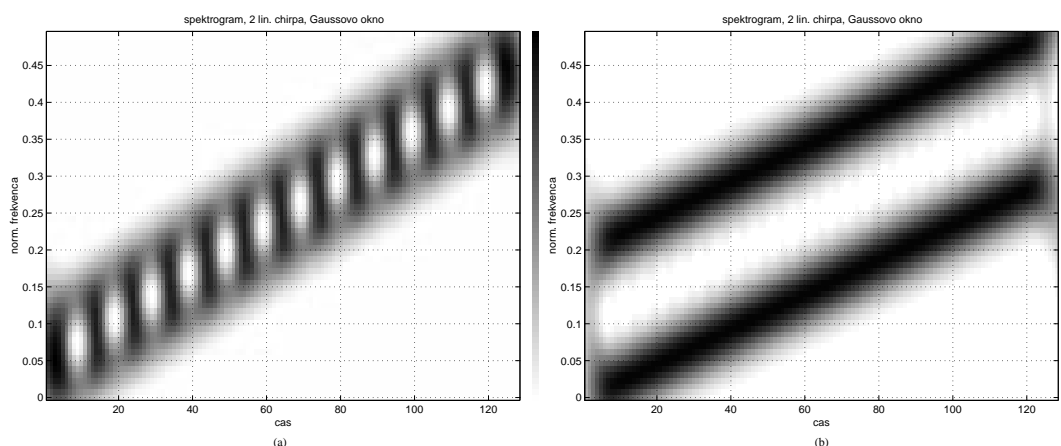
V naslednjem primeru si bomo ogledali pojav interference. Vzemimo signal sestavljen iz dveh linearno moduliranih signalov (cvrkov), ki sta sorazmerno relativno skupaj v časovno–frekvenčni ravnini. Torej

$$x(t) = x_1(t) + x_2(t),$$

kjer je  $x_i(t) = \exp[j2\pi t(f_{1,i} + \frac{f_{2,i} - f_{1,i}}{N} t)/f_s]$  in  $f_s$  je frekvenca vzorčenja,  $f_{1,i}$  spodnja meja modulacije,  $f_{2,i}$  pa zgornja meja modulacije.

Vzemimo torej signal  $x$  na 128 točkah s podatki  $f_{1,1} = 0$ ,  $f_{2,1} = 0.4$  in  $f_{1,2} = 0.1$ ,  $f_{2,2} = 0.5$ . V drugem primeru pa malce razmaknimo FM komponente: signal  $x$  s podatki  $f_{1,1} = 0$ ,  $f_{2,1} = 0.3$  in  $f_{1,2} = 0.2$ ,  $f_{2,2} = 0.5$ .

Na sliki 2.12 lahko vidimo, da imamo v prvem primeru zelo izrazit pojav interference, ker so komponente signala blizu skupaj, v drugem primeru pa interferenčnih členov ni (razen na robovih).



Slika 2.12: Spektrograma signala sestavljenega iz linearno moduliranih signalov. Na levi strani (a) imamo izrazit pojav interference, desno (b) pa ne. Obakrat smo uporabili Gaussovo okno, dolžine 23 točk.

### 2.3.1.2 Skalogram

Podobno kot spektrogram definiramo tudi *skalogram* kot kvadrat absolutne vrednosti valčne transformacije:

$$SW_x(t, f) = \left| \int_{-\infty}^{+\infty} x(s) \psi_{t,a}^*(s) ds \right|^2. \quad (2.16)$$

Podobno kot pri spektrogramu, lahko pokažemo, da je

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} SW_x(t, f) dt \frac{da}{a^2} = E_x,$$

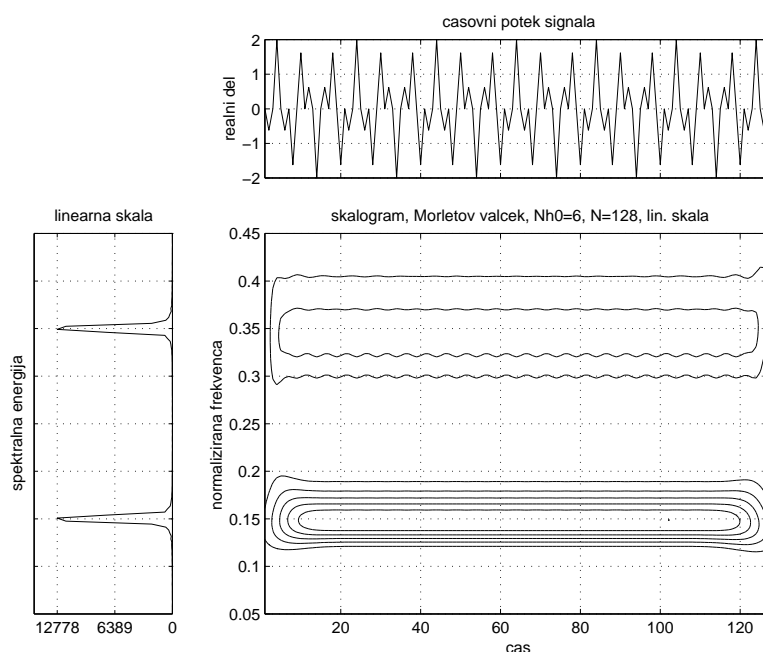
le da imamo tu opravka s skalo  $a$ , ki pa je, kot smo že pri CWT transformaciji pokazali, direktno povezana s pojmom frekvence. Takšnim predstavitev pravimo *afine predstavitve*, ki jih bomo kasneje natančneje obravnavali.

Analogno kot pri CWT imamo tudi tu opravka s Heisenberg-Gaborjevo neenakostjo (princip nedoločenosti), za razliko s spektrogramom pa je tu ločljivost po frekvenci in času odvisna od frekvence.

Za primer si pogledajmo sliko 2.13 skalograma signala sestavljenega iz dveh sinusnih valovanj<sup>1</sup>. Tu lepo vidimo, da je frekvenčna ločljivost odvisna od frekvence: ločljivost je večja pri višjih frekvencah  $f$ .

Podobno kot pri spektrogramu je tudi s pojavom interference. Križni členi so večji, bližje kot so posamezne komponente signala (bolj se prekrivajo skalogrami posameznih komponent). To pa pomeni, da je tudi tu interferenca vezana na ločljivost predstavitve.

<sup>1</sup>S pojmom sinusno valovanje ali tudi kompleksni sinusoid označujemo signale oblike  $x(t) = e^{2\pi f_0 t}$ .



Slika 2.13: Primer skalograma z Morletovimi valčki,  $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{j\omega_0 t} e^{-\frac{t^2}{2}}$ , na signalu sestavljenem iz dveh sinusnih valovanj.

### 2.3.2 Cohenov razred

Kvadratnih predstavitev, ki zadoščajo pogojem (2.12), (2.13) ali (2.14) je več. Zato za predstavitve  $\rho_x$  ponavadi zahtevamo še kakšne dodatne omejitve glede na namen in nadaljnjo analizo ter obdelavo signala.

Ena izmed pomembnih lastnosti, ki jih lahko pripišemo energijski predstavitvi signala, je, da *ohranja premike po času in frekvenci* [Cohen-66]. Takšne predstavitve tvorijo *Cohenov razred* porazdelitev.

Spektrogram, ki smo ga predstavili v prejšnjem poglavju, spada med takšne predstavitve, saj zadošča pogoju (2.12), ohranja translacije po času in frekvenci, ne zadošča pa robnima pogojema (2.13) in (2.14).

V nadaljevanju si bomo ogledali nekaj Cohenovih predstavitev, med njimi najpomembnejšo *Wigener–Villejevo porazdelitev*, ki je osnova za vse ostale predstavitve.

#### 2.3.2.1 Wigener–Villejeva porazdelitev

##### Definicija

*Wigener–Villejeva porazdelitev*<sup>2</sup> (*WVD*) je definirana kot

$$W_x(t, f) = \int_{-\infty}^{+\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-j2\pi f\tau} d\tau, \quad (2.17)$$

oziroma

$$W_x(t, f) = \int_{-\infty}^{+\infty} X(f + \xi/2)X^*(f - \xi/2)e^{j2\pi\xi t} d\xi.$$

Porazdelitev WVD ima veliko matematičnih lastnosti, ki jih bom opisal v naslednjem podpoglavju. Če poudarimo le najpomembnejše: WVD je realna funkcija časa in frekvence, velja namreč

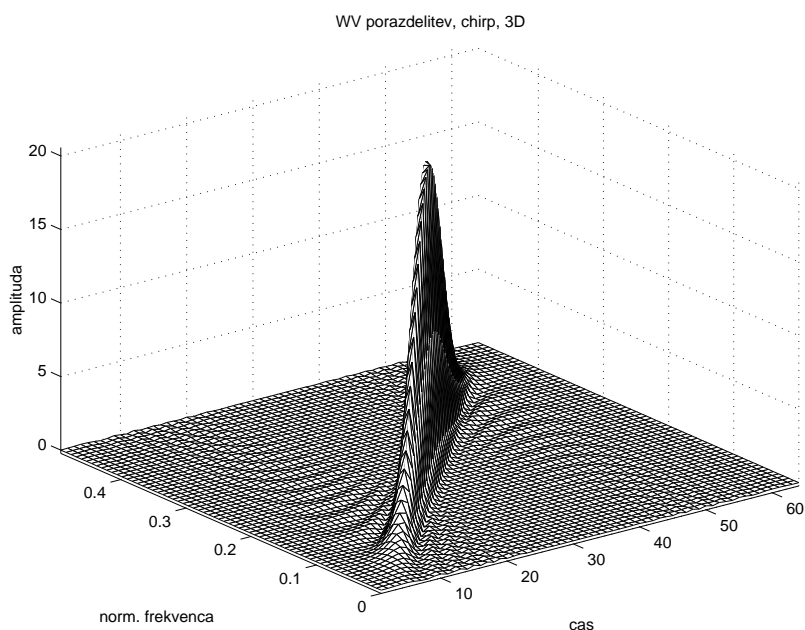
$$W_x(t, f) = W_x^*(t, f),$$

iz česar neposredno sledi realnost predstavitve. Ostale pomembne lastnosti so še ohranjanje premikov po času in frekvenci ter robni pogoji.

Kot primer si pogledjmo že omenjen amplitudno Gaussovo in linearno frekvenčno moduliran signal t.i. cvrk (ang. chirp):

$$x(t) = \left(\frac{\alpha}{\pi}\right)^{1/4} \exp\left[-\frac{\alpha}{2}t^2 + j\beta t^2\right],$$

s trenutno frekvenco  $f_x(t) = 2\beta t$ .



Slika 2.14: Slika WVD cvrka.

Izračunajmo WVD

$$W_x(t, f) = \exp[-\alpha t^2] \sqrt{\frac{\alpha}{\pi}} \int_{-\infty}^{+\infty} \exp\left[-\frac{\alpha}{4}\tau^2\right] \exp[-j(f-2\beta t)\tau] d\tau = 2 \exp\left[-\left(\alpha t^2 + \frac{1}{\alpha}(f-2\beta t)^2\right)\right],$$

<sup>2</sup>Pri časovno–frekvenčni analizi signalov se v primeru kvadratnih predstavitev mešata pojma predstavitev (ang. representation) in porazdelitev (ang. distribution). V prvem primeru mislimo sliko porazdelitve, v drugem pa zgolj porazdelitev energije. Ker je pri verjetnosti funkcija porazdelitve verjetnosti nenegativna, realna s ploščino 1, v primeru porazdelitev energije pa to ni vedno res, je bolj smiselno, če govorimo o predstavitev. Vsekakor sta pri takšni analizi signalov v primeru kvadratnih predstavitev pojma ekvivalentna.

Predstavitev cvrka  $W_x(t, f)$  je prikazana na sliki 2.14. Iz slike 2.14 je razvidna tudi skoraj popolna časovno–frekvenčna ločljivost.

### Lastnosti

Podali bomo nekaj lastnosti WVD, ki jih lahko najdemo v [Hlawatsch-92], večino izpeljav in dokazov pa v [Qian-96] in [Cohen-95].

1. *energijska porazdelitev*: z integriranjem po času in frekvenci  $W_x(t, f)$  dobimo energijo signala  $x$ :

$$E_x = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_x(t, f) dt df.$$

2. *robna pogoja*: z integriranjem  $W_x(t, f)$  po posameznih spremenljivkah dobimo energijo spektra in trenutno moč signala  $x(t)$ :

$$\begin{aligned} \int_{-\infty}^{+\infty} W_x(t, f) dt &= |X(f)|^2, \\ \int_{-\infty}^{+\infty} W_x(t, f) df &= |x(t)|^2. \end{aligned}$$

3. *realna funkcija*:

$$W_x(t, f) \in \mathbb{R}, \quad \forall t, f.$$

4. *ohranitev premikov*: WVD ohranja premike po času in frekvenci:

$$\begin{aligned} y(t) = x(t - t_0) &\Rightarrow W_y(t, f) = W_x(t - t_0, f), \\ y(t) = x(t) \exp[j2\pi f_0 t] &\Rightarrow W_y(t, f) = W_x(t, f - f_0). \end{aligned}$$

5. *ohranitev raztegov*: WVD ohranja tudi raztege:

$$y(t) = \sqrt{k}x(kt); \quad k > 0 \quad \Rightarrow \quad W_y(t, f) = W_x(kt, \frac{f}{k}).$$

Ta lastnost skupaj s premiki po času definira *afine predstavitve*, tako da WVD spada tudi v ta razred predstavitev.

6. *ohranitev konvolucije*: konvolucija signalov  $h$  in  $x$  pomeni konvolucijo ustreznih WVD predstavitev:

$$y(t) = \int_{-\infty}^{+\infty} h(t - s)x(s)ds \quad \Rightarrow \quad W_y(t, f) = \int_{-\infty}^{+\infty} W_h(t - s, f)W_x(s, f)ds..$$

7. *produkt v konvolucijo*: to je dualna lastnost prejšnje. Če moduliramo signal  $x$  (množimo) s funkcijo  $m$ , to pomeni konvolucijo po frekvenci ustreznih WVD:

$$y(t) = m(t)x(t) \quad \Rightarrow \quad W_y(t, f) = \int_{-\infty}^{+\infty} W_m(t, f - \xi)W_x(s, \xi)d\xi.$$



8. *ohranitev nosilcev*: če ima signal kompakten nosilec v času (v frekvenci), potem ima ustrezni WVD tudi isti kompaktni nosilec v času (v frekvenci):

$$\begin{aligned} x(t) = 0, \quad |t| > T &\Rightarrow W_x(t, f) = 0, \quad |t| > T, \\ X(f) = 0, \quad |f| > B &\Rightarrow W_x(t, f) = 0, \quad |f| > B. \end{aligned}$$

9. *unitarnost*: transformacija WVD ohranja skalarni produkt po času:

$$\left| \int_{-\infty}^{+\infty} x(t)y^*(t)dt \right|^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_x(t, f)W_y^*(t, f)dtdf.$$

To enačbo poznamo tudi pod imenom *Moyalova formula* [Moyal-49]. Pomembna pa je predvsem za teoretične izpeljave, saj je nesmiselno, da bi računali produkte WV preslikav namesto skalarnega produkta signalov.

10. *trenutna frekvenca*: trenutno frekvenco signala lahko izračunamo kot prvi moment po frekvenci WVD:

$$f_x(t) = \frac{\int_{-\infty}^{+\infty} fW_{x_a}(t, f)df}{\int_{-\infty}^{+\infty} W_{x_a}(t, f)df},$$

kjer je  $x_a$  pridružen analitični signal danega signala  $x$ .

11. *skupinska zakasnitev*: dualno lastnost skupinsko zakasnitev signala tako izračunamo kot prvi moment po času WVD:

$$t_x(f) = \frac{\int_{-\infty}^{+\infty} tW_{x_a}(t, f)dt}{\int_{-\infty}^{+\infty} W_{x_a}(t, f)dt}.$$

12. *popolna ločljivost linearno moduliranih signalov (lokalizacija cvrkov)*:

$$x(t) = e^{j2\pi\nu_x(t)t}, \text{ kjer je } \nu_x(t) = \nu_0 + 2\beta t \Rightarrow W_x(t, f) = \delta(f - (\nu_0 + \beta t)).$$

To smo si že ogledali v primeru na sliki 2.14.

## Interferenca

Ker spada WVD med kvadratne predstavitve, velja tudi tu, podobno kot pri spektrogramu, načelo kvadratne superpozicije:

$$W_{x+y}(t, f) = W_x(t, f) + W_y(t, f) + 2\Re\{W_{x,y}(t, f)\},$$

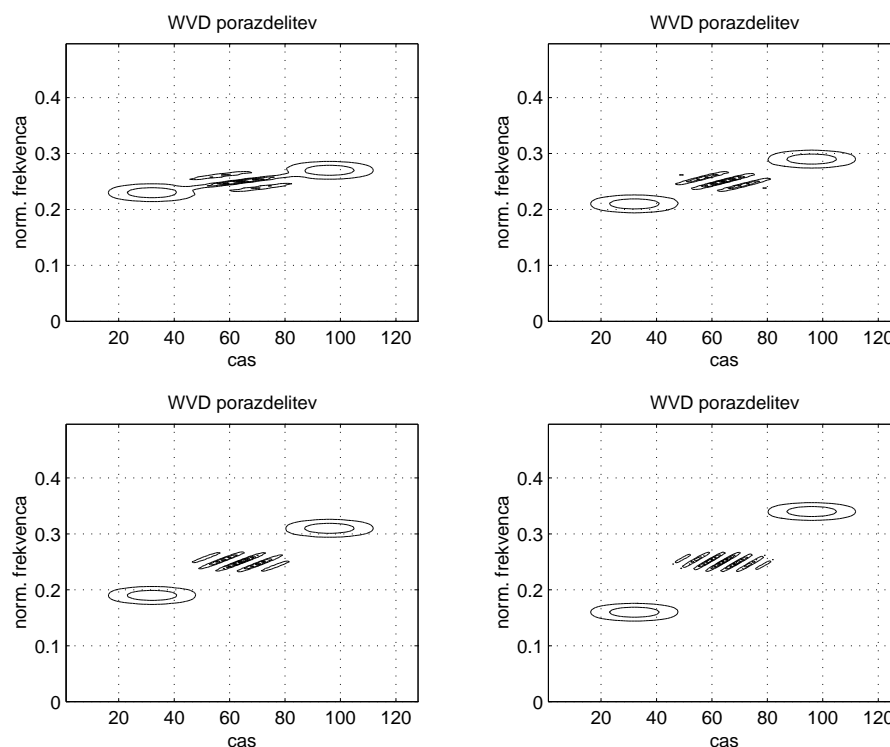
kjer predstavlja

$$W_{x,y}(t, f) = \int_{-\infty}^{+\infty} x(t + \tau/2)y^*(t - \tau/2)e^{-j2\pi f\tau}d\tau$$

križno WV predstavitev signalov  $x$  in  $y$ . Posplošitev tega načela na  $N$  komponent signala je preprosta: v poštev pridejo vse možne kombinacije,  $\binom{N}{2}$ , križnih WVD.

Za razliko od spektrograma, pa tu velikost interferenčnih členov ni odvisna od razmaknjenosti posameznih komponent v časovno–frekvenčni ravnini. Členi se pojavljajo v vsakem primeru. Ravno to pa je največji problem WVD, saj na sliki takšne predstavitve tako ne moremo ločiti interferenčnih pojavov od tistih delov, ki dejansko opisujejo potek signala. Izkaže se, da je pojav interference posledica lastnosti WVD (zlasti robnih pogojev, trenutne frekvence, skupinske zakasnitve, lokalizacije cvrkov, unitarnosti, ...) [Loughlin-93]. Tako da moramo pri kvadratnih predstavitvah vedno tehtati med upoštevanjem lastnosti in pojavom interference.

Kljub temu pa lahko nekaj povemo o *geometriji interference* med posameznimi komponentami v signalu [Auger-97a]. Pri transformaciji WVD se v splošnem interference pojavi v okolici geometrične sredine med dvema točkama v časovno–frekvenčni ravnini. Sami interferenčni pasovi pa so vzporedni daljici, ki veže ti dve točki.



Slika 2.15: Pojav interference med dvema točkama v časovno–frekvenčni ravnini. Pri WVD se interferenčni členi manifestirajo v geometričnem središču med dvema točkama, pasovi pa so vzporedni daljici, ki ju povezuje [Auger-97a].

To lahko vidimo na sliki 2.15, kjer smo obravnavali dva t.i. Gaussova atoma<sup>3</sup>, ki smo ju medsebojno razmikali. Tu lahko opazujemo geometrijske lastnosti interferenčnih pasov. Prav tako pa se z oddaljenostjo frekvenčnih komponent spreminja tudi število pasov na sliki.

<sup>3</sup>Gaussovo amplitudno modelirani signali  $x(t) = e^{-\pi(\frac{t-t_0}{T})^2}$  s središčem v točki  $t_0$  in sorazmerno kratkim trajanjem  $T$ .

### Psevdo Wigner–Villejeva porazdelitev

Pri WVD predstavitvi (2.17) gre dejansko za integriranje količine

$$q_x(t, \tau) = x(t + \tau/2)x^*(t - \tau/2)$$

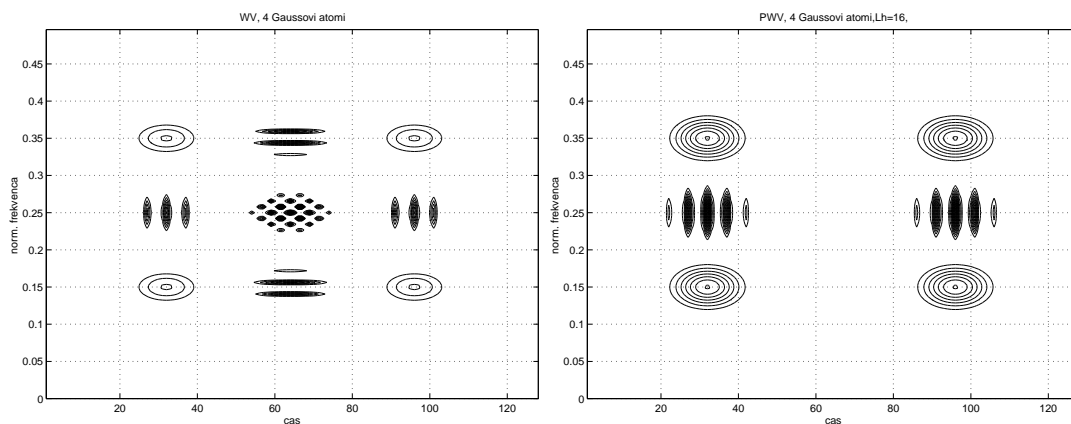
za  $\tau = -\infty$  do  $\tau = +\infty$ . Pri praktični obdelavi in analizi signalov pa to ne moremo zagotoviti, zato po zgledu kratkočasovne Fourierjeve analize količino  $q_x$  množimo z ustreznim oknom ter tako interval integriranja ustrezno zmanjšamo. Tako dobimo novo porazdelitev

$$PW_x(t, f) = \int_{-\infty}^{+\infty} h(\tau)x(t + \tau/2)x^*(t - \tau/2)e^{-j2\pi f\tau} d\tau, \quad (2.18)$$

kjer je  $h(t)$  ustrezno regularno okno. Tako dobimo novo predstavitev *psevdo Wigner–Villejevo porazdelitev (PWVD)*. Definicijo (2.18) si lahko razlagamo tudi kot glajenje (konvolucija) po frekvenci WVD predstavitve, velja namreč:

$$PW_x(t, f) = \int_{-\infty}^{+\infty} H(f - \xi)W_x(t, \xi)d\xi,$$

kjer je  $H(f)$  Fourierjeva transformacija okna  $h(t)$ . Z glajenjem WVD dosežemo dvoje: zmanjšamo vpliv interference in izgubimo nekaj pomembnih lastnosti WVD-ja (slika 2.16).



Slika 2.16: Slika WVD in PWD štirih Gaussovih atomov lociranih na ogliščih pravokotnika v časovno–frekvenčni ravnini. (a) WVD porazdelitev poleg slik posameznih atomov imamo še 6 področji interference (dva v sredini se prekrivata). (b) PWVD porazdelitev: zaradi glajenja interference ni tako izrazita (predvsem po frekvenčni smeri), frekvenčna ločljivost pa je slabša (atomi so v smeri frekvence bolj razmazani, tudi interferenčni členi so odebeljeni).

Na sliki 2.16 lahko vidimo, da PWD zmanjša interferenčne pasove pravokotne na smer frekvence (zaradi glajenja po frekvenci), medtem ko pasovi pravokotni glede na časovno os ostanejo takšni kot pri WVD. Po drugi strani pa PWD ne ustreza robnim pogojem, unitarnosti, ne ohranja nosilca po frekvenci; zato se zmanjša tudi frekvenčna ločljivost.

### 2.3.2.2 Cohenov razred

#### Definicija

Kot smo že omenili, zajema Cohenov razred predstavitve, ki ohranjajo *premake po času in frekvenci*. Ta lastnost je zelo pomembna pri analizi signalov, saj sama translacija tako ne vpliva na vrednost predstavitve, temveč samo na ustrezen premik funkcije predstavitve.

Za takšne energijske predstavitve, pa se dá pokazati [Cohen-95], da jih lahko zapišemo kot

$$C_x(t, f; k) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{j2\pi\xi(s-t)} k(\xi, \tau) x(s + \tau/2) x^*(s - \tau/2) e^{-j2\pi f\tau} d\xi ds d\tau,$$

kjer je jedro  $k(\xi, \tau)$  dvodimenzionalna funkcija, ki jo imenujemo *parametrizacijska funkcija*. Torej predstavitve, ki jih lahko zapišemo na gornji način z ustreznimi parametrizacijskimi funkcijami, tvorijo Cohenov razred porazdelitev. Gornji pogoj lahko prevedemo tudi na

$$C_x(t, f; \Pi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(s - t, \xi - f) W_x(s, \xi) ds d\xi, \quad (2.19)$$

kjer je

$$\Pi(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k(\xi, \tau) e^{-j2\pi(f\tau + \xi t)} dt df$$

dvodimenzionalna Fourierjeva transformacija parametrizacijske funkcije  $k$ .

Primer take predstavitve je WVD, kjer je  $\Pi(t, f) = \delta(t)\delta(f)$  dvodimenzionalni Diracov impulz, oziroma  $k(\xi, \tau) = 1$ .

Zvezo (2.19) lahko interpretiramo tudi kot glajenje (po času in frekvenci) WVD porazdelitve: to pa pomeni, da takšne predstavitve v bistvu izvirajo iz WVD, le da jo ustrezno gladijo in s tem ustrezno manjšajo vpliv interferenčnih členov.

Posplošitev znanih energijskih predstavitev na Cohenov razred pomeni orodje za načrtovanje novih predstavitev. Tako lahko dobimo nove predstavitve, s tem da preprosto podamo parametrizacijsko funkcijo  $k$ . Pokaže se, da z ustrezno izbiro jeder  $k$  lahko opišemo vse predstavitve iz Cohenovega razreda [Cohen-95], hkrati pa lahko modeliramo tudi nove predstavitve glede na namen uporabe in tip signalov, ki jih obdelujemo.

#### Glajenje v obeh smereh

Tu bomo pokazali, da tudi spektrogram spada v Cohenov razred. Z upoštevanjem Moyalove formule (lastnost 9 na str. 24) in lastnosti 7 (str. 24) lahko zapišemo spektrogram kot

$$S_x(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_h(s - t, \xi - f) W_x(s, \xi) ds d\xi \quad (2.20)$$

To pa pomeni, da je spektrogram tudi predstavitev Cohenovega razreda, kjer je  $\Pi(s, \xi) = W_h(s, \xi)$  WVD okenske funkcije  $h$ . Tu gre v bistvu za glajenje WVD predstavitve

sočasno v obeh smereh (po času in frekvenci). Funkcija glajenja je odvisna od okna: kratkočasovno okno pomeni manjše glajenje po času in večje po frekvenci, kar pri spektrogramu povzroči že znano boljše časovno in slabšo frekvenčno ločljivost in obratno.

### Ločeno glajenje

V prejšnjem primeru smo nadzirali glajenje z okensko funkcijo  $h$ . S tem pa nekako izgublamo nadzor glajenja po posameznih smereh v časovno–frekvenčni ravnini. Da bi povečali kontrolo nad glajenjem po posameznih smereh, lahko zapišemo funkcijo glajenja kot produkt funkcij po posameznih smereh:

$$\Pi(t, f) = g(t)H(-f),$$

kjer je  $H(f)$  Fourierjeva transformacija okna  $h(t)$ . Tako dobimo novo predstavitev:

$$SPW_x(t, f) = \int_{-\infty}^{+\infty} h(\tau) \int_{-\infty}^{+\infty} g(s-t)x(s+\tau/2)x^*(s-\tau/2)ds e^{-j2\pi f\tau} d\tau,$$

ki jo poznamo pod imenom *zglajena pseudo Wigner–Villejeva porazdelitev (SPWD)*. Za razliko od spektrograma imamo tu kontrolo nad glajenjem in s tem ločljivostjo po času in frekvenci: bolj kot gladimo po času in/ali frekvenci, slabšo ločljivost dobimo po času in/ali frekvenci, medtem ko pri spektrogramu z boljšanjem ene ločljivosti slabšamo drugo.

Kako pa je z interferenco? Že prej smo nakazali namen glajenja v zmanjšanju interference posameznih predstavitev. To je seveda odvisno od namena uporabe časovno–frekvenčnih slik signalov. Če zahtevamo ločljivost, se moramo sprijazniti (vsaj v primeru Cohenovih predstavitev) z bolj izrazito interferenco, sicer pa žrtvujemo nekaj natančnosti na račun “lepše” slike. Tu lahko gladimo v obeh smereh, po času in frekvenci, kar se pokaže tudi pri odstranjevanju interference. Za razliko od PWVD tu lahko uspešno odstranimo interferenčne pasove tako vzdolž frekvenčne osi kot tudi vzdolž časovne, hkrati pa izgublamo ločljivost v obeh smereh.

Na sliki 2.17 lahko primerjamo interferenco in ločljivost WVD, PWVD in SPWVD. V primeru signala na sliki 2.17 sestavljenega iz vsote Gaussovega amplitudno medeliranega signala in kompleksnega sinusoida (konstantna frekvenca 0.15 normalizirane) se pojavljajo interferenčne “lise” vzdolž časovne osi, zato si s PWVD ne moremo pomagati. V tem primeru sintetičnega signala lahko žrtvujemo nekaj ločljivosti na račun bolj jasne slike.

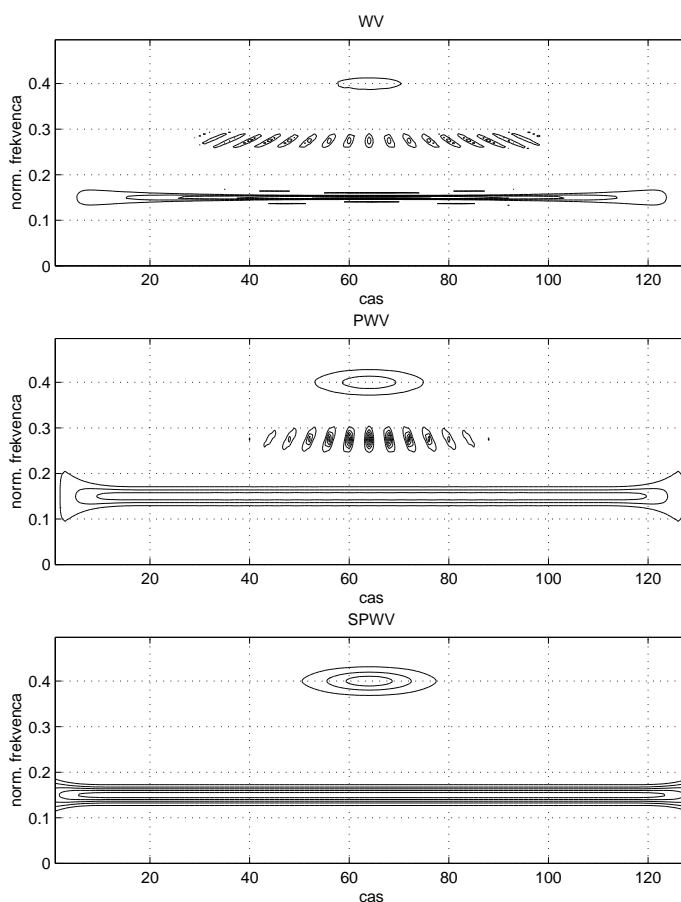
### 2.3.2.3 Večlična funkcija

#### Definicija in lastnosti

Podobno kot smo definirali pri trenutni frekvenci signala njeno dualno količino, skupinsko zakasnitev, lahko to storimo tudi tu.

Definirajmo nov način predstavitev kot

$$A_x(\xi, \tau) = \int_{-\infty}^{+\infty} x(s+\tau/2)x^*(s-\tau/2)e^{-j2\pi\xi s} ds. \quad (2.21)$$



Slika 2.17: Primerjava slik sintetičnega signala z WVD, PWVD in SPWVD.

Funkciji  $A_x(\xi, \tau)$  pravimo *simetrična večlična funkcija* (ang. *ambiguity function, AF*). Z njo merimo korelacije signala v časovno–frekvenčni ravnini, se pravi podobnost med signalom  $x$  in premaknjenimi verzijami signala v časovno–frekvenčni prostoru. V primeru prejšnjih predstavitev so nam spremenljivke  $t$  in  $f$  predstavljale dejanske (absolutne) vrednosti po času in frekvenci, tu pa nam  $\tau$  in  $\xi$  predstavljajo *zamike* (med signali) po času in frekvenci (ang. zamik po času - delay, zamik po frekvenci - doppler), torej relativne vrednosti.

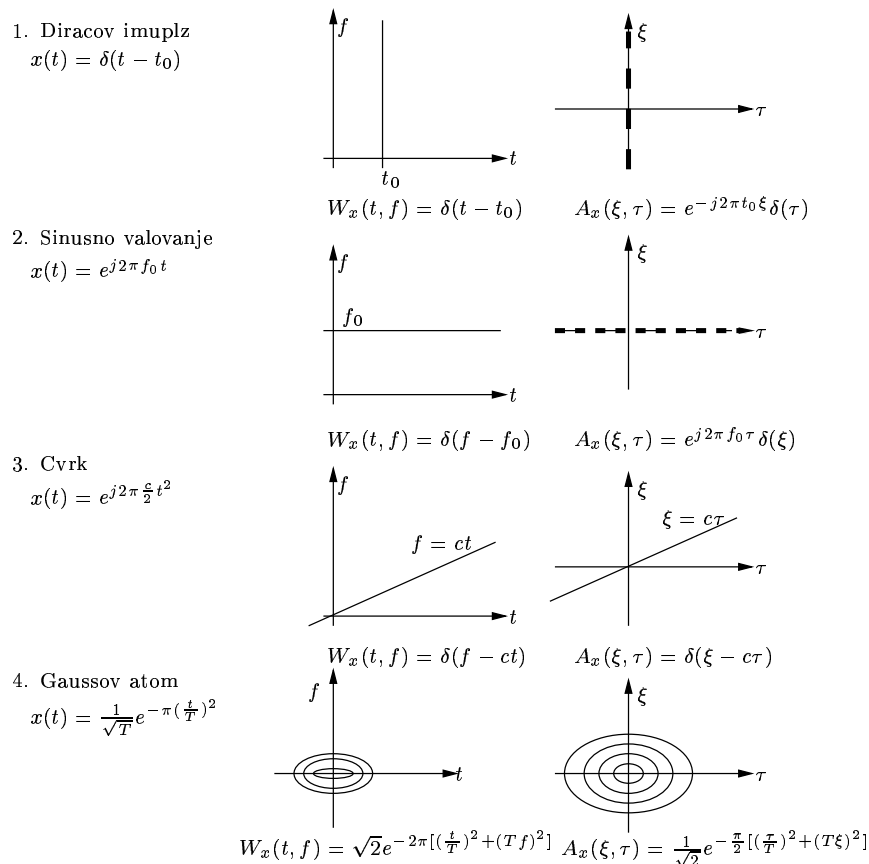
Večlična funkcija je v splošnem kompleksna funkcija in zadošča pogoju simetrije (liha kompleksna funkcija)

$$A_x(\xi, \tau) = A_x^*(-\xi, -\tau).$$

Tako definirano večlično funkcijo (obstajajo tudi nesimetrične oblike AF) lahko dobimo neposredno iz WVD predstavitve, če izvedemo dvodimenzionalno Fourierjevo transformacijo (2D-FT):

$$A_x(\xi, \tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_x(\tau, f) e^{j2\pi(f\tau - \xi t)} dt df. \quad (2.22)$$

Zveza (2.22) predstavlja alternativno definicijo večlične funkcije, pomeni pa tudi, da je AF dualna količina WVD v smislu dvodimenzionalne Fourierjeve transformacije.



Slika 2.18: Primerjava slik WVD in AF preprostih sintetiziranih signalov [Hlawatsch-92].

Na sliki 2.18 lahko primerjamo obnašanje večlične funkcije z Wigner-Villejevo porazdelitvijo na primeru preprostih sintetiziranih signalov.

Dualnost AF lahko s pridom izkoristimo tudi pri dokazovanju lastnosti te predstavitve. Poglejmo nekaj najpomembnejših [Hlawatsch-92]:

- *robna pogoja:*

Časovno avtokorelacijo lahko izračunamo kot

$$r_x(\tau) = A_x(0, \tau),$$

spektralno pa kot

$$R_x(\xi) = A_x(\xi, 0).$$

- *energija:*

Energija signala je vrednost  $A_x(\xi, \tau)$  v koordinatnem izhodišču ravnine  $(\xi, \tau)$  in velja:

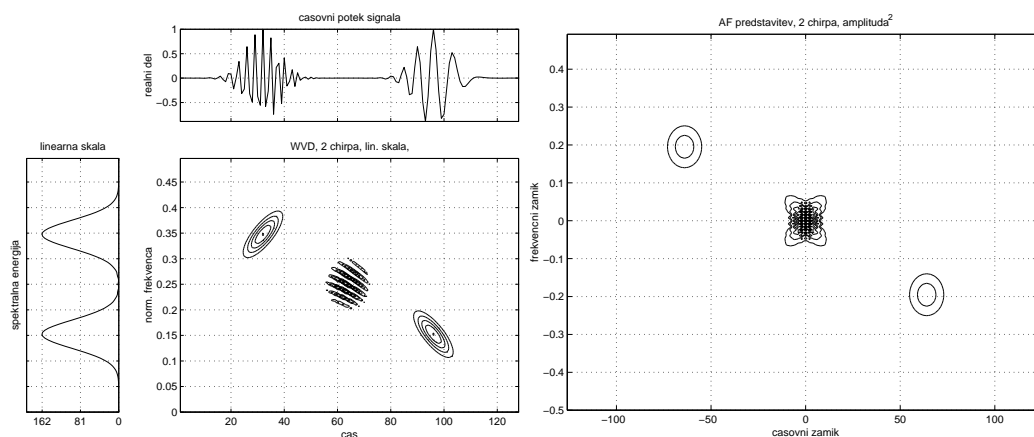
$$|A_x(\xi, \tau)| \leq A_x(0, 0) = E_x, \quad \forall \xi, \tau.$$

- *neobčutljivost na premike po času in frekvenci:*

$$y(t) = x(t - t_0) e^{j2\pi f_0 t} \Rightarrow A_y(\xi, \tau) = A_x(\xi, \tau) e^{j2\pi(f_0 \tau - t_0 \xi)}.$$

- *interferenca:*

V primeru signala sestavljenega iz več različnih komponent, se komponente z AF funkcijo izražajo v glavnem okoli koordinatnega izhodišča, interferenčni členi pa stran od izhodišča. Razdalja področij interference od izhodišča v AF ravnini je proporcionalna razdalji med komponentami signalov v običajni časovno–frekvenčni ravnini  $(t, f)$ .



Slika 2.19: Sliki signala sestavljenega iz dveh cvrkov: (a) WVD predstavitev (b) predstavitev z AF funkcijo.

Kot primer si lahko ogledamo primerjavo izražave signala sestavljenega iz dveh komponent linearnih cvrkov. Na sliki 2.19(a) pri WVD so interferenčni pasovi locirani v geometrijskem središču med obema signaloma, pri AF (slika 2.19(b)) pa se signal izrazi v okolici središča ravnine  $(\xi, \tau)$ , interferenčni členi pa stran od središča.

### Alternativna definicija Cohenovega razreda

Pri geometriji interference AF funkcije smo lahko ugotovili, da se signal dejansko izraža samo v okolici, interferenčni členi pa stran od koordinatnega središča. Zato interferenco lahko odpravimo, če na dani AF predstavitvi upoštevamo samo okolico središča (izvajamo v bistvu dvodimenzionalno filtriranje z nizkim filtrom) in potem z inverzno 2D-Fourierjevo transformacijo izračunamo nazaj prvotno WVD predstavitev. To pa ni nič drugega kot to, da AF funkcijo množimo z ustreznim jedrom in izvedemo inverzno 2D-FT. Tako lahko zapišemo definicijo (2.19) kot:

$$C_x(t, f; k) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k(\xi, \tau) A_x(\xi, \tau) e^{-j2\pi(f\tau + \xi t)} d\xi d\tau. \quad (2.23)$$

Parametrizacijsko funkcijo  $k$  (jedro) si lahko razlagamo kot neko utežno funkcijo dane AF predstavitve, ki jo lahko prilagajamo različnim potrebam in namenom obdelave signalov. Najpogosteje seveda z njo odpravljamo interferenčne člene. Jedro  $k(\xi, \tau)$  v tem primeru določimo tako (nekakšno masko), da odpravimo področja interference za določen tip signalov, to pa pomeni dejansko glajenje (s konvolucijo) prvotne predstavitve s funkcijo  $\Pi(t, f)$ , ki je 2D-FT funkcije jedra. Poleg tega lahko izberemo tudi



takšno jedro, da bo dobljena predstavitev ustrezala določenim matematičnim lastnostim. Povezavo med izbiro jedra in lastnostmi ponazarja tabela 2.1.

Tabela 2.1: Povezava med lastnostmi jedra večlične funkcije in lastnostmi (str. 24) tako dobljenih predstavitev.

lastnosti predstavitve	jedro $k(\xi, \tau)$
ohranjanje premikov po času	jedro neodvisno od časa $t$
ohranjanje premikov po frekvenci	jedro neodvisno od frekvence $f$
realna funkcija	$k(\xi, \tau) = k(-\xi, -\tau)$
robni pogoj po času	$k(\xi, 0) = 1$
robni pogoj po frekvenci	$k(0, \tau) = 1$
trenutna frekvenca	$k(\xi, 0) = 1$ in $\frac{\partial}{\partial \tau} k(\xi, \tau) _{\tau=0} = 0$
skupinska zakasnitev	$k(0, \tau) = 1$ in $\frac{\partial}{\partial \xi} k(\xi, \tau) _{\xi=0} = 0$
pozitivna funkcija	$k(\xi, \tau)$ je večlična funkcija okna $h(t)$

Poglejmo si še nekaj že prej opisanih predstavitev prevedenih v AF ravnino. WVD je 2D-FT AF, torej je jedro  $k(\xi, \tau) = 1, \forall \xi, \tau$ . Pri spektrogramu je  $k(\xi, \tau) = A_h^*(\xi, \tau)$ , kar pomeni, da je glajenje odvisno od okenske funkcije  $h$  (v to smo se lahko prepričali že v prejšnjih poglavjih). Pri SPWD pa je jedro  $k(\xi, \tau) = G(\xi)h(\tau)$  funkcija dveh ločenih funkcij, s katerima lahko ločeno po času in frekvenci nastavljam ločljivost in občutljivost na interferenco.

### 2.3.2.4 Ostale znane predstavitve

Tu si bomo ogledali še nekaj znanih energijskih predstavitev, ki spadajo v Cohenov razred porazdelitev.

#### Richaczkova in Margenau-Hillova predstavitev

Idejo v izvedbi *Richaczkove porazdelitve* [Richaczek-68] lahko najdemo na povsem drugem področju fizike.

Denimo, da preučujemo spreminjanje napetosti po času. S Fourierjevo transformacijo to lahko zapišemo kot

$$V(t) = \frac{1}{\sqrt{2\pi}} \int V_\omega e^{j\omega t} d\omega.$$

$V_\omega e^{j\omega t}$  si lahko predstavljamo kot napetost pri določeni frekvenci. Če predpostavimo, da je upornost 1, potem po Ohmovem zakonu izračunamo tok  $i_\omega = V_\omega e^{j\omega t}$ . Skupni tok v frekvenčnem razponu  $\Delta\omega$  je tako

$$i(t) = \int_\omega^{\omega+\Delta\omega} i_\omega(t) d\omega = \frac{1}{\sqrt{2\pi}} \int_\omega^{\omega+\Delta\omega} V_\omega e^{j\omega t} d\omega.$$

Moč je definirana kot produkt napetosti in toka, torej  $V(t)i^*(t)$ . Energijo v časovnem intervalu  $\Delta t$  in frekvenčnem  $\Delta\omega$  pa izračunamo kot

$$E(t, \omega) = \int_t^{t+\Delta t} V(t)i^*(t) dt = \frac{1}{\sqrt{2\pi}} \int_t^{t+\Delta t} \int_\omega^{\omega+\Delta\omega} V_\omega^* V(t) e^{-j\omega t} d\omega dt.$$

Porazdelitev energije v točki  $(t, \omega)$  potem izračunamo kot

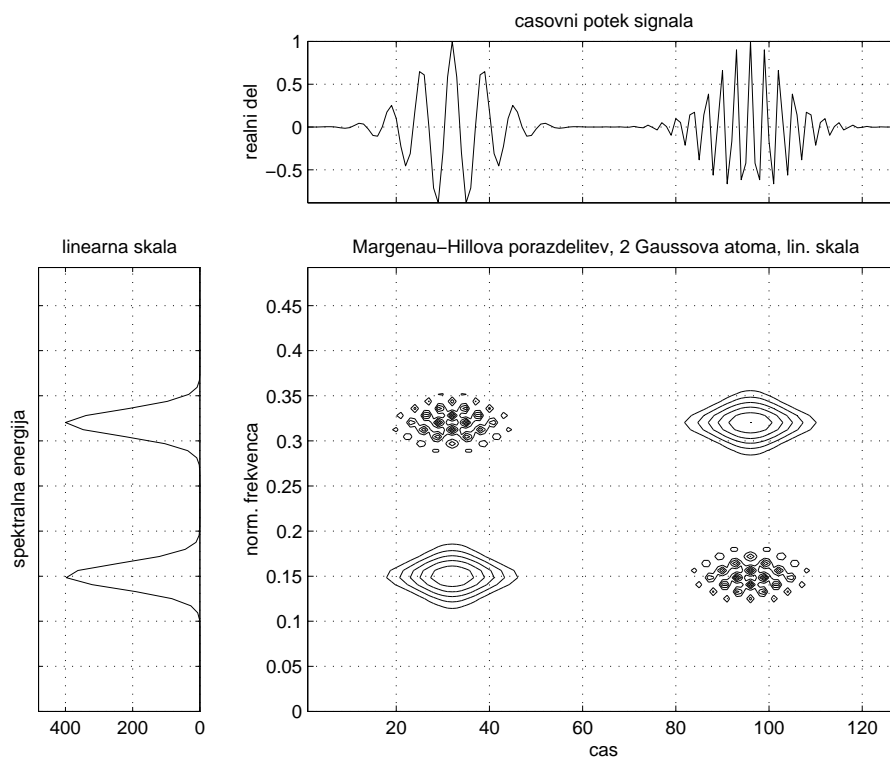
$$e(t, \omega) = \lim_{\Delta t, \Delta \omega \rightarrow 0} \frac{E(t, \omega)}{\Delta t \Delta \omega} = \frac{1}{\sqrt{2\pi}} V_{\omega}^* V(t) e^{-j\omega t}.$$

Če zadnji izračun prevedemo v našo notacijo, definiramo *Richaczko* energijsko porazdelitev

$$R_x(t, f) = x(t) X^*(f) e^{-j2\pi f t}. \quad (2.24)$$

Predstavitev spada v Cohenov razred. V to se lahko prepričamo, ko pokažemo, da je to v bistvu glajena AF funkcija z jedrom  $k(\xi, \tau) = e^{-j\pi\xi\tau}$ . To je kompleksna predstavitev, ki ustreza lastnostim (1–2) in (4–11) (lastnosti na str. 24). Lahko pa uporabimo samo realni del te predstavitve, ki prav tako spada v Cohenov razred (jedro je  $k(\xi, \tau) = \cos(\pi\xi\tau)$ ), in jo poznamo pod imenom *Margenau–Hillova porazdelitev*. Lastnosti te predstavitve so (1–5), (8) in (10–11), glej lastnosti na str. 24. Na podoben način kot pri WVD lahko definiramo tudi glajene verzije obeh predstavitev, znani pod imeni *psevdo-Richaczkova porazdelitev* in *psevdo-Margenau–Hillova porazdelitev*.

Geometrija interference je tu drugačna kot pri Wigener–Villejevi predstavitvi: interferenčni členi, ki pripadajo točkama  $(t_1, f_1)$  in  $(t_2, f_2)$  v časovno–frekvenčni ravnini se pokažejo v točkah  $(t_1, f_2)$  in  $(t_2, f_1)$ . To si lahko ogledamo na sliki 2.20.



Slika 2.20: Slika prikazuje pojav interference pri Margenau–Hillovi predstavitvi. Gre za drugačno geometrijo interference kot pri WVD. Signal je sestavljen iz dveh Gaussovih atomov.

Zato moramo biti previdni pri uporabi Richaczkove (ali Margenau–Hillove) predstavitve pri večkomponentnih signalih, kjer nastopajo posamezne komponente na istem mestu bodisi po času ali/in frekvenci.

### Pageova predstavitev

Motivacija za to energijsko porazdelitev predstavlja konstrukcija t.i. *kavzalne gostote energije*, kjer v bistvu merimo spektralno energijo signala pred časom  $t$ , ali

$$E_x^-(t, f) = \left| \int_{-\infty}^t x(u) e^{-j2\pi fu} du \right|^2.$$

*Pageova porazdelitev* je tako definirana podobno kot prej z odvodom te energije

$$P_x(t, f) = \frac{d}{dt} E_x^-(t, f) = 2\Re \left\{ x(t) \left( \int_{-\infty}^t x(u) e^{-j2\pi fu} du \right)^* e^{-j2\pi ft} \right\}. \quad (2.25)$$

Spada med predstavitve Cohenovega razreda z jedrom  $k(\xi, \tau) = e^{-j\pi\xi|\tau|}$ , ustreza lastnostim (1–5) in (7–10) na str. 24.

Prav tako obstaja njena zglajena verzija t.i. *pseudo-Pageova predstavitev*.

### Vezano glajenje WVD predstavitev

Tu bomo obravnavali še nekaj Cohenovih predstavitev, kjer je jedro funkcija produkta časovnega in frekvenčnega zamika, torej

$$k(\xi, \tau) = \Phi(\tau\xi). \quad (2.26)$$

Za funkcijo  $\Phi$  predpostavimo, da je padajoča (v vseh smereh stran od 0) in  $\Phi(0) = 1$ . Poleg robnih pogojev s tem dosežemo še nekaj: ker je funkcija padajoča, to pomeni da imamo opravka z nizkoprepustnim filtrom, kar po (2.23) ni nič drugega kot glajenje z odpravljanjem interference. Zato takšnim predstavitvam pravimo *porazdelitve reducirane interference* (ang. *reduced interference distributions*). Poleg že omenjenih Richaczkove in Margenau–Hillove spadajo sem še:

- *Choi–Williamsova porazdelitev (CW)*

Dobimo je, če za  $\Phi$  izberemo funkcijo Gaussove porazdelitve:

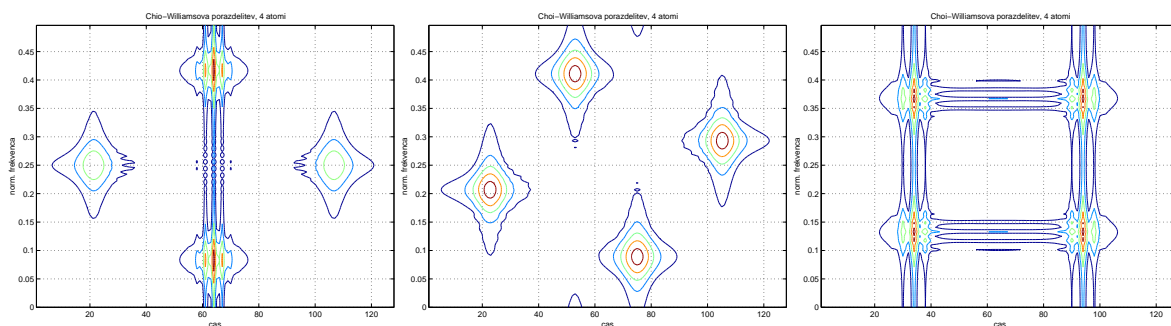
$$k(\xi, \tau) = \exp \left[ -\frac{(\pi\xi\tau)^2}{2\sigma^2} \right].$$

Tako dobimo Choi–Williamsovo [Choi-89] predstavitev

$$CW_x(t, f) = \sqrt{\frac{2}{\pi}} \int \int_{-\infty}^{+\infty} \frac{\sigma}{|\tau|} e^{-2\sigma^2(s-t)^2/\tau^2} x(s + \tau/2) x^*(s - \tau/2) e^{-j2\pi f\tau} ds d\tau. \quad (2.27)$$

Zanjo velja, če  $\sigma \rightarrow +\infty$ , dobimo WVD, če pa gre  $\sigma$  proti 0 manjšamo interferenco in slabšamo ločljivost. Ta predstavitev zadošča lastnostim (1–5) in (10–11). Ta predstavitev ima to lastnost, da se interferenčni členi, ki nastanejo med posameznimi komponentami, razpršijo po celotni časovno–frekvenčni ravnini. Geometrija interference pa je precej odvisna od tipa signalov.

Pri signalih, kjer nastopajo posamezne komponente sočasno bodisi po frekvenci ali po času, se interferenca ojača, sicer pa je odvisna od položaja posameznih komponent v časovno–frekvenčni ravnini (slika 2.21).



Slika 2.21: Pojav interference pri različnih položajih štirih Gaussovih atomov. Uporabili smo CW pri  $\sigma = 1$ .

- *Born–Jordanova in Zhao–Atlas–Marksova porazdelitev*

Tu izberemo jedro

$$k(\xi, \tau) = \frac{\sin(\pi\xi\tau)}{\pi\xi\tau}.$$

To porodi *Born–Jordanovo porazdelitev* [Cohen-66] definirano kot

$$BJ_x(t, f) = \int_{-\infty}^{+\infty} \frac{1}{|\tau|} \left[ \int_{t-|\tau|/2}^{t+|\tau|/2} x(s + \tau/2)x^*(s - \tau/2)ds \right] e^{-j2\pi f\tau} d\tau. \quad (2.28)$$

Ta predstavitev zadošča lastnostim (1–5), (8) in (10–11).

Če gladimo Born–Jordanovo porazdelitev po frekvenci, dobimo *Zhao–Atlas–Marksovo porazdelitev* [Zhao-90]:

$$ZAM_x(t, f) = \int_{-\infty}^{+\infty} h(\tau) \left[ \int_{t-|\tau|/2}^{t+|\tau|/2} x(s + \tau/2)x^*(s - \tau/2)ds \right] e^{-j2\pi f\tau} d\tau. \quad (2.29)$$

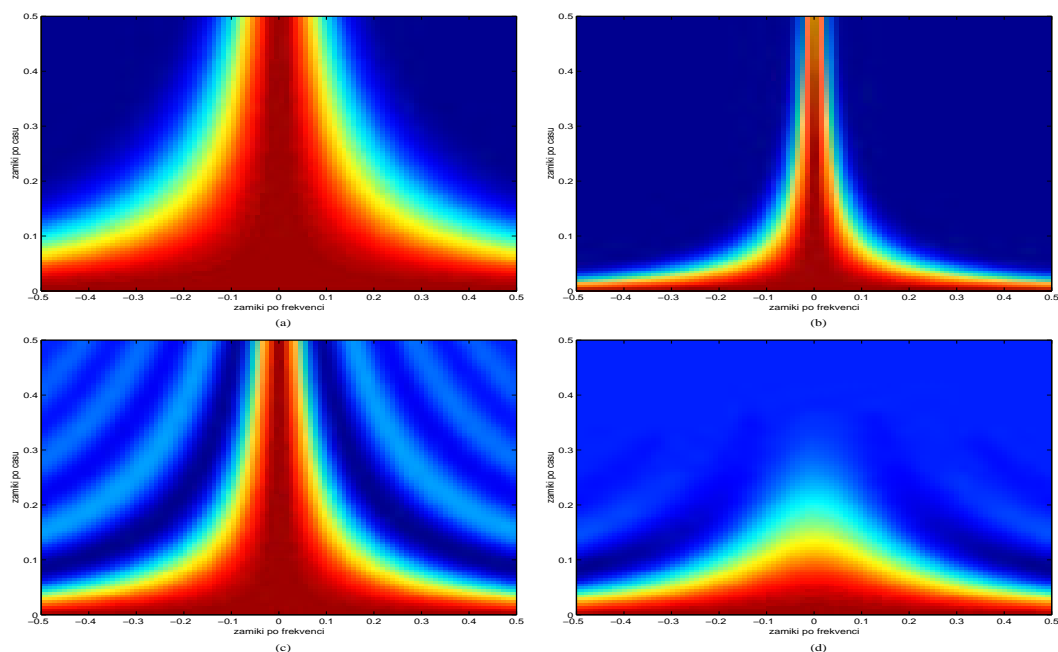
Bistvena lastnost jedra  $k(\xi, \tau) = \frac{\sin(\pi\xi\tau)}{\pi\xi\tau}$  je ta, da interferenčne člene, ki nastopajo med frekvencami sinusnih valovanj (vodoravne črte v časovno–frekvenčni ravnini) potiska k dejanskim slikam valovanj. Torej tu ne razpršimo interference, ampak jo skrijemo k dejanskim spektralnim komponentam. Takšne predstavitve so dobre, ko imamo opravka s signali, ki so v časovno–frekvenčni ravnini vzporedni vzdolž frekvenčne osi (npr. govorni signali).

V primeru ZAM porazdelitve si najpogosteje izbiramo

$$h(\tau) = \frac{1}{\tau} \exp(-\alpha\tau^2). \quad (2.30)$$

Razlika med delovanji jeder v AF ravnini v primeru CW in ZAM porazdelitev je prikazana na sliki 2.22.

Primerjavo med naštetimi predstavitvami bomo preučevali na primeru govornih signalov v poglavju 4.



Slika 2.22: Slike jeder v AF ravnini (tu gre za množenje jedra z osnovno predstavitvijo): (a) CW, manjša  $\sigma$ , (b) CW, večja  $\sigma$ , (c) ZAM, manjša  $\alpha$ , (d) ZAM, večja  $\alpha$ .

### 2.3.3 Afine predstavitve

Kot smo lahko videli v primeru zvezne valčne transformacije imamo lahko predstavitve, ki ne ohranjajo premikov po času in frekvenca, ampak ohranjajo premike in raztege po času. V primeru zvezne valčne transformacije smo tako dobili ločljivost po frekvenca odvisno od opazovane frekvence.

Energijske predstavitve, ki *ohranjajo premike in raztege po času*, tvorijo nov razred predstavitev - *afine predstavitve (porazdelitve)*.

V nadaljevanju si bomo ogledali osnovne značilnosti in lastnosti takšnih preslikav. Ker takšnih predstavitev v svojem magistrskem delu nisem uporabljal za obdelavo govornih signalov, jih ne bom tako podrobno predstavil kot Cohenov razred porazdelitev. Podrobnejšo razlago lahko najdemo v [Bertrand-91], [Hlawatsch-92], [Shenoy-95].

#### 2.3.3.1 Definicija

Najprej si oglejmo t.i. *afine transformacije* na signalu  $x(t)$  (tu predpostavljamo da je signal analitičen, glej (2.3)). To so preslikave oblike:

$$x(t) \rightarrow x_{a',b'}(t) = \frac{1}{\sqrt{|a'|}} x\left(\frac{t-b'}{a'}\right)$$

ali v frekvenčnem prostoru:

$$X(f) \rightarrow X_{a',b'}(f) = \sqrt{|a'|} e^{-j2\pi f b'} X(a'f).$$

Preslikavam, ki ohranjajo affine transformacije, pravimo *afine preslikave*, v našem primeru energijskih porazdelitev jim bomo rekli *afine predstavitve*. Torej  $\Omega_x(t, a)$  je afina predstavitev, če velja:

$$\Omega_{x_{a', b'}}(t, a) = \Omega_x\left(\frac{t - b'}{a'}, \frac{a}{a'}\right).$$

Izkaže se [Rioul-92], da takšno predstavitev lahko vedno zapišemo kot

$$\Omega_x(t, a; \Pi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi\left(\frac{s - t}{a}, a\xi\right) W_x(s, \xi) ds d\xi, \quad (2.31)$$

kjer lahko zopet razumemo  $\Pi(t, f)$  kot funkcijo glajenja Wigner–Villejeve porazdelitve. Da zagotovimo ohranitev energije signala s takšnimi predstavitvami, moramo zahtevati še

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(t, f) dt \frac{df}{|f|} = 1.$$

Primer takšne predstavitve je seveda sama Wigner–Villejeva porazdelitev. To lahko pokažemo tako, da v izrazu (2.31) zamenjamo skalo  $a$  z

$$a = \frac{f_0}{f}$$

in za  $\Pi$  potem vzamemo

$$\Pi(t, f) = \delta(t)\delta(f - f_0).$$

Alternativno definicijo afinih predstavitev lahko povežemo z večlično funkcijo (2.21). V tem primeru velja

$$\Omega_x(t, a; \Phi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(a\xi, \tau/a) A_x(\xi, \tau) e^{-j2\pi\xi t} d\xi d\tau, \quad (2.32)$$

kjer lahko razumemo  $\Phi(\xi, \tau)$  kot utežno funkcijo, za katero velja

$$\Phi(\xi, \tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi(t, f) e^{j2\pi(f\tau + \xi t)} dt df.$$

Obstaja pa še ena alternativa definiciji (2.31) z uporabo t.i. *bifrekvenčnega jedra*  $\Psi(f, \nu)$

$$\Omega_x(t, a; \Psi) = \frac{1}{|a|} \int \int_{-\infty}^{+\infty} \Psi(f, \nu) e^{-j2\pi\xi t} X\left(\frac{\nu - \frac{f}{2}}{a}\right) X^*\left(\frac{\nu + \frac{f}{2}}{a}\right) e^{-j2\pi f t/a} df d\nu, \quad (2.33)$$

kjer je

$$\Psi(f, \nu) = \int_{-\infty}^{+\infty} \Pi(t, \nu) e^{-j2\pi f t} dt.$$

Uporabo vseh treh ekvivalentnih opisov afinih predstavitev lahko s pridom uporabljamo pri dokazovanju lastnosti takšnih predstavitev.

### 2.3.3.2 Lastnosti

Tudi pri afinih porazdelitvah (tako kot pri Cohenovem razredu) načrtujemo predstavitve tako, da zadoščajo še dodatnim pogojem. Ogleдали si bomo nekaj željenih lastnosti takšnih predstavitev:

1. *ohranitev energije*: tu želimo, da lahko energijo signala  $x$  dobimo z integriranjem predstavitve po času in skali:

$$E_x = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi) dt \frac{da}{a^2},$$

2. *robna pogoja*: gostoto spektralne energije in trenutno moč signala, lahko izračunamo z integriranjem predstavitve po posameznih spremenljivkah:

$$\int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi) dt = \left| X\left(\frac{f_0}{a}\right) \right|^2, \quad \int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi) \frac{da}{a^2} = |x(t)|^2,$$

3. *realna funkcija*:

$$\Omega_x(t, a; \Pi) \in \mathbb{R}, \quad \forall t, a,$$

4. *lokalizacija po času*:

$$X(f) = \frac{1}{\sqrt{f}} e^{-j2\pi f t_0} U(f) \quad \Rightarrow \quad \Omega_x\left(t, \frac{f_0}{f}; \Pi\right) = f \delta(t - t_0) U(f),$$

kjer je  $U(f)$  Heavisideova stopnica.

5. *unitarnost*: tu mislimo predvsem na ohranjanje skalarnega produkta v prostoru časa in skale:

$$\left| \int_{-\infty}^{+\infty} x(t) y^*(t) dt \right|^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi) \Omega_y^*(t, a; \Pi) dt \frac{da}{a^2},$$

6. *skupinska zakasnitev*: tu želimo skupinsko zakasnitev signala izračunati kot prvi moment po času:

$$t_x\left(\frac{f_0}{a}\right) = \frac{\int_{-\infty}^{+\infty} t \Omega_x(t, a; \Pi) dt}{\int_{-\infty}^{+\infty} \Omega_x(t, a; \Pi) dt}.$$

Vsekakor prilagajamo afine predstavitve namenu obdelave in tipom signalov, zato smo tu podali le nekaj željenih lastnosti takšnih predstavitev.

### 2.3.3.3 Afine predstavitev

V nadaljevanju bom samo naštel nekaj porazdelitev iz tega razreda skupaj z referencami, ki podrobneje opisujejo vsako predstavitev.

Najbolj znana predstavitev te skupine je *skalogram*, ki smo si ga že ogledali, na podoben način kot pri Cohenovem razredu lahko tvorimo tudi zglajene afine predstavitev: *pseudo afino Wigenerjevo* [Rioul-91] in *zglajeno pseudo afino Wigenerjevo porazdelitev* [Rioul-92]. Veliko podskupino tvorijo predstavitev, ki jih načrtujemo z bifrekvenčnim jedrom (2.33), med katere spadajo: *Bertrandova porazdelitev* [Bertrand-91], *D-Flandrinova porazdelitev* [Flandrin-89], *aktivna in pasivna Unterbergerjeva porazdelitev* [Bertrand-91] ter njihove *glajene verzije*. Vse predstavitev si lahko ogledamo v dodatku A magistrskega dela.

## Zaključek

V tem poglavju smo si ogledali časovno–frekvenčne predstavitev. Najprej smo predstavili nekaj osnovnih motivacij za uporabo časovno–frekvenčnih predstavitev predvsem pri obdelavi in analizi nestacionarnih večkomponentnih signalov.

V prvem delu smo si ogledali predstavitev, ki temeljijo na razstavljanju signala po osnovnih elementarnih funkcijah - atomih. Take predstavitev smo poimenovali atomarne predstavitev, najbolj znani med njimi sta kratkočasovna Fourierjeva in valčna transformacija. Tu smo se predvsem ukvarjali z ločljivostjo in lokalizacijo v časovno–frekvenčni ravnini.

Prehod med atomarnimi in energijskimi predstavitvami predstavljata spektrogram in skalogram, ki ju dobimo iz kratkočasovne Fourierjeve transformacije oz. valčne transformacije. Prvi spada v razred Cohenovih energijskih predstavitev oz. porazdelitev, drugi pa ima lastnosti afinih predstavitev.

V nadaljevanju smo si tako ogledali Cohenov razred porazdelitev, ki zajema vse tiste energijske predstavitev, ki ohranjajo premike po času in frekvenci. Najpomembnejša med vsemi predstavitvami je Wigener–Villejeva porazdelitev, ki ima veliko matematičnih lastnosti. Ker energijske porazdelitve spadajo med kvadratne porazdelitve, velja tu pravilo kvadratne superpozicije, ki povzroči pojav interference med posameznimi komponentami v signalu. Interferenco poskušamo zmanjšati z glajenjem predstavitev v obeh smereh, po času in frekvenci. Tako dobimo t.i. glajene verzije osnovnih predstavitev, ki pa ne zadoščajo vsem željenim matematičnim lastnostim, kar ima za posledico slabšo ločljivost posameznih komponent v signalu. Cohen [Cohen-66] je posplošil opis takšnih predstavitev, kar nam olajša načrtovanje novih predstavitev z željenimi lastnostmi oz. lastnostmi pogojenimi z različnimi tipi signalov. Podali smo tudi povezavo z večlično funkcijo v prostoru zamikov po času in frekvenci.

Na koncu pa smo si na kratko še ogledali glavne značilnosti afinih predstavitev, ki tvorijo drugi večji razred kvadratnih (energijskih) predstavitev. Zanje je značilno, da ohranjajo premike in raztege po času, ne pa premikov po frekvenci. Tako v bistvu gledamo na te predstavitev kot funkcije časa in skale.

Definicije večine predstavitev so zbrane v dodatku A magistrskega dela.



---

# 3 Diskretne časovno–frekvenčne predstavitev

---

## 3.1 Vzorčenje zveznih predstavitev

## 3.2 Vzorčenje s pomočjo operatorske teorije

---

V poglavju 2 smo si ogledali in opisali lastnosti časovno–frekvenčnih predstavitev v primeru zveznih signalov. Seveda pa imamo pri obdelavi in analizi signalov opraviti v glavnem z njihovimi vzorčenimi verzijami. Zato je seveda to teorijo potrebno udejaniti predvsem na vzorčenih verzijah signalov.

Pri tem naletimo na vsaj dva večja problema. V poglavju 2 smo vedno predpostavili, da imamo na voljo signal v času od  $-\infty$  do  $+\infty$ , kar pa je v realnih aplikacijah nemogoče doseči. Ta predpostavka nam je olajšala izračun posameznih transformacij, hkrati pa smo lahko zaradi tega pokazali kar nekaj lastnosti takšnih predstavitev. Seveda pri diskretizaciji signalov iz realnega sveta ne moremo pričakovati takšnih predpostavk. Zato se v glavnem omejimo na oknjene verzije posameznih predstavitev, kar ima za posledico izgubo nekaterih pomembnih matematičnih lastnosti zveznih verzij takšnih predstavitev in s tem povezano slabšo ločljivost in lokalizacijo v časovno–frekvenčnem prostoru, kot smo si lahko ogledali že v prejšnjem poglavju. V tem primeru v bistvu z vzorčenimi verzijami lahko samo ocenjujemo zvezne časovno–frekvenčne predstavitev. Drugi problem pri vzorčenju je vzorčenje samo. Kot bomo lahko videli, bomo postavili tu dodatne pogoje za vzorčenje, če hočemo dobiti verodostojne ocene zveznih predstavitev.

Vse to je neposredno povezano z diskretizacijo časovno–frekvenčnih predstavitev, obstajajo pa še druge možnosti, kjer se ne ukvarjamo z direktno diskretizacijo predstavitev in njenimi problemi, ampak že vnaprej predpostavimo vzorčen signal in iščemo diskretne časovno–frekvenčne predstavitev na takšnem signalu. Pri tem iskanju se omejimo na takšne, ki so sorodne diskretiziranim verzijam zveznih predstavitev. Tu si pomagamo z operatorsko teorijo.

V nadaljevanju se bomo najprej osredotočili na diskretizacijo zveznih predstavitev, kjer bomo na primeru Wigner–Villejeve predstavitev nakazali probleme takšnega pristopa, kasneje pa si bomo ogledali operatorsko teorijo in njeno zvezo z diskretnimi časovno–frekvenčnimi predstavitvami.

### 3.1 Vzorčenje zveznih predstavitev

Pod pojmom *vzorčenje zveznih časovno–frekvenčnih predstavitev* mislimo predvsem neposredno prevedbo zveznih predstavitev k diskretnim različicam v primeru vzorčenih signalov.

Diskretizacijo atomarnih (linearnih) predstavitev, ki so zelo razširjene pri analizi in obdelavi signalov, smo si že ogledali v prejšnjem poglavju, kjer smo predstavili diskretizacijo valčne in krakočasovne Fourierjeve transformacije ter z njo v zvezi tudi Gaborjeve transformacije. Iz teh diskretizacij tudi neposredno sledijo vzorčenja dveh energijskih predstavitev: spektrograma in skalograma.

V tem delu si bomo tako pogledali vzorčenje preostalih predstavitev na primeru Wigner–Villejeve porazdelitve, kjer se bomo ukvarjali tudi s problemi, ki so posledica diskretizacije.

#### 3.1.1 Diskretna Wigner–Villejeva porazdelitev

Tu bomo izpeljali diskretno verzijo Wigner–Villejeve porazdelitve, ki predstavlja predvsem oceno zvezne porazdelitve.

Če v definiciji WVD (2.17) zamenjamo  $u = \tau/2$  in zapišemo  $\omega = 2\pi f$ , dobimo:

$$W_x(t, \omega) = 2 \int_{-\infty}^{+\infty} x(t+u)x^*(t-u)e^{-j2\omega u} du. \quad (3.1)$$

Zapišimo še vzorčeni signal  $x(t)$  kot  $x[n] = x(nT)$ . Z uporabo trapeznega pravila lahko integral iz (3.1) zapišemo kot

$$W_x[n, \omega] = 2T \sum_k x[n+k]x^*[n-k]e^{-j2\omega kT}. \quad (3.2)$$

Tako smo dobili Wigner–Villejevo porazdelitev *diskretno po času in zvezno po frekvenci* (DTWVD). Zanj velja:

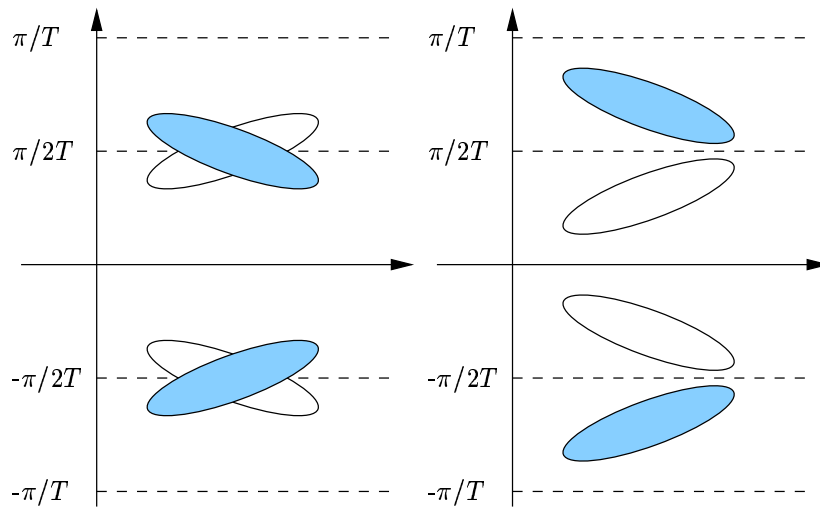
$$W_x[n, \omega + \frac{\pi}{T}] = W_x[n, \omega], \quad (3.3)$$

kar pomeni, da je DTWVD periodična s periodo  $\frac{\pi}{T}$  (npr. pri spektrogramu imamo periodo v takem primeru  $\frac{\pi}{2T}$ ). Po Shannonovi teoriji vzorčenja lahko DTWVD opisuje frekvenc večjih od  $\frac{\pi}{2T}$ . V primeru v signalu frekvence večje od  $\pi/2T$  pride do pojava *prekrivanja* (ang. *aliasing*) [Gyergyek-91, str. 563], slika 3.1.

Prvi način, da bi odpravili ta problem, je, da podvojimo vzorčenje signala. Najbolj enostavno je, če izvedemo interpolacijo predhodnega vzorčenja, slika 3.2.

To dosežemo tako, da podvojimo število odtipkov in sicer tako, da dodamo med vsak odtipek novo točko z vrednostjo 0. Potem pa izvedemo nizkoprepustno filtriranje na takšnem signalu, da odpravimo napake interpolacije. Postopek, prikazan na sliki 3.2, lahko zapišemo kot:

$$y[m] = \sum_{n=0}^N x[n]\gamma[m-2n], \quad (3.4)$$



Slika 3.1: Slika prikazuje pojav prekrivanja pri DTWVD. Levo je DTWVD s prekrivanjem, desno pa slika brez prekrivanja.

kjer  $\gamma[m]$  opisuje nizkoprepustni filter in je  $y[m]$  dvojno prevzorčeni signal. V tem primeru se tako čas vzorčenja  $T$  zmanjša za polovico, torej na  $T/2$ . Tako iz (3.2) dobimo

$$W_x[m, \omega] = 2 \frac{T}{2} \sum_k y[m+k] y^*[m-k] e^{-j2\omega k \frac{T}{2}}, \quad (3.5)$$

s periodo  $\frac{2\pi}{T}$ , kar je bil naš namen.

Nadalje zapišimo  $\theta = \omega T/2$ , brez škode za splošnost lahko predpostavimo  $T = 2$  in tako (3.5) poenostavimo v

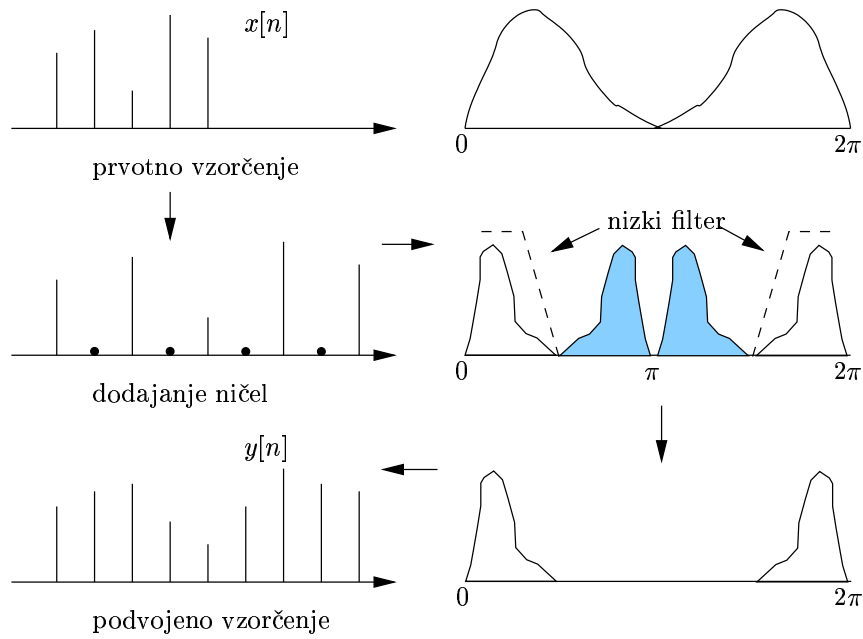
$$W_x[m, \theta] = 2 \sum_{k=-\infty}^{\infty} y[m+k] y^*[m-k] e^{-j2\theta k}. \quad (3.6)$$

S tem smo se tako znebili problema vzorčenja danega signala v primeru Wigner-Villejeve predstavitve, še vedno pa zahtevamo v enačbi (3.6) neskončen (po času) signal. Rešitev za to je že v prejšnjem poglavju opisana psevdo Wigner-Villejeva porazdelitev, ki uvede dodatno okno Wignerjevi predstavitvi (z namenom glajenja po frekvenci v prejšnjem poglavju). To lahko v tem primeru zapišemo kot

$$W_x[m, \theta] = 2 \sum_{k=-\infty}^{\infty} w[k] y[m+k] y^*[m-k] e^{-j2\theta k}. \quad (3.7)$$

V nadaljevanju bomo izbrali pravokotno okno, da bomo poenostavili izpeljavo, in sicer

$$w[k] = \begin{cases} 1, & |k| < 2L, \\ 0, & \text{sicer.} \end{cases}$$



Slika 3.2: Podvojevanje frekvence vzorčenja signala z interpolacijo.

Če to upoštevamo v (3.7), izpeljemo:

$$\begin{aligned}
 W_x[m, \theta] &= 2 \sum_{k=-(2L-1)}^{2L-1} y[m+k]y^*[m-k]e^{-j2\theta k} \\
 &= 2 \sum_{k=-(2L-1)}^0 y[m+k]y^*[m-k]e^{-j2\theta k} \\
 &\quad + 2 \sum_{k=0}^{2L-1} y[m+k]y^*[m-k]e^{-j2\theta k} - 2y[m]y^*[m] \\
 &= 4\Re \left\{ \sum_{k=0}^{2L-1} y[m+k]y^*[m-k]e^{-j2\theta k} \right\} - 2y[m]y^*[m].
 \end{aligned}$$

Če izvedemo še diskretizacijo po frekvenci  $\theta$ , dobimo

$$DW_x[m, l] = 4\Re \left\{ \sum_{k=0}^{2L-1} y[m+k]y^*[m-k]e^{-j\frac{4\pi ln}{2L}} \right\} - 2y[m]y^*[m] \quad (3.8)$$

za  $0 \leq l < 2L$ . Formula (3.8) definira *diskretno Wigner-Villejevo porazdelitev (DWVD)*. Zanj velja

$$DW_x[m, l] = DW[m, l + iL] \quad \text{za } i = 0, \pm 1, \pm 2, \dots$$

To pomeni, da pri računanju (3.8) s FFT ( $2L$  točkovno) lahko predpostavimo

$$y[m+k]y^*[m-k] = 0 \quad \text{za } |n| > L,$$

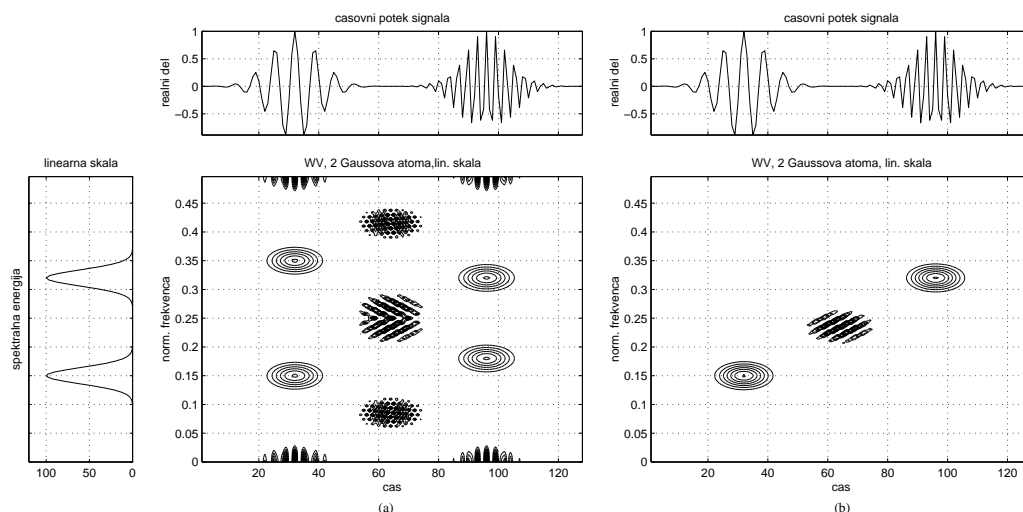
saj v nasprotnem dobimo drugo polovico členov enakih prvi polovici. S to predpostavko se (3.8) poenostavi v

$$DW_x[m, l] = 4\Re \left\{ \sum_{k=0}^{L-1} y[m+k]y^*[m-k]e^{-j\frac{2\pi ln}{L}} \right\} - 2y[m]y^*[m] \quad (3.9)$$

za  $0 \leq l < L$ . S tem se znebimo redundantnih členov pri računanju s FFT metodami, maksimalna frekvenca, ki jo lahko brez težav prekrivanja ločimo v signalu, pa je  $\pi/T$ . Formulo (3.9) lahko razumemo kot oceno diskretne Wigener–Villejeve porazdelitve.

Tako vzorčena predstavitev je zelo “blizu” zvezni verziji predstavitve. Največji problem tega pristopa je predhodno nadvzorčenje signala, kar ni vedno zaželeno (zaradi napak).

Temu pa se lahko v primeru realnih signalov izognemo z uporabo Hilbertove transformacije. Tako tvorjen analitičen signal, ki smo ga definirali v poglavju 2.1.3.1, ima to pomembno lastnost, da ohranja (podvojuje) samo pozitivne spektralne komponente, medtem ko uničuje negativne. Tako dobimo samo eno polovico spektra in s tem problem prekrivanja odpade. Hilbertovo transformacijo lahko razumemo v nekem smislu kot filtriranje z nizkoprepustnim filtrom, kar je podobno kot v primeru nadvzorčenja signala z interpolacijo. To lahko izkoristimo, da uporabimo direktne diskretne verzije predstavitev na analitičnih signalih. Dejstvo pa je, da porazdelitev energije analitičnega signala lahko drugačna od dejanske, kljub temu da sta energiji dejanskega in analitičnega signala neposredno povezani (zveza (2.4)). Takšen pristop diskretizacije odpade v primeru kompleksnih signalov, npr. pri radarskih aplikacijah [Qian-96].



Slika 3.3: Slika dveh Gaussovih atomov z diskretno verzijo WVD: (a) realni signal, (b) analitičen signal tvorjen iz realnega signala.

Ta pojav lahko opazujemo na sliki 3.3, kjer imamo signal sestavljen iz dveh Gaussovih atomov. Enkrat vzamemo realni del tega signala, drugič pa njegovo analitično verzijo. V prvem primeru lahko vidimo, da imamo namesto dveh atomov v časovno–frekvenčni ravnini kar štiri, kar je posledica “slabega” vzorčenja in s tem pojava prekrivanja za diskretno verzijo WVD. Tudi interferenčnih območij je zaradi tega več. Pojavijo se tudi interferenčni členi, ki so posledica negativnih spektralnih komponent (na zgornjem in

spodnjem robu slike 3.3(a)) V drugem primeru analitičnega signala pa je slika 3.3(b) veliko bolj “čista” in prikazuje dejanski potek signala v časovno–frekvenčni ravnini.

## 3.2 Vzorčenje s pomočjo operatorske teorije

Tu si bomo ogledali diskretizacijo Cohenovega razreda predstavitev s pomočjo operatorske teorije.

Leon Cohen je v svojem delu [Cohen-66] posplošil izražave časovno–frekvenčnih predstavitev, ki ohranjajo premike po času in frekvenci, kar smo si ogledali v prejšnjem poglavju. V knjigi [Cohen-95] pa je pokazal, kako lahko takšne predstavitve opišemo tudi s t.i. *karakterističnimi funkcijami* z uporabo operatorskih metod, ki sta jih prva predstavila Moyal [Moyal-49] in Ville [Ville-48].

Skupna značilnost teh metod je, da gledamo na kvadratne časovno–frekvenčne predstavitve kot na porazdelitvene funkcije energije v prostoru časa in frekvence, kar smo si že ogledali v predhodnih poglavjih (od tod tudi ime energijske porazdelitve). Na ta način lahko uporabimo teorijo verjetnostnega računa, kakor bomo pokazali v nadaljevanju.

Tak zapis časovno–frekvenčnih predstavitev pa bomo uporabili predvsem na vzorčenih signalih in tako neposredno iz vzorčenega signala dobili diskretno časovno–frekvenčno predstavitev.

### 3.2.1 Karakteristične funkcije zveznih predstavitev

Pri Cohenovi metodi [Cohen-95] izražave kvadratnih časovno–frekvenčnih predstavitev s karakterističnimi funkcijami moramo najprej definirati operatorja, ki delujeta na signal  $x(t)$  po času, operator  $\mathcal{T}$ , in po frekvenci, operator  $\mathcal{W}$ . Definiciji sta implicitni in sicer:

$$\int_{-\infty}^{+\infty} x^*(t)\mathcal{W}x(t)dt = \int_{-\infty}^{+\infty} X^*(f)fX(f)df, \quad (3.10)$$

$$\mathcal{T}x(t) = tx(t). \quad (3.11)$$

Definiciji operatorjev (3.10) in (3.11) si lahko razlagamo kot izračun povprečne frekvence v signalu oziroma povprečnega časa signala  $x(t)$ , če razumemo količini  $x^*(t)x(t)$  in  $X^*(f)X(f)$  kot porazdelitvi energije v signalu, kot smo to že počeli v poglavju 2.1.2 na str. 6.

EksPLICITNA rešitev v primeru (3.10) je [Cohen-95]:

$$\mathcal{W}x(t) = -j\frac{d}{dt}x(t). \quad (3.12)$$

### 3.2.1.1 Karakteristične funkcije

Z uporabo časovnega in frekvenčnega operatorja skupaj lahko definiramo časovno-frekvenčne predstavitve v primeru zveznega signala. To bomo storili s pomočjo *karakterističnih funkcij*.

Karakteristično funkcijo operatorja  $\mathcal{A}$  podamo z

$$M(\alpha) = \langle \exp(j\alpha\mathcal{A}) \rangle = \int_{-\infty}^{+\infty} s^*(t) \exp(j\alpha\mathcal{A})s(t)dt. \quad (3.13)$$

Od tu pa lahko izračunamo porazdelitev zvezne spremenljivke  $a$ , ki jo opiše operator  $\mathcal{A}$ , kot

$$P(a) = \int_{-\infty}^{+\infty} M(\alpha) \exp(-ja\alpha)d\alpha. \quad (3.14)$$

V primeru časovno-frekvenčnih energijskih porazdelitev,  $P(t, f)$ , imamo opravka z dvema zveznima spremenljivkama. Takšno predstavitev ustrezno definira dvodimenzionalna karakteristična funkcija:

$$M(\eta, \tau) = \langle \exp(j\mathcal{T}\eta + j\mathcal{W}\tau) \rangle, \quad (3.15)$$

iz katere z dvodimenzionalno Fourierjevo transformacijo izračunamo časovno-frekvenčno predstavitev:

$$P(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} M(\eta, \tau) \exp(-j\eta t) \exp(-j\tau f) d\eta d\tau. \quad (3.16)$$

Tu je potrebno opozoriti, da imamo opravka z operatorji  $\mathcal{T}$  in  $\mathcal{W}$ , ki na splošno med sabo ne komutirajo, zato tudi ne smemo v karakteristični funkciji (3.15) poljubno zamenjavati vrstni red operatorjev, kot to delamo v primeru spremenljivk. Izkaže se namreč, da različni vrstni redi operatorjev definirajo različne karakteristične funkcije in s tem različne predstavitve, npr.  $\langle \exp(j\mathcal{T}\eta + j\mathcal{W}\tau) \rangle$ ,  $\langle \exp(j\mathcal{T}\eta) \exp(j\mathcal{W}\tau) \rangle$ ,  $\langle \exp(\frac{j}{2}\mathcal{W}\tau) \exp(j\mathcal{T}\eta) \exp(j\mathcal{W}\tau) \rangle$  opisujejo različne časovno-frekvenčne predstavitve.

Cohen [Cohen-95, str. 152–158] je nadalje pokazal, da so karakteristične funkcije definirane z različnimi vrstnimi redi operatorjev ekvivalentne, če v definiciji (3.15) uvedemo ustrezno funkcijo  $\phi(\eta, \tau)$ , ki jo imenujemo jedro. Temu pravilu pravimo *pravilo ujemanja* (*ang. correspondence rule*). To pomeni, da s karakteristično funkcijo

$$M(\eta, \tau) = \phi(\eta, \tau) \int_{-\infty}^{+\infty} s^*(t) \exp(j\mathcal{T}\eta) \exp(j\mathcal{W}\tau) s(t) dt \quad (3.17)$$

lahko opišemo vse možne kombinacije nastopa operatorjev v karakteristični funkciji. Z izbiro (določitvijo) jedra  $\phi(\eta, \tau)$  in z izbiro vrstnega reda operatorjev pa tako lahko enolično definiramo časovno-frekvenčne predstavitve iz Cohenovega razreda, ki smo jih zapisali s (3.16).

Tudi tu si lahko razlagamo jedro  $\phi(\eta, \tau)$ , analogno kot parametrizacijsko funkcijo (2.19) v prejšnjem poglavju pri Cohenovih porazdelitvah, kot funkcijo glajenja dane predstavitve.

### 3.2.2 Prevedba v diskreten prostor

Tu bomo podobno kot v prejšnjem podrazdelku zapisali predstavitve s karakterističnimi funkcijami, le da bomo tu privzeli, da je signal vzorčen.

Na voljo imamo vzorčen signal  $x[n]$  s periodo  $N$ . Diskretno Fourierjevo transformacijo signala zapišimo kot  $X[k]$ . Naš cilj je torej definirati diskretno časovno–frekvenčno predstavitev  $P[n, k]$  neposredno iz diskretno časovno podanega signala  $x[n]$  in ustrezne frekvenčne slike  $X[k]$ .

Najprej prevedimo zvezna operatorja iz (3.10) in (3.11) v njuni diskretni različici  $\mathcal{K}$  in  $\mathcal{L}$  tako, da zapišemo implicitni definiciji za diskreten primer

$$\sum_{n=0}^{N-1} x^*[n] \mathcal{K}x[n] = \sum_{k=0}^{N-1} X^*[k] kX[k], \quad (3.18)$$

$$\mathcal{L}x[n] = nx[n]. \quad (3.19)$$

Operatorja  $\mathcal{K}$  in  $\mathcal{L}$  sta diskretna in linearna, tako da ju lahko interpretiramo kot matriki in operacije iz izpeljav (3.18) in (3.19) tako obravnavamo kot matrične enačbe. Zato najprej zapišimo niz  $\{x[0], x[1], \dots, x[N-1]\}$  kot vektor  $\mathbf{x}$ . Operacijo konjugiranja v tem primeru predstavimo kot Hermitsko transponiranje vektorja in jo označimo kot  $\mathbf{x}^*$ . Operatorja  $\mathcal{K}$  in  $\mathcal{L}$  pa predstavimo z matrikama  $\mathbf{K}$  in  $\mathbf{L}$ .

V tem primeru lahko rešimo matrični enačbi (3.18) in (3.19). Iz (3.19) neposredno sledi, da je matrika  $\mathbf{L}$  diagonalna z elementi  $\{0, 1, \dots, N-1\}$  po diagonalni.

Rešitev matrične enačbe (3.18) pa je [McLaughlin-97]:

$$\mathcal{K} = o[n] \otimes x[n], \quad \text{kjer je} \quad (3.20)$$

$$o[n] = \begin{cases} \frac{(j/2)(-1)^n}{\sin(\pi n/N)}, & n \neq 0, \\ 0, & n = 0 \end{cases}, \quad (3.21)$$

kjer pomeni  $\otimes$  krožno konvolucijo. Matrika  $\mathbf{K}$  je torej krožna matrika oblike:

$$\mathbf{K} = \begin{bmatrix} o[0] & o[1] & \cdots & o[N-1] \\ o[1] & \cdots & o[N-1] & o[0] \\ \vdots & \vdots & \vdots & \vdots \\ o[N-1] & o[0] & \cdots & o[N-2] \end{bmatrix} \quad (3.22)$$

Izračunane diskretne operatorje bomo uporabili naprej za izračun diskretnih karakterističnih funkcij. Še prej pa si bomo ogledali nekaj lastnosti takšnih operatorjev, ki jih bomo potrebovali v nadaljevanju.



### 3.2.3 Lastnosti operatorjev

#### Premiki

Pri zveznih operatorjih je Cohen pokazal, da velja [Cohen-66]:

$$e^{j\mathcal{W}\tau}x(t) = x(t + \tau), \quad (3.23)$$

kjer je  $\mathcal{W}$  zvezni operator definiran s (3.10). V primeru diskretnega operatorja  $\mathbf{K}$  lahko pokažemo podobno:

$$e^{j\mathbf{K}m}s[n] = \sum_{l=0}^{\infty} \frac{(j\mathbf{K}m)^l}{l!} s[n] = \sum_{l=0}^{\infty} \frac{(jm)^l}{l!} (\mathbf{K})^l s[n]. \quad (3.24)$$

Izračunajmo diskretno Fourierjevo transformacijo izraza (3.24):

$$DFT\{e^{j\mathbf{K}m}s[n]\} = \sum_{l=0}^{\infty} \frac{(jm)^l}{l!} k^l S[k] = \sum_{l=0}^{\infty} \frac{(jmk)^l}{l!} S[k] = e^{jmk} S[k].$$

Če izračunamo nazaj inverzno DFT, dobimo iskano lastnost:

$$e^{j\mathbf{K}m}s[n] = s[n + m]. \quad (3.25)$$

#### Komutatorji

Komutator dveh operatorjev je definiran kot  $\mathcal{C} = [A, B] = [AB - BA]$ . Pri teoriji operatorjev nam pomenijo konstantne vrednosti komutatorjev pomembne lastnosti takšnih operatorjev. V primeru, da je  $\mathcal{C} = 0$ , to pomeni, da sta operatorja med sabo komutativna, torej lahko poljubno spreminjamo vrstni red operatorjev.

V zveznem primeru naših operatorjev lahko izračunamo

$$\begin{aligned} (\mathcal{T}\mathcal{W} - \mathcal{W}\mathcal{T})x(t) &= (t(-j\frac{d}{dt}) - (-j\frac{d}{dt})t)x(t) \\ &= -jtx'(t) + jx(t) + jtx'(t) \\ &= jx(t), \end{aligned} \quad (3.26)$$

kar pomeni, da je  $[\mathcal{T}, \mathcal{W}] = j$  oziroma  $[\mathcal{W}, \mathcal{T}] = -[\mathcal{T}, \mathcal{W}] = -j$ . Vrednost komutatorja teh dveh operatorjev je torej konstantna, kar bomo izkoristili kasneje pri izračunu Wigener-Villejeve zvezne porazdelitve.

V diskretnem primeru pa konstantne vrednosti operatorja v tem primeru ne moremo zagotoviti:

$$\begin{aligned} (\mathbf{L}\mathbf{K} - \mathbf{K}\mathbf{L})x[n] &= n(o[n] \otimes x[n]) - o[n] \otimes nx[n] \\ &= n \sum_p o[p]x[n-p] - \sum_p o[p](n-p)x[n-p] \\ &= no[n] \otimes x[n], \end{aligned} \quad (3.27)$$

kar pomeni, da je  $[\mathbf{L}, \mathbf{K}] = no[n]$  oziroma  $[\mathbf{K}, \mathbf{L}] = -no[n]$ .

*Krožne matrike*

V jeziku linearne algebre so definirane krožne matrike kot

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 & \cdots & h_N \\ h_2 & \cdots & h_N & h_1 \\ \vdots & \vdots & \vdots & \vdots \\ h_N & h_1 & \cdots & h_{N-1} \end{bmatrix}$$

Vsaka takšna matrika  $\mathbf{H}$  se dá diagonalizirati. Torej  $\mathbf{F}\mathbf{H}\mathbf{F}^* = \Lambda$ , kjer je  $\mathbf{F}$   $N \times N$  matrika z elementi  $\exp(-j\frac{2\pi k}{N}n)$ . To pa ni nič drugega kot matrika diskretne Fourierjeve transformacije (vsoto lahko razumemo kot produkt matrike in vektorja sestavljenega iz signala).

V primeru (3.22) to pomeni, da lahko krožno matriko diagonaliziramo in dobimo [Atlas-97, Narayanan-96]:

$$\mathbf{F}\mathbf{K}\mathbf{F}^* = \text{diag}\{0, 1, \dots, \frac{N-1}{2}, -\frac{N-1}{2}, \dots, -1\} \quad (3.28)$$

Tudi to lastnost bomo uporabili pri izpeljavi diskretnih časovno–frekvenčnih predstavitev.

Še več lastnosti zveznih in diskretnih operatorjev pa lahko najdemo v članku [McLaughlin-97].

### 3.2.4 Izpeljava diskretnih časovno–frekvenčnih predstavitev

V diskretnem primeru lahko zapišemo karakteristično funkcijo (3.15) z diskretnima operatorjema (3.18) in (3.19) kot

$$M[\eta, \tau] = \langle \exp(j2\pi\mathbf{L}\eta + j2\pi\mathbf{K}\tau) \rangle. \quad (3.29)$$

Podobno kot v zveznem primeru imamo tudi tu možne različne rede operatorjev, ki pa jih lahko opišemo z izbiro enega samega vrstnega reda in uvedbo funkcije jedra  $\phi[\eta, \tau]$ , ki je v tem primeru seveda diskretna funkcija. Prav ustrezna izbira vrstnega reda operatorjev pa nam omogoča izpeljavo diskretnih različic časovno–frekvenčnih zveznih predstavitev.

V nadaljevanju si bomo ogledali izpeljavo diskretne Richaczkove predstavitve in probleme, ki nastanejo, če hočemo na tak način izpeljati Wigener–Villejevo porazdelitev.

#### 3.2.4.1 Diskretna Richaczkova predstavitvev

Če uporabimo karakteristično funkcijo  $\langle \exp(j2\pi\mathbf{L}\eta) \exp(j2\pi\mathbf{K}\tau) \rangle$ , potem lahko zapišemo

$$M[\eta, \tau] = \langle e^{j\mathbf{L}\eta} e^{j\mathbf{K}\tau} \rangle = \sum_n x^*[n] e^{j\mathbf{L}\eta} e^{j\mathbf{K}\tau} x[n].$$

Tu konstanto  $2\pi$  obravnavamo v ustreznih spremenljivkah in jo v nadaljevanju ne bomo več pisali. Ob uporabi lastnosti (3.25) poenostavimo  $M[\eta, \tau]$  v

$$M[\eta, \tau] = \sum_n x^*[n] e^{jL\eta} x[n + \tau]. \quad (3.30)$$

Če prevedemo zapis časovno–frekvenčnih predstavitev (3.16) v zveznem primeru v diskretno različico, dobimo

$$P[n, k] = \sum_\tau \sum_\eta M[\eta, \tau] e^{-j\tau k} e^{-j\eta n}. \quad (3.31)$$

V definicijo diskretne porazdelitve (3.31) nadalje vstavimo karakteristično funkcijo iz (3.30):

$$P[n, k] = \sum_\tau \sum_\eta \left( \sum_u x^*[u] e^{jL\eta} x[u + \tau] \right) e^{-j\tau k} e^{-j\eta n}.$$

Operator  $e^{jL\eta}$  se po času prevede v  $e^{jn\eta}$ . Če to upoštevamo, dobimo

$$\begin{aligned} P[n, k] &= \sum_\tau \sum_\eta \left( \sum_u x^*[u] e^{jn\eta} x[u + \tau] \right) e^{-j\tau k} e^{-j\eta n} \\ &= \sum_\tau \sum_u x^*[n] x[u + \tau] \delta[n - u] e^{-j\tau k} \\ &= \sum_\tau x^*[n] x[n + \tau] e^{-j\tau k} \\ &= x^*[n] e^{jn k} X[k]. \end{aligned} \quad (3.32)$$

Tako smo dobili diskretno (konjungirano) verzijo Richaczkove porazdelitve (2.24).

### 3.2.4.2 Wigner–Villejeva porazdelitev

Poglejmo si še primer, kako lahko s pomočjo karakterističnih funkcij izračunamo Wigner–Villejevo porazdelitev.

Najprej bomo izpeljali zvezno Wignerjevo porazdelitev. Če si izberemo drugačen vrstni red operatorjev v karakteristični funkciji, in sicer  $\langle e^{j\mathcal{T}\eta + j\mathcal{W}\tau} \rangle$ , zapišemo karakteristično funkcijo kot:

$$M(\eta, \tau) = \int_{-\infty}^{+\infty} x^*(t) e^{j\mathcal{T}\eta + j\mathcal{W}\tau} x(t) dt. \quad (3.33)$$

V tem primeru si pomagamo z izrekom [Wilcox-67], ki pravi, da je

$$\begin{aligned} e^{j\mathcal{A}\alpha + j\mathcal{B}\beta} &= e^{-\frac{1}{2}\alpha\beta[\mathcal{A}, \mathcal{B}]} e^{j\mathcal{B}\beta} e^{j\mathcal{A}\alpha} \\ &= e^{\frac{1}{2}\alpha\beta[\mathcal{A}, \mathcal{B}]} e^{j\mathcal{A}\alpha} e^{j\mathcal{B}\beta} \end{aligned} \quad (3.34)$$

natanko tedaj, ko velja

$$[[\mathcal{A}, \mathcal{B}], \mathcal{A}] = [[\mathcal{A}, \mathcal{B}], \mathcal{B}] = 0. \quad (3.35)$$

V našem primeru je komutator  $[\mathcal{T}, \mathcal{W}] = j$  iz česar sledi, da je pogoj (3.35) izpolnjen in tako lahko vsoto operatorjev spremenimo v produkt (kompozitum) operatorjev:

$$e^{j\mathcal{T}\eta + j\mathcal{W}\tau} = e^{j\frac{1}{2}\eta\tau} e^{j\mathcal{T}\eta} e^{j\mathcal{W}\tau}$$

Ob upoštevanju tega preuredimo (3.33) v

$$M(\eta, \tau) = \int_{-\infty}^{+\infty} x^*(t) e^{j\frac{1}{2}\eta\tau} e^{j\mathcal{T}\eta} e^{j\mathcal{W}\tau} x(t) dt. \quad (3.36)$$

To lahko nadalje poenostavimo ob upoštevanju lastnosti premika zveznega operatorja (3.23) in dejstva, da je  $\mathcal{T} = t$ , v

$$M(\eta, \tau) = \int_{-\infty}^{+\infty} x^*(t) e^{j\frac{1}{2}\eta\tau} e^{jt\eta} x(t + \tau) dt. \quad (3.37)$$

Z uvedbo nove spremenljivke  $u = t + \frac{\tau}{2}$  zvezo (3.37) preuredimo v:

$$M(\eta, \tau) = \int_{-\infty}^{+\infty} x^*(u - \frac{\tau}{2}) e^{ju\eta} x(u + \frac{\tau}{2}) du, \quad (3.38)$$

kar ni nič drugega kot definicija zvezne *večlične funkcije* iz (2.21).

Z dvodimenzionalno Fourierjevo transformacijo iz karakteristične funkcije (3.38) izračunamo Wigner–Villejevo porazdelitev:

$$\begin{aligned} P(t, f) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^*(u - \frac{\tau}{2}) e^{ju\eta} x(u + \frac{\tau}{2}) e^{-j\eta t} e^{-j\tau f} d\eta d\tau du \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^*(u - \frac{\tau}{2}) \delta(u - t) x(u + \frac{\tau}{2}) e^{-j\tau f} d\tau du \\ &= \int_{-\infty}^{+\infty} x^*(t - \frac{\tau}{2}) x(t + \frac{\tau}{2}) e^{-j\tau f} d\tau. \end{aligned} \quad (3.39)$$

S tem smo nakazali pot, kako s pomočjo karakterističnih funkcij izračunamo energijske predstavitve.

Poglejmo si še, kaj se zgodi v diskretnem primeru. Tu imamo že takoj na začetku težave, saj komutatorji diskretnih operatorjev ne zadoščajo pogoju (3.35) izreka. To pa zaradi tega, ker  $[\mathbf{L}, \mathbf{K}]$  ni konstanta, ampak operator, kar ima za posledico, da je tudi

$$[[\mathbf{L}, \mathbf{K}], \mathbf{L}] = n^2 o[n]$$

nov operator, ki seveda ni enak 0. Zato na ta način ne moremo izpeljati diskretne Wigner–Villejeve porazdelitve.

### 3.2.4.3 Splošen zapis diskretnih predstavitev

Če vse skupaj povzamemo in povežemo v enoten okvir, lahko zapišemo diskretne časovno-frekvenčne predstavitve s pomočjo karakterističnih funkcij tako, da najprej definiramo diskretno karakteristično funkcijo

$$M[\eta, \tau] = \sum_{n=0}^{N-1} x^*[n] \exp\left(\frac{j2\pi\eta\mathbf{L}}{N} + \frac{j2\pi\tau\mathbf{K}}{N}\right) x[n]. \quad (3.40)$$

Z uporabo pravila ujemanja [Cohen-95, str. 152–158] lahko poljubno mešamo vrstni red operatorjev, vendar moramo pri tem uvesti posebno funkcijo  $\phi[\eta, \tau]$ , ki jo imenujemo jedro. Tako lahko karakteristično funkcijo zapišemo drugače kot

$$M[\eta, \tau] = \sum_{n=0}^{N-1} x^*[n] \phi[\eta, \tau] \exp\left(\frac{j2\pi\eta\mathbf{L}}{N}\right) \exp\left(\frac{j2\pi\tau\mathbf{K}}{N}\right) x[n]. \quad (3.41)$$

Če privzamemo takšen vrstni red operatorjev, lahko enačbo (3.41) obravnavamo na dva načina.

Pri prvem načinu gre za poenostavitev enačbe (3.41), kar smo že počeli v primeru diskretne Richaczckove predstavitve. Vse skupaj se poenostavi v

$$M[\eta, \tau] = \phi[\eta, \tau] \sum_{n=0}^{N-1} x^*[n] x[n + \tau] \exp\left(\frac{j2\pi n\eta}{N}\right) \quad (3.42)$$

in

$$P[n, k] = \sum_{\eta=0}^{N-1} \exp\left(\frac{-j2\pi\eta n}{N}\right) \sum_{\tau=0}^{N-1} \exp\left(\frac{-j2\pi\tau k}{N}\right) M[\eta, \tau] \quad (3.43)$$

Drugi način izpeljave enačbe (3.41) pa je, če si vse skupaj predstavljamo kot produkte matrik z vektorji. V tem primeru [Atlas-97] predstavimo operatorja  $\mathbf{L}$  in  $\mathbf{K}$  z ustreznima matrikama. Eksponenciranje matrik pa izvedemo tako, da najprej diagonaliziramo dani matriki, potem pa eksponenciramo elemente diagonalne matrike. Matrika  $\mathbf{L}$  je že diagonalna, tako da tu kar eksponenciramo diagonalne elemente. Matriko  $\mathbf{K}$  pa smo že diagonalizirali v primeru krožnih matrik (3.28). Tako lahko zapišemo karakteristično funkcijo iz (3.41) kot

$$M[\eta, \tau] = \mathbf{x}^* \phi \lambda_1 \mathbf{F} \lambda_2 \mathbf{F}^* \mathbf{x}, \quad (3.44)$$

kjer pomeni  $*$  Hermitsko konjugiranje,  $\lambda_1$  diagonalna matrika operatorja  $\exp\left(\frac{j2\pi\eta\mathbf{L}}{N}\right)$ ,  $\lambda_2$  pa diagonalna matrika operatorja  $\exp\left(\frac{j2\pi\tau\mathbf{K}}{N}\right)$  in  $\mathbf{F}$  DFT matrika.

Če izračunamo še dvodimenzionalno diskretno Fourierjevo transformacijo, dobimo predstavitev v matrični obliki zapisano kot [Atlas-97]:

$$P[n, k] = \mathbf{x}^* (\Phi * \mathbf{F}) \mathbf{F}^* \mathbf{x}, \quad (3.45)$$

kjer operacija  $*$  predstavlja produkt matrik po elementih in je  $\Phi$  dvodimenzionalna DFT jedra  $\phi$ .

Na ta način lahko opišemo diskretne verzije Richaczckove in Margenau–Hillove porazdelitve ter spektrograma.

---

## Zaključek

---

V tem poglavju smo podali dva načina pretvorbe zveznih časovno–frekvenčnih predstavitev v diskretne.

Pri prvem načinu neposredne prevedbe zveznih predstavitev v njihove diskretne različice, smo v primeru diskretne Wigner–Villejeve predstavitve naleteli na problem podvzorčenja in s tem povezanega pojava prekrivanja. To smo poskušali odpraviti z prevzorčenjem signala, v primeru realnega signala pa se dá to rešiti z uporabo analitičnega signala.

V drugem primeru smo se ukvarjali samo z diskretnimi predstavitvami ob predpostavki, da imamo na voljo samo vzorčni signal. Vpeljali smo pojem karakteristične funkcije, ki smo ga povzeli po operatorski teoriji, uveljavljeni predvsem na področju kvantne fizike. Tako smo si ogledali primer izpeljave diskretne Richaczkovske predstavitve in probleme pri izpeljavi Wigner–Villejeve porazdelitve. Bolj do izraza je tu prišel tudi pojem jedra, ki smo ga spoznali že v prejšnjem poglavju.

---

# 4 Analiza in razpoznavanje govornega signala

---

4.1 Časovno–frekvenčne slike govornega signala

4.2 Razpoznavanje govora

---

V tem poglavju se bomo posvetili obdelavi in analizi govornega signala z različnimi časovno–frekvenčnimi predstavitvami.

V prvem delu bomo na primeru krajšega govornega signala preučevali časovno–frekvenčne slike signala, pridobljene z različnimi predstavitvami. Ugotavljali bomo razlike in posebnosti posameznih predstavitev ter preučevali uporabnost predstavitev za nadaljnjo analizo in obdelavo govornega signala.

Drugi del bo posvečen razpoznavanju govora. Predstavili bomo različne postopke pridobivanja značilk na podlagi različnih časovno–frekvenčnih predstavitev. Poskušali bomo pridobivati in določati značilke primerne za razpoznavanje in primerjali rezultate razpoznavanja ob uporabi različnih predstavitev.

## 4.1 Časovno–frekvenčne slike govornega signala

V tem poglavju se bomo posvetili prikazovanju govornega signala z različnimi časovno–frekvenčnimi predstavitvami. Podali bomo primerjavo med slikami različnih predstavitev in poskušali določiti najbolj ustrezne predstavitve za nadaljnjo analizo govornega signala predvsem za razpoznavanje govora.

### 4.1.1 Slike časovno–frekvenčnih predstavitev

Pri prikazu časovno–frekvenčne predstavitve signala v bistvu rišemo graf funkcije časa in frekvence. To lahko opišemo na več načinov. Vsekakor nam energijske časovno–frekvenčne predstavitve poskušajo opisati potek energije po času in frekvenci. Načeloma si to lahko razlagamo tudi tako, da nam nek izsek predstavitve podaja lokalno informacijo o porazdelitvi energije.

Slika oziroma graf takšne predstavitve nam tako podaja informacijo o spremembah in velikosti porazdelitve energije (odseka) signala v časovno–frekvenčnem prostoru. Na podlagi te informacije lahko sklepamo o lastnostih obravnavanega signala. Zato je potrebno poleg predstavitev posebno pozornost nameniti tudi prikazu le-teh.

Grafi predstavitev so tridimenzionalne funkcije. Zato jih lahko predstavimo na več načinov: bodisi s ploskvami v tridimenzionalnem prostoru ali pa s projekcijami 3D prostora v ravnino z uporabo različnih tehnik.

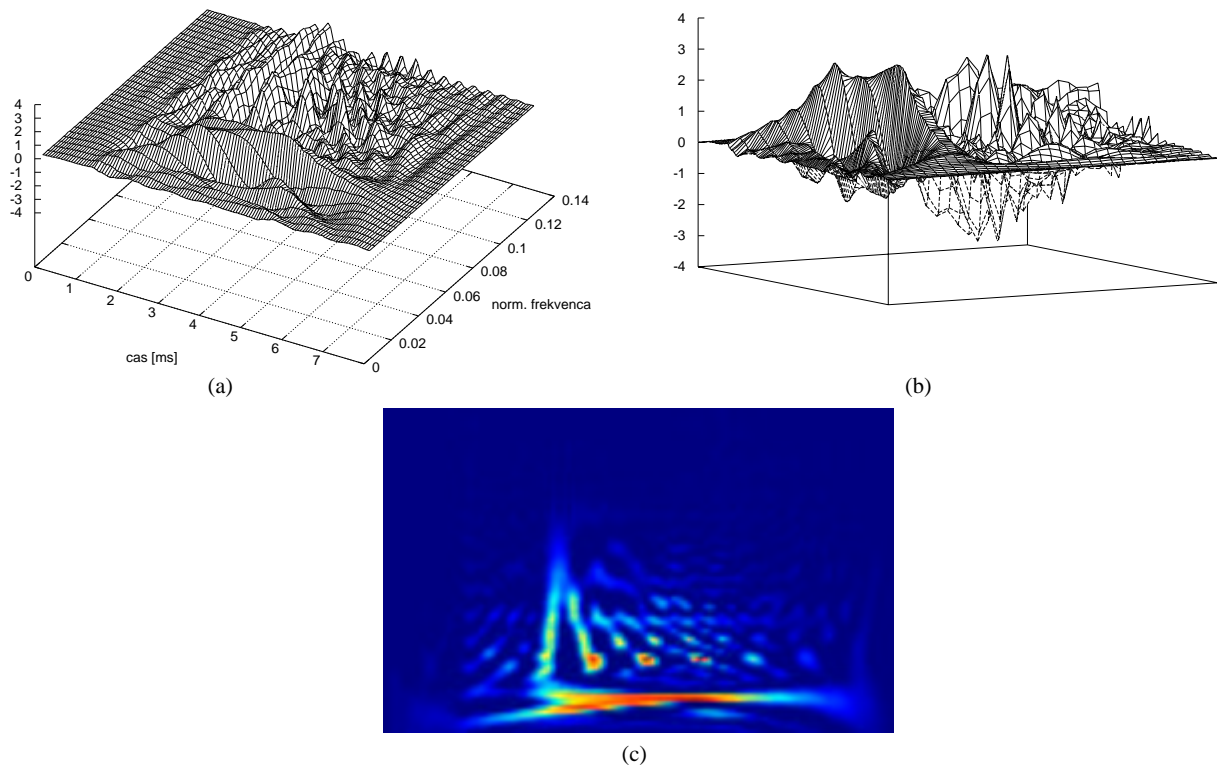
Najbolj razširjeni in uporabljeni 2D tehniki v primeru časovno–frekvenčnih predstavitev sta slika kontur in uporaba barvnih palet. V prvem primeru povezujemo s sklenjenimi barvnimi krivuljami točke z enakimi vrednostmi v ravnini, v drugem pa nam intenzivnost barve (barvni odtenek) opisuje posamezne vrednosti časovno–frekvenčne predstavitve. V obeh primerih pa moramo seveda podati še legendo. V prvem primeru povemo koliko krivulj smo izbrali in katere vrednosti nam te opisujejo. V drugem pa v bistvu podamo preslikavo med vrednostmi časovno–frekvenčne predstavitve in odtenki barv v barvni paleti. Na sliki 4.1 sta prikazani dve tehniki prikaza časovno–frekvenčne predstavitve.

S 3D sliko predstavitve v linearni skali (primer slike 4.1 (a)) brez težav predstavimo porazdelitev energije v prostoru, vendar imamo zaradi problemov pri prikazu 3D grafa težave pri razumevanju in interpretiranju slike. Pri 2D sliki predstavitve v linearni skali lahko zelo dobro lociramo spremembe porazdelitve, težave pa nastopijo v preslikavi med vrednostmi in odtenki barvne palete oziroma izbiri števil kontur. Pri logaritemski skali imamo še problem z zalogo vrednosti časovno–frekvenčne predstavitve.

Energijske predstavitve imajo različne lastnosti. V primeru prikaza predstavitev moramo upoštevati tiste, ki nam lahko pokvarijo ali deformirajo slike predstavitev.

Prva takšna problematična lastnost predstavitev je, da so lahko nekatere kompleksne. Takšne so vse omenjene točkovne predstavitve: kratkočasovna Fourierjeva in valčna transformacija ter Gaborjeva transformacija. Prvi dve imata ustrezni prevedbi s spektrogramom in skalogramom v energijske predstavitve, s slikami katerih nimamo težav.





Slika 4.1: Različne slike Wigner–Villejeve porazdelitve ene osnovne periode fonema /e/: (a) 3D slika WVD v linearni skali (ptičja perspektiva), (b) 3D slika WVD v linearni skali (stranski pogled), (c) 2D slika WVD v linearni skali.

Kompleksna energijska predstavitev pa je še Richaczkova, kjer rišemo samo realni del, ki ima ime Margenau–Hillova predstavitev.

Drugi problem je, da vse predstavitve niso pozitivne, kar bi sicer pričakovali od predstavitev, ki opisujejo porazdelitev energije. To je posledica načela kvadratne superpozicije in z njim povezanih interferenčnih členov, kar predstavlja resno oviro za razvoj aplikacij za analizo in obdelavo signalov, ki bi temeljili na takšnih predstavitvah. Narejenih je bilo kar nekaj korakov v smeri pridobivanja nenegativnih časovno–frekvenčnih predstavitev.

V primeru linearne skale to ne predstavlja ovire za 3D slike predstavitev (primer na sliki 4.1(b), kjer se lepo vidijo negativni deli, medtem ko se na sliki 4.1(a) ne). Pri 2D slikah pa negativne vrednosti pri linearni skali že deformirajo sliko, ki tako postane neberljiva. To je posledica preslikave med vrednostmi predstavitev in barvno paleto. Denimo, da predstavljajo barvno paleto odtenki iz intervala  $[0, 1]$ , zaloga vrednosti časovno frekvenčne predstavitev pa je na intervalu  $[tfr_{min}, tfr_{max}]$ . Pri tem privzemimo, da je  $tfr_{min} < 0$ . Linearna preslikava barvnih odtenkov skrči (oz. razširi) naš interval  $[tfr_{min}, tfr_{max}]$  na interval  $[0, 1]$ . To pa pomeni, da večji kot je  $|tfr_{min}|$ , slabša bo slika, saj bo le nekaj barvnih odtenkov (tistih okoli 1) opisalo skoraj vse (pozitivne) vrednosti časovno–frekvenčne predstavitev. Rešitev je več. Omenimo le dve. Najbolj preprosto je, da se kar znebimo negativnih vrednosti predstavitev (postavimo jih na 0), saj so to le vrednosti interferenčnih členov in kvečjemu “kvarijo” sliko si-

gnala. Druga možnost pa je da ne uporabimo linearne preslikave skrčitve, ampak izbiramo odtenke v neenakomernih razmikih, npr. z uporabo logaritemske skale. V obeh primerih na slikah dejansko zmanjšujemo vpliv interferenčnih členov, kar je tudi smiselno, ne pa v primeru, ko preučujemo geometrijo posameznih interferenc.

V primeru logaritemske skale pa pomeni negativnost predstavitve še večji problem. Pri govornem signalu je smiselno prikazovati porazdelitev energije v logaritemskem merilu. Tu se zgledujemo po definiciji glasnosti

$$G = 10 \log_{10} \frac{j}{j_0},$$

kjer je predstavlja  $j_0$  najmanjšo jakost, ki jo uho še zazna. V primeru časovno-frekvenčnih predstavitev govornih signalov privzamemo, da je  $j_0 = 1$  in tako dejansko rišemo in analiziramo logaritem porazdelitve. V primeru strogo pozitivnih predstavitev to seveda ni problem. Pojavi pa se vprašanje, kako vse skupaj interpretirati v primeru negativnih vrednosti predstavitve.

V primeru slik govornega signala smo ta problem rešili z uvedbo praga, od kje naprej je smiselno gledati porazdelitev energije. Torej, če je zaloga vrednosti časovno-frekvenčne predstavitve danega signala na intervalu  $[tfr_{min}, tfr_{max}]$ , definiramo prag  $p$  kot delež zaloge vrednosti, ki jo bomo opazovali. To pomeni, da dejansko rišemo novo predstavitev, ki jo definiramo na naslednji način. Naj bo  $TFR_x(t, f)$  dejanska predstavitev signala  $x$  in  $tfr_{max} = \max_{(t,f)} TFR_x(t, f)$ . Spremenjeno predstavitev tako definiramo kot

$$TFR_x^{(p)}(t, f) = \begin{cases} p \cdot tfr_{max}, & \text{če } TFR_x(t, f) \leq p \cdot tfr_{max}, \\ TFR_x(t, f), & \text{sicer.} \end{cases} \quad (4.1)$$

S popravljeno predstavitvijo  $TFR_x^{(p)}(t, f)$  tako nimamo težav z logaritmiranjem. Ravno tako pa jo lahko uporabimo tudi pri 2D slikah z linearno skalo. Pri slikah govornih signalov smo dejansko uporabili za prikaz  $\log_{10}(TFR_x^{(p)}(t, f))$ . Množenje z 10 lahko pri 2D slikah z barvnimi paletami zanemarimo, saj ne vpliva na preslikavo zaloge vrednosti v barvno paleto.

Seveda tako popravljena predstavitev ne ustreza vsem lastnostim dejanske predstavitve, za prikaz pa je nadvse primerna, čeprav z njo tako ne moremo v celoti preučevati geometrije interferenčnih členov.

### 4.1.2 Primerjava predstavitev

Tu si bomo ogledali slike različnih časovno-frekvenčnih predstavitev na primeru krajšega odseka govornega signala. Podali bomo primerjavo in analizo slik.

Za govorni signal smo si izbrali signal dveh osnovnih period fonema /e/ moškega govorca. Signal je bil vzorčen s frekvenco 16 kHz, izsek 350 točk tako predstavlja približno 22 ms signala.

Takšen govorni signal, ki predstavlja dve osnovni periodi vokala /e/, si lahko razlagamo tudi kot odziv vokalnega trakta na kvazi-periodično vzbujanje, ki ga povzročimo

z našimi glasilkami pri tvorbi zvonečih glasov [Rabiner-93]. Resonančnim frekvencam vokalnega trakta pravimo *formantne frekvence (formanti)* in so odvisne od posameznega glasu.

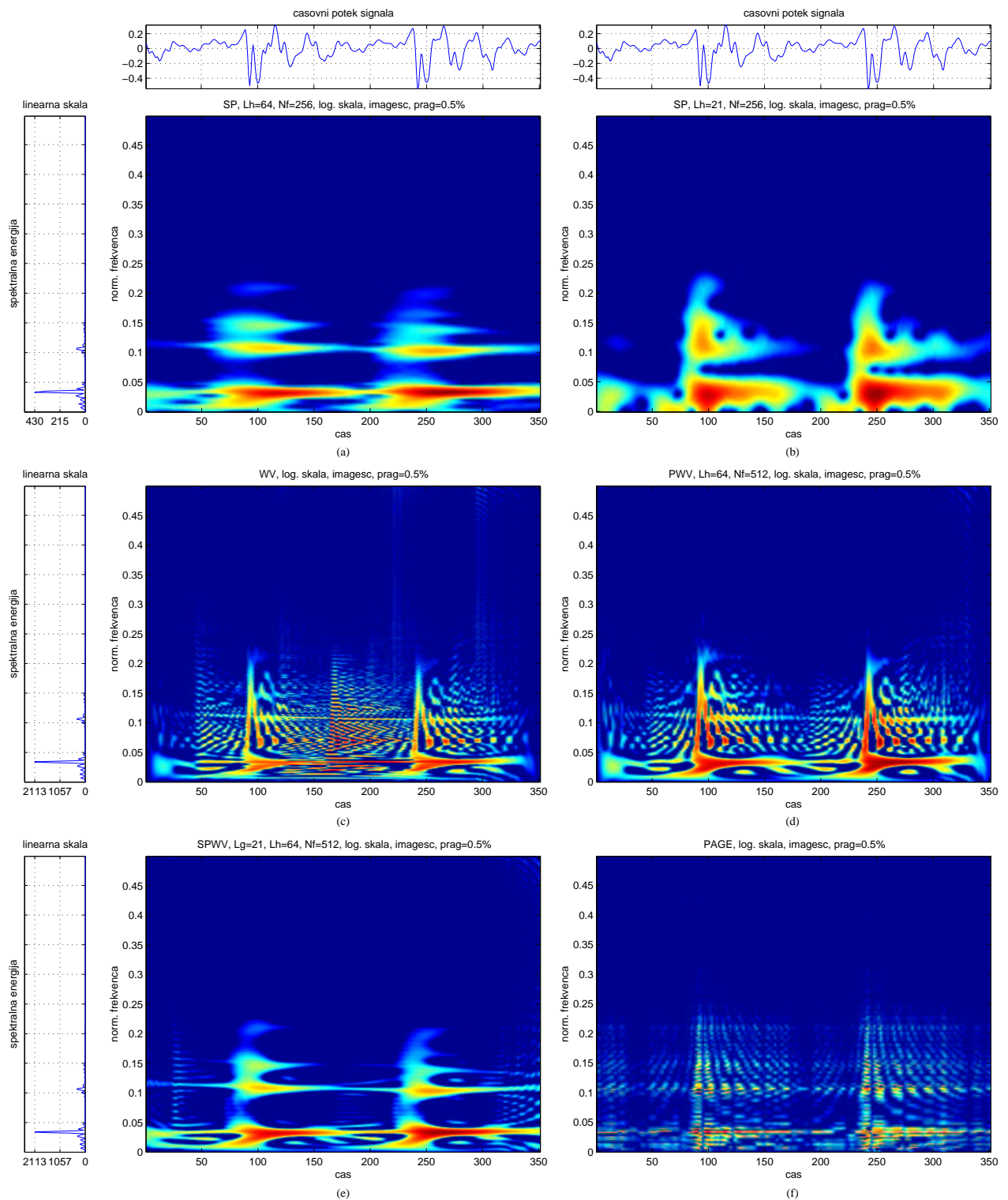
S slikami časovno–frekvenčnih predstavitev lahko opazujemo nastanek in potek posameznih formantnih frekvenc takšnega govornega signala v odvisnosti od frekvence vzbujanja, ki ji pravimo tudi *osnovna frekvenca ali frekvenca osnovnega tona*. V našem primeru opazujemo slike dveh osnovnih period zvonečega glasu /e/.

Na sliki 4.2(a) in (b) lahko primerjamo dva spektrograma analiziranega signala. V obeh primerih smo uporabili *Hammingovo okno* (pri obdelavi govornih signalov se to okno najbolj uporablja [Rabiner-78, str. 250–282]): v prvem primeru, (a), je bilo okno daljše (64 točk), v drugem, (b), pa krajše (21 točk).

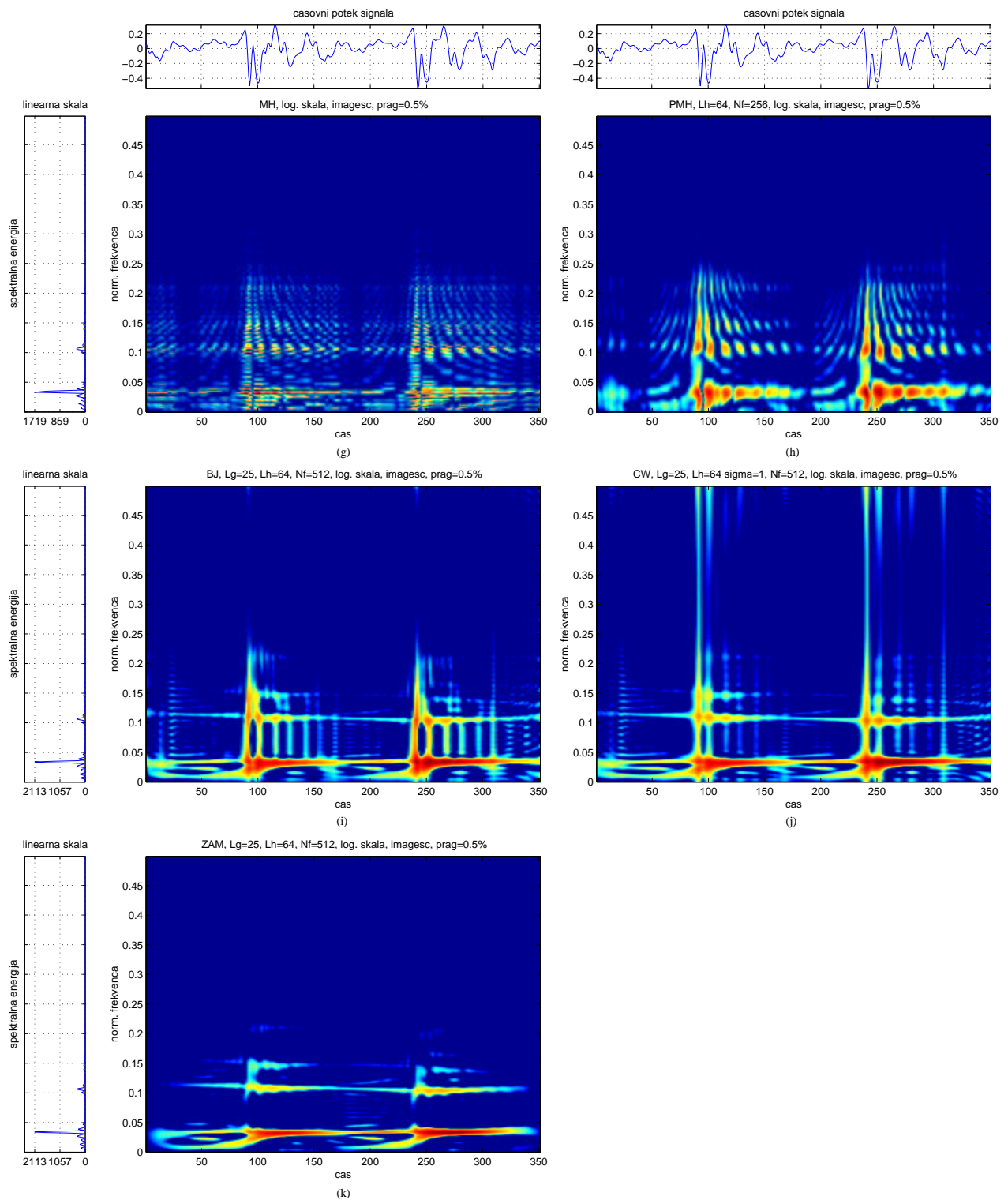
Ker je spektrogram že precej uveljavljen pri prikazu in interpretaciji časovno–frekvenčne analize signalov, bomo kar na primeru slike 4.2(a) komentirali vsebino slike. Na sliki 4.2(a) lahko opazimo 3 formantne frekvence, ki jih opisujejo lise na položajih približno 0.04 normalizirane frekvence (640 Hz), 0.12 (pribl. 1900 Hz) in približno v pasu okoli 0.15 (2400 Hz). Četrto formantno frekvenco lahko zasledimo še v pasu okoli 0.22 normalizirane frekvence (3500 Hz). Očitno je, da je prva formantna frekvenca spektralno dominantna, kar lahko opazimo tudi na spektralni sliki (ob strani) signala, kar se v barvni skali izrazi kot bolj intenzivna barva. Spektralna energija posameznih formantov pa potem pada do še komaj zaznavne četrte formantne frekvence (vse svetlejši odtenki iz barvne palete). Tu moramo poudariti, da uporabljamo logaritemsko skalo in popravljeno predstavitev definirano s (4.1) s pragom 0.5%. Podobno iz slike 4.2(a), vendar težje kot na sliki 4.2(b), lahko razberemo tudi periodo osnovnega tona govorca, ki znaša v našem primeru okoli 150 točk ali približno 9 ms, kar pomeni osnovno frekvenco okoli 100 Hz (to lahko vidimo tudi neposredno iz signala).

Sam spektrogram je nenegativna časovno–frekvenčna energijska predstavitev in zaradi svoje konstrukcije na sliki ne pričakujemo motečih interferenčnih členov, ki onemogočajo boljšo analizo signala. Pomanjkljivost pa je seveda v tem, da ne moremo sočasno doseči natančnejše frekvenčne in časovne lokalizacije posameznih komponent signala. To se lepo vidi iz slik 4.2(a) in (b), kjer imamo v prvem primeru dobro frekvenčno ločljivost (boljšo lokalizacijo po frekvenci) in slabo časovno (razmazanost na sliki (a)), v drugem pa ravno obratno: območja intenzivnejše barve (energije signala) se po času skrčijo po frekvenci pa razmažejo (ne ločimo več formantov tako dobro). Pomanjkljivost spektrograma je tudi v tem, da ne ustreza lastnostim robnih pogojev (lastnost 2, poglavje 2.3.2.1 na str. 24), to pa pomeni, da se npr. t.i. hipni glasovi (zaporniki, ...) na slikah spektrograma razmažejo [Loughlin-93].

Na tem mestu moramo poudariti, da smo pri vseh nadaljnjih predstavitev uporabili *analitične verzije* originalnega signala zaradi že opisanih težav pri vzorčenju in analizi zveznega signala. To pomeni, da smo dejansko računali časovno–frekvenčne predstavitve *Hilbertove transformacije* signala (definicija (2.3), poglavje 2.1.3.1). S tem pa nismo bistveno pokvarili oziroma spremenili časovno–frekvenčne slike porazdelitve energije v signalu, kar se lahko prepričamo iz preostalih slik. Pri spektrogramu bi lahko dejansko vzeli tudi analitično verzijo realnega signala, vendar pri tej predstavitvi nimamo težav z interferenco ob Nyquistovi frekvenci vzorčenja.



Slika 4.2: Slike različnih predstavitev dveh osnovnih period fonema /e/: (a) spektrogram, dolgo okno, (b) spektrogram, kratko okno, (c) Wigner–Villejeva predstavitev, (d) psevdo Wigner–Villejeva predstavitev, (e) zglajena psevdo Wigner–Villejeva predstavitev, (f) Pageova predstavitev.



Slika 4.3: Slike različnih predstavitev dveh osnovnih period fonema /e/ (nadaljevanje slike 4.2): (g) Margenau–Hillova predstavitev (h) psevdo Margenau–Hillova predstavitev, (i) Born–Jordanova predstavitev (zglajena verzija), (j) Choi–Williamsova predstavitev (zglajena verzija), (k) Zhao–Atlas–Marksova predstavitev (zglajena verzija).

Slike 4.2(c), (d) in (e) prikazujejo Wigener–Villejeve različice predstavitev, ki zaradi številnih matematičnih lastnosti zelo natančno po času in frekvenci opišejo porazdelitve energije v signalu. Na žalost pa je zaradi teh lastnosti slika tudi znatno pokvarjena, pojavi se namreč pojav izrazite interference. Interferenčni členi se v primeru Wigener–Villejeve predstavitve (WVD) (slika 4.2(c)) pojavijo na polovici osnovne periode fonema, vzdolž frekvenčne osi pa nastopi interferenca med posameznimi formantnimi frekvencami. Kljub temu, da so na sliki prikazani samo pozitivni členi in smo negativne interferenčne člene zanemarili, nam ti zelo kvarijo sliko. Vseeno lahko opazimo najboljšo (v primerjavi z ostalimi) časovno in frekvenčno lokalizacijo posameznih dogodkov v časovno–frekvenčni ravnini.

Z glajenjem po frekvenci (4.2(d)) v primeru psevdo Wigener–Villejeve predstavitve (PWVD) smo odpravili interferenčne člene na sredini osnovne periode (interferenčni členi na sredini so izginili), poslabšali pa smo frekvenčno ločljivost. Tu smo za glajenje po frekvenci uporabili Hammingovo okno dolžine 64 točk.

Z dodatnim glajenjem po času (4.2(e)) pri zglajeni psevdo Wigener–Villejevi predstavitvi (SPWVD) smo odpravili interferenčne člene vzdolž frekvenčne osi. Tu smo za glajenje po času uporabili Hammnigovo okno dolžine 21 točk. S tem so postale vidne vse formantne frekvence, dodatno pa smo poslabšali časovno lokalizacijo. V primerjavi s slikama spektrograma s SPWVD natančneje opišemo porazdelitev energije.

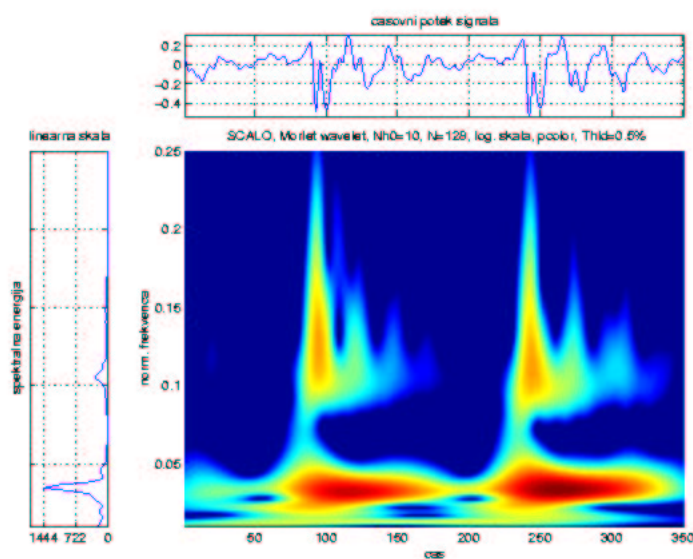
Drugo skupino predstavitev prikazujejo slike 4.2(f), 4.3(g) in 4.3(h). Ideje izvedbe predstavitev se tu razlikujejo kot pri WVD. Pageova predstavitev (PD, slika 4.2(f)) temelji na principu kavzalnosti, opisuje namreč porazdelitev energije na podlagi akumulacije energije do določenega trenutka v času. Margenau–Hillova predstavitev (MHD) (4.3(g)) je izpeljanka Richaczkove porazdelitve, kjer gre tudi za drugačen pristop opisovanja energije signala. Podobnost principov modeliranja obeh predstavitev se kaže v zrnatosti slik 4.2(f) in 4.3(g). Tudi tu sklepamo kompromis med natančnostjo predstavitve in interferenco. Kot lahko vidimo, imamo že v osnovnih verzijah PD in MHD manj interferenčnih členov kot pri WVD, hkrati pa lahko zelo natančno določimo formantne frekvence iz slik 4.2(f) in 4.3(g). Na sliki 4.3(h) pa je prikazana glajena (po frekvenci) verzija Margenau–Hillove predstavitve (PMHD). Slika je bolj čista kot v primeru 4.3(g), formantne frekvence so sorazmerno dobro vidne, časovna ločljivost je dobra, slabša pa frekvenčna. Tu smo za glajenje po frekvenci uporabili Hammingovo okno dolžine 64 točk.

Preostale tri predstavitve, Born–Jordanovo (BJD, 4.3(i)), Choi–Williamsovo (CWD, 4.3(j)) in Zhao–Atlas–Marksovo (ZAM, 4.3(k)), dobimo na podlagi konstrukcije jeder iz večlične funkcije. Tu izbiramo takšna jedra, da ohranimo čim več lastnosti predstavitve in hkrati čimbolj zmanjšamo vpliv interference. Tu so predstavljene zglajene verzije predstavitev, kjer smo izvajali glajenje s Hammingovim oknom po času dolžine 25 točk, po frekvenci pa 64 točk. Kljub temu lahko opazujemo posledice izbire osnovnih jeder za konstrukcijo predstavitev. Tudi interferenca se tu drugače porazdeljuje kot npr. pri WVD. Pri Born–Jordanovi porazdelitvi (slika 4.3(i)) nastopajo interference med formanti pravokotno na časovno os, kar je posledica jedra (definicija (2.28)). Pri Choi–Williamsovi porazdelitvi lahko opazimo navpične črte po celotni sliki, ki so prav tako posledica izbire jedra pri izbrani  $\sigma = 1$  (glej sliko 2.22). Zelo čisto sliko dobimo

z zglajeno verzijo Zhao–Atlas–Marksove porazdelitve (slika 4.3(k)). Lastnost jedra te predstavitve je, da interferenčne člene, ki nastajajo vzdolž frekvenčne osi (vzporedno s časovno osjo) vlečemo skupaj v same spektralne komponente, ki se pojavijo v signalu. Ocena energije se lokalno rahlo pokvari, interference pa se tako delno znebimo. V našem primeru smo uporabljali jedro (2.30) pri  $\alpha = 1$ . Kot lahko vidimo iz slike 4.3(k), ZAM predstavitev zelo dobro in natančno opiše poteke posameznih formantnih frekvenc (podobno kot SPWD) in zaradi tega predstavlja resno alternativo spektrogramu pri analizi in obdelavi govornih signalov.

V magistrskem delu smo se bolj posvetili primerjavi energijskih predstavitev, ki spadajo v *Cohenov razred*, manj pa afnim predstavitvam. Razlogov je več, poglobljen pa je, da smo proučevali vpliv predstavitev na razpoznavanje govora. In Cohenov razred porazdelitev predstavlja bolj naravno in očitno povezavo med časovno–frekvenčno sliko in izbiro značilk. Vsekakor pa tudi afine predstavitve, kjer se ukvarjamo predvsem s časom in skalo, predstavljajo resen izziv za obdelavo, analizo in predvsem razpoznavanje govora. Primere afinih predstavitev si lahko ogledamo v dodatku A magistrskega dela.

Tu si oglejmo samo en primer takšne predstavitve istega govornega odseka fonema /e/. Na sliki 4.4 je prikazan *skalogram* govornega signala.



Slika 4.4: Slika skalograma dveh osnovnih period fonema /e/.

Za osnovo smo si izbrali Morletove valčke. Analiza je bila izvedena na analitični verziji signala v frekvenčnem pasu od 0.01 do 0.25 normalizirane frekvence v diadični skali.

Iz slike 4.4 je jasno razvidna časovno–frekvenčna ločljivost tako značilna za afine predstavitve: časovna ločljivost s frekvenco narašča, frekvenčna natančnost pa pada.

Kot smo lahko razbrali iz slik posameznih predstavitev, moramo vedno tehtati med natančnostjo in interferenco. Seveda ni nujno, da je dovolj natančna slika tista, ki nam dá dovolj informacije za nadaljnjo obdelavo in analizo. Tudi interferenca lahko pove nekaj o sami razporeditvi spektralnih komponent in jo tako lahko upoštevamo kot informacijo v nadaljevanju. Iz slik se je potrdilo, da predstavlja ZAM predstavitev

resno konkurenco spektrogramu pri obdelavi in analizi govornega signala. Slike danih predstavitev nam tako podajajo nek občutek, kaj lahko od posamezne predstavitve pričakujemo in katero lahko uporabimo za nadaljnjo obdelavo signala.

V nadaljevanju si bomo ogledali še en način, kako izboljšati časovno–frekvenčno ločljivost danih predstavitev.

### 4.1.3 Popravljenе predstavitev

Kot smo lahko videli v prejšnjem poglavju, je berljivost slik nestacionarnih signalov z različnimi predstavitvami v veliki meri odvisna od pojava interference. V prejšnjem poglavju smo poskušali odpraviti te člene direktno z glajenjem po času in frekvenci osnovnih predstavitev, vendar smo s tem izgubili natančnost predstavitve. Tu pa si bomo na primeru govornega signala ogledali še en posreden postopek odpravljanja interference in hkrati izboljšanja natančnosti predstavitve. Izboljšanim predstavitvam pridobljenih na podlagi takšnih metod (*metode realokacije, angl. reassignments methods*) pravimo *realocirane predstavitev*. Tu se naslanjamo predvsem na delo F. Augerja in ostalih [Auger-95].

#### 4.1.3.1 Realocirani spektrogram

Na primeru spektrograma bomo razložili idejo teh metod. Sprva je bil v nadaljevanju opisani postopek namenjen izboljšavi natančnosti spektrograma. To pa zato, ker ima spektrogram veliko dobrih lastnosti: je nenegativna porazdelitev, ima zelo šibke interferenčne člene in ga lahko zelo hitro izračunamo. Omejen pa je s principom nedoločenosti, zato ne moremo doseči poljubne natančnosti predstavitve sočasno po času in frekvenci.

Če še enkrat zapišemo spektrogram kot dvodimenzionalno konvolucijo Wigner–Villejeve porazdelitve okna in obravnavanega signala:

$$S_x(t, f; h) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x(s, \xi) W_h(t - s, f - \xi) ds d\xi, \quad (4.2)$$

lahko ugotovimo, da pri spektrogramu dejansko seštevamo lokalne prispevke Wigner–Villejeve (WV) porazdelitve signala znotraj področja okna  $W_h(t - s, f - \xi)$  s središčem v točki  $(t, f)$  časovno–frekvenčne ravnine. Ti prispevki ne predstavljajo nič drugega kot uteženo povprečje WV predstavitve znotraj območja delovanja okna, ki je simetrično glede na točko  $(t, f)$ . Ta točka torej predstavlja geometrično sredino WVD okna.

Osnovna ideja izboljšave je v tem, da si lahko izberemo tudi kakšno drugo točko  $(\hat{t}, \hat{f})$  znotraj območja delovanja okna, okrog katere akumuliramo prispevke za predstavitev. Ena izmed možnosti je, da izberemo za točko *masno središče*.

To je intuitivno še najbližje temu, kar v resnici počnemo pri spektrogramu. Lokalno porazdelitev energije  $W_x(s, \xi) W_h(t - s, f - \xi)$  namreč lahko obravnavamo v jeziku fizike (mehanike) kot porazdelitev mase sistema, ki ga tvori območje delovanja okna. Zatorej je bolj smiselno obravnavati porazdelitve okoli masnega središča in ne geometrijskega.



Če to razmišljanje še matematično osmislimo, gre vse skupaj nekako takole: vse vrednosti spektrograma, izračunane okoli točk  $(t, f)$ , premaknemo v točke  $(\hat{t}, \hat{f})$ , ki predstavljajo masna središča porazdelitve energije signala okoli točke  $(t, f)$ :

$$\hat{t}(x; t, f) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s W_x(s, \xi) W_h(t - s, f - \xi) ds d\xi}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x(s, \xi) W_h(t - s, f - \xi) ds d\xi} \quad (4.3)$$

$$\hat{f}(x; t, f) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi W_x(s, \xi) W_h(t - s, f - \xi) ds d\xi}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x(s, \xi) W_h(t - s, f - \xi) ds d\xi}. \quad (4.4)$$

Realocirano verzijo spektrograma okoli točk  $(t', f')$  pa izračunamo kot vsoto prispevkov spektrograma glede na točke  $(\hat{t}, \hat{f})$ :

$$S_x^{(r)}(t', f'; h) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_x(t, f; h) \delta(t' - \hat{t}(x; t, f)) \delta(f' - \hat{f}(x; t, f)) dt df \quad (4.5)$$

Najbolj zanimiva lastnost popravljenega spektrograma je, da za izračun porazdelitve v tem primeru potrebujemo tudi informacijo o fazi kratkočasovne Fourierjeve transformacije in ne samo kvadrat amplitude, kot je to pri navadnem spektrogramu. To se vidi iz drugega zapisa točk  $\hat{t}$  in  $\hat{f}$  [Auger-95].

Kljub temu, da takšna predstavitev ni več bilinearna, še vedno ohranja premike po času in frekvenci in kar je v našem primeru bistveno, popolno lokalizira linearne cvrke (lastnost 12 pri WVD, poglavje 2.3.2.1 na str. 24). To pa pomeni, da tako s popravljenjo verzijo lahko izboljšamo ločljivost predstavitve.

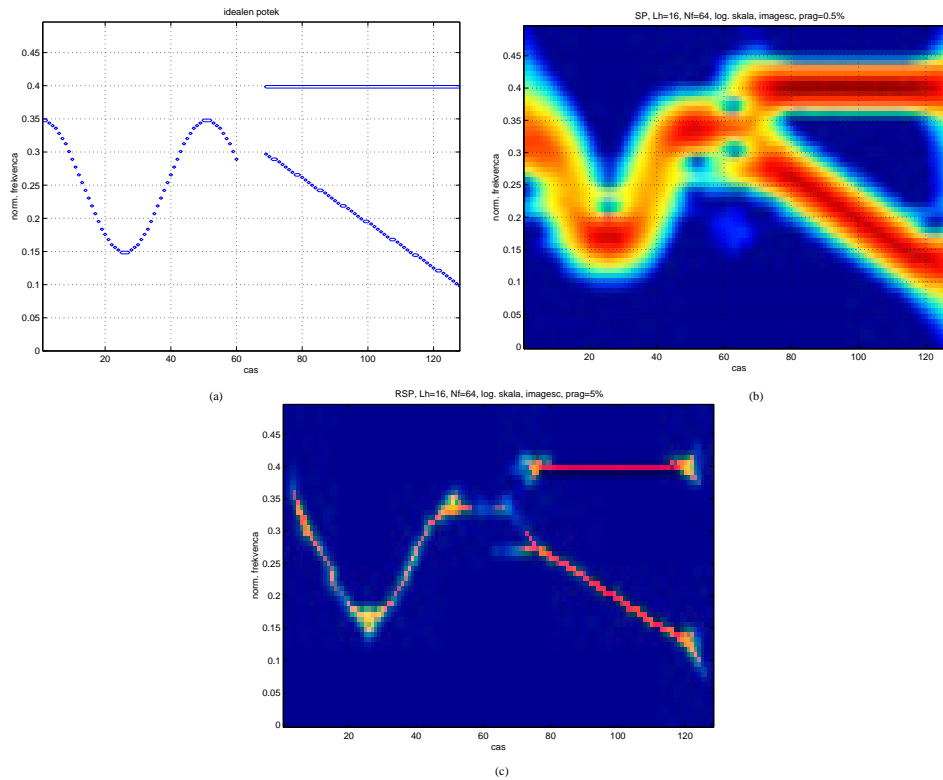
V to se lahko prepričamo na primeru sintetiziranih signalov na sliki 4.5. Tu imamo primer večkomponentnega nestacionarnega signala sestavljenega iz sinusno modulirane komponente (prvi del), sinusnega valovanja in linearne cvrke (v drugem delu). Idealni potek v časovno-frekvenčnem prostoru je prikazan na sliki 4.5(a). Desno zgoraj imamo primer navadnega spektrograma (slika 4.5(b)), spodaj pa realocirano verzijo (slika 4.5(c)). S časovno-frekvenčno ločljivostjo popravljenega spektrograma se skoraj približamo idealnemu poteku, z navadnim spektrogramom pa je vse skupaj bolj razmazano v obeh smereh.

#### 4.1.3.2 Realokacija predstavitev Cohenovega razreda

Idejo opisano pri spektrogramu lahko posplošimo tudi na preostale predstavitve. Namesto zveze (4.2) zapišemo splošno verzijo predstavitev Cohenovega razreda s

$$C_x(t, f; \Pi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Pi(t - s, f - \xi) W_x(s, \xi) ds d\xi. \quad (4.6)$$

Če v definicijah (4.3), (4.4) in (4.5) zamenjamo jedro  $W_h(s, \xi)$  s splošnim jedrom  $\Pi(s, \xi)$ ,



Slika 4.5: Primerjava spektrogramov: (a) idealen potek signala v časovno-frekvenčni ravnini, (b) navaden spektrogram, (c) realocirani spektrogram.

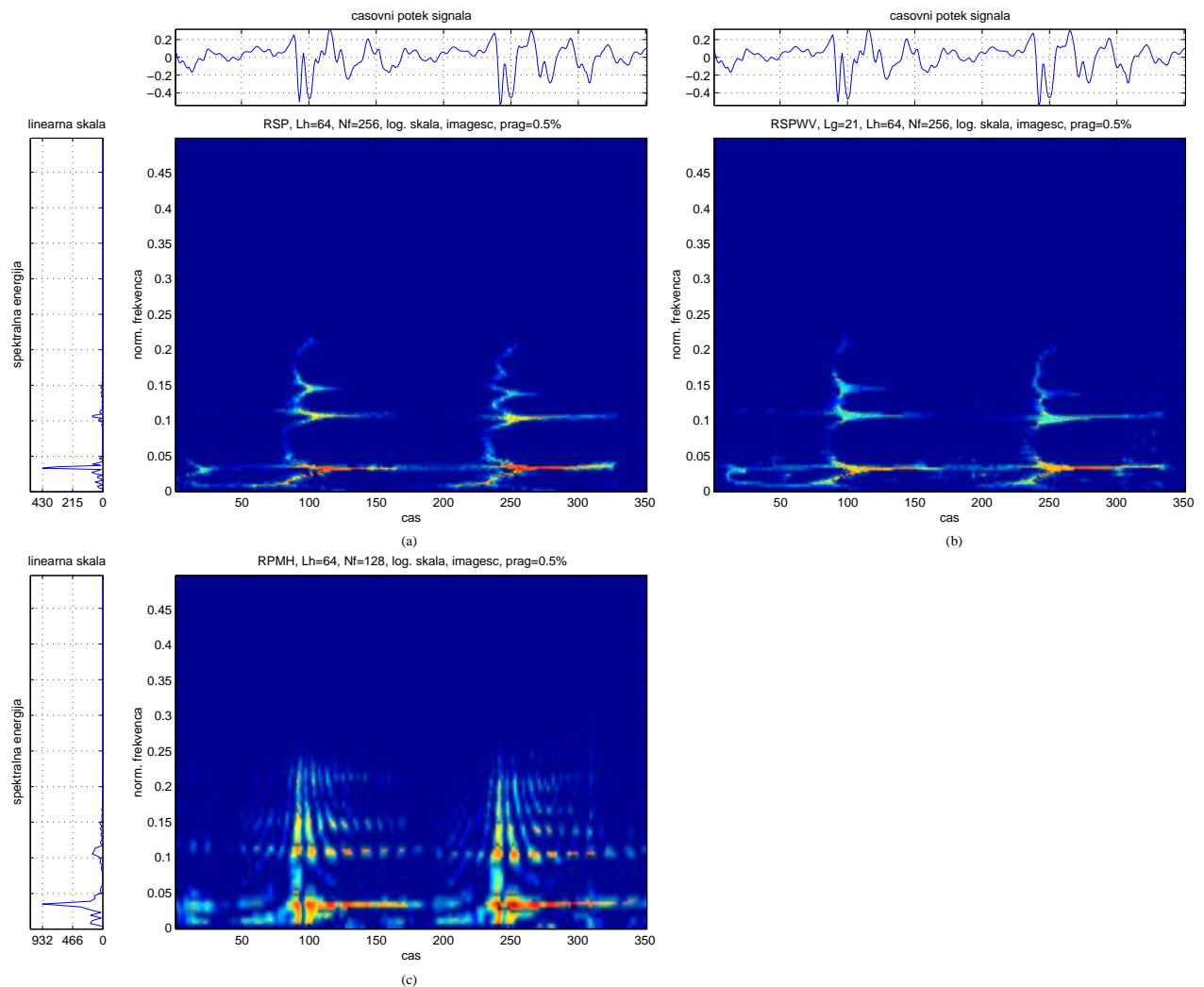
lahko definiramo nove realocirane verzije predstavitev Cohenovega razreda kot:

$$\begin{aligned}\hat{t}(x; t, f) &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s \Pi(t-s, f-\xi) W_x(s, \xi) ds d\xi}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Pi(t-s, f-\xi) W_x(s, \xi) ds d\xi}, \\ \hat{f}(x; t, f) &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi \Pi(t-s, f-\xi) W_x(s, \xi) ds d\xi}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Pi(t-s, f-\xi) W_x(s, \xi) ds d\xi}, \\ C_x^{(r)}(t', f'; \Pi) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_x(t, f; \Pi) \delta(t' - \hat{t}(x; t, f)) \delta(f' - \hat{f}(x; t, f)) dt df\end{aligned}$$

Na splošno tudi tu izboljšamo slike predstavitev, se pravi zmanjšamo vpliv interferenčnih členov in povečamo natančnost lokalizacije spektralnih komponent v časovno-frekvenčni ravnini. Vse popravljene verzije ustrezajo lastnosti 12 pri WVD (poglavje 2.3.2.1 na str. 24).

Podobno lahko posplošimo idejo tudi na afine predstavitve. Lahko pa izvedemo samo delne popravke, in sicer samo po času ali pa po frekvenci, odvisno od tega, katera informacija nas zanima in je tudi pomembna pri samem signalu.

Slike posameznih realociranih verzij predstavitev si lahko ogledamo na primeru dveh osnovnih period fonema /e/, slika 4.6. Uporabljeni so bili enaki parametri pridobivanja slik kot pri originalnih verzijah, slike 4.2 (a), (e) in 4.3 (h). Na sliki 4.6(a) je prikazana porazdelitev energije, kot jo opiše realocirani spektrogram. V primerjavi z originalno verzijo (slika 4.2(a)) lahko ugotovimo boljše ločljivost tako v času, še bolj očitno pa



Slika 4.6: Slike različnih realociranih predstavitev dveh osnovnih period fonema /e/: (a) realocirani spektrogram, (b) realocirana zglajena psevdo Wigner–Villejeva predstavitev, (c) realocirana psevdo Margenau–Hillova predstavitev.

po frekvenci. Ravno tako je tudi pri ostalih dveh slikah realocirane zglajene psevdo Wigner–Villejeve (RSPWVD, slika 4.6(b)) in psevdo Margenau–Hillove predstavitve (RPMH, slika 4.6(c)). Če na sliki 4.6 primerjamo popravljene predstavitve med sabo, lahko ugotovimo, da sta izboljšani spektrogram in SPWVD skoraj identična, kar je posledica idealne lokalizacije linearnih cvrkov, pri RPMH pa je še vedno prisoten vpliv osnovne verzije, čeprav je lokalizacija (predvsem po frekvenci) boljše in vpliv interference manjši.

Metode popravljanja osnovnih verzij predstavitev so pri analizi govora že uspešno uporabili pri določevanju formantnih frekvenc iz govornega signala [Plante-95], pri ocenjevanju osnovne frekvence govora [Ainsworth-98] in pri izboljšavi razpoznavanja vokalov [Yang-98]. Kljub vsemu pa takšne predstavitve ponujajo še veliko odprtih vprašanj in novih možnosti uspešne uporabe pri različnih nalogah obdelave in analize govora.

## 4.2 Razpoznavanje govora

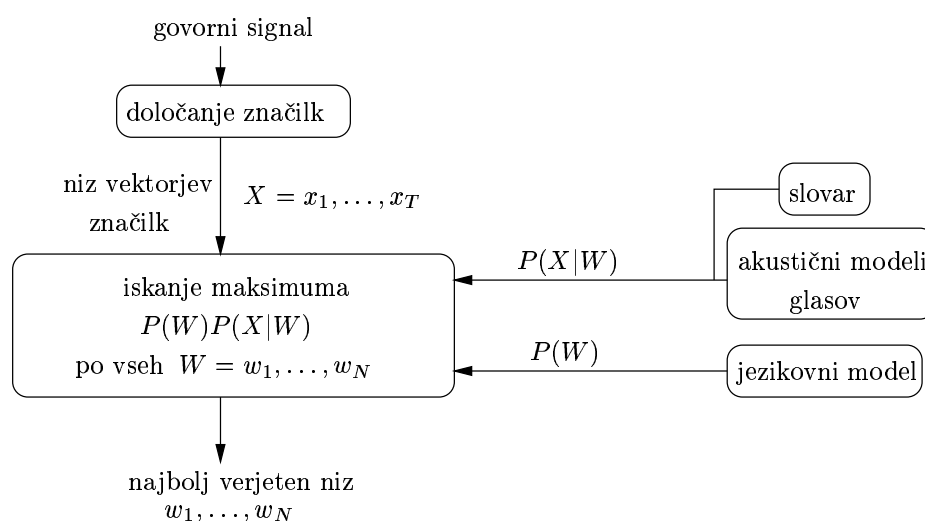
Medtem ko smo se v prejšnjem razdelku posvečali predvsem slikam govornega signala pridobljenimi z različnimi predstavitvami, bomo v tem razdelku uporabili predstavitev za pridobivanje različnih značilk govornega signala primernih za razpoznavanje govora. Naš namen ni bil toliko poiskati “najboljše” značilke v smislu postopkov pridobivanja značilk, ampak smo se podobno kot pri slikah posvetili primerjavi različnih časovno–frekvenčnih predstavitev v primeru razpoznavanja govora. Hoteli smo poiskati alternative obstoječemu principu pridobivanja značilk, ki temelji na spektrogramu.

Govorni signal je zelo težaven za časovno–frekvenčno analizo. Vsebuje tako izrazite nestacionarnosti, kratke, hipne spremembe (npr. odpore pri zapornikih), kot tudi daljše stacionarne dele (npr. vokali), [Mihelič-93]. Zaradi tega predstavlja veliki izziv časovno–frekvenčnim predstavitvam. Potrebna je namreč smiselna izbira predstavitev, ki bi dovolj natančno opisala potek spektralnih komponent po času, hkrati pa ne bi smela imeti stranskih efektov. Tu mislimo predvsem na interferenčne člene, ki dejansko pokvarijo pravo sliko (govornega) signala.

V nadaljevanju bomo najprej opisali sistem, ki smo ga uporabili za razpoznavanje, in govorno zbirko, na kateri smo izvajali poskuse razpoznavanja. Kasneje pa bomo predstavili nekaj rezultatov razpoznavanja z različnimi časovno–frekvenčnimi predstavitvami. Opisali bomo metode uporabe teh predstavitev in probleme, ki smo jih pri tem reševali.

### 4.2.1 Razpoznavalnik govora

Sistem, ki smo ga uporabili za naše poskuse, je zelo razširjen in uveljavljen pri razpoznavanju govora. Shema sistema je prikazana na sliki 4.7.



Slika 4.7: Shema razpoznavalnika govora, ki temelji na Bayesovem odločitvenem pravilu.

Pri razpoznavanju govora, ki temelji na statističnem modeliranju, gre za iskanje zaporedja govornih enot  $W = w_1, \dots, w_N$  (besede, podbesedne enote, fonemi, ...), ki

je najbolj verjetno glede na zaporedje akustičnih dogodkov (vektorji značilnik)  $X = x_1, \dots, x_T$  pridobljenih v času  $t = 1, \dots, T$ . Torej iščemo  $\arg \max_W P(W|X)$ . Z uporabo Bayesovega pravila prevedemo  $P(W|X) = P(W)P(X|W)/P(X)$ . Verjetnost  $P(W)$  ocenjujemo z jezikovnim modelom. Pogojno verjetnost  $P(X|W)$  pa opišemo z akustičnimi modeli. Pri modeliranju govornih enot najpogosteje uporabljamo prikrite Markovove modele.  $P(X)$  je verjetnost zaporedja akustičnih dogodkov in v primeru, ko so vektorji značilnik neodvisni od obravnavanih govornih enot  $w_i$ , jo lahko zanemarimo.

Za ocenjevanje parametrov sistema in tudi za testiranje, potrebujemo ustrezno količino pravilno označenih in skrbno izbranih podatkov. Pri razpoznavanju govora so podatki zajeti v govorni zbirki.

#### 4.2.1.1 Govorna zbirka K211d

Zbirka, ki smo jo uporabili pri naših analizah, je govorna zbirka K211d. Podrobneje je opisana v doktorski disertaciji S. Dobriška [Dobrišek-01]. Tu bomo podali samo nekaj osnovnih podatkov pomembnih za nadaljevanje.

Govorna zbirka K211d vsebuje ročno razčlenjene in označene posnetke posamezno izgovorjenih besed. Zbirka je namenjena predvsem analizi nekaterih glasov in sklopov glasov ter primerna za izvajanje hitrih preizkusov in ovrednotenih postopkov izboljšav posameznih komponent pri modeliranju razpoznavalnika glasov in parametrizacije govornega signala. Nastala je v okviru raziskovalnega dela Laboratorija za umetno zaznavanje, sisteme in kibernetiko in predstavlja poleg večje zbirke GOPOLIS [Dobrišek-96] in zbirke posnetkov televizijskih in radijskih vremenskih oddaj, VNTV [Žibert-00], temelj za razvoj aplikacij, ki slonijo na razpoznavanju slovenskega govora.

Zbirka K211d vsebuje statistično uravnoteženo število glasov in sorazmerno večje število glasov, ki so v govorjenem slovenskem jeziku bolj redki. To pa predstavlja skupaj z dejstvom, da so posnetki ročno označeni, idealno osnovo za preizkušanje novih značilnik primernih za razpoznavanje.

#### Statistika zbirke

Zbirka vsebuje nekaj manj kot uro posnetega materiala. Deset govorcev (5 žensk in 5 moških) je izgovorilo ločeno 251 besed. Ta korpus besed predstavlja hkrati tudi besednjak zbirke. Govor je posnet s kakovostnim mikrofonom. Signal je vzorčen s šestnajstbitnim analogno digitalnim pretvornikom z linearnim kvantizatorjem amplitude in s frekvenco vzorčenja 16 kHz. Posnetek vsake posamezno izgovorjene besede vsebuje še okoli 200 milisekund začetnega in končnega premora.

Besednjak je opremljen z različnimi fonetičnimi prepisi. Za razpoznavanje v nadaljevanju smo uporabili širši fonetični prepis, ki ga sestavljajo alofoni podani v tabeli 4.1.

Tu je potrebno poudariti alofona [U] in [I], ki predstavljata dvoglasniške sklope samoglasnikov s fonemoma /v/ in /j/. Smiselnost uporabe teh alofonov je utemeljena v [Dobrišek-01]. Poleg vseh alofonov sta v zbirki dodana še simbola *sis* in *sie*,

Tabela 4.1: Alofoni širšega fonetičnega prepisa. Zapis v MRPA [Wells-92] formatu.

skupina glasov	alofoni
samoglasniki	@ E O a e i o u
zvočniki	j I l r v w W U m n
nezvočniki	p b f t d s z ts S Z tS k g x

ki označujeta tišino pred in za izgovorjeno besedo.

V širših fonetičnih prepisih besed je vse skupaj 17.107 glasov. Za učni del zbirke smo uporabili prve tri ženske in moške govorce, preostale štiri govorce pa za testni del. S tem smo dosegli razpoznavanje, ki je od govorca neodvisno.

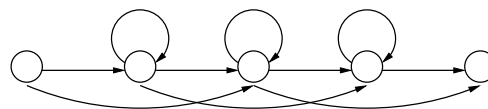
#### 4.2.1.2 Kako smo učili

Tu bomo govorili o modelu in o določevanju parametrov tega modela, ki smo ga uporabili za osnovo razpoznavanja testnega dela zbirke K211d. Opisali bomo tiste dele in postopke razpoznavalnika, ki jih pri preizkušanju različnih časovno-frekvenčnih predstavitev nismo spreminjali. Za medsebojno primerjavo rezultatov smo pazili, da smo uporabili enotne postopke za učenje modelov in da smo uporabljali enake modele.

Za statističen model razpoznavalnika (slika 4.7) smo uporabili danes najbolj uveljavljene in razširjene prikrite Markovove modele, PMM [Jelinek-98]. Modeliranje razpoznavalnika govora smo izvedli s pomočjo orodja HTK [Young-00].

Izvedli smo modeliranje dveh vrst govornih enot: modelirali smo monofonske in difonske enote. Pri monofonih gre za modeliranje jeder govornih enot, medtem ko pri difonskih govornih enotah modeliramo prehode med govornimi enotami [Dobrišek-01]. To je smiselno, če hočemo preizkušati različne značilke govornega signala. Tako namreč lahko ugotovimo, katere dele govornih odsekov bolje opisujejo izbrane značilke.

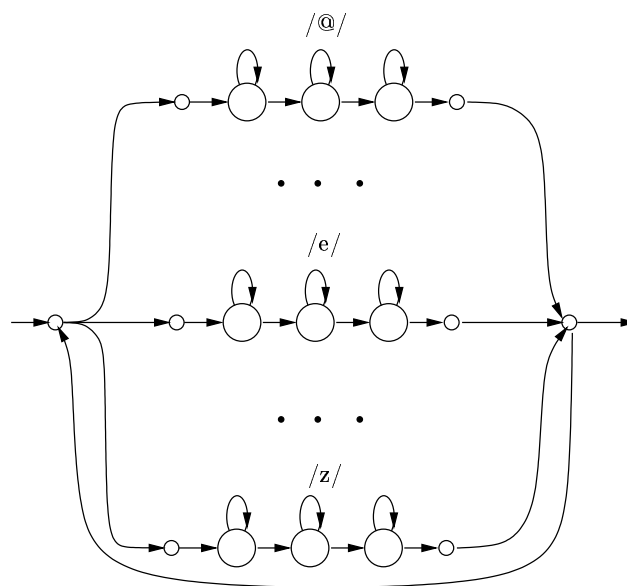
Ne glede na izbiro govornih enot pa smo uporabljali nespremenjeno topologijo PM modelov. Odločili smo se za klasičen levo desni model [Pavešić-00, str. 515–517], prikazan na sliki 4.8.



Slika 4.8: Levo desni PMM.

Uporabljali smo modele s tremi stanji. Porazdelitev akustičnih značilk pa smo modelirali z linearno kombinacijo Gaussovih porazdelitev verjetnosti z diagonalnimi kovariančnimi matrikami.

Poudariti moramo, da imamo tu v mislih razpoznavalnik glasov in ne besed, zato je tu topologija modela razpoznavalnika malce različna kot pri razpoznavanju besed. Topologija modela je prikazana na sliki 4.9.



Slika 4.9: Topologija PMM za razpoznavanje glasov.

V samem postopku gradnje razpoznavalnika nismo eksplicitno upoštevali jezikovnega modela za glasove. Dejstvo pa je, da se jezikovni model “skriva” v sami zbirki, saj uporabljamo tako pri učenju kot pri razpoznavanju iste besede (drugih govorcev), se pravi ista zaporedja posameznih glasov.

Parametre modela PMM smo ocenjevali na podlagi pridobljenih značilk iz učnega dela zbirke K211d. Tu smo uporabili vsakič drugačne značilke govornih odsekov pridobljene na podlagi različnih časovno–frekvenčnih predstavitev, ki jih bomo natančneje opisali v naslednjem razdelku. Poudariti pa moramo, da smo pridobivali vse tipe značilk iz izbranih vnaprej določenih fiksnih odsekov govornega signala (okvirjev), torej pri določanju značilk nismo upoštevali predhodne segmentacije govornega signala po govornih enotah.

Učenje je potekalo v več fazah in na standarden način. V prvem koraku smo inicializirali verjetnosti prehodov iz stanja v stanje (matrike prehodov) ter določili začetne približke funkcijam gostot verjetnosti, torej vektorje povprečnih vrednosti in variance v diagonalnih kovariančnih matrikah v stanjih PM modelov. To smo izvedli na podlagi ročno označenih posnetkov pri monofonskih modelih, pri difonskih pa smo uporabili segmentacijo pridobljeno iz sintetizatorja slovenskega govora [Gros-97]. Prvo ocenjevanje je potekalo s postopkom *Viterbijevega učenja* [Jelinek-98, str.25] z orodjem HInit [Young-00].

Pri tem postopku izvajamo v bistvu prileganje vektorjev značilk nizu modelov govornih enot (ki jih modeliramo) z iskanjem najbolj verjetne poti skozi PM model. Ceno poti skozi graf določajo verjetnosti prehodov med stanji in verjetnosti oddanih simbolov v stanjih opisanih s porazdelitvami (gostotami) verjetnosti. Najboljša pot nam tako pove, kateri vektorji značilk so bili najbolj verjetno oddani iz posameznih stanj avtomatov (PMM) v vseh modelih. Vektorje nato pripišemo najbolj verjetnim stanjem in iz vseh vektorjev, pripisanih nekemu stanju pri prehodu govorne zbirke, izpeljemo nove vrednosti parametrov vsakega stanja posebej.

Podobno deluje tudi *Baum–Welchov postopek* [Jelinek-98, str.27], ki smo ga uporabili za nadaljnje ocenjevanje parametrov. Sprememba je v tem, da nam parametre modela ne določajo samo vektorji značilnik, ki so šli skozi ta model, ampak vsi vektorji. Pri tem s pomočjo rekurzivnih formul določamo uteži posameznih vektorjev v stanjih, ki tako pomenijo verjetnost, da je bil posamezen vektor oddan iz tega stanja. V bistvu gre tu za mehčanje uteži po stanjih, ki nastopajo pri Viterbijevem učenju (pri Viterbiju imamo uteži 0 ali 1, tu pa verjetnosti, torej od 0 do 1). Ta postopek smo izvedli z orodjem HRest [Young-00].

Za dokončno ocenjevanje smo uporabili različico tega postopka (*angl. embedded training*), ki smo ga izvedli z orodjem HERest [Young-00].

V našem primeru smo izvajali učenje v desetih iteracijah. Učili pa smo tudi modele, kjer smo v stanjih modelirali porazdelitve vektorjev značilnik s kombinacijo več normalnih porazdelitev. Po vsaki dodani porazdelitvi smo vsakič doučevali model z desetimi ponovitvami postopka učenja. Za razpoznavanje smo uporabljali modele s kombinacijo ene, dveh in treh porazdelitev na stanje.

Vezave parametrov nismo izvajali in tudi potrebe po nevidenih govornih enotah (tako pri monofonih kot difonih) ni bilo zaradi narave govorne zbirke.

Model in postopki pomenijo standarden pristop določevanja statističnega modela razpoznavalnika govora. Ker je bil naš namen medsebojno primerjati pridobljene značilke, tu nismo eksperimentirali z različno paleto možnosti, ki nam jo tako modeliranje ponuja.

#### 4.2.1.3 Kako smo testirali

Testiranje značilnik je potekalo na testnem delu zbirke K211d, ki ga tvorijo štirje govorcji (2ž+2m). Ker smo za učenje in razpoznavanje uporabili različne govorce, lahko govorimo o razpoznavanju, ki je od govorca neodvisno.

Testiranje smo izvajali na različnih modelih z različnimi govornimi enotami. Za govorne enote smo enkrat uporabili monofonske modele, drugič pa difonske. Rezultate smo primerjali za modele s kombinacijo ene, dveh ali treh porazdelitev verjetnosti na stanje po modelu ob nespremenjeni topologiji modelov. Različne so bile le značilke govornega signala. Vedno smo pazili, da smo analizirali med seboj primerljive značilke govornega signala. Se pravi, da jih je bilo (približno) enako in da smo imeli modele z istim številom parametrov. Pri tem smo bolj spremljali primerjavo med posameznimi predstavitevami kot pa sam rezultat razpoznavanja.

### 4.2.2 Razpoznavanje govora z različnimi predstavitevami

V tem delu se bomo posvetili pridobivanju značilnik govornega signala primernih za razpoznavanje govora. Tu bomo predvsem preizkusili različne časovno–frekvenčne predstavitve in na primeru govorne zbirke K211d ugotavljali primernost teh predstavitev in z njimi povezanih postopkov pridobivanja značilnik za razpoznavanje govora.

Zaradi številnih ugodnih lastnosti je spektrogram prevladujoča časovno–frekvenčna



predstavitev za analizo in nadaljnjo obdelavo govornega signala, ki ga štejemo med kvazi stacionarne, torej v bistvu nestacionarne signale s krajšimi odseki stacionarnosti. Ravno to pa je glavni argument uporabe spektrograma, katerega delovanje je nadvse podobno - prikazuje namreč "lokalni" spekter signala. Zatorej, če je signal lokalno stacionaren, je takšna predstavitev smiselna. Poleg tega spektrogram zelo malo pokvari časovno-frekvenčno sliko z motečimi interferenčnimi členi. Kljub temu pa ima kar nekaj pomanjkljivosti. Tu mislimo predvsem na princip nedoločenosti, ki povzroči omejeno natančnost bodisi v eni ali drugi smeri. Postavlja se tudi vprašanje izbire okna. Pri obdelavi govora se je nekako uveljavilo okno dolžine od 20 ms do 30 ms [Picone-93]. To pa pomeni, da vedno enako opisujemo signal ne glede na njegove lastnosti.

Zaradi tega se nam je zdelo smiselno poskusiti uporabo drugih časovno-frekvenčnih predstavitev pri razpoznavanju govora.

Potrebno je izpostaviti še en problem, zakaj ostale časovno-frekvenčne predstavitve niso tako razširjene pri analizi signalov. To je problem diskretizacije predstavitev in večja časovna zahtevnost algoritmov izvedbe teh predstavitev. O diskretizaciji smo že govorili v poglavju 3. Bistveno pa je, da diskretne verzije ne ohranjajo vseh lastnosti zveznih predstavitev. To pa pomeni še dodatne težave teh predstavitev. En pristop k dokazovanju in izvedbi diskretnih različic je neposredna izpeljava diskretnih predstavitev s pomočjo matrik in karakterističnih funkcij, ki smo ga opisali v poglavju 3.

Druga težava je v izvedbi teh predstavitev. Izkaže se namreč, da če bi računali predstavitve za vsako točko signala po času in frekvenci, je s stališča predstavitve to mogoče in nadvse preprosto izvedljivo, toda zaradi ogromne količine podatkov, ki bi jih morali izračunati, je to s trenutnimi sistemi računalnikov časovno in prostorsko neizvedljivo. Tako bi namreč morali za primer 250 ms realnega signala, vzorčenega s 16 kHz, obdelati približno 8 milijonov točk ( $N(N-1)/2$ ,  $N$  dolžina vzorčenega odseka signala), kar je dejansko neizvedljivo in tudi za razpoznavanje povsem nesmiselno. V nadaljevanju si bomo ogledali, kako smo "popravljali" predstavitve in kako smo izvedli izločanje značilk, da bi jih lahko uporabili za razpoznavanje govornih enot slovenskega govora.

Pri določanju značilk smo upoštevali nekaj smiselnih predpostavk in metod, ki so že uveljavljene pri pridobivanju značilk za razpoznavanje. Predstavili bomo te postopke, ki smo jih pri različnih časovno-frekvenčnih predstavitvah vsakič vključili v proces določitve značilk iz govornega signala.

#### 4.2.2.1 Skupne značilnosti postopka pridobivanja značilk

Predpostavimo, da imamo vzorčeni govorni signal  $s[n] = s[nT]$ , kjer je  $T = \frac{1}{f_s}$ ,  $f_s$  je frekvenca vzorčenja. V našem primeru je bila  $f_s = 16$  kHz.

Signal najprej *filtriramo* z visokoprepustnim filtrom (FIR filter) s prenosno funkcijo  $H_{pre} = 1 + a_{pre}z^{-1}$ , s katerim poudarimo spekter govornega signala nad 1 kHz. Razlogi za izvedbo filtriranja so v fizioloških lastnostih govoril in tudi slušnega organa človeka [Markel-80, Zwicker-90]. Filtriranje lahko v primeru diskretnega signala  $s[n]$

zapišemo z diferenčno enačbo prvega reda:

$$s'[n] = s[n] + a_{pre}s[n-1]. \quad (4.7)$$

$a_{pre}$  izbiramo iz intervala  $[-1, -0.4]$ . V našem primeru smo izbrali  $a_{pre} = -0.97$ . Dodatno lahko izvedemo še *ničenje srednjih vrednosti* signala (odštevanje povprečja od signala), vendar v primeru posnetkov iz naše baze to ni bilo potrebno.

Nadalje označimo diskretno časovno–frekvenčno predstavitev s  $TFR[n, k]$ , kjer označujemo z  $(n, k)$  točke po času in frekvenci v časovno–frekvenčni ravnini. Zaenkrat bomo pustili odprto, kako računamo takšno predstavitev in na kako dolgih odsekih govornega signala. Predpostavimo pa, da imamo signal dolg  $N$  točk.

V naslednjem koraku smo izvedli *preslikavo frekvenčne skale*. Tu upoštevamo lastnosti človeškega slušnega organa, in sicer neenakomerno, približno logaritemsko frekvenčno občutljivost [Zwicker-90, str.94]. Obstaja nekaj približkov takšnih preslikav, s katerimi se približamo frekvenčni občutljivosti človeškega slušnega organa. Prva takšna je t.i. frekvenčna *bark lestvica*, ki jo dobimo s preslikavo [Picone-93, str.10]

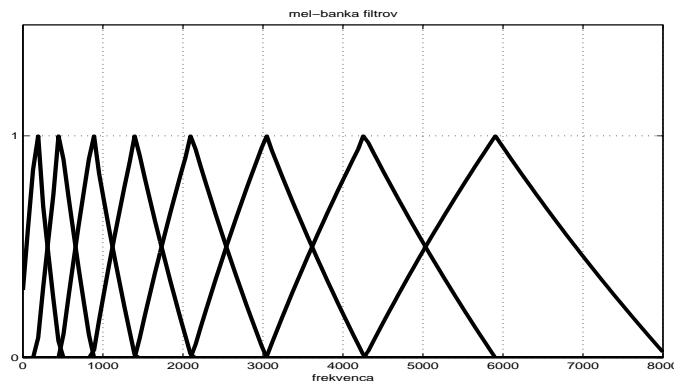
$$bark = 13 \arctan\left(\frac{0.76f}{1000}\right) + 3.5 \arctan\left(\frac{f^2}{7500^2}\right), \quad (4.8)$$

kjer je  $f$  dejanska frekvenca. Naslednji približek slušnega zaznavanja je *melodična lestvica*, podana s preslikavo [Picone-93, str.10]

$$mel = 2595 \log_{10}(1 + f/700), \quad (4.9)$$

ki jo pogosto modeliramo tudi tako, da vzamemo linearno skalo od 0 do 1000 Hz in logaritemsko nad 1000 Hz.

Neenakomerno frekvenčno ločljivost spektralnih meritev dosežemo z uporabo vrste filtrov razporejenih s srednjimi frekvencami frekvenčnih odzivov po lestvicah opisanih s (4.8) ali (4.9). Širine frekvenčnih pasov filtrov niso enake, ponavadi jih določimo s pomočjo bark lestvice (4.8), [Picone-93, str.10]. Primer osmih trikotnih filtrov razporejenih po melodični skali, ki smo jih uporabljali v naših primerjalnih analizah, je na sliki 4.10.



Slika 4.10: Razporeditev trikotnih filtrov po melodični skali.

Ponavadi so filtri razporejeni po celotni frekvenčni skali od 0 do polovične Nyquistove frekvence, lahko pa jih tudi omejimo na manjše območje, tako da definiramo spodnjo,

$f_L$ , in zgornjo,  $f_H$ , frekvenčno mejo. V našem primeru smo izbrali  $f_L = 50$  Hz,  $f_H = 7500$  Hz. Tu nismo eksperimentirali z različnimi izbirami razporeditve filtrov, čeprav se je v primeru slovenskega govora celo bolje izkazala linearna razporeditev filtrov pri od govorca odvisnem razpoznavanju govora [Mihelič-91, str.70–71]. Ker smo imeli od govorca neodvisno razpoznavanje, smo ostali pri melodični razporeditvi.

Filtriranje z mel-banko filtrov (slika 4.10) si lahko razlagamo tudi kot uteženo povprečenje energije (po frekvenci) opisane s  $TFR[n, k]$  znotraj območja delovanja posameznega filtra. Ta povprečja nadalje še logaritmiramo in jih označimo z  $m[q, n]$ . V matematičnem jeziku lahko ta postopek opišemo z naslednjimi koraki. Zapišimo mel-banko filtrov z matriko  $M = M[q, k]_{q=1, k=1}^{Q, K}$ , kjer je  $K$  število frekvenčnih meritev (točk po frekvenci),  $Q$  pa število filtrov. Vsaka vrstica matrike  $M$  predstavlja en trikotni filter. Tudi časovno–frekvenčno predstavitev  $TFR[n, k]$  lahko razumemo kot matriko. Zapišimo  $D = TFR^T$ . V matriki  $D$  opisuje  $n$ -ti stolpec porazdelitev energije po frekvenci v času (točki)  $n$ . Tu upoštevamo, da imamo  $N$  časovnih in  $K$  frekvenčnih točk v diskretno podani časovno–frekvenčni ravnini. Operacijo filtriranja lahko razumemo kot matrično množenje in če izvedemo še logaritmiranje, kjer razumemo operacijo  $\log(\cdot)$  kot logaritmiranje posameznih elementov matrike, dobimo

$$M_{TFR} = \log(M \cdot D). \quad (4.10)$$

Elementi  $M_{TFR} \in \mathbb{R}^{Q \times N}$  so povprečki  $m[q, n]$ ,  $q = 1, \dots, Q$ ,  $n = 1, \dots, N$ . Potihem smo že privzeli, da imamo opravka z realnimi časovno–frekvenčnimi predstavitvami. Predstavitev  $M_{TFR}[q, n]$  dejansko predstavlja logaritme energije frekvenčnih pasov signala  $s[n]$  v točkah  $n$  opisane z različnimi časovno–frekvenčnimi predstavitvami.

Nadaljevanje nas privede do koeficientov melodičnega kepstra (MFCC) pridobljenih iz različnih časovno–frekvenčnih predstavitev. Dobimo jih, če izvedemo *diskretno kosinusno transformacijo* na členih  $m[q, n]$ :

$$c[i, n] = \sqrt{\frac{2}{Q}} \sum_{q=1}^Q m[q, n] \cos\left(\frac{\pi i}{Q}(q - 0.5)\right). \quad (4.11)$$

Vpeljavo transformacije (4.11) lahko razložimo kot aproksimacijo logaritmov energije posameznih frekvenčnih pasov v prostoru, ki ga razpenjajo kosinusne temeljne funkcije, ki so zelo blizu lastnim funkcijam, pridobljenim iz Karhunen–Loevejeve transformacije [Mihelič-91, Malayath-00]. Drugi razlog pa je v tem, da s transformacijo (4.11) dekoleriramo točke  $M_{TFR}$ , kar je nujno potrebno, če pridobljeno informacijo uporabljamo za razpoznavanje s PM modeli modeliranimi z diagonalnimi kovariančnimi matrikami.

Tu je treba izpostaviti dejstvo, da ni smiselno izvajati teh transformacij za vsako točko pa času  $n$ . Zaradi tega pogosto nadomestimo  $m[q, n]$  s krajšimi odseki po času, se pravi v nekem smislu gladimo po času z različnimi prijemi, ki jih bomo predstavili v nadaljevanju, z namenom redukcije podatkov.

Če sledimo standardnemu pristopu določanja značilk (trenutno najbolj uspešnih) pri razpoznavanju govora, moramo vpeljati še odvode koeficientov kepstra. Aproksimacij

za odvode je več. Omenimo samo eno, in sicer:

$$\dot{c}[i, n] = \sum_{j=-N_d}^{N_d} j c[i, n + j], \quad (4.12)$$

kjer je  $2N_d + 1$  dolžina odseka po času s središčem v točki  $n$ , kjer aproksimiramo odvod  $\dot{c}[i, n]$ . Izpustili smo normalizacijski faktor, ki je vedno enak, če gledamo vedno enako dolge odseke. Aproksimacija (4.12) predstavlja iskanje koeficienta regresijske premice nad podatki  $c[i, n]$ . Pri višjih odvodih je postopek aproksimacije analogen, le da za podatke vzamemo aproksimacije prvih odvodov. Jasno je, da postopek lahko ponovimo za odvode višjih redov, vendar so se pri razpoznavanju govora uveljavili prvi in drugi odvodi [Furui-86].

Ob tem moramo poudariti še pomembno lastnost odvajanja. Ta operacija namreč poudarja šum oziroma napake meritev. Zato je smiselno računati odvode že na dovolj zglajenih predstavitev signala, torej na takšnih podatkih, kjer smo že dovolj odstranili napake meritev. V standardnih postopkih se izvaja odvajanje v zadnji fazi, torej šele na nivoju koeficientov melodičnega kepstra. Lahko pa bi ga izvajali že pred izvedbo kosinusne transformacije (zaradi linearnosti transformacije bi bilo vseeno). V našem primeru smo izvajali tudi takšne posege v določanje značilk, pri tem pa smo predhodno dodatno gladili območja. Uporabljali smo tudi drugačne ocene odvodov.

Kot dodatno značilko za razpoznavanje smo določali še kratkočasovno energijo signala. V bistvu smo ocenjevali logaritem energije signala kot

$$E_n = \log \sum_{j=-N_e}^{N_e} (s'[n + j])^2. \quad (4.13)$$

Pri tem predstavlja  $E_n$  logaritem energije odseka signala dolžine  $2N_e + 1$  v času  $n$ .

Če povzamemo, smo se pri konstrukciji značilk v veliki meri naslanjali na postopek pridobivanja značilk kot koeficientov melodičnega kepstra govornega signala in pridruženih odvodov ter energije. Pri tem smo eksperimentirali z različnimi časovno-frekvenčnimi predstavitvami, z različnimi postopki redukcije informacije znotraj predstavitev in različnimi izbirami odsekov predstavitev, na podlagi katerih smo potem določali koeficiente kepstra. Dodatno smo še ocenjevali odvode pred izvedbo kosinusne transformacije in to primerjali s standardnim pristopom.

V nadaljevanju si bomo najprej ogledali klasično uporabo spektrograma pri določanju značilk. Tu bomo podrobneje opisali postopek izpeljave značilk, na katerega se bomo sklicevali pri izvedbi drugih predstavitev. Primerjali bomo rezultate razpoznavanja na govorni zbirki K211d pri različnih izvedbah Wigener–Villejeve predstavitev, psevdo Wigener–Villejeve, psevdo Margenau–Hillove in Zhao–Atlas–Marksove porazdelitve.

Postopke izračunov predstavitev smo izvedli z uporabo Matlabovega orodja za časovno-frekvenčne predstavitev [Auger-97] in kombinacijo programov, pisanih v programskem jeziku C, v okolju GNU-Octave [Octave], učenje PM modelov pa z uporabo orodja HTK [Young-00].

### 4.2.2.2 Spektrogram

Pri razpoznavanju govora se kot časovno–frekvenčno predstavitev skoraj izključno uporablja spektrogram. Zato smo jo v našem primeru vzeli za referenčno v primerjavi z ostalimi predstavitvami.

Upoštevali smo standarden pristop izračunavanja značilk. Frekvenca vzorčenja,  $f_s$ , je bila v našem primeru 16 kHz. Opišimo postopek v naslednjih korakih:

1. *izvajanje visokoprepustnega filtriranja*

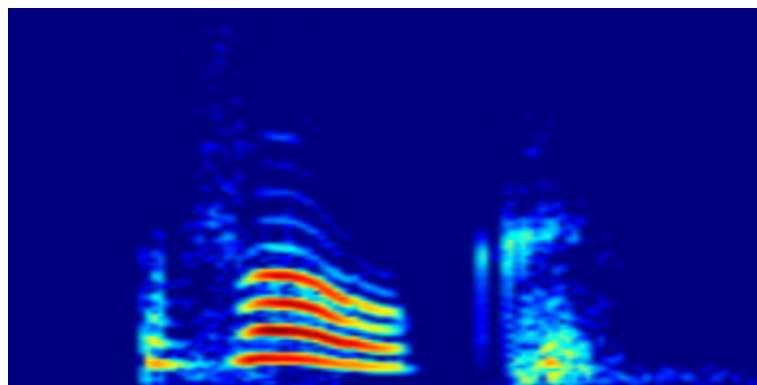
$$s'[n] = s[n] + a_{pre}s[n - 1].$$

V našem primeru je bil  $a_{pre} = -0.97$ .

2. *izračun banke filtrov* razporejenih po melodični skali, ki jih lahko zapišemo v matriki  $M = [mf[q, n]]_{q=1, n=1}^{Q, N}$ . Tu smo izbrali  $Q = 32$  filtrov razporejenih po melodični skali (4.9) med frekvenca  $f_L = 50$  Hz in  $f_H = 7500$  Hz.
3. *izračun časovno–frekvenčne predstavitve* spektrograma  $TFR[n, k] = SPEC[n, k]$ . Za analizo smo uporabili Hammingovo okno dolžine 25 ms,  $w_h = 25$  ms, frekvenčno “sliko” pa smo računali na vsakih 10 ms, se pravi, da smo uporabljali preskok med okvirji dolžine  $t_f = 10$  ms, kar pomeni 100 slik na sekundo ali  $f_r = 100$  Hz. Izbira preskoka skupaj z dolžino in tipom okna predstavlja običajen pristop pri analizi in obdelavi govornega signala.

Tako smo dejansko obravnavali signal po izsekih dolžine  $N_w = 400$  točk.  $SPEC[n, k]$  se tako prevede v  $D = SPEC[n(i), k(j)]$ , kjer je  $n(i) = i \cdot (t_f f_s)$ ,  $i = 0, 1, \dots, [N/(t_f f_s)]$  ( $[N/(t_f f_s)]$  je celi del) in  $k(j) = j$ ,  $j = 0, 1, \dots, N_w/2$ . Okno je bilo vedno simetrično glede na  $n(i)$ .

To je malce drugače zapisana izvedba računanja spektrograma na podlagi kratkočasovne Fourierjeve transformacije na odsekih dolžine  $N_w$  signala  $s[n]$  dolgega  $N$  točk. Uvedba takšne interpretacije je potrebna, da ostanemo konsistentni s prejšnjim razdelkom in v nadaljevanju z drugimi časovno–frekvenčnimi predstavitvami.



Slika 4.11: Spektrogram besede /vOUk/.

Spektrogram govornega signala besede /vOUk/ ženske govorke je prikazan na sliki 4.11.

## 4. izvedba filtriranja z mel-banko filtrov

$$M_{SPEC} = \log(M \cdot D).$$

Elementi matrike  $M_{SPEC}$  so  $m[q, n(i)]$ ,  $q = 1, \dots, Q$ . Operacija  $\log(\cdot)$  se izvaja po elementih matrike.

## 5. izračun koeficientov melodičnega kepstra (4.11):

$$c[l, n(i)] = \sqrt{\frac{2}{Q}} \sum_{q=1}^Q m[q, n(i)] \cos\left(\frac{\pi l}{Q}(q - 0.5)\right), \quad l = 1, \dots, N_c.$$

V našem primeru je bil  $N_c = 12$ . Poudariti moramo, da nismo upoštevali ničtega koeficienta kosinusne transformacije.

## 6. izračun prvih in drugih odvodov z oceno koeficientov regresijskih premic

$$d[l, n(i)] = \frac{\sum_{\tau=1}^{\Theta} (c[l, n(i) + \tau] - c[l, n(i) - \tau])}{2 \sum_{\tau=1}^{\Theta} \tau^2}. \quad (4.14)$$

Za druge odvode  $a[l, n(i)]$  smo izvedli isto formulo (4.14) na  $d[l, n(i)]$ . Pri prvih in drugih odvodih smo upoštevali oceno pri  $\Theta = 2$ .

## 7. oceno kratkočasovne energije določimo iz (4.13)

$$E_n = \log \sum_{m=-N_w/2}^{N_w/2} (s'[n + m])^2.$$

Iz zveze (4.14) izračunamo še prve in druge odvode energije.

V našem primeru smo tako na vsakih 10 ms določili 12 statičnih značilk - koeficientov melodičnega kepstra, ki skupaj z energijo ter prvimi in drugimi odvodi tvorijo vektor 39 akustičnih značilk.

Učenje modelov iz učnega dela zbirke K211d je potekalo po že opisanem postopku (poglavje 4.2.1.2).

Rezultati razpoznavanja alofonov (tabela 4.1) na testnem delu zbirke, ki smo jih modelirali z monofonskimi in difonskimi govornimi enotami, so prikazani v tabeli 4.2.

V tabeli 4.2 lahko primerjamo rezultate razpoznavanja na testni bazi K211d med monofoni in difoni. Pri monofonskih modelih naraščajo rezultati razpoznavanja s povečevanjem števila porazdelitev na model, hkrati pa padajo rezultati pri difonih. To si lahko razlagamo s premajhno učno bazo, saj se z večanjem kombinacij gostot verjetnosti na stanje povečuje število parametrov modela. Zato smo v nadaljevanju predstavitev opuščali izvedbo razpoznavanja z difonskimi enotami s kombinacijo več funkcij gostot na stanje in upoštevali samo rezultate z eno.

Primerjavo z ostalimi predstavitevami bomo podali v nadaljevanju.

---

<sup>1</sup>Natančnost razpoznavanja (ang. accuracy) definiramo kot  $A = \frac{N - (N_z + N_v + N_i)}{N}$ , kjer je  $N$  število razpoznavanih enot,  $N_z$ ,  $N_v$ ,  $N_i$  pa zaporedoma število napak zamenjave, vrivanj in izbrisov govornih enot. Pravilnost razpoznavanja (ang. correctness) definiramo kot  $C = \frac{N - (N_z + N_i)}{N}$ , torej brez števila vrivanj ( $N_v$ ).

Tabela 4.2: Rezultati razpoznavanja z MFCC značilkami z uporabo PMM s 3 stanji na model.

		1 porazdelitev	2 porazdelitvi	3 porazdelitve
monofoni	pravilnost	75.09%	78.18%	80.59%
	natančnost <sup>1</sup>	72.58%	75.83%	78.27%
difoni	pravilnost	75.61%	73.29%	70.93%
	natančnost	73.29%	71.23%	68.68%

#### 4.2.2.3 Wigener–Villejeva predstavitev

Pri Wigener–Villejevi predstavitvi (WVD) bomo opisali postopek pridobivanja značilk in probleme, ki pri tem nastopijo. Podobno smo težave odpravljali tudi pri drugih predstavitev v nadaljevanju.

Zaradi načina konstrukcije WVD ne moremo tako očitno razkosati signal na posamezne odseke in potem izvajati iste operacije kot pri spektrogramu. Pojavljajo se tudi moteči interferenčni členi z negativnimi vrednostmi, kar nas prisili v drugačno načrtovanje izvedbe pridobivanja značilk.

Pojdimo po korakih postopka, ki smo ga uporabili pri spektrogramu. Prva dva koraka postopka sta enaka. Tudi tu smo izbrali  $Q = 32$  filtrov razporejenih po melodični skali med 50 in 7500 Hz.

Tretji korak pa je problematičen. Povsem nesmiselno in računsko prezahtevno bi bilo računati  $TFR[n, k] = WVD_s[n, k]$  za vsako točko  $n$  signala  $s[n]$ . Poleg tega se pojavi še problem podvzorčenja pri uporabi diskretne WVD. Tega problema se rešimo tako, da uporabimo pri izračunu z WVD analitično verzijo signala, ki ga dobimo z uporabo Hilbertove transformacije (poglavje 2.1.3.1). Torej

$$s_a[n] = \text{hilbert}(s'[n]), \quad (4.15)$$

kjer je preslikava  $\text{hilbert}(s[n]) = s[n] + jHT(s[n])$ .

Da se znebimo negativnih delov predstavitve uporabimo s pragom  $p$  spremenjene predstavitve (4.1), torej  $WVD_s^{(p)}[n, k]$ . Prag smo poljubno spreminjali.

Problem zmanjševanja števila podatkov predstavitve smo reševali na več načinov. Prvi način, predstavitev imenujmo WVD1, je analogen kot pri spektrogramu. Samo na nekaj enakomerno oddaljenih mestih predstavitve odčitamo spektralno meritev. Preskok definirajmo s  $t_f$ , ki je v našem primeru znašal  $t_f = 10$  ms. Tako smo upoštevali samo točke predstavitve v razmaku 10 ms ali  $N_f = 160$  točk ( $N_f = t_f f_s$ ). Formalno to zapišemo kot  $WVD1_{s_a}^{(p)}[n(i), k] = WVD_{s_a}^{(p)}[n(i), k]$ , kjer je  $n(i) = i \cdot N_f$ ,  $i = 0, 1, \dots, [N/N_f]$  in  $k = 0, 1, \dots, K - 1$ ,  $K$  je število odtipkov po frekvenci (v našem primeru je  $K = 256$ ).

Pri tem moramo poudariti, da imamo opravka s kompleksnim signalom.

Kljub očitni podobnosti s spektrogramom, imamo tu veliko razliko. Pri spektrogramu gre namreč za princip okna in preskoka, kjer je preskok manjši od dolžine okna, s tem pa dosežemo sorazmerno gladko prehajanje iz enega preskoka na drugi. Tu pa lahko govorimo samo o preskoku brez prekrivanja. Zato se pojavijo nezveznosti, ki so posledica občutljivosti predstavitve in interferenčnih členov. Očitno je namreč, da na odčitane vrednosti pri preskokih vplivajo “nevidne” vrednosti med preskoki.

Da bi odpravili to pomanjkljivost, smo se odločili posnemati spektrogram v celoti. Uvedli smo namreč postopek prekrivanja, da bi s tem omogočili gladke prehode med odčitanimi vrednostmi.

To izvedemo na naslednji način. Ne izračunamo predstavitve celotnega signala  $s_a[n]$ , ampak samo nekega krajšega odseka, torej:

$$x_{N_f, N_w}[n] = s_a[n] \Big|_{[iN_f, iN_f + N_w]}, \quad (4.16)$$

kjer je  $N_f$  preskok in  $N_w$  dolžina okvirja, s katerim izrežemo del signala. To ni nič drugega kot princip okna in preskoka, ki ga dobro poznamo pri kratkočasovni Fourierjevi transformaciji in posredno tudi pri spektrogramu. Seveda privzamemo, da je  $N_f < N_w$ . V primeru WVD smo imeli okvirje dolžine 32 ms,  $N_w = 512$ , s preskoki  $N_f = 160$  točk ali 10 ms. Na signalu  $x_{N_f, N_w}[n]$  izračunamo WVD, nato pa še izvedemo povprečenje vzdolž časovne osi. S tem izgubljammo časovno informacijo, kar je konec koncev tudi naš namen redukcije podatkov. Še vedno pa v primeru WVD ohranjamo “idealno” frekvenčno ločljivost. Povprečenje je lahko tudi uteženo, kar lahko razumemo kot (grobno) glajenje po času z ustreznim oknom. Če to še formalno zapišemo, dobimo

$$WVDf^{(p)}[n', k] = WV D_{x_{N_f, N_w}}^{(p)}[n', k], \quad n' = iN_f, iN_f + 1, \dots, iN_f + N_w$$

in potem izvedemo povprečenje (glajenje) po času, tako da dobimo vektor za trenutek  $n(i) = iN_f$ ,  $i = 0, 1, \dots, [N/N_f]$ :

$$wv_{n(i)}[k] = \sum_{n'=iN_f}^{iN_f+N_w} h[n'] WV Df^{(p)}[n', k],$$

kjer je  $h[n']$  normirana utežna funkcija. Primer take funkcije je  $h[n'] = \frac{1}{N_w}$ , iz katere dobimo navadno povprečje, lahko pa izberemo tudi kakšno znano normirano okno. V našem primeru smo uporabili še Hammingovo okno.

Skupno predstavitev dobimo z združevanjem vektorjev v stolpce matrike predstavitve

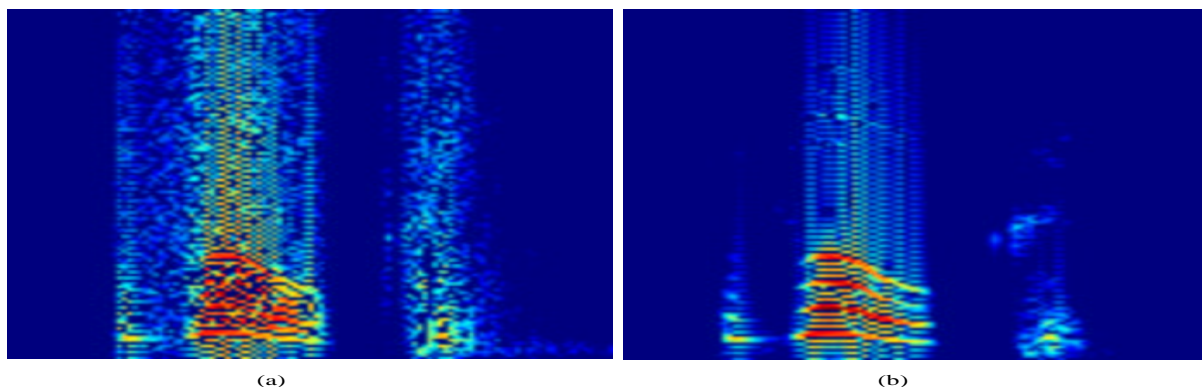
$$MWVD[n(i), k] = wv_{n(i)}[k].$$

Predstavitev  $MWVD[n(i), k]$ , kjer smo uporabili za glajenje navadno povprečenje, poimenujemo  $WVD2[n(i), k]$ , uporaba Hammingovega okna kot utežnega faktorja pa nam dá porazdelitev  $WVD3[n(i), k]$ .

Razlika med WVD1 brez prekrivanja in WVD3 s prekrivanjem je lepo prikazana na sliki 4.12 daljšega govornega signala besede /vOUk/ ženske govorce.

Slika 4.12(a) z WVD1 je zelo šumna. Pojav interference je izrazit, tudi razpoznavanje pričakujemo zato slabše. Slika z WVD3 pa je precej boljša. Interferenčni členi so





Slika 4.12: WVD predstavitvi besede /vOUk/: (a) WVD1, (b) WVD3 z uporabo Hammingovega okna pri  $N_w = 512$  in  $N_f = 160$ .

razporejeni samo med spektralnimi komponentami. Gre namreč za grobo glajenje po času. V primerjavi s sliko spektrograma 4.11 je slika 4.12(b) z WVD3 bolj natančna, vendar ima interferenčne člene. Opazimo lahko tudi že omenjen efekt razmazane slike v zadnjem delu pri spektrogramu (pri pridihu alofona /k/), pri WVD pa ne. To je zato, ker spektrogram ne ustreza lastnosti robnih pogojev in se zato hipna sprememba razmaže po sliki [Loughlin-93], WVD pa s tem nima težav.

Izvedba korakov 4, 5 in 6 ter izračun energije so enaki kot pri spektrogramu. Tudi tu smo uporabili 39 značilk in nismo eksperimentirali z ocenami odvodov.

Rezultati razpoznavanja z WVD1, WVD2 in WVD3 so prikazani v tabeli 4.3.

V tabeli 4.3 opazimo bistven napredek razpoznavanja ob uporabi značilk z WVD2 in WVD3, kar se je pokazalo tudi pri ostalih predstavitev v nadaljevanju. Neposreden pristop z odčitavanjem na vsakih  $N_f$  korakov pri WVD1 nam dá bistveno slabše rezultate (več kot 10% absolutno), pri difonih pa je razlika še bolj očitna. To je posledica izrazitih točk nezveznosti pri prvi verziji v primerjavi z drugima dvema, kar je lepo vidno na sliki 4.12. V prvem primeru nam namreč interferenčni členi “zameglijo” časovno–frekvenčno sliko in posledično poslabšajo razpoznavanje. Rezultati razpoznavanja z difoni so tudi nekoliko višji v primerjavi z monofoni, kar si lahko razlagamo z natančnejšimi meritvami signala v časovno–frekvenčni ravnini. Ugotovimo lahko tudi, da ni bistvene razlike, če za grobo glajenje uporabljamo pravokotno okno (WVD2) ali pa Hammingovo (WVD3). Rezultati z uporabo Hammingovega okna so nekoliko višji, vendar ne pri difonih. Opozoriti moramo še, da je razpoznavanje zelo občutljivo tudi na izbiro praga  $p$  pri popravljenih predstavitev. Z višanjem praga se razpoznavanje slabša. Podobno se zgodi, če je prag prenizek. V našem primeru in tudi v nadaljevanju smo uporabljali pragove  $p = 0.005\% - 0.05\%$ .

#### 4.2.2.4 Psevdo Wigner–Villejeva predstavitev

V primeru psevdo Wigner–Villejeve predstavitve (PWVD) smo v glavnem ponovili postopke, ki smo jih opisali pri WVD, razlika je le v uporabi dodatnega okna glajenja po frekvenci, kar je posledica definicije in konstrukcije PWVD. Za glajenje po frekvenci

smo uporabili Hammingovo okno.

Vse ostalo je bilo nespremenjeno. Tudi tu smo izvedli verzije pridobivanja značiln, ki smo jih poimenovali PWVD1, PWVD2 in PWVD3. PWVD1 je analogna WVD1, torej gre za izbiro vektorjev (po frekvenci) na vsakih 10 ms iz slike PWV predstavitve celotnega signala definirani z oknom glajenja po frekvenci dolžine 256 točk, pri čemer je bil  $K = 256$ .

V drugem primeru PWVD2 in PWVD3 smo uporabili princip grobega glajenja po času, se pravi računanja predstavitev izsekov signala s prekrivanjem. Okvir in korak sta bila enaka kot pri WVD, torej  $N_w = 512$  in  $N_f = 160$ . Osnovno okno finega glajenja po frekvenci je bilo isto kot prej, torej Hammingovo dolžine  $N_w/4 = 128$  točk, za "grob" glajenje po času pa smo uporabili v primeru PWVD2 pravokotno okno, pri PWVD3 pa Hanningovo okno. Tako lahko razumemo PWVD2 in PWVD3 kot zglajeno psevdo WV predstavitev (SPWVD).

Rezultati razpoznavanja za vse tri verzije so podani v tabeli 4.4.

Podobno kot pri WVD je tudi pri PWVD. Če primerjamo rezultate razpoznavanja pri WVD (tabela 4.3) s PWVD (tabela 4.4), lahko ugotovimo, da imamo skoraj identične rezultate pri posameznih verzijah predstavitev. Opazimo lahko izboljšavo razpoznavanja difonov s PWVD1 v primerjavi z WVD1. Tudi tu lahko ugotovimo bistven preskok razpoznavanja med PWVD1 ter verzijama PWVD2 in PWVD3. Občutna razlika v razpoznavanju se tu pokaže tudi med verzijama PWVD2 in PWVD3. Rezultati razpoznavanja so pri verziji PWVD3 (grobno glajenje s Hanningovim oknom) približno za

Tabela 4.3: Rezultati razpoznavanja z WV predstavitvijo (3 stanja na model).

		1 porazdelitev	2 porazdelitvi	3 porazdelitve
<b>monofoni</b>	pravilnost	61.70 %	65.42 %	66.55 %
	WVD1	natančnost	56.25 %	62.51 %
<b>difoni</b>	pravilnost	43.50 %	-	-
	WVD1	natančnost	34.22 %	-
<b>monofoni</b>	pravilnost	71.99 %	76.64 %	78.48 %
	WVD2	natančnost	67.51 %	72.90 %
<b>difoni</b>	pravilnost	73.98 %	-	-
	WVD2	natančnost	69.89 %	-
<b>monofoni</b>	pravilnost	71.88 %	76.04 %	77.79 %
	WVD3	natančnost	68.09 %	73.16 %
<b>difoni</b>	pravilnost	73.71 %	-	-
	WVD3	natančnost	69.59 %	-

Tabela 4.4: Rezultati razpoznavanja s PWV predstavitvijo (3 stanja na model).

		1 porazdelitev	2 porazdelitvi	3 porazdelitve
<b>monofoni</b> PWVD1	pravilnost	61.84 %	65.45 %	67.20 %
	natančnost	55.70 %	62.79 %	65.10 %
<b>difoni</b> PWVD1	pravilnost	56.36 %	-	-
	natančnost	50.11 %	-	-
<b>monofoni</b> PWVD2	pravilnost	71.63 %	76.14 %	78.11 %
	natančnost	67.51 %	72.85 %	74.80 %
<b>difoni</b> PWVD2	pravilnost	73.71 %	-	-
	natančnost	69.88 %	-	-
<b>monofoni</b> PWVD3	pravilnost	73.16 %	77.23 %	78.67 %
	natančnost	70.20 %	74.72 %	76.10 %
<b>difoni</b> PWVD3	pravilnost	75.43 %	-	-
	natančnost	71.90 %	-	-

2% (absolutno) višji. Pri difonih se zgodi podobno kot v primeru WVD. Rezultati so nekoliko višji, če si izberemo za razpoznavanje prehode med glasovi v primerjavi z glasovnimi jedri.

#### 4.2.2.5 Psevdo Margenau–Hillova predstavitev

Postopki so tu povsem identični kot pri PWVD. Izvedli smo dve verziji pridobivanja značilk, PMHD1 in PMHD2.

V primeru PMHD1 računamo značilke na isti način kot pri PWVD1, le predstavitev je druga in Hammingovo okno glajenja po frekvenci smo izbrali dolžine 100 točk pri  $K = 256$ .

V primeru PMHD2 smo uporabili isti postopek kot pri PWVD2, se pravi pravokotno okno grobega glajenja po času in Hammingovo okno finega glajenja po frekvenci.

Slike PMHD1 in PMHD2 so podobne kot v primeru WVD (slika 4.12).

Rezultati so prikazani v tabeli 4.5.

PMH predstavitev je po načinu in ideji konstrukcije povsem različna kot verzije WV predstavitev, čeprav vse spadajo v Cohenov razred. Verzija PMH je izpeljana iz kompleksne Richaczkove predstavitve. Rezultati razpoznavanja (tabela 4.5) so tu kar za nekaj procentov (absolutno) višji kot pri WVD ali PWVD. To velja tako za PMHD1 kot za PMHD2. To si lahko razložimo, da imamo tu drugačno geometrijo interferenc,

Tabela 4.5: Rezultati razpoznavanja s PMH predstavitvijo (3 stanja na model).

		1 porazdelitev	2 porazdelitvi	3 porazdelitve
<b>monofoni</b>	pravilnost	62.97 %	68.70 %	71.43 %
	PMH1	natančnost	58.62 %	64.53 %
<b>difoni</b>	pravilnost	59.84 %	-	-
	PMH1	natančnost	57.06 %	-
<b>monofoni</b>	pravilnost	75.25 %	78.15 %	80.13 %
	PMH2	natančnost	72.62 %	75.77 %
<b>difoni</b>	pravilnost	76.15 %	-	-
	PMH2	natančnost	73.19 %	-

ki tudi niso tako izrazite kot pri WV verzijah časovno–frekvenčnih predstavitev. Pri PMHD2 lahko opazimo tudi višji procent natančnega razpoznavanja pri modeliranju z difoni kot pri monofonih, kar bi bil lahko eden izmed pokazateljev, da je PMHD potencialno dobra predstavitev za modeliranje prehodov med glasovi.

#### 4.2.2.6 Zhao–Atlas–Marksova predstavitev

Zhao–Atlas–Marksova porazdelitev je zaradi številnih dobrih lastnosti primerna za obdelavo govornih signalov. Tej predstavitvi zaradi stožčaste oblike jedra, ki ga uporabljamo pri konstrukciji, pravimo tudi *predstavitev stožčastega jedra* (*ang. cone-shaped kernel distribution*), zato jo bomo krajše označevali s CKD.

S CKD predstavitvijo smo eksperimentirali največ, saj smo z opazovanjem različnih situacij govornega signala v časovno–frekvenčni ravnini ugotovili primernost predstavitve za analizo in v našem primeru tudi za razpoznavanje govora. Diskretno verzijo CKD predstavitve smo izvedli s postopkom opisanim v [Pitton-95]. Pri takšni konstrukciji smo uporabljali navadne signale in ne analitičnih različic.

Prvi poskus smo izvajali podobno kot v primeru WVD1, PWVD1 in PMHD1. Poimenovali smo ga CKD1. Tu gre za odčitavanje vrednosti vzdolž frekvenčne osi na vsakih 10 ms. Pri tem smo tu in tudi v nadaljevanju uporabljali  $K = 256$ . V primeru CKD1 smo eksperimentirali z različnimi preskoki ( $t_f = 10$  ms, 5 ms) in številom mel-filtrov ( $Q = 32, 40$ ). Kakšnih bistvenih izboljšav pa nismo dosegli, zato bomo razumeli pod imenom CKD1 predstavitev s preskokom  $t_f = 10$  ms ( $N_f = 160$ ) in številom filtrov  $Q = 32$ .

Druga verzija CKD z imenom CKD2 je bila izvedena podobno kot pri prejšnjih predstavitev WVD2, PWVD2 in PMHD2. Za povprečenje po času smo uporabljali pravokotno utežno okno dolžine  $N_w = 512$  točk (32 ms) in preskokom  $N_f = 160$  točk ( $t_f = 10$  ms). Vse ostalo je bilo isto kot pri prejšnjih verzijah.

V nadaljevanju smo eksperimentirali še z odvajanjem pred izvedbo kosinusne transformacije. V bistvu smo tako iskali nove možnosti izbire značilke dane časovno–frekvenčne predstavitve. Zaradi linearnosti kosinusne transformacije je za odvajanje z diferencami vseeno ali ga izvajamo pred ali po transformaciji. Boljše pa je odvajanje po transformaciji, ker v bistvu vzamemo samo nekaj prvih koeficientov transformacije in tako na zglajeni predstavitvi računamo odvode. V prvem primeru smo tako izvajali odvajanje neposredno na sliki časovno–frekvenčnega odseka signala, dolžine  $N_w = 1024$  točk (64 ms) s preskokom  $N_f = 160$ . Na odsekih dolžine  $N_w$  smo računali CKD in izvedli filtriranje po melodični skali z banko trikotnih filtrov. Tako smo dobili sliko dimenzije  $N_w \times Q = 1024 \times 32$  točk. Sliko smo razrezali po času na  $I = 5$  enakih odsekov. Znotraj frekvenčnih pasov smo po časovnih odsekih izvedli povprečenje. Tako smo dobili  $I \times Q = 5 \times 32$  povprečij  $mp[i, q]$ . Centralna povprečja ( $mp[3, q]$ ) smo uporabili za oceno statičnih značilke govornega signala, z ostalimi pa smo ocenjevali odvode. Prvi odvod je bil tako

$$d[q] = \frac{mp[4, q] - mp[2, q]}{h_{42}},$$

kjer je  $h_{42}$  dolžina med 4. in 2. odsekom. Za drugi odvod pa smo uporabili druge difference, torej

$$d1[q] = \frac{mp[2, q] - mp[1, q]}{h_{21}}, \quad d2[q] = \frac{mp[5, q] - mp[4, q]}{h_{54}},$$

$$a[q] = \frac{d2[q] - d1[q]}{h},$$

kjer je  $h$  dolžina enega odseka.  $d[q]$  in  $a[q]$  sta predstavljal(a) aproksimaciji prvega in drugega odvoda s simetričnimi diferencami. Nato smo izvedli kosinusno transformacijo (korak 5) ločeno na  $mp[3, q]$ ,  $d[q]$  in  $a[q]$ . Energijo smo računali v okolici odseka 3. Tako smo dobili 39 značilke. To verzijo predstavitve smo poimenovali CKD3.

Drugi način modeliranja prehodov posameznih spektralnih komponent na sliki časovno–frekvenčne predstavitve CKD smo izvedli z verzijama CKD4 in CKD5. Tu smo ravno tako kot prej računali CKD in nato še filtrirali na odseku  $N_w = 1024$  s preskokom  $N_f = 160$ . Nato pa smo sliko dimenzije  $N_w \times Q$  razsekali po času na  $I = 17$  enakih delov. Podobno kot prej smo izvedli povprečenje in dobili  $mp[i, q]$ ,  $i = 1, \dots, I$ ,  $q = 1, \dots, Q$ .

V primeru CKD4 smo podatke  $mp[i, q]$  za vsak  $q$  aproksimirali po metodi najmanjših kvadratov s funkcijo

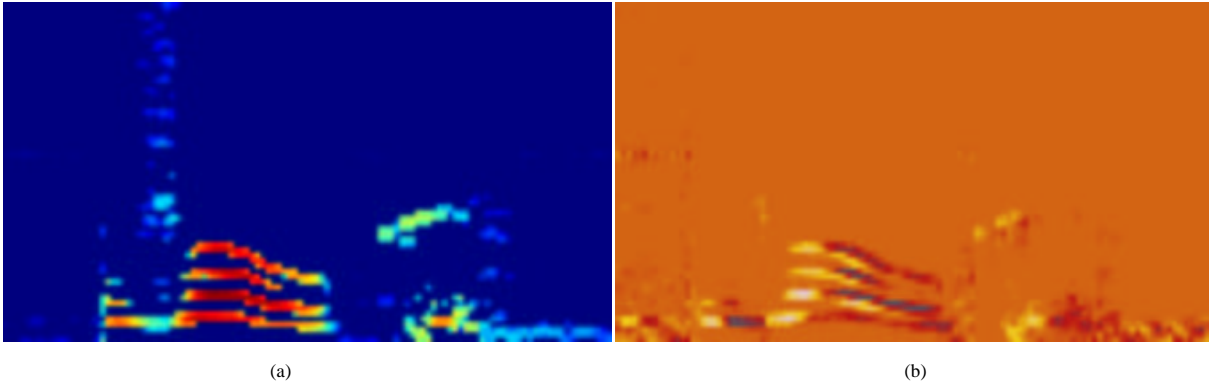
$$f_q(x) = a_1^{(q)} \arctan(x) + a_0^{(q)}. \quad (4.17)$$

Iskali smo torej takšna  $a_0^{(q)}$  in  $a_1^{(q)}$ , da bo vsota

$$\sum_{i=1}^I |mp[i, q] - f_q(x_i)|^2 \quad (4.18)$$

minimalna. Pri tem smo izbrali  $x_i$  neenakomerno iz intervala  $[-2.5, 2.5]$ . S tem smo poskušali smiselno zajeti prehode v predstavitvi. Lahko bi izvedli tudi aproksimacijo s funkcijo  $f_q(x) = a_1 \arctan(a_2 x) + a_0$ , vendar je parameter  $a_2$  zelo občutljiv na nekaj osrednjih vrednosti  $mp[i, q]$ , pa tudi reševanje (4.18) bi se spremenilo v reševanje nelinearnega sistema. Parameter  $a_0^{(q)}$  si lahko razlagamo kot povprečenje izseka,  $a_1^{(q)}$  pa

kot približek za prvi odvod, saj je  $a_1^{(q)} \approx \frac{m_D[q] - m_L[q]}{\pi}$ , kjer sta  $m_D[q]$  povprečje desne polovice slike,  $m_L[q]$  pa leve v frekvenčnem pasu  $q$ .



Slika 4.13: Predstavitve besede /vOUk/: (a) porazdelitev energije s CKD, (b) slika prvega odvoda predstavitve CKD.

Na sliki 4.13 lahko opazujemo potek energije govornega signala besede /vOUk/ ženske govorka (slika 4.13(a)) in sliko 4.13(b) prvega odvoda izračunanega na podlagi CKD4 (parameter  $a_1^{(q)}$ ). Svetlejši deli (rumene barve) so deli naraščanja energije v signalu, temno rdeči in sivi deli pa deli padanja. Oranžna barva pomeni nespremenjeno energijo v časovno–frekvenčni ravnini. Tu je potrebno omeniti pojav ojačevanja šuma pri operaciji odvajanja. V začetnem in končnem delu na sliki 4.13(b) odvoda lahko opazimo razmazane dele, pa čeprav imamo tu območja tišine začetnega in končnega premora. Ravno to je težava takšnega pristopa modeliranja odvodov pred dokončnim glajenjem s kosinusno transformacijo.

Na koncu smo izvedli še kosinusno transformacijo ločeno na  $a_0^{(q)}$  in  $a_1^{(q)}$ ,  $q = 1, \dots, Q$ . Tako smo za vsak odsek dobili 24 značilik, ki smo jim dodali še logaritem energije.

Pri drugi izvedbi, CKD5, pa smo se zgledovali po principu določanja značilik iz večločljivostnega spektra [Dobrišek-97, Dobrišek-01]. Tu smo  $mp[i, q]$  aproksimirali po metodi najmanjših kvadratov (4.18) s funkcijo

$$g_q(x) = a_0^{(q)} + a_1^{(q)} \cos\left(\frac{x\pi}{I-1}\right) + a_2^{(q)} \sin\left(\frac{x\pi}{I-1}\right). \quad (4.19)$$

Pri tem smo izbrali  $x_i = 0, 1, \dots, I-1$ . Interpretacija koeficientov  $a_i^{(q)}$  je podobna kot v prejšnjem primeru.

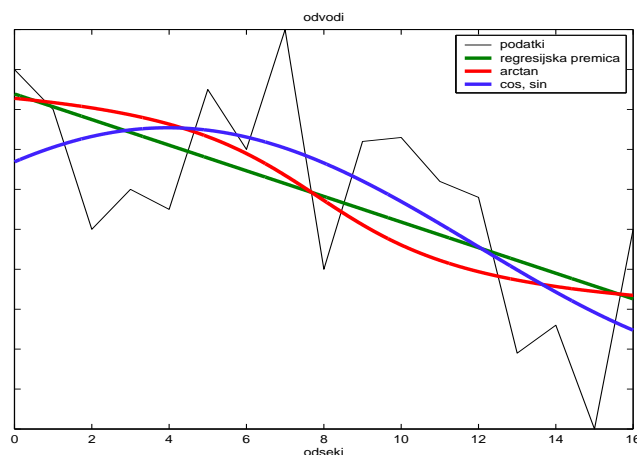
Na sliki 4.14 si lahko pogledamo aproksimacije podatkov  $mp[i, q]$  z  $f_q(x)$ ,  $g_q(x)$  in regresijsko premico.

Podobno kot v primeru CKD4 smo tudi tu izvedli kosinusno transformacijo (korak 5) ločeno na  $a_0^{(q)}$ ,  $a_1^{(q)}$  in  $a_2^{(q)}$   $q = 1, \dots, Q$ . Tako smo dobili 36 značilik za vsak odsek, ki smo mu pridružili še logaritem energije.

Izvedli smo še modeliranje odvodov s funkcijo  $g_q(x)$  po uporabi kosinusne transformacije. Se pravi najprej smo tako kot v primerih CKD4 in CKD5 razkosali sliko na enake dele, potem izvedli kosinusno transformacijo in nato na koeficientih kepstra izvajali

Tabela 4.6: Rezultati razpoznavanja s CKD predstavitevjo (3 stanja na model).

		1 porazdelitev	2 porazdelitvi	3 porazdelitve
<b>monofoni</b> CKD1	pravilnost	70.49 %	73.78 %	75.35 %
	natančnost	66.59 %	71.10 %	73.20 %
<b>difoni</b> CKD1	pravilnost	68.01 %	-	-
	natančnost	65.54 %	-	-
<b>monofoni</b> CKD2	pravilnost	75.46 %	79.28 %	80.35 %
	natančnost	72.25 %	76.74 %	77.97 %
<b>difoni</b> CKD2	pravilnost	76.38 %	-	-
	natančnost	73.27 %	-	-
<b>monofoni</b> CKD3	pravilnost	72.97 %	76.20 %	77.26 %
	natančnost	71.06 %	74.17 %	74.92 %
<b>difoni</b> CKD3	pravilnost	69.14 %	-	-
	natančnost	67.16 %	-	-
<b>monofoni</b> CKD4	pravilnost	70.17 %	72.34 %	73.01 %
	natančnost	66.08 %	68.94 %	70.18 %
<b>difoni</b> CKD4	pravilnost	66.12 %	-	-
	natančnost	63.71 %	-	-
<b>monofoni</b> CKD5	pravilnost	71.15 %	75.18 %	77.01 %
	natančnost	69.03 %	73.51 %	74.52 %
<b>difoni</b> CKD5	pravilnost	68.34 %	-	-
	natančnost	66.27 %	-	-
<b>monofoni</b> CKD6	pravilnost	73.92 %	78.21 %	80.19 %
	natančnost	71.65 %	76.36 %	78.19 %
<b>difoni</b> CKD6	pravilnost	76.73 %	-	-
	natančnost	73.83 %	-	-



Slika 4.14: Tri različni pristopi modeliranja prehodov po posameznih frekvenčnih pasovih predstavitve CKD.

aproksimacijo po metodi najmanjših kvadratov s funkcijo  $g_q(x)$ . To predstavitev smo poimenovali CKD6.

Primerjavo rezultatov predstavitev CKD1, CKD2, CKD3, CKD4, CKD5 in CKD6 si lahko ogledamo v tabeli 4.6.

CKD predstavitev se je izkazala za zelo primerno pri razpoznavanju govora. Rezultati razpoznavanj z različnimi verzijami, podani v tabeli 4.6, nam to lahko potrdijo. Že prva verzija CKD1 je najboljša med vsemi preizkušanimi verzijami s podobnimi konstrukcijami (razen spektrograma). CKD2 se je podobno kot pri prejšnjih predstavitvah izkazala za bistveno boljšo v primerjavi s CKD1, čeprav preskok v rezultatih ni tako izrazit kot v prejšnjih primerih. Pri CKD2 lahko podobno kot pri PMHD2 in PWVD2 ugotovimo nekaj boljše razpoznavanje glasovnih prehodov (difoni) kot pa glasovnih jeder (monofoni).

Predstavitve CKD3, CKD4 in CKD5 nam dajo nižje rezultate razpoznavanja kot CKD2. Pri teh verzijah smo izvajali odsvetlovanje pred kosinusno transformacijo in s tem smo, kot je razvidno iz tabele 4.6, bistveno pokvarili rezultate. Zanimivo je, da so pri CKD4 in CKD5 rezultati celo nižji pri difonih kot pri CKD1. Tu moramo še enkrat opozoriti, da imamo v primeru CKD4 nekaj manj značilk (brez drugih odvodov), kar se tudi pozna pri rezultatih, predvsem v primerjavi s CKD3 in CKD5. Tako smo samo potrdili dejstvo, da odsvetlovanje poudarja šum in s tem napake meritev s CKD predstavitevjo, kar v primeru razpoznavanja pokvari rezultate.

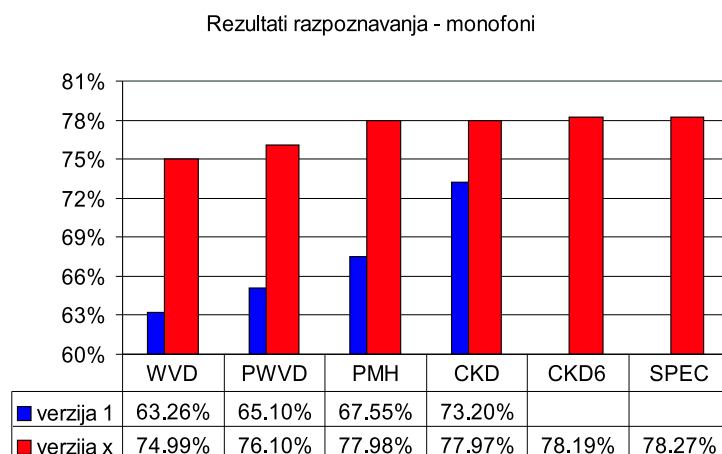
Modeliranje odsvetlovanj s funkcijo (4.19) po glajenju s kosinusno transformacijo se je v primeru CKD6 izkazalo za bistveno boljše kot za modeliranje pred transformacijo. Z verzijo CKD6 smo med CKD predstavitevami dosegli celo najboljše rezultate, tako v primeru monofonov kot difonov.



#### 4.2.2.7 Primerjava rezultatov

V prejšnjem delu smo primerjali rezultate razpoznavanja govornih enot pri različnih postopkih modeliranja znotraj posameznih časovno–frekvenčnih predstavitev. Tu pa se bomo posvetili primerjavi rezultatov med predstavitvami.

Rezultati razpoznavanja (natančnost) pri monofonih v primeru modeliranja s tremi funkcijami gostot na stanje so prikazani na sliki 4.15. Tu smo primerjali razpoznavanje s posameznimi predstavitvami: WVD, PWVD, PMH, CKD in spektrogramom SPEC. Z modro barvo (na sliki 4.15) so obarvane prve verzije posameznih predstavitev (WVD1, PWVD1, PMH1, CKD1), torej tiste, kjer smo uporabljali samo preskoke brez prekrivanj. Z rdečo barvo so označene najboljše verzije (v smislu razpoznavanja) posameznih predstavitev: WVD3, PWVD3, PMH2, CKD2. V zadnjem delu sta še najboljša predstavitev izmed CKD, verzija CKD6, in spektrogram SPEC.

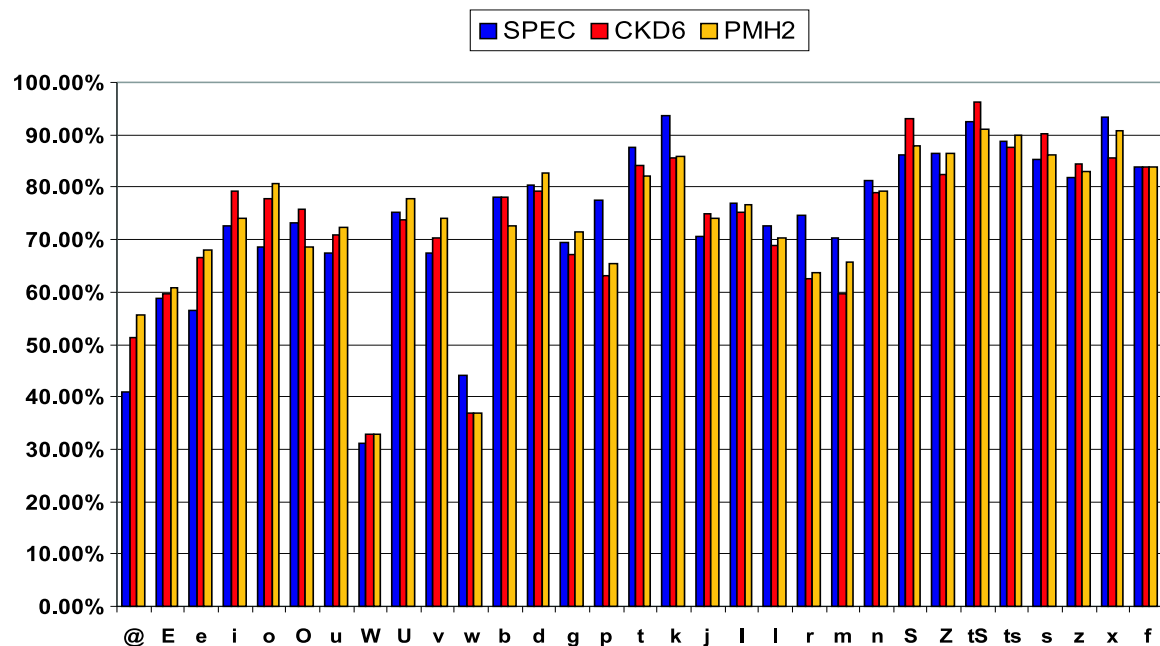


Slika 4.15: Rezultati razpoznavanja monofonov z različnimi predstavitvami.

Na prikazu 4.15 je lepo viden velik napredek razpoznavanja z verzijami, kjer smo uporabili princip prekrivanja in preskokov. Z WVD in PWVD predstavitevama dobimo nekaj slabše rezultate v primerjavi z ostalimi. To si lahko razlagamo s tem, da imamo pri teh porazdelitvah veliko interferenčnih členov, ki očitno povzročijo slabše razpoznavanje. Rezultati s PMH2, CKD6 (CKD2) in SPEC so skoraj identični. Tako lahko zaključimo, da je v primeru monofonov določanje značilik iz teh treh (štirih) porazdelitev enakovredno, izkaže se celo, da imamo v nekaterih primerih (pri modeliranju dveh funkcij gostot na stanje) razpoznavanje malenkost boljše pri PMH2 in CKD6 kot pri SPEC. Potrebno je še poudariti, da je tudi verzija CKD1 najboljša izmed vseh verzij 1 pri ostalih porazdelitvah. Pri tem je potrebno omeniti, da smo pri WVD, PWVD in PMH uporabljali analitičen signal, pri CKD verzijah in SPEC pa zaradi izvedbe postopka navadno različico signala.

Smiselno si je ogledati tudi porazdelitev napak po posameznih obravnavanih fonemih v primeru PMH2, CKD6 in SPEC. V ta namen smo pripravili dva prikaza. Pri prvem gre za prikaz razpoznavanja po posameznih govornih enotah, pri drugem pa prikazujemo primerjave razporeditev napak za posamezne predstavitve. Pri tem smo za ugotavljanje napak uporabljali popravljene verzije Levenshteinove razdalje [Pavešić-00, str. 444]

opisane v doktorski disertaciji [Dobrišek-01]. Z njimi namreč dobimo boljše (smiselne) razporeditve napak razpoznavanja.

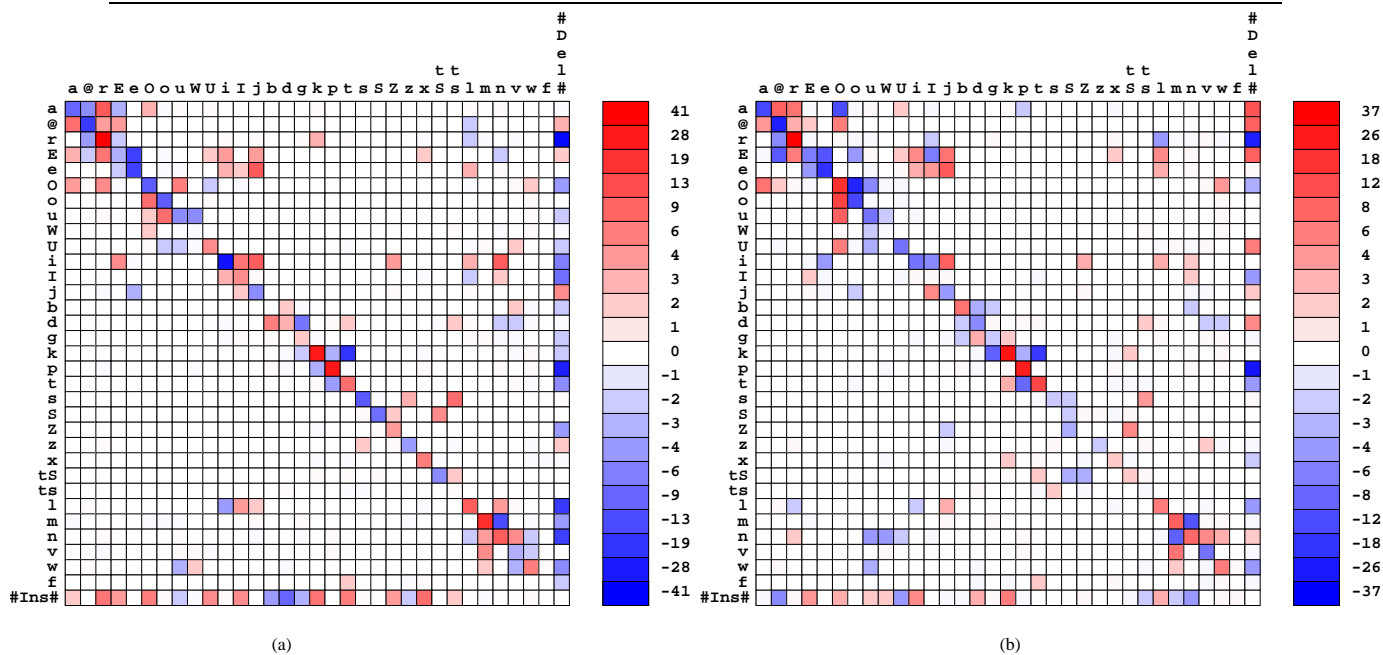


Slika 4.16: Rezultati razpoznavanja s SPEC, CKD6 in PMH2 predstavitvami posameznih monofonov govorne zbirke K211d.

Slika 4.16 prikazuje rezultate razpoznavanja po posameznih fonemih govorne zbirke K211d v primeru razpoznavanja z značilkami pridobljenimi iz SPEC, CKD6 in PMH2. V splošnem lahko ugotovimo boljše razpoznavanje vokalov /@/, /E/, /e/, /i/, /o/, /u/ s CKD6 in PMH2 v primerjavi s SPEC. Razlog za to je gotovo v bolj natančnem slikanju formantnih frekvenc brez dodatnih interferenc v primerjavi s spektrogramom. Nekaj boljše razpoznavanje s CKD6 je tudi pri šumniku /S/ in zlitniku /tS/. Spektrogram pa je boljši pri razpoznavanju zapornikov /p/, /t/, /k/ in ustnika /r/. V vseh ostalih primerih se izkažejo predstavitve za enakovredne, kar je posledica zelo podobnih metod modeliranja značilk.

V drugem primeru 4.17 smo preučevali razporeditev napak po posameznih glasovih. Primerjali smo SPEC in CKD6 (slika 4.17(a)) ter SPEC in PMH2 (slika 4.17(b)). Tu smo odštevali matrice zamenjav posameznih predstavitev. Na sliki 4.17(a) smo odštevali SPEC in CKD6, na sliki 4.17(b) pa SPEC in PMH2. Matriko zamenjav za posamezne predstavitve smo dobili na podlagi že prej omenjene verzije postopka poravnave nizov govornih odsekov. Pozitivne vrednosti (rdeči odtenki) v taki matriki pomenijo večje število zamenjav (v primeru #Del# izbrisov, #Ins# vrivanj) prve predstavitve v primerjavi z drugo in obratno, modri odtenki pomenijo večje število ustreznih operacij druge predstavitve v primerjavi s prvo. Rdeči in modri odtenki so izbrani v logaritemskem merilu (ločeno) in razporejeni po isti skali.

Iz slike 4.17(a), podobno kot iz prikaza 4.16, lahko za CKD6 ugotovimo boljše razpoznavanje vokalov in slabše nekaterih zapornikov v primerjavi s SPEC. Hkrati smo pri CKD6 imeli več izbrisov in manj vrivanj kot pri SPEC. Iz slike 4.17(a) lahko zelo dobro opazujemo porazdelitev napak pri posameznih fonemih, tako npr. za vokale /i/,



Slika 4.17: Primerjava porazdelitve napak razpoznavanja: (a) SPEC - CKD6, (b) SPEC - PMH2.

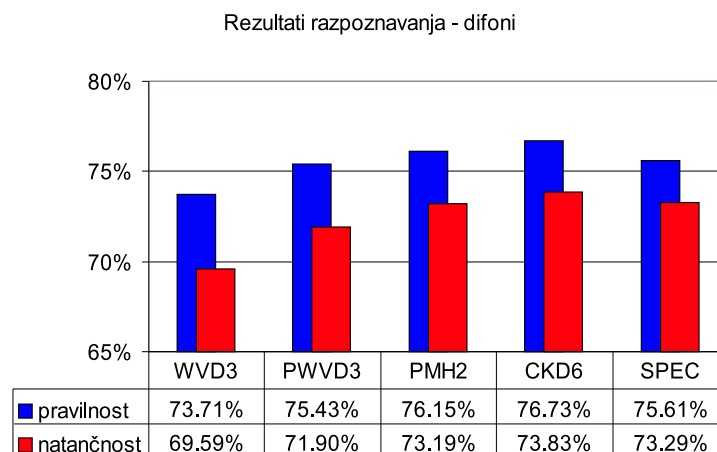
/O/ in /o/ ter za sičnik /s/ in šumnik /S/ lahko ugotovimo skoraj popolno razpoznavanje (razen vrivanj) s CKD6 v primerjavi s SPEC in obratno dobro razpoznavanje zapornikov /k/, /p/, /t/, ustnika /l/, /r/, in nosnika /m/ s SPEC. Za CKD6 lahko opazimo pogosto zamenjevanje /k/ s /t/ in /m/ z /n/, pri SPEC pa je precej napak pri vokalih.

Na sliki 4.17(b) lahko ugotovimo podobno kot v primeru 4.17(a). Tudi tu imamo boljše razpoznavanje vokalov s PMH2 in slabše nekaterih zapornikov. V splošnem imamo nekaj boljše razpoznavanje posameznih fonemov kot pri SPEC (več modrih kot rdečih odtenkov na diagonali matrike 4.17(b)), hkrati imamo tudi približno enako število izbrisov in vrivanj za razliko s CKD6 (slika 4.17(a)). Napake ostajajo tako približno enakomerno razporejene po matriki zamenjav. To je posledica modeliranja značilnk, ki je podobno v obeh primerih.

Kot smo ugotavljali že pri posameznih predstavitev, smo z vsemi praviloma boljše razpoznavali glasovne prehode modelirane z difoni kot pa glasovna jedra (monofoni).

Na sliki 4.18 so prikazani rezultati razpoznavanja difonov z uporabo značilnk pridobljenih z WVD3, PWVD3, PMH2, CKD6 in SPEC. Tu smo modelirali razpoznavalnik s samo eno porazdelitvijo na stanje modela difona. S histogrami na sliki 4.18 so prikazani rezultati pravilnosti in natančnosti razpoznavanja, ki so višji v primerjavi z istimi verzijami uporabljenih pri monofonih. Ugotovimo lahko tudi boljše razpoznavanje difonov s PMH2 in CKD6 kot pri SPEC (pri CKD6 za približno 1% relativno). To si lahko v primeru CKD6 razlagamo z drugačnim modeliranjem odvodov (z aproksimacijo s sinusom in kosinusom).

V celoti gledano smo s CKD6 in PMH2 dosegli primerljive oziroma ponekod boljše rezultate razpoznavanja kot pri SPEC v obeh primerih: z monofoni in difoni. WVD in



Slika 4.18: Rezultati razpoznavanja difonov s posameznimi predstavitvami.

PWVD izkazujeta nekaj manjšo zanesljivost razpoznavanja predvsem zaradi interferenc, ki se pojavljajo med posameznimi spektralnimi komponentami. Pomembna je tudi ugotovitev, da je bistveno boljše določanje značilk s prekrivanjem segmentov kot sam razrez časovno-frekvenčne slike. Primerljivost rezultatov pri različnih predstavitvah je posledica dejstva, da smo uporabili podobne postopke določanja značilk, ki temeljijo na koeficientih melodičnega kepstra.

## Zaključek

V tem poglavju smo preučevali uporabo časovno-frekvenčnih predstavitev na primeru govornih signalov. V prvem delu smo ocenjevali in ugotavljali, katere predstavitve so smiselne za analizo in nadaljnjo obdelavo govornih signalov. Opazovali smo vpliv interference in načine, kako se jo smiselno znebiti tako, da ne pokvarimo bistveno časovno-frekvenčne slike govornega signala. Za referenčne slike smo vedno jemali slike narejene na podlagi spektrograma.

V drugem delu pa smo ugotavljali vpliv časovno-frekvenčnih predstavitev na razpoznavanje govora. Poskuse smo izvajali s statističnim modelom razpoznavalnika zgrajenega s PM modeli na bazi K211d. Predvsem smo se ukvarjali z izbiro in določitvijo značilk, ki bi bile optimalne glede na uporabljeni model. Tu smo eksperimentirali z različnimi verzijami časovno-frekvenčnih predstavitev, pri katerih smo se naslanjali na danes najbolj uveljavljen postopek pridobivanja značilk - koeficientov melodičnega kepstra. V primeru Zhao-Atlas-Marksove predstavitve smo eksperimentirali tudi z modeliranjem dinamičnih značilk, kjer smo ugotavljali vplive različnih metod in načinov ocene dinamike.

Na podlagi rezultatov razpoznavanja smo tako ugotovili, da so bili naši postopki določevanja značilk približno enakovredni. Še posebno je potrebno izpostaviti pseudo Margenau-Hillovo in že omenjeno Zhao-Atlas-Marksovo porazdelitev, s katerima smo v nekaterih primerih celo presegli rezultate razpoznavanja na podlagi spektrograma. V vseh obravnavanih predstavitvah smo ugotavljali boljše razpoznavanje ob uporabi modelov glasovnih prehodov kot pa samih glasovnih jeder.

---

# 5 Zaključek

---

5.1 Sklepne ugotovitve

5.2 Nadaljnje delo

---

V tem poglavju bomo podali glavne ugotovitve in sklepe magistrskega dela. Osredotočili se bomo predvsem na časovno–frekvenčne izražave govornega signala v primeru razpoznavanja govora. Hkrati pa bomo podali nekaj razmišljanj za nadaljevanje dela na področju razpoznavanja in obdelave govornih signalov.

## 5.1 Sklepne ugotovitve

---

V magistrskem delu smo se ukvarjali s časovno–frekvenčnimi predstavitvami signalov s poudarkom na govornih signalih.

Ti predstavljajo zaradi svoje narave in lastnosti resen izziv za analizo in nadaljnjo obdelavo. Govorni signal je zelo kompleksen, opisuje tako dolge sorazmerno stacionarne dele kot tudi hipne, kratke glasove.

Spektrogram kot časovno–frekvenčna predstavitev je vrsto let služil skoraj kot edina možna predstavitev govora v časovno–frekvenčnem prostoru. Zaradi svojih številnih ugodnih lastnosti, kot so nenegativnost porazdelitve, razumljivost izvajanja operacij in izvedba s pomočjo hitre Fourierjeve transformacije, je še danes vodilna predstavitev v vrsti aplikacij, ki obravnavajo in zajemajo značilnosti govornih signalov. Njegova glavna pomanjkljivost pa je njegova enakomerna občutljivost povezana s principom nedoločenosti; povsem enako namreč tako merimo daljše stacionarne dele in krajše, hipne dogodke v signalu.

V magistrskem delu smo ustrezno umestili spektrogram v skupino predstavitev z imenom Cohenov razred. Ogledali smo si še druge predstavitve s tega razreda, ki spadajo med energijske porazdelitve. Manjši poudarek pa smo namenili drugi večji skupini - afnim predstavitev. Tako smo spoznali in podrobneje proučili delovanje Wigener–Villejeve, psevdo Wigener–Villejeve, Margenau–Hillove, ki je izvedba Richaczkove predstavitve, in njene glajene verzije. Še posebej pa smo se ukvarjali z Zhao–Atlas–Marksovo porazdelitvijo, ki ima vrsto ugodnih lastnosti za analizo in razpoznavanje govornega signala.

Obdelali smo probleme diskretizacije teh predstavitev. Posebej smo na primeru Wigener–Villejeve predstavitve izpostavili problem podvzorčenja ter predstavili neposreden pristop diskretizacije, ki je pomemben predvsem pri dokazovanju lastnosti diskretnih predstavitev.

V poglavju 4 smo se posvetili govornemu signalu. Predstavili smo nekaj metod prikaza časovno–frekvenčnih predstavitev, kjer smo vpeljali princip prikazovanja s prekrivanjem. Tega smo potem uspešno aplicirali v razpoznavanje govora. S takšnimi verzijami predstavitev smo v povprečju znatno popravili razpoznavanje govora. S tem pa smo tako vpeljali paleto novih možnosti za izračun in izbiro značilk govornega signala. S temi verzijami smo se zelo približali rezultatom razpoznavanja s spektrogramom, v nekaterih primerih smo ga celo presegli. To je pomemben dosežek, saj pri teh predstavitvah nismo omejeni z izbiro dolžine in tipa okna, predstavitev tudi niso obremenjene s principom nedoločenosti in kot take predstavljajo resno alternativo spektrogramu, predvsem v smislu pridobivanja novih, drugačnih značilk.

Namen magistrskega dela ni bil samo izboljšati rezultate razpoznavanja govora z različnimi izbirami značilk, ampak predvsem predstaviti različne možnosti obravnave signala in s tem povezanih problemov.

Pri razpoznavanju govora smo se pri izbiri značilk preko različnih časovno–frekvenčnih predstavitev držali standardnih že uveljavljenih postopkov pridobivanja značilk. Tako

smo upoštevali znane lastnosti narave govornega signala in človeškega slušnega zaznavanja. To je seveda smiselno, če bi tudi pri gradnji modelov razpoznavalnika govora simulirali “model” človeškega razpoznavanja govora, v nasprotnem pa to ni nujno potrebno. V magistrskem delu se s tem nismo posebej ukvarjali, čeprav bi bilo pri akustičnem modelirannju govora potrebno gledati na sistem kot celoto. To pomeni, da bi morali iskati optimalne značilke glede na model razpoznavalnika oziroma bi morali smiselno prilagajati modele glede na trenutne lastnosti značilk pridobljenih iz signalov.

V naslednjem razdelku bomo podali še nekaj smernic za nadaljnje delo v zvezi z obdelavo in razpoznavanjem govornega signala.

## 5.2 Nadaljnje delo

---

V poglavju 4 magistrskega dela smo se posvečali predvsem pridobivanju značilk za obdelavo govornega signala na podlagi že obstoječih postopkov izbire značilk, spreminjali smo le časovno–frekvenčne predstavitve govornega signala.

Tako smo v prejšnjem poglavju predvsem razmišljali o naravi govornih signalov, ki smo jih zato tudi tako spreminjali, da bi v nekem smislu posnemali človekovo slušno zaznavanje. Hkrati smo imeli v mislih tudi naravo govornega signala. Tako smo upoštevali kvazi stacionarnost signala in tudi postopke modeliranja in izbire značilk prilagodili tem dejstvom.

Na časovno–frekvenčne predstavitve pa lahko gledamo tudi kot na *slike* signala v dvodimenzionalnem prostoru časa in frekvence. V tem primeru smo zelo odvisni od izbire časovno–frekvenčne predstavitve. Slike posameznih tipov signalov morajo biti namreč dovolj dobre, da lahko ločujemo med razredi signalov. To pa pomeni, da lahko prevedemo problem razpoznavanja govornih signalov na problem razpoznavanja slik govornih signalov. Pri tem moramo upoštevati obliko, se pravi posamezne elemente in njihovo razporeditev na sliki. Hkrati pa moramo upoštevati tudi njihovo velikost, se pravi inteziteto barve na slikah z barvnimi lestvicami, kot smo jih opisovali v prejšnjem poglavju. Zato je razpoznavanje posameznih elementov takšnih slik precej zahteven problem za avtomatično obdelavo. Poleg tega moramo imeti še vedno v mislih signale oziroma razrede signalov, ki jih poskušamo medsebojno ločevati, zato ne moremo samo razpoznavati posameznih objektov (spektralnih komponent, ki se pojavljajo v določenih trenutkih v času) na slikah, ampak moramo najprej ali pa sočasno določiti samo tiste, ki so za ločevanje med posameznimi razredi signalov najbolj optimalni. Tu pridejo v poštev tudi interferenčni členi, ki nosijo informacijo o spremembah faze signala.

V primeru obdelave slik je pomembna tudi izbira modela razpoznavalnika. Sploh ni nujno, da bi gradili razpoznavalnike za razpoznavanje slik govornih signalov na način kot je danes v veljavi (s PM modeli).

V nadaljevanju bomo podali nekaj razmišljanj obdelave časovno–frekvenčne slike (govornih) signalov.

### 5.2.1 Renyijeva entropija

Denimo, da nas zanima informacija, iz koliko elementarnih signalov je sestavljen signal v danem trenutku. Se pravi, zanima nas, koliko elementarnih objektov (signalov) je skritih v sliki časovno–frekvenčne predstavitve.

Pri tem si moramo seveda zagotoviti dobro predstavitev in mero, s katero bomo merili to informacijo. Časovno–frekvenčne predstavitve obravnavamo kot *porazdelitve* energije signala. Zato bi lahko informacijo merili s Shannonovo mero entropije [Pavešić-97, str. 24], definirano kot

$$I_x = - \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx,$$

kjer je  $f(x)$  gostota verjetnosti spremenljivke  $x$ . To mero pa v našem primeru ne moremo uporabiti, saj so nekatere energijske porazdelitve tudi negativne, zato uporabimo posplošeno mero znano kot *Renyijevo entropijo*, podano kot

$$R_x^\alpha = \frac{1}{1-\alpha} \log_2 \int_{-\infty}^{\infty} f^\alpha(x) dx.$$

Iz Renyijeve mere entropije prvega reda, ko gre  $\alpha \rightarrow 1$ , se dá izpeljati Shannonovo entropijo. Renyijevo entropijo tretjega reda pa lahko uporabimo v zvezi s časovno–frekvenčnimi predstavitvami  $C_x(t, f)$ :

$$R_C^3 = -\frac{1}{2} \log_2 \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_x^3(t, f) dt df \right\}.$$

Rezultat zgornje mere je podan v bitih in nam pove naslednje: če predpostavimo, da nam en elementaren signal prisoten v signalu  $x(t)$  pomeni 0 bitov informacije ( $2^0$ ), potem nam dva elementarna signala podata en bit informacije ( $2^1$ ), štirje signali dva bita informacije ( $2^2$ ) in tako naprej.

Tako informacijo pa že lahko štejemo za značilko pri razpoznavanju signalov.

### 5.2.2 Houghova transformacija

Če nas zanimajo linearno frekvenčno modulirani signali, jih lahko na slikah v časovno–frekvenčni ravnini opazujemo kot premice. Tako lahko prevedemo problem iskanja linearnih cvrkov, v iskanje premic na slikah, kar je pri razpoznavanju slik znan problem. To dosežemo z uporabo Houghove transformacije [Pavešić-00, str.83], ki jo lahko v polarnih koordinatah zapišemo kot

$$x \cos \theta + y \sin \theta = \rho.$$

Bistvo te transformacije je, da se točke slike, ki so razporejene v okolici neke premice na sliki, v parametričnem prostoru  $(\rho, \theta)$  preslikajo v okolico ene same točke, ki tako predstavlja lokalni maksimum. Z izbiro ustreznega praga, tako lahko štejemo in ugotavljamo smer premic na dani sliki.



Houghovo transformacijo lahko v primeru časovno–frekvenčnih predstavitev prevedemo v t.i. *Wigener–Houghovo transformacijo* [Auger-97a]. Zapišimo signal kot linearni cvrk z dodanim belim Gaussovimi šumom  $n(t)$  s trajanjem  $T$ :

$$x(t) = e^{j2\pi(\nu t + \beta/2t^2)} + n(t).$$

Z izpeljavo Wigener–Villejeve predstavitev, ki idealno ločuje linearne cvrke, na signalu  $x(t)$ , dobimo Wigener–Houghovo transformacijo:

$$\begin{aligned} WH_x(\nu, \beta) &= \int_T W_x(t, \nu + \beta t) dt \\ &= \int_{-\infty}^{\infty} \int_T x(t + \tau/2) x^*(t - \tau/2) e^{-j2\pi(\nu t + \beta)\tau} dt d\tau \end{aligned}$$

Ideja je tu analogna kot pri običajni Houghovi transformaciji. Točke, ki opisujejo linearne cvrke na slikah WV predstavitev, bodo v prostoru  $(\nu, \beta)$  razporejene samo okoli enega vrha. Ob uvedbi praga tako dobimo informacijo, koliko in kje na sliki lahko iščemo linearno modulirane signale.

### 5.2.3 Učenje jedra predstavitev

Kot zadnjo idejo za določevanje značilk naj omenimo še učenje predstavitev za razpoznavanje.

V poglavju 3.2.4.3 smo si ogledali izpeljavo diskretne verzije večlične funkcije, s katero lahko definiramo diskretno Richaczko porazdelitev. Z vpeljavo jedra  $\phi[\eta, \tau]$  pa lahko definiramo poljubne predstavitev, ki izhajajo iz Richaczkove porazdelitve:

$$\begin{aligned} A[\eta, \tau] &= \sum_{n=0}^{N-1} x^*[n] x[n + \tau] \exp\left(\frac{j2\pi n\eta}{N}\right), \\ M[\eta, \tau] &= \phi[\eta, \tau] A[\eta, \tau] \end{aligned}$$

in

$$P[n, k] = \sum_{\eta=0}^{N-1} \exp\left(\frac{-j2\pi\eta n}{N}\right) \sum_{\tau=0}^{N-1} \exp\left(\frac{-j2\pi\tau k}{N}\right) M[\eta, \tau].$$

Jedro  $\phi[\eta, \tau]$  lahko izbiramo na različne načine in potrebe. Najbolj pogosto se določa jedra, ki gladijo osnovne predstavitev, tako da izničujejo interferenčne člene. Pri tem pa upoštevamo še dodatne lastnosti jeder, da z njimi ne pokvarimo lastnosti predstavitev.

Druga možnost pa je učenje jeder, ki se lahko prilagajajo tipu signalov, ki jih obdelujemo [Baraniuk-90]. Takim predstavitev pravimo *optimalne predstavitev* in jih računamo z reševanjem linearnih programov. Te so seveda optimalne glede na tip signalov, ki jih analiziramo. V ozadju je še vedno namen izboljšati sliko predstavitev. Podoben princip pa lahko uporabimo, če hočemo določevati jedra, s katerimi bi lahko optimalno ločevali med posameznimi razredi signalov [Gillespie-01]. Tu lahko interferenčni členi pomenijo dodatno informacijo za iskanje jedra separacije.

Formalno lahko to zapišemo kot iskanje jedra z maksimizacijo  $L_2$  norme med dvema predstavitevama v ravnini  $(\eta, \tau)$ :

$$\hat{\phi}[\eta, \tau] = \arg \max_{\phi[\eta, \tau]} \sum_{\eta, \tau} |\phi[\eta, \tau](A_1[\eta, \tau] - A_2[\eta, \tau])|^2,$$

kjer sta  $A_1[\eta, \tau]$  in  $A_2[\eta, \tau]$  tipični predstavitvi prvega in drugega razreda. Enolično rešitev gornjega optimizacijskega problema zagotovimo s pogojem, da ima jedro končno energijo:

$$\sum_{\eta, \tau} |\phi[\eta, \tau]|^2 = 1.$$

Takšno določevanje jedra je smiselno, če imamo časovno poravnane signale, ki so si zelo podobni, vendar jih moramo vseeno ločevati.

V primeru govornega signala bi lahko to izvedli kot dodatne značilke za fino ločevanje posameznih skupin fonemov, ki se pogosto zamenjujejo v procesu razpoznavanja.

~

Tako smo podali še nekaj možnosti pridobivanja značilk iz slike dane časovno–frekvenčne predstavitve, ki bi jih lahko smiselno integrirali v postopke določevanja in izbire značilk za uspešno razpoznavanje govornih signalov.

---

## Viri in literatura

- [Ainsworth-98] W. A. Ainsworth, C. R. Day, G. Meyer. *Improving pitch estimation with short duration speech samples*. In Proc. ICSLP, Sydney, Australia. str. 1363–1366. 1998.
- [Atlas-97] L. Atlas, J. Droppo, and J. McLaughlin. *Optimizing time–frequency distributions via operator theory*. Proceedings of the SPIE. Volume 3162. 1997.
- [Auger-95] F. Auger, P. Flandrin. *Improving the readability of time–frequency and time–scale representations by the reassignment method*. IEEE Trans. on Signal Processing. Volume 43, št. 5, str. 1068–1089. 1995.
- [Auger-97] F. Auger, P. Flandrin, P. Gonçalves, O. Lemoine. *Time–frequency toolbox for use with MATLAB*. Reference guide. 1997.
- [Auger-97a] F. Auger, P. Flandrin, P. Gonçalves, O. Lemoine. *Time–frequency toolbox for use with MATLAB*. Tutorial. 1997.
- [Baraniuk-90] R. G. Baraniuk in D. L. Jones. *Optimal kernels for time–frequency analysis*. Proc. SPIE. Volume 1348, str. 181 – 187. 1990.
- [Bertrand-91] J. Bertrand, P. Bertrand. *Affine time–frequency distributions*. Time–frequency analysis, urednik B. Boashash. Longman–Cheshire, Melbourne, str. 1–23. 1991.
- [Choi-89] H. I. Choi, W. J. Williams. *Improved Time–Frequency Representation of Multicomponent Signals Using Exponential Kernels*. IEEE Transactions on Acoust., Speech, Signal Processing. Volume ASSP-37, št. 6, str. 862–871. 1989.
- [Cohen-66] L. Cohen. *Generalized phase–spaced distribution functions*. Journal of Math. Phys.. Volume 7, str. 781–806. 1966.
- [Cohen-95] L. Cohen. *Time–Frequency Analysis*. Prentice Hall Signal Processing Series. 1995.
- [Daubechies-90] I. Daubechies. *The Wavelet Transform, Time–Frequency Localization and Signal Analysis*. IEEE Transactions on Information Theory. Volume 36, No. 5, str. 961–1005. 1990.

- [Daubechies-92] I. Daubechies. *Ten Lectures on Wavelets*. Philadelphia: SIAM. 1992.
- [Dobrišek-96] S. Dobrišek, J. Gros, F. Mihelič, K. Pepelnjak in I. Ipšič. *GOPOLIS: Slovenian Speech Database of Spoken Flight Information Queries*. Proceedings of the 3rd Slovenian-German and 2nd SDRV Workshop on Speech and Image Understanding, str. 37–46. Ljubljana. 1996.
- [Dobrišek-97] S. Dobrišek, F. Mihelič in N. Pavešič. *A Multiresolutionally Oriented Approach for Determination of Cepstral Features in Speech Recognition*. Proceedings of the 5'th European Conference On Speech Communication and Technology. Volume 3, str. 1367 – 1370. Rhodes, Greece. 1997.
- [Dobrišek-01] S. Dobrišek. *Analiza in razpoznavanje glasov v govornem signalu*. Doktorska disertacija. Univerza v Ljubljani, Fakulteta za elektrotehniko. Ljubljana. 2001.
- [Flandrin-89] P. Flandrin. *Some aspects of non-stationary signal processing with ephassis on time-frequency and time-scale methods*. Wavelets: Time-Frequency Methods and Phase Space, str. 68–98. Springer-Verlag 1989.
- [Furui-86] S. Furui. *Speaker Independent Isolated Word Recogniser Using Dynamic Features of Speech Spectrum*. IEEE Transaction on Acoustic, Speech and Signal Processing. Volume 34, No 1, str. 52 – 59. 1986.
- [Gabor-46] D. Gabor. *Theory of Communication*. J. IEE (London). Volume 93(III), str. 429–457. 1946.
- [Gillespie-01] B. W. Gillespie in L. E. Atlas. *Optimizing Time-Frequency Kernels for Classification*. IEEE Trans. on Signal Processing. Volume 49, št. 3, str. 485–496. 2001.
- [Gros-97] J. Gros. *Sistem za sintezo slovenskega govora*. Doktorska disertacija. Univerza v Ljubljani, Fakulteta za elektrotehniko. Ljubljana. 1997.
- [Gyergyek-91] L. Gyergyek. *Teorija signalov in obdelava signalov*. Založba FE, Fakulteta za elektrotehniko, Ljubljana. 1991.
- [Harris-78] F. Harris. *On the use of windows for harmonic analysis with the discrete Fourier transform*. In Proceedings IEEE. Volume 66, str. 51–83. 1978.
- [Hlawatsch-91] F. Hlawatsch. *Time-Frequency Methods for Signal Processing*. V Technical Report 1291-0001. Departement of Electrical Engineering, University of Rhode Island. 1991.
- [Hlawatsch-92] F. Hlawatsch in G.F. Boudreaux-Bartels. *Linear and Quadratic Time-Frequency Signal Representations*. IEEE Signal Processing Magazine. Volume 9, št. 2, str. 21–67. 1992.

- [Ipšič-96] I. Ipšič. *Razpoznavanje besed v vezanem govoru*. Doktorska disertacija. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. Ljubljana. 1996.
- [Jelinek-98] F. Jelinek. *Statistical Methods for Speech Recognition*. The MIT Press. Cambridge, Massachusetts, London, England. 1998.
- [Loughlin-93] P. J. Loughlin, J. W. Pitton, L. E. Atlas. *Bilinear Time-Frequency Representations: New Insights and Properties*. IEEE Transactions on Signal Processing. Volume 41, št. 2. 1993.
- [Markel-80] J. Markel in A. H. Gray, Jr. *Linear Prediction of Speech*. Springer-Verlag, New York. New York. 1980.
- [Marušič-98] B. Marušič. *Kodiranje slik z valčno transformacijo*. Magistrsko delo. Fakulteta za elektrotehniko, Univerza v Ljubljani. Ljubljana. 1998.
- [Malayath-00] N. Malayath. *Data-Driven Methods for Extracting Features from Speech*. Doctoral dissertation. Departement of Electrical and Computer Engineering. Oregon Graduate Institute of Science and Technology. 2000.
- [McLaughlin-97] J. McLaughlin, L. E. Atlas. *Applications of Operator Theory to Time-Frequency Analysis and Classification*. sprejeto v IEEE Transactions on Signal Processing. 1997.
- [Mihelič-91] F. Mihelič, *Akustično fonetična pretvorba slovenskega govora*. Doktorska disertacija. Fakulteta za elektrotehniko in računalništvo, Univerza v Ljubljani. Ljubljana. 1991.
- [Mihelič-93] F. Mihelič in B. Lozej. *Dolžine trajanja in glasnost izgovorjave slovenskih fonemov*. Zbornik delavnice: Jezik tako in drugače. str. 441 – 451. Ljubljana. 1993.
- [Moyal-49] J. E. Moyal. *Quantum mechanics as a statistical theory*. Proc. Camb. Phil. Soc. Volume 45, str. 99–124. 1949.
- [Narayanan-96] S.B. Narayanan, J. McLaughlin, L. Atlas, and J. Droppo. *An operator theory approach to discrete time-frequency distributions*. In Proc. of the IEEE Int. Symp. on Time-Frequency and Time-Scale Analysis, str. 521–524. 1996.
- [Octave] J. W. Eaton. *GNU-Octave*. <http://www.octave.org>.
- [Papo-77] A. Papoulis. *Signal Analysis*. McGraw-Hill Book Co. New York. 1996.
- [Pavešič-97] N. Pavešič. *Informacija in kodi*. Založba FE in FRI, Fakulteta za elektrotehniko, Univerza v Ljubljani. 1997.
- [Pavešič-00] N. Pavešič. *Razpoznavanje vzorcev*. Založba FE, Fakulteta za elektrotehniko, Ljubljana. 2000.

- [Picone-93] J. Picone. *Signal Modeling Techniques In Speech Recognition*. Proceedings of the IEEE. 1993.
- [Pitton-95] J. W. Pitton in L. E. Atlas. *Discrete-Time Implementation of the Cone-Kernel Time-Frequency Representation*. IEEE Trans. on Signal Processing. Volume 43, št. 8, str. 1996–1998. 1995.
- [Plante-95] F. Plante, W. A. Ainsworth. *Formant tracking using reassigned spectrum*. In Proc. of EUROSPEECH, Madrid. str. 741–744. 1995.
- [Qian-96] S. Qian in D. Chen. *Joint Time-Frequency Analysis: Methods and Applications*. Prentice-Hall PTR, NJ. 1996.
- [Qian-99] S. Qian in D. Chen. *Joint Time-Frequency Analysis*. IEEE Signal Processing Magazine. Volume 16, št. 2, str. 53–67. 1999.
- [Rabiner-78] L. R. Rabiner in R. W. Schafer. *Digital Processing of Speech Signals*. Prentice-Hall. Englewood Cliffs, NJ. 1978.
- [Rabiner-93] L. R. Rabiner in B-H. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall. Englewood Cliffs, NJ. 1993.
- [Richaczek-68] A. W. Richaczek. *Signal Energy Distribution in Time and Frequency*. IEEE Transactions on Information Theory. Volume 14, str. 369–374. 1968.
- [Rioul-91] O. Rioul in M. Vetterli. *Wavelets and Signal Processing*. IEEE Signal Processing Magazine. Volume 8, št. 4, str. 14–38. 1991.
- [Rioul-92] O. Rioul in P. Flandrin. *Time-scale distributions: A general class extending wavelet transform*. IEEE Transactions on Signal Processing. Volume 40, št. 7, str. 1746–1757. 1992.
- [Shenoy-95] R. G. Shenoy, T. W. Parks. *Wide-band ambiguity functions and affine Wigner distributions*. Signal Processing. Volume 41, št. 1, str. 339–363. 1995.
- [Ville-48] J. Ville. *Theorie et applications de la notion de signal analytique*. str. 61–74. 1948.
- [Wells-92] J. Wells, W. Barry, M. Grice, A. Fourcin in D. Gibbon. *Standard Computer Compatible Transcriptions*. ESPRIT Project 2589 (SAM). SAM-UCL-037. 1992.
- [Wilcox-67] R. M. Wilcox. *Exponential Operators and Parameter Differentiation in Quantum Physics*. Journal of Math. Phys. Volume 8, str. 962–982. 1967.
- [Yang-98] D. Yang, W. A. Ainsworth, G. Meyer. *Vowel separation using the reassigned amplitude-modulation spectrum*. In Proc. ICSLP, Sydney, Australia. str. 947–950. 1998.

- [Young-00] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Vatchev in P. Woodland. *The HTK Book*. Cambridge University Engineering Department, Cambridge, United Kingdom. 2000.
- [Zhao-90] Y. Zhao, L. E. Atlas, R. J. Marks. *The use of cone-shaped kernels for generalized time–frequency representations of nonstationary signals*. IEEE Transactions on Acoust., Speech, Signal Processing. Volume 38, str. 1084–1091. 1990.
- [Zwicker-90] E. Zwicker in H. Fastl. *Psychoacoustics: Facts and Models*. Springer-Verlag. Berlin, Heidelberg. Germany. 1990.
- [Žibert-00] J. Žibert, F. Mihelič. *Slovenian Weather Forecast Speech Database*. In Proceedings of SOFTCOM 2000, International Conference on Software, Telecommunications and Computer Networks, str 199–206. Split. 2000.

---

# A Pregled časovno–frekvenčnih predstavitev

---

V dodatku so zbrane znane časovno–frekvenčne predstavitve signalov povzete po [Hlawatsch-92]. To so zvezne predstavitve, ki jih delimo na linearne oz. atomarne predstavitve in kvadratne oz. energijske predstavitve. Pri energijskih predstavitev se uporablja izraz tudi energijske porazdelitve. Obstajajo pa še predstavitve, ki ne spadajo v noben razred.

## A.1 Linearne časovno–frekvenčne predstavitve

---

1. kratkočasovna Fourierjeva transformacija (STFT):

$$STFT_x(t, f; h) = \int_{-\infty}^{\infty} x(\tau)h^*(\tau - t)e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{\infty} X(\xi)H^*(\xi - f)e^{j2\pi t(\xi - f)} d\xi,$$

2. zvezna valčna transformacija (CWT):

$$CWT_x(t, f; \psi) = \int_{-\infty}^{+\infty} x(s)\psi_{t,a}^*(s)ds = \int_{-\infty}^{+\infty} X(\xi)\sqrt{|f_0/f|}\Psi_{t,f}^*\left(\frac{f_0}{f}\xi\right)e^{j2\pi t\xi}d\xi,$$

3. Gaborjeva transformacija, Gaborjevi koeficienti:

$$x(t) = \sum_n \sum_m G_x[n, m]g_{n,m}(t).$$

## A.2 Kvadratne časovno–frekvenčne predstavitve

---

Kvadratne predstavitve lahko razdelimo še glede na lastnosti, ki jih imajo. Delimo jih na dva večja razreda: Cohenov razred in afine predstavitve.

Pri časovno–frekvenčni analizi signalov se v primeru kvadratnih predstavitev mešata pojma predstavitev (ang. representation) in porazdelitev (ang. distribution). V prvem primeru mislimo sliko porazdelitve, v drugem pa zgolj porazdelitev energije. Ker je pri verjetnosti funkcija porazdelitve verjetnosti nenegativna, realna s ploščino 1, v primeru porazdelitev energije pa to ni vedno res, je bolj smiselno, če govorimo o predstavitev. Vsekakor sta pri takšni časovno–frekvenčni analizi signalov v primeru kvadratnih predstavitev pojma ekvivalentna.



1. *spektrogram*:

$$SPEC_x(t, f; h) = |STFT_x(t, f; h)|^2 = \left| \int_{-\infty}^{\infty} x(\tau) h^*(\tau - t) e^{-j2\pi f\tau} d\tau \right|^2,$$

2. *skalogram*:

$$SCAL_x(t, f; \psi) = |CWT_x(t, f; \psi)|^2 = \left| \int_{-\infty}^{+\infty} x(s) \psi_{t,a}^*(s) ds \right|^2,$$

3. *Wigener-Villejeva porazdelitev*:

$$W_x(t, f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{+\infty} X\left(f + \frac{\xi}{2}\right) X^*\left(f - \frac{\xi}{2}\right) e^{j2\pi t\xi} d\xi,$$

4. *psevdo Wigener-Villejeva porazdelitev*:

$$PW_x(t, f) = \int_{-\infty}^{+\infty} h(\tau) x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{+\infty} H(f - \xi) W_x(t, \xi) d\xi,$$

5. *zglajena psevdo Wigener-Villejeva porazdelitev*:

$$SPW_x(t, f) = \int_{-\infty}^{+\infty} h(\tau) \int_{-\infty}^{+\infty} g(s - t) x(s + \tau/2) x^*(s - \tau/2) ds e^{-j2\pi f\tau} d\tau,$$

6. *večlična funkcija (AF)*:

$$A_x(\xi, \tau) = \int_{-\infty}^{+\infty} x(s + \tau/2) x^*(s - \tau/2) e^{-j2\pi\xi s} ds = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_x(\tau, f) e^{j2\pi(f\tau - \xi t)} dt df,$$

7. *Richaczkova porazdelitev*:

$$R_x(t, f) = x(t) X^*(f) e^{-j2\pi ft},$$

8. *Margenau-Hillova porazdelitev*:

$$MH_x(t, f) = \Re\{x(t) X^*(f) e^{-j2\pi ft}\},$$

9. *Pageova porazdelitev*:

$$P_x(t, f) = \frac{d}{dt} E_x^-(t, f) = 2\Re\left\{x(t) \left(\int_{-\infty}^t x(u) e^{-j2\pi fu} du\right)^* e^{-j2\pi ft}\right\},$$

10. *Choi-Williamsova porazdelitev*:

$$CW_x(t, f) = \sqrt{\frac{2}{\pi}} \int \int_{-\infty}^{+\infty} \frac{\sigma}{|\tau|} e^{-2\sigma^2(s-t)^2/\tau^2} x(s + \tau/2) x^*(s - \tau/2) e^{-j2\pi f\tau} ds d\tau,$$

11. *Born–Jordanova porazdelitev:*

$$BJ_x(t, f) = \int_{-\infty}^{+\infty} \frac{1}{|\tau|} \left[ \int_{t-|\tau|/2}^{t+|\tau|/2} x(s + \tau/2)x^*(s - \tau/2)ds \right] e^{-j2\pi f\tau} d\tau,$$

12. *Zhao–Atlas–Marksova porazdelitev:*

$$ZAM_x(t, f) = \int_{-\infty}^{+\infty} h(\tau) \left[ \int_{t-|\tau|/2}^{t+|\tau|/2} x(s + \tau/2)x^*(s - \tau/2)ds \right] e^{-j2\pi f\tau} d\tau,$$

13. *afina zglačena pseudo Wignerjeva porazdelitev:*

$$ASPW_x(t, a) = \frac{1}{a} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h\left(\frac{\tau}{a}\right) g\left(\frac{s-t}{a}\right) x\left(s + \frac{\tau}{2}\right) x^*\left(s - \frac{\tau}{2}\right) ds d\tau$$

14. *Bertrandova porazdelitev:*

$$BED_x(t, f) = f \int_{-\infty}^{+\infty} X(f\lambda(u)e^{(u/2)})X^*(f\lambda(-u)e^{(-u/2)})\lambda(u)e^{j2\pi t f u} du,$$

kjer je  $\lambda(u) = \frac{u/2}{\sinh(u/2)}$ ,

15. *D-Flandrinova porazdelitev:*

$$FD_x(t, f) = f \int_{-\infty}^{+\infty} X\left[f\left(1 + \frac{u}{4}\right)^2\right] X^*\left[f\left(1 - \frac{u}{4}\right)^2\right] \left[\left(1 - \frac{u}{4}\right)^2\right] e^{j2\pi t f u} du,$$

16. *aktivna Unterbergerjeva porazdelitev:*

$$AUD_x(t, f) = f \int_0^{\infty} X(fu)X^*\left(\frac{f}{u}\right) \left(1 + \frac{1}{u^2}\right) e^{j2\pi t f(u-1/u)} du,$$

17. *pasivna Unterbergerjeva porazdelitev:*

$$PUD_x(t, f) = f \int_0^{\infty} X(fu)X^*\left(\frac{f}{u}\right) \frac{1}{u} e^{j2\pi t f(u-1/u)} du,$$

18. *Butterworthova porazdelitev:*

$$BUD_x(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(\tau, \nu) A_x(\tau, \nu) e^{j2\pi(t\nu - f\tau)} d\tau d\nu,$$

kjer je  $\psi(\tau, \nu) = \frac{1}{1 + \left(\frac{\tau}{\tau_0}\right)^{2M} \left(\frac{\nu}{\nu_0}\right)^{2N}}$ ,

19. *glajena eksponentna porazdelitev:*

$$GED_x(t, f) = \int_{-\infty}^{\infty} \int_{\nu} \psi(\tau, \nu) A_x(\tau, \nu) e^{j2\pi(t\nu - f\tau)} d\tau d\nu,$$

kjer je  $\psi(\tau, \nu) = \exp\left[-\left(\frac{\tau}{\tau_0}\right)^{2M} \left(\frac{\nu}{\nu_0}\right)^{2N}\right]$ ,

20. *posplošena Wigner–Villejeva porazdelitev:*

$$GWD_x^{(\alpha)}(t, f) = \int_{-\infty}^{+\infty} x\left(t + \left(\frac{1}{2} - \alpha\right)\tau\right) x^*\left(t - \left(\frac{1}{2} - \alpha\right)\tau\right) e^{-j2\pi f\tau} d\tau$$

21. *Levinova porazdelitev:*

$$LD_x(t, f) = -\frac{d}{dt} \left| \int_t^{\infty} e^{-j2\pi f\tau} d\tau \right|^2 = 2\Re \left\{ x^*(t) e^{j2\pi ft} \int_t^{\infty} x(\tau) e^{-j2\pi f\tau} d\tau \right\}$$

22. *realni del posplošene Wigner–Villejeve porazdelitve:*

$$RGWD_x^{(\alpha)}(t, f) = \Re \{ GWD_x^{(\alpha)}(t, f) \}$$

### A.3 Ostale časovno–frekvenčne predstavitve

Obstaja še nekaj predstavitev, ki ne zadoščajo pogoju linearnosti in/ali bilinearnosti:

1. *prilagajanje Gaussovega jedra tipu signalov:*

$$RGD_x(t, f) = \int_{\tau} \int_{\nu} \psi_x(\tau, \nu) A_x(\tau, \nu) e^{j2\pi(t\nu - f\tau)} d\tau d\nu,$$

$$\text{kjer je } \psi_x(\tau, \nu) = \exp \left[ -\frac{(\tau/\tau_0)^2 + (\nu/\nu_0)^2}{2\sigma_x^2(\Theta)} \right], \Theta = \arctan \frac{\nu/\nu_0}{\tau/\tau_0}$$

2. *Cohenova nenegativna porazdelitev:*

$$CND_x^{(c, \rho)}(t, f) = \frac{|x(t)|^2 |X(f)|^2}{E_x} [1 + c\rho(\xi_x(t), \eta_x(f))],$$

$$\text{kjer je } \xi_x(t) = \frac{1}{E_x} \int_{-\infty}^t |x(\tau)|^2 d\tau, \eta_x(f) = \frac{1}{E_x} \int_{-\infty}^f |X(f')|^2 df', E_x = \int_t |x(t)|^2 dt$$

---

# Slovar izrazov

V magistrskem delu smo uporabili precej novih strokovnih izrazov. Za boljše razumevanje in lažji pregled smo tu podali slovar slovenskih in ustreznih angleških izrazov. Še enkrat poudarimo, da smo prevajali porazdelitve (distributions) kot predstavitev.

<b>afine predstavitev</b>	affine distributions
<b>aktivna Unterbergerjeva predstavitev</b>	active Unterberger distribution
<b>analitičen signal</b>	analytic signal
<b>atomarne predstavitev</b>	atomic decompositions
<b>Bertrandova predstavitev</b>	Bertrand distribution
<b>bifrekvenčno jedro</b>	bi-frequency kernel
<b>Born–Jordanova predstavitev</b>	Born–Jordan distribution
<b>Butterworthova predstavitev</b>	Butterworth distribution
<b>Choi–Williamsova predstavitev</b>	Choi–Williams distribution
<b>Cohenov razred predstavitev</b>	distributions of Cohen’s class
<b>časovna širina</b>	time spreading, standard deviation in time
<b>časovno–frekvenčna ločljivost</b>	time–frequency resolution
<b>D–Flandrinova predstavitev</b>	D–Flandrin distribution
<b>diadična skala</b>	diadic scale
<b>diskretna kratkočasovna Fourierjeva transformacija</b>	discrete short–time Fourier transform
<b>diskretna valčna transformacija</b>	discrete wavelet transform
<b>energijske porazdelitve</b>	energy distributions
<b>Fourierjeva transformacija</b>	Fourier transform
<b>frekvenčna širina</b>	frequency spreading, standard deviation in frequency

---

Gaborjeva transformacija	Gabor transform
Gaborjevi koeficienti	Gabor coefficients
Gaborjevi nastavki	Gabor logons
glajena eksponentna predstavitev	generalized exponential distribution
Hilbertova transformacija	Hilbert transform
idealna lokalizacija linearnih cvrkov	perfect localization on linear chirp signals
interferenca	interference
interferenčni členi	interference terms
izhodiščni valček	mother wavelet
jedro	kernel function
karakteristična funkcija	characteristic function
kratkočasovna Fourierjeva transformacija	short-time Fourier transform
križni členi	cross terms
kvadratna superpozicija	quadratic superposition principle
kvadratne časovno-frekvenčne predstavitve	quadratic time-frequency representations
Levinova predstavitev	Levin distribution
linearne časovno-frekvenčne predstavitve	linear time-frequency representations
Margenau-Hillova predstavitev	Margenau-Hill distribution
načelo kvadratne superpozicije	quadratic superposition principle
ohranitev konvolucije	compatibility with filterings
ohranitev nosilcev	support conservation
ohranitev premika po času	time translation covariance
ohranitev premika po frekvenci	frequency translation covariance
ohranitev raztegov	dilation covariance
Pageova predstavitev	Page distribution
parametrizacijska funkcija	parametric function
pasivna Unterbergerjeva predstavitev	passive Unterberger distribution

---

---

pojav prekrivanja	aliasing
porazdelitve reducirane interference	reduced interference distributions
posplošena Wigner–Villejeva predstavitev	generalized Wigner–Ville distribution
povprečen čas	average time
povprečna frekvenca	average frequency
pravilo ujemanja	correspondence rule
princip nedoločenosti	uncertainty principle, Heisenberg–Gabor inequality
produkt v konvolucijo	compatibility with modulations
psevdo afina Wigner–Villejeva predstavitev	pseudo affine Wigner–Ville distribution
psevdo Margenau–Hillova predstavitev	pseudo Margenau–Hill distribution
psevdo Pageova predstavitev	pseudo Page distribution
psevdo Wigner–Villejeva predstavitev	pseudo Wigner–Ville distribution
realocirane predstavitve	reassigned distributions
Richaczekova predstavitev	Richaczek distribution
robni pogoj	marginal property
signal cvrk	chirp signal
skalogram	scalogram
skupinska zakasnitev	group delay
spektrogram	spectrogram
trenutna amplituda	instantaneous amplitude
trenutna frekvenca	instantaneous frequency
unitarnost	unitarity
valčna transformacija	wavelet transform
večlična funkcija	ambiguity function
vzorčenje po diadični skali	diadic sampling
Wigner–Villejeva predstavitev	Wigner–Ville distribution
zamik po času	delay

---

---

<b>zamik po frekvenci</b>	doppler
<b>zglajena psevdno afina Wigener–Villejeva predstavitev</b>	smoothed pseudo affine Wigener–Ville distribution
<b>zglajena psevdno Wigener–Villejeva predstavitev</b>	smoothed pseudo Wigener–Ville distribution
<b>Zhao–Atlas–Marksova predstavitev</b>	Zhao–Atlas–Marks distribution, cone-shaped kernel distribution
<b>zvezna valčna transformacija</b>	continuous wavelet transform





# Izjava

Podpisani Janez Žibert izjavljam, da sem avtor magistrskega dela z naslovom  
*Časovno-frekvenčne predstavitve govornih signalov.*

Janez Žibert