



UNIVERSITETI POLITEKNIK I TIRANËS
FAKULTETI INXHINIERISË MATEMATIKE DHE INXHINIERISË FIZIKE
DEPARTAMENTI I INXHINIERISË MATEMATIKE

Doktoratë

Një vlerësim i performancës së Bankës së Filtrave Mel dhe Gamaton përkundrejt FFT

Punoi :

Saimir TOLA

Udhëheqësi Shkencor:

Prof. Asoc. Ligor NIKOLLA

TIRANË, 2016

METODAT E ANALIZËS SPEKTRALE PËR ASR

Mbrojtur me datë : përpara jurisë :

- | | |
|---|---------|
| 1.Prof. Dr. Agron TATO | Kryetar |
| 2.Prof. Asoc. Shkëlqim KUKA | Anëtar |
| 3.Prof. Dr. Fatmir HOXHA | Anëtar |
| 4.Prof. Dr. Xhezair TELITI | Anëtar |
| 5.Prof. Asoc. Eglantina KALLUÇI (XHAJA) | Anëtar |

METODAT E ANALIZËS SPEKTRALE PËR ASR

Përmbajtja

Kapitulli 1

1.1 Hyrje.....	4
1.2 Rishikimi i Literaturës	5
1.3 Përpunimi/Proçesimi i zërit	8
1.4 Formacioni Spektral	11
1.5 Analiza Spektrale	14
1.6 Frekuenca bazë	16

Kapitulli 2

2.1 Koefiçentët Kepstral	16
2.2 Koefiçentë e Parashikimit Linear LP	27
2.3 Amplituda e bankave të filtrave të derivuar nga LP.....	28
2.4 Koefiçentët Kepstral LP	29

Kapitulli 3

3.1 Banka e Filtrave Dixhitale	32
3.2 Banka e Filtrit Mel	35
3.3 Koefiçentët Kepstral të filtrave Mel.....	42
3.4 Banka e Filtrit Gamaton	47
3.5 Banka e filtrit Gamachirp	51

Kapitulli 4

4.1 Baza e të dhënave Aurora.....	54
4.2 Transformimet Furie	54

METODAT E ANALIZËS SPEKTRALE PËR ASR

4.3 Transformimi Diskrete Furie	55
4.4 Transformimi i Shpejtë Furie	56
4.4.1 Shpejtësia	57
4.5 Transformimet Furie Sparse (i Rrallë)	59
4.6 Transformet Wavelet (Valore).....	60

Kapitulli 5

5.1 Përmirësimi i sistemeve ASR	62
5.2 Aplikacionet dhe mjetet ASR.....	79
5.3 Konkluzionet.....	87
5.4 Referencat	89

Kapitulli 1

1.1 Hyrje

Shumë besojnë që aftësia e njeriut për të folur shënon dhe agimin e civilizimit njerëzor. Aplikacionet e Njohjes Automatike të zërit (ASR)¹, përdoren për të transformuar të folurin në tekst. Në 50 vitet e fundit shkalla e inovacionit teknologjik në këtë drejtim është zhvilluar jashtë mase. Sistemet e njohjes së të folurit përdoren në një sërë aplikacionesh si psh: Direct Voice Input (DVI) në mjetet ajore; Voice Users Interfaces (VUI) në telefonatat e komanduara me zë, apo në këmbimin e telefonatave; Domotic Appliance Control (DAC) futje të dhënash; krijim raportesh; dhe procesim nga të folurit në tekst. Nëpërmjet algoritmeve Big Data dhe GPU-ve të cilat mund të procesojnë detyra shumë intensive matematikore, sot jemi në pararojë për sa i përket saktësisë në performancën e këtyre sistemeve.

Ky punim analizon zhvillimin e sistemeve të Njohjes Automatike të Zërit kundrejt metodave të ndryshme të analizës spektrale, duke u bazuar në një vlerësim kritik të performancës së bankave të filtrave Mel dhe Gamatone² përkundrejt FFT.

Kontributi akademik i këtij projekti, lidhet me origjinalitetin e rekomandimeve, pasi ato bazohen në një sërë të dhënash eksperimentale. Kërkimi shkencor bazohet në kuptimin e efektivitetit të sistemeve ekzistuese të Njohjes

¹ Që tani e tutje, ASR do t'i referohemi si sistemet e njohjes së zërit pasi ky term, edhe pse jo shumë korrekt nga ana e konceptit përdoret gjerësisht për t'iu referuar këtyre aplikacioneve.

² Banka Filtrash spektrale

METODAT E ANALIZËS SPEKTRALE PËR ASR

Automatike të Zërit (ASR), dhe fokusohet në mënyrë specifike në performancën e bankave të filtrave MEL dhe Gamatone duke i krahasuar ato me vlerat spektrale të FFT.

Proçesi i njohjes së zërit përfishin ndarjen e sinjalit audio nga të dhëna të tjera që konsiderohen të panevojshme për proçesin, si zhurma në sfond, dhe luhatjet emocionale të tonit të zërit. Në të folurën njerëzore, forma e aparatit të së folurit duke përfshirë këtu dhe gjuhën dhe dhëmbët, ndikon në tonin dhe teksturën e sinjalit audio që emetohet, në formën e zërit. Forma e aparatit folës ose ndryshe kutia e zërit, përfaqësohet në shtresën e spektrit të fuqisë afatshkurtër dhe sistemet ASR përdoren për të paraqitur këtë shtresë në mënyrë sa më korrekte.

1.2 Rishikimi i literaturës/Konceptet themelore

Proçesi i njohjes së zërit fillon me parametrizimin e sinjalit analog. Sinjali analog, është një sinjal i vazhdueshëm që përmban variabla sasiore që ndryshojnë me kohën .

Metodat dhe algoritmat e njohjes së zërit kanë evoluar me kalimin e kohës. Algoritmat e sotëm kanë parametra që imitojnë sjelljen njerëzore dhe proçeset audio dhe janë zhvilluar me objektivin për të maksimizuar performancën e njohjes. Njohja e dikurshme e zërit mbështetej në teknologji të varura nga mikrofonët.

Me kalimin e kohës, objektivi i kësaj shkence ishte që të bëheshim të pavarur nga mikrofonët. Gjithsesi, parametrizimi i teknologjive të pavarura nga mikrofonët bazohej në teknologjitë e shkuara edhe pasi ndodhi tranzicioni. Për sa i përket parametrave, ndryshimi thelbësor ishte të dallonte nevojat e dy formave të ASR. Së pari, duhej të dallohej nëse ishin të pavarur nga mikrofoni dhe pastaj dizenjoheshin për shkrime gjuhësore që nuk ndryshojnë pavarësisht krijuesit të sekuencës gjuhësore. Për këtë arsye, parametrat përfaqësues të zërit të folësit janë më pak të pëlqyeshme se energjitë spektrale principale të tingullit (Picone, 1993). Sistemet (ASR)³ janë zhvilluar shumë që

³ Automatic Speech Recognition

METODAT E ANALIZËS SPEKTRALE PËR ASR

pas aplikimit të modeleve akustike të bazuara në Rrjetat e Neuroneve apo DNN-AM⁴ (Hinton et al 2012)

Me energji spektrale do të kuptojmë :

$$E_s = (x(t), x(t)) = \int_{-\infty}^{+\infty} |x(t)|^2 dt \quad (1.1) \quad \text{për rastin e sinjalit të vazhdueshëm në kohë}$$

$$E_s = (x(n), x(n)) = \sum_{-\infty}^{+\infty} |x(n)|^2 \quad (1.2) \quad \text{për rastin e sinjalit me kohë diskrete}$$

Sistemet ASR mund të klasifikohen sipas fjalorit të njohjes apo masës së leksikut (Whitaker, 2001). ASR i klasifikojmë ASR të vogla deri në një 1000 fjalë; sisteme të mesme deri në 10K fjalë; sisteme të mëdha deri në 100K, dhe sisteme premium me një fjalor mbi 100,000 fjalë të cilat janë të përshtatshme për ASR të gjuhëve me shumë fjalë të parregullta dhe me tendencë të ngjitjes së fjalëve si Turqishtja; për shembull kemi ASR 800 K për Arabishten. Ka gjithashtu dhe sisteme me fjalor të pakufizuar të cilat përpiqen të modelojnë të gjitha fjalët potenciale të një gjuhe (Besacier et al, 2014).

ASR dizenjohen duke përdorur një sërë teknikash. Si përsëmbull: Modelet e Fshehura Markov (HMM) (Young 2008), Përthyerja Dinamike Kohore (DTW) ose Programimi Dinamik (Jing2010), Rrjetet Dinamike Bayesiane (DBN) (Stephenson, 2002), Makinat Vektoriale Suportuese (SVM) (Solera-Urena, 2007). Gjithashtu gjejmë përdorim dhe modelet hibride të ASR të cilat do të jenë në fokus të kësaj teme. (Trentin 2001). Gjithashtu Rrjetet Artificiale të Neuroneve (ANN) gjejmë përdorim në ASR për të kryer nëndetyra në lidhje me objektivin final. Kemi ANN që përfshijnë një shtresë të fshehur individuale ose dhe rrjeta të shumëfishta të fshehura dhe kanë nënfunksion në ASR si p.sh. modelimin akustik dhe modelimin e gjuhës (Arisoy 2012).

Ashtu siç theksohet dhe nga Young, sistemi standard ASR dizenjohet me një arkitekturë të tillë që përdor modelin stokastik HMM të prezantuar në figurën 1. Ky model ka të integruar tre elementë kryesorë (2008): Modelimin Akustik (akustik-fonetik) , Modelimin Leksikor (shqiptimin leksikun/fjalorin) dhe Modelimin e Gjuhës.

⁴ Deep Neural Network Acoustic Model

METODAT E ANALIZËS SPEKTRALE PËR ASR

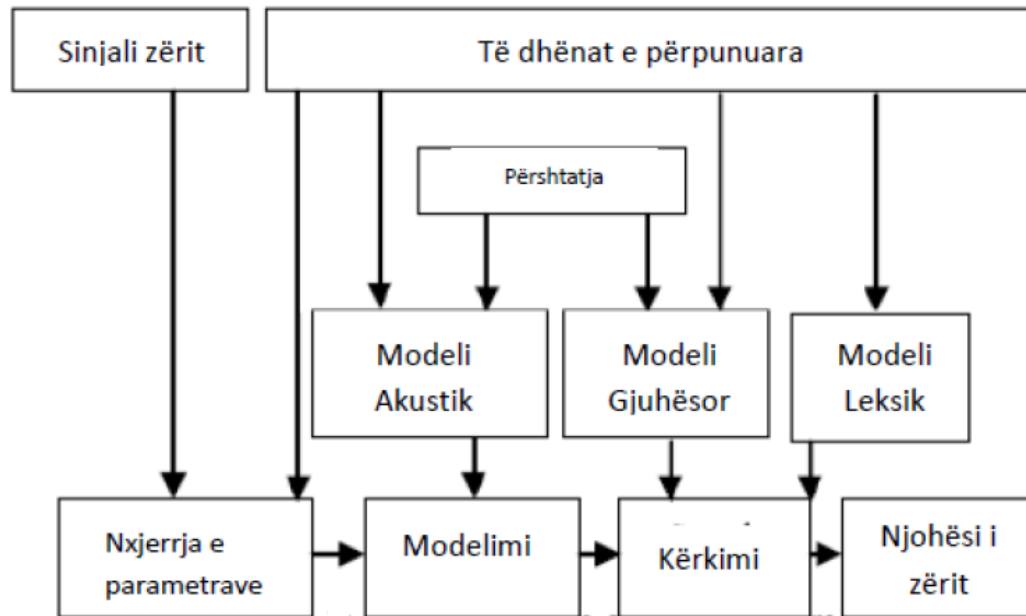


Fig. 1: Modulet që përbëjnë një sistem ASR (Sharma & Sharma, 2013)

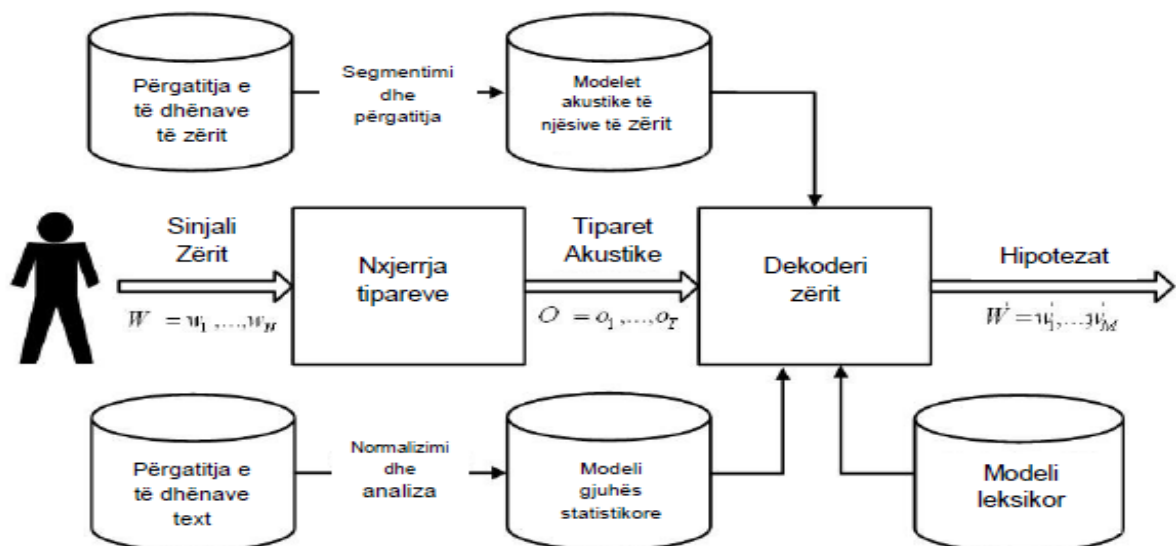


Fig. 2: Arkitektura e sitemeve ASR (Bresacier et al, 2014)

Modelimi akustik na mundëson përkthimin e sinjaleve audio duke diskriminuar klasat e njësive bazë të zërit që janë të pavarura nga konteksti i fjalisë si për shembull, fjalët me një germë, rrokje, ose që varen nga konteksti si alofone⁵, trifone⁶, pentafove. Një tjetër funksion është përcaktimi i ndryshueshmërisë relevante të zërit sipas mikrofonëve, kanaleve dhe mjedisit.

Vektorët e tipareve të sinjaleve të zërit (p.sh., melfrekuenca, koeficientët keprstral (MFCC), koeficientët e parashikimit linear (LPC), Koeficientët e Parashikimit Linear të Perceptimit (PLP), tiparet e ngushticave (ML) etj), të gjitha këto nxirren nga sinjali akustik për reduktim dimensional dhe modelim probabilitar (Besacier etj 2014)

1.3 Përpunimi i Zërit

Zemra e Sistemeve Njohës të zërit është njohja e modelit. Njohja e zërit kur shikohet sipas njohjes së modelit, përbëhet nga dy metoda; Modelimi i Sinjalit dhe Përpunimi i Rrjetave. Në Modelimin e Sinjalit, sekuenca të provave të folurës asimilohen në vektore organizues duke përdorur hapësirën probabilitare për të përshkruar një ngjarje. Përpunimi në Rrjet, kontrollon gjetjen e sekuençës me probabilitetin më të madh për këto ngjarje, duke marrë parasysh kufizimet e mundshme sintaksore.

Sistemi i modelimit të sinjalit formohet nga katër nën-ndarje : dhënia e formës spektrale, analiza spektrale, transformimi parametrik dhe modelimi statistikor. Nga këto nëndarje, tre të parat mund të shikohen si paraprirëse të nën-ndarjes katër, e cila është më komplekse për nga natyra. Vetë modelimi statistikor ndahet në Sistemin e Modelimit Statistikor dhe Sistemit të njohjes së zërit (Picone, 1993).

⁵ Disa mënyra të ndryshme për të shqiptuar një germë/fonemë

⁶ Sekuencë tre fonemash. Trefonet janë shumë të dobishme në modelet e procesimit të gjuhës natyrale , pasi mund të përcaktojnë kontekstet e ndryshme në të cilën një fonemë mund të shfaqet në një gjuhë të veçantë

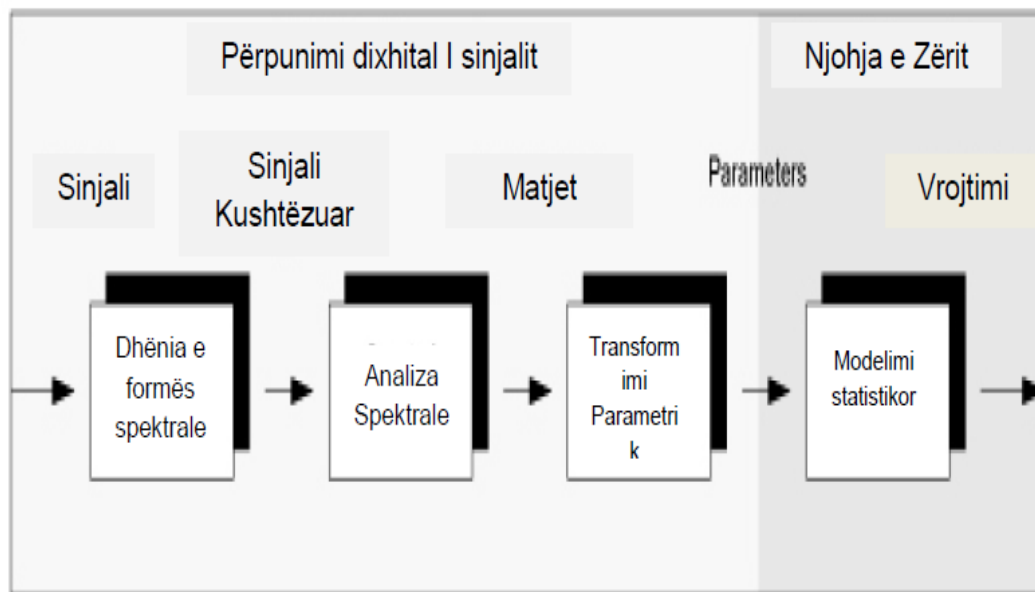


Fig. 3: Proçesi i Njohjes së Zërit (Picone, 1993)

Le të përcaktojmë tre forcat kryesore që përcaktojnë dizejimin dhe zhvillimin e sistemeve të Modelimit të Sinjalit. Së pari, përcaktimi i parametrave që përfaqësojnë aspekte kryesore të sinjalit folës që po kërkojmë. Preferohen ato parametra që janë analogë me ato të përdorura nga sistemi njerëzor i të dëgjuarit. Parametrat që imitojnë sistemet e të dëgjuarit quhen parametra perceptual-kuptimplotë. Pastaj, parametrat që dëshirohen janë ato që janë plastikë me luhatjet në kanale, folës dhe transduktor. Natyra e plasticitetit quhet problemi i fortësisë apo i mosvariancës. Parametrat që kapin dinamikën spektrale apo variancat në spektër me kohën u bënë më të pëlqyerat. Ky u quajt si problem i korrelacionit temporal (kohor).

Teknikat e modelimit Markov, futet si një sistem me fuqinë për të modeluar nga ana statistikore rrjedhën kohore të sinjalit duke përdorur parametra që përfshinin si matje absolute dhe diferenciale të spektrit të sinjalit. Modelet Markov, janë shëndërruar në një standard për industrinë. Modelimi i sinjalit tashmë kërkon shumë më pak kohë përpunimi në një aplikim tipik të madh të përpunimit të së folurit.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Zhvillimi teknologjik si Njësia e Përpunimit Grafik (GPU) kanë ndryshuar potencialin e përdorimit të modeleve Markov. GPU përdor procesorë me një çip të vetëm që bëjnë të mundur plotësimin e detyrave intensive matematikore që do ta tejkalonin një CPU të zakonshme. Tashmë, koha e përpunimit nëpërmjet metodave të ndryshme të modelimit të sinjalit zë një përqindje të neglizhueshme të kohës totale të përpunimit. Fokusi është transferuar drejt mbajtjes së performancës së lartë dhe përmirësimit të potencialeve të sistemit në të gjithë gamën e aplikimeve të mundshme të ASR.

Vendosja e Parametrave që përcaktojnë qartë sinjalin mund të llogaritet thjesht në hardware fikse dhe mund të përpunohen nëpërmjet teknikave të kuantizimit⁷ të menjëhershëm. Konsideratat në lidhje me memorien përgjithësisht tejkalojnë çdo përfitim të vogël që mund të merret nga përdorimi i një modeli të ri sinjalesh. Kur marrim në konsideratë, zhvillimin historik të ASR, plastika në lidhje me zhurmën në sfond ka qënë një forcë madhore në përcaktimin e dizenjës së modeleve të sinjaleve.

Kështu që, shumë nga modelet e sinjaleve që përdoren sot, janë rezultat i kërkimeve që përfshinin zhurmat e ambjentit. Për shembull, i tillë është kontrolli zanor i instrumentave ushtarake që përfshijnë njohjen e zërit në kabinën e pilotit, dhe kontrollin zanor të telefonit duke përfshirë transaksionet automatike të telefonit.

Me rritjen e sofistikimit dhe elegancës së ASR Sistemi i njohjes në vetvete tani kontribon më shumë në problemin e fuqisë së zhurmës në modelin e sinjalit. Për këtë arsye, kemi veshtirësi në izolimin e përmirësimeve të algoritmit të modelimit të sinjalit. Për më tepër, modelet e sinjaleve që mund të jenë shumë të mira për një aplikim mund të mos jenë aq të mira për një tjetër. Për shembull, në njohjen e pavarur të zërit, që është e përpunuar për një mjedis të caktuar, si leximi i vazhdueshëm dixhital për aplikacionet në telekomunikacion, metodat specifika të variancës statistikore të kanalit dhe të mikrofonëve mund të parashikohen me siguri të lartë dhe llogariten si gjerësia valore e kanalit.

⁷ Kuantizim, i referohet procesit të kufizimit të vlerave të munshme (madhësisë/sasisë) në një bashkësi diskrete vlerash, nëpërmjet rregullave të mekanikës kuantike.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Ndërkohë një strukturë tipike e pavarur ASR ka nevojë për një minimum prej 50 mikrofonësh (Barnard,2009), por transmetimet radio apo regjistrimet në forumet e biznesit mund të dominohen nga dymbëdhjetë apo më pak mikrofonë.

Në aplikimet identifikues me apo pa mikrofon, të mësuarit e karakteristikave unike të përdoruesit dhe të akustikës së mjedisit të përdoruesit është vendimtare. Edhe pse duket sikur ky dallim kërkon metoda të ndryshme të modelimit të sinjalit, shumica e atyre që kemi diskutuar në këtë punim punojnë mjaftueshëm mirë në të dy tipet e aplikacioneve.

1.4 Formacioni Spektral

Formacioni spektral është sistemi që konverton sinjalin akustik analog në një sinjal dixhital duke përdorur transformimin A/D. Shpesh herë ky sistem përdor dhe filtra, duke identifikuar kështu dhe komponentë të rëndësishëm të frekuencës.

Formacioni spektral përbëhet nga dy operacione bazë. I pari është përkthimi A/D, që në thelb përkthen sinjalin që emetohet nga presioni akustik në një sinjal dixhital; dhe Filtrimi Dixhital. Filtrimi Dixhital fokusohet në komponentët thelbësor që përbëjnë sinjalin akustik. Ashtu sic theksohet në punimet e Picone , në vend që të merremi me zgjedhjen e një frekuence provë dhe të studiojmë implikimet që ajo sjell; zgjedhja si provë e frekuencës së përshtatshme luan një rol kyç në modelimin e sinjalit.

Mikrofoni i përdorur në procesin e përkthimit A/D zakonisht ka dhe efekte anësore të padëshirueshme si zhurmat e frekuencës së linjës; humbja e informacionit me frekuencë të lartë dhe të ulët, dhe shtrëmbërimi jolinear. Konvertuesi A/D sjell shtrëmbërimet e veta në formën e përgjigjes me një frekuencë jo të njëjtë me atë në hyrje dhe funksionin e transferimit jo linear hyrje/dalje, dhe variablin e shmangies DC (1993). Në mënyrë që të konvertojmë një sinjal analog akustik në një sinjal dixhital, si fillim duhet të filtrojmë sinjalin në mënyrë që të reduktojmë informacionet mbi frekuencën. Më pas sinjali mund të kampionohet.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Në teorinë e sinjalit me kampionim do të kuptojmë reduktimin e një sinjali nga të vazhdueshëm në diskret. Një shembull i përgjithshëm është konvertimi i një sinjali valor (një sinjal i vazhdueshëm) në një sekuencë kampionësh (sinjali diskret) figura A.

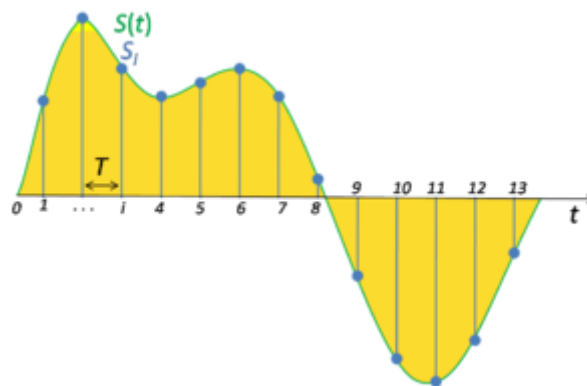


Figura A. Konvertimi i një sinjali valorë në një sekuencë kampionësh

Një kampion është një vlerë ose një set vlerash në një pikë në një kohë apo hapësirë të caktuar. Një kampionues është një proces i cili ekstrakt/nxjerr kampione nga një sinjal i vazhduar.

Në teori, një frekuencë përfaqësuese prej Y Hz, kërkon një frekuencë kampionimi prej $2Y$ Hz. Veshi mund të perceptojë një bandë dinamike (amplitudë) prej 20 bit. Zakonisht çdo kampion ruhet në 16 bit, ndërsa për bisedat telefonike përdoren kampione nga 8-12 bit. Intensiteti i zërit nuk shpërndahet njësoj në rangun e frekuencave. Ai bie zakonisht me 6dB për oktavë. Shpesh sinjali përforcohet për të kompensuar këtë dukuri (Rosell, 2006).

Dobësia e shpejtë e frekuencave të larta dhe të ulëta shpesh shkakton probleme për algoritmat e analizës parametrike spektrale të vijën më pas. Për shkak të kufizimeve të frekuencave në përgjigje në kanalet e telekomunikacionit dhe përdorimin e gjërë të kampioneve 8kHz në telefoninë dixhitale. Frekuenca më e zakonshme e kampionimit për sinjalin e komunikimit telefonik është 8kHz. Në vitet '90, me zhvillimin e rrjeteve dixhitale me bandë të gjërë, lindën aplikacione telekomunikimi që përdornin hyrje audio të një cilësie më të lartë dhe jepnin qartësi më të lartë në frekuencë dhe më të shpejtë në kohë.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Qëllimi kryesor i procesit të dixhitalizimit është të prodhojë një përfaqësi kampionësh të sinjalit zanor me një normë të lartë SNR (raporti Sinjal/Zhurmë). Sistemet e telekomunikacionit sot japin vazhdimisht shkallë më të larta të SNR-ve në aplikacionet e njohjes së zërit, të cilat janë më se të mjaftueshme për të arritur një performancë të lartë. Është shumë e rëndësishme të monitorohen luhatjet në transduktorë⁸, kanale dhe zhurmën e ambjentit; pasi secila mund të ndikojë si problematikë në performancë. Pasi konvertimi i sinjalit përfundon, hapi i fundit i post filtrimit përfundon me ekzekutimin e filtrit Finite Impulse Response (FIR). Zakonisht përdoret dhe një filter dixhital me koeficient 1 që njihet si filtri Pree-emphasis.

Dy avantazhet më të njohura në përdorimin e këtij filtri janë: së pari, seksionet audio të sinjaleve të zërit zakonisht kanë një gradient spektral negativ ose dobësim afër 20dB për dekadë, për shkak të sistemit njerëzor të prodhimit të zërit. Filtri Pree-emphasis shërben për të mënjeluar gradientit natyral para Analizës Spektrale gjë që përmirëson dhe efikasitetin e saj. Avantazhi i dytë vjen nga fakti që të dëgjuarit njerëzor, është më i ndjeshëm për zonën mbi spektrin 1kHz. Kjo është zona e spektrit që përforcohet nga filtri Pree-emphasis. Kështu, filtri lejon algoritmin e analizës spektrale të modelojë aspektet më të rëndësishme nga ana perceptive të zërit.

Filtri pree-emphasis është një filtër i cili qëllimisht modifikon karakteristikat e amplitudës përkundrejt frekuencës së sinjalit për të reduktuar efektet negative të zhurmave. Komponentet me frekuencë të lartë të sinjalit ‘theksohen’ për të prodhuar një index modulimi më të barabartë për frekuencën e spektrit të transmetuar.

$$\text{Filtri Pree-emphasis} \quad y(n) = ax(n) - bx(n-1) \quad (1.2)$$

$$\text{Minimum dhe Maximumi në dalje} : V_{\min}=(a-b)V_{\text{kulm}} \quad V_{\max}=(a+b)V_{\text{kulm}} \quad (1.3)$$

Përkufizimi i filtrit Pre-emphasis :

$$\text{pre - emphasi} = \left(1 - \frac{V_{\min}}{V_{\max}}\right) = \left(1 - \frac{(a-b)}{(a+b)}\right) = \frac{2}{1 + \frac{a}{b}} \quad (1.4)$$

⁸ Paisje që konvertojnë sinjale fizike në sinjalë elektrikë dhe e kundërta.

METODAT E ANALIZËS SPEKTRALE PËR ASR

$$\text{Përgjigja e frekuencës në } \pi : H(\theta) = a - b \exp(-j\theta) \rightarrow H(\pi) = a + b \quad (1.5)$$

Ku a dhe b janë koeficientët e filtrit :

$$a = 1 - \frac{1}{2} \text{pre-emphasis} \quad \text{dhe} \quad b = \frac{1}{2} \text{pre-emphasis} \quad (1.6)$$

Gjithashtu, studimet kanë treguar që filtri Pree-emphasis gjithashtu rrit frekuencat mbi 5kHz, një zonë në të cilën sistemi dëgjimore bëhet shumë i pandjeshëm. Sidoqoftë, frekuencat mbi 5kHz dobësohen në mënyrë të natyrshme nga sistemi i prodhimit të zërit dhe zakonisht u jepet një peshë më e vogël në sistemet tipike të njohjes së zërit (Piccone 1993).

Algoritmat Pree-emphasis që pasuan më tej janë shumë më kompleksë. Një algoritëm i rëndësishëm është ‘Adaptive Pree-emphasis’, në të cilën gradienti spektral zbutet automatikisht përpara Analizës Spektrale. Kjo metodë përdor algoritma të cilët përdorin filtra formues të cilët reduktojnë ato zona të zërit që kanë shumë zhurma. Në vitet 90' klasifikimi i algoritmave e folura/zhurmë bazohej në faktin si përdoreshin filtrat adaptues. Sidoqoftë, asnjë nga këto metoda nuk kanë pasur sukses të gjërë në Aplikacionet e Njohjes së Zërit. Në fakt, kohë e fundit, këto lloj aplikacionesh e kanë eliminuar komplet fazën e Pre-emphasis dhe e kanë zëvendësuar me gradientin spektral si pjesë të modelit statistikor të njohjes e zërit.

1.5 Analiza spektrale

Analiza spektrale përfshin analizën e spektrit me qëllim kapjen e aspekteve të spikatura të sinjalit zanor/tingull. Algoritmat e Analizës Spektrale të përdorura në sistemet e sotme të njohjes së zërit, përfshijnë gjashtë segmente kryesore. Figura 1 tregon teknikat për gjenerimin e këtyre algoritmave. Proçeset e implementuara në qarqet analoge janë banka e parë e filtrave që u përdor si fillim. Metodot e parashikimit linear u prezantuan në 1970, dhe ishte teknika më e preferuar në vitet 80'. Tani, teknikat e parashikimit dhe Transformimit Linear Furie, përdoren gjërësisht dhe përdoren në aplikime të shumta të përpunimit të zërit.

Për të analizuar spektrin e sinjalit kampion ndahet në fraksione ose intervale të shkurtra

METODAT E ANALIZËS SPEKTRALE PËR ASR

kohore. Sidoqoftë, është e nevojshme që fraksionet e sinjalit të jenë mjaftueshëm të gjatë që të përmbajnë të paktën një cikël të plotë të frekuencës më të ulët për të cilën janë të interesuar përdoruesit. Gjatësia normale e fraksionit të sinjalit është 20-25ms, që përshijnë disa perioda të glottisit⁹. Zakonisht sekuencat janë të tilla që mbivendosen kështu që qendra të shtrihet deri në 10 ms distancë. Çdo sekuencë shumëzohet me një dritare konike kështu që vlerat afër kufijve të bëhen zero. Kjo parandalon ndërprerje në skaje që mund të ndikojnë rezultatin e përpunimit të mëtejshëm.

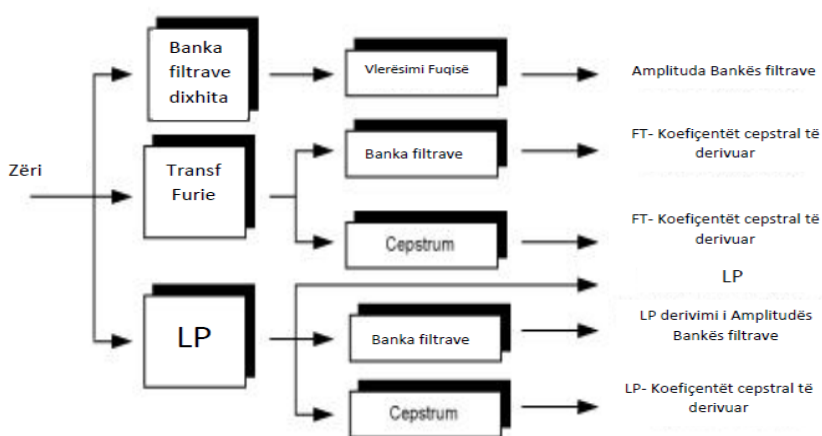


Fig. 4: Teknikat e gjenerimit të algoritmeve të Analizës Spektrale (Piccone 1993)

Sinjali i zërit përfaqësohet nga një vektor tipar/karakteristik për secilën sekuencë. Teknikat për ekstraktimin e këtyre vektorëve përshkruhen në seksionet e mëposhtme, dhe mund të kombinohen në mënyra të ndryshme (2006). Sipas Piccone, tipet e masave spektrale të përdorura në sistemet e njohjes së zërit ndahen në dy klasa: fuqisë që mat fuqinë spektrale bruto apo temporale të sinjalit; dhe amplituda spektrale që mat fuqinë mbi intervale të vecanta frekuence në spektër. Një parametër i zakonshëm i krijuar për njohjen e zërit do të përfshijë secilën nga këto matje.

Në vitet '90 pati një interes në rritje në frekuencën themelore për përdorim si tipar ritmik (prosodic), për përdorim në njohjen e gjuhëve intonative (p.sh. Kineze) ose gjuhëve që kanë disa komponentë intonativë (p.sh. Japoneze), dhe si matje të identitetit

⁹ the part of the larynx consisting of the vocal cords and the slit-like opening between them. It affects voice modulation through expansion or contraction

METODAT E ANALIZËS SPEKTRALE PËR ASR

apo autencitetit të folësit. Le t'i shikojmë shkurtimisht llogaritjet e frekuencës bazë dhe të fuqisë, e pastaj të fokusohemi në vlerësimin e amplitudës spektrale.

1.6 Frekuenca bazë

Frekuenca bazë përcaktohet si frekuenca në të cilën kordat e zërit dridhen gjatë kohës kur njeriu flet. Frekuenca bazë (F_0) për shumë kohë ka qënë një parameter që merr rrisht të lartë për të vlerësuar në mënyrë të besueshme sinjalin e zërit. Në të shkuarën, sistemet e njohjes së zërit e kanë neglizhuar këtë parametër për disa arsye, të cilat përfshijnë: barrën e madhe të llogaritjeve për të marrë një vlerësim të saktë. Me procesorët GPU sot kjo pjesë nuk përbën më problem. Ka pasur një shqetësim në të shkuarën që vlerësimi i pabesueshëm mund të ishte një pengesë për të arritur performancë të lartë si dhe vështirësitë për të karakterizuar ndërveprimet kompleksë ndërmjet F_0 dhe fenomenit të tipareve të komponentëve vokale siç janë intonacioni apo theksimi i fjalëve (Piccone 1993).

Një spektër dy dimensional është në thelb një fotografi e shpejtë e spektrit të zërit në një moment kohor. Ky “moment” kohor është gjithmonë një dritare me gjatësi më të madhe sesa një kampion në të folurin e dixhitalizuar dhe ndodhet në qendër mbi pikën e analizës, në shumicën e rasteve një spektër dy dimensional akustik do të paraqesë amplitudën në aksin vertikal (Y) dhe frekuencën në aksin horizontal (X). Shumicën e rasteve aksi i amplitudës do të jetë në deciBel (dB). Aksi i frekuencës do të jetë në Hertz (Hz) ose kilohertz (kHz). Spektri në shumicën e rasteve është një analizë FFT apo LPC.

Kapitulli 2

2.1 Koeficientët Kepstral

Koeficientët kepralë u përdorën për herë të parë në fillim të viteve '70. Një koeficient kepral është një lloj operacioni mbivendosës që i referohet shumëzimit, funksionit eksponencial dhe shumës. Një CC (koeficient kepral) është një metodë homomorfeke e përpunimit të sinjalit dhe i përket bashkësisë së proceseve jo lineare të ndërtuara mbi parimin e përgjithshëm të mbivendosjes. Bankat e filtrave, dhe teknikat e transformimit Furie që diskutuam më sipër janë sisteme lineare.

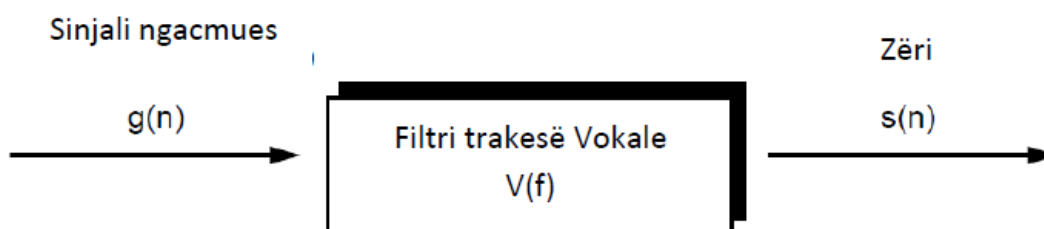


Fig. 4: Motivimi për përpunimin Homomorfik (Piccone, 1993)

Termi matematikor “cepstrum” u ligjërua për herë të parë në vitin 1963 në artikullin e Bogert, Healey dhe Tukey dhe përdoret për të përshkruar një transformim Furie të madhësisë logaritmike të spektrit (IEEE). Sipas Piccone, sistemet homomorfeke janë konsideruar si të dobishme për përpunimin e zërit pasi ata ofrojnë një metodologji për të ndarë sinjalin dalës nga forma e traktit vocal. Njohja automatike e zërit është ndërtuar me qëllim përdorimin e një sërë teknikash për të modeluar formën e traktit vokal apo karakteristikat e tij. Spektri i përbërë audio i të folurës njerëzore ashtu sic matet nga transformimi Furie përfshin sinjalin dalës që filtrohet nga një filtër linear që përfaqëson formën e traktit vocal dhe varion me kohën. Ky proces ndarës përshkruhet si dekompozim¹⁰. Kështu që në teori, sinjali dalës dhe forma e traktit vokal janë të

¹⁰ simplification of a complex signal (as instrumental data) usually by removal of instrument noise

METODAT E ANALIZËS SPEKTRALE PËR ASR

mbivendosura dhe mund të ndahen duke përdorur teknika standarte të përpunimit të sinjalit .

Hapi i parë në llogaritjen e Kepstrumit është llogaritja e madhësisë spektrale log. Pastaj përdoret një transformim Furie në spektrin log. Barazimi i mëposhtëm përshkruan këtë proces, ku $c(n)$ përcaktohet si kepstrum.

$$C(n) = \frac{1}{N_s} \sum_{k=0}^{N_s-1} \log_{10} |S_{avg}(k)| e^{j \frac{2\pi}{N_s} kn}, \quad 0 \leq n \leq N_s - 1 \quad (2.1)$$

Ekuacioni i mësipërm mund të identifikohet si transformimi Invers Diskret Fourier për spektrin log. Me premisën që spektri madhësisë log është një funksion me të vërtetë i vazhdueshëm, ky barazim mund të thjeshtohet:

$$C(n) = \frac{2}{N_s} \sum_{k=1}^{N_s} \frac{s_{avg}(I(k)) \cos 2\pi kn}{N_s} \quad (2.2)$$

Sipas Piccone, ky përkufizim mund të modifikohet më mire për të qënë një kepstrum në hapësirën Mel. Kjo arrihet duke përdorur Transformimin Furie në frekuenca të pozicionuara në mënyrë të përshtatshme. Duke qënë se kepstrum llogaritet me metoda dhe algoritme jolinearë; duket sikur është i ndjeshëm ndaj zhurmave të ambientit dhe shtrëmbërimeve të sinjalit. Për këtë arsye, parametrat kepral që arrihen nëpërmjet vlerësimeve spektrale me rezolucion të lartë konsiderohen si më të mirë kur përpunojmë zë me zhurma të ambientit.

2.2 Koeficientët e Parashikimit Linear (LPC)

Koeficientët e parashikimit linear janë një metodë që synojnë të modelojnë spektrin e një sinjali zanor. Ky është një proces autoregresiv që themelohet mbi premisën që vlerat e shkuara kanë impakt në vlerat e ardhme. Sipas Piccone, modelet parametrike si LPC, kanë impakt në përpunimin e zërit që prej viteve '70. Ai tregon që në fund të viteve '70 algoritmat që i përshtateshin në parametra spektrit u përdorën në pothuajse çdo sistem

METODAT E ANALIZËS SPEKTRALE PËR ASR

përpunimi të zërit. Këto modele u përdorën në aplikime të njohjes, kompresimit apo dhe të verifikimit. Popullariteti i LPC ka rënë në aplikacionet e njohjes. Përpara se të përdroreshin koeficientët Kepstral të Frekuencës Mel (MFCC), LPC dhe koeficientët Kepstralë të parashikimit linear (LPCC) ishin dhe përbërësit kryesorë për sistemet ASR. Sipas (Atal 1974), LPCC është një teknikë klasike ekstraktimi për tiparet akustike. Teknika si LPC motivuan evolucionin në teknika të fuqishme të modelimit statistikor si MFCC. Modelet e tipit LCP përdorin modelimin parametrik të bazuara në teknikën e parashikimit linear (LP) ose të paktën teorinë e katrorëve mesatarë më të vegjël.

Ka një literaturë shumë të gjërë për LP si p.sh. Marple (1987). Algoritmi bazë për LP që është algoritmi i katrorëve mesatarë më të vegjël gjendet në shumë fusha të ndryshme si analiza e serive kohore për aplikimet financiare; metodat e entropise maksimale; fizika kuantike, gjeofizika; filtrat adaptues dhe vlerësimi spektral në përpunimin e sinjalit (Piccone1993).

Një avantazh i modelit LP është që modelet e sakta janë autoregresive dhe në këtë mënyrë mund të arrijnë parashikimin e vlerave të ardhme të sinjalit nisur nga vlerat e shkuara. Cilësia e modelit të përdorur përcaktohet në bazë të madhësisë së gabimit.

Sipas Piccone, ekzistojnë tre mënyra të ndryshme për të llogaritur modelet LP: metodat e kovariances të bazuara në matricën e kovariancës e cila përcaktohet si metoda e katrorëve më të vegjël; metodat e autokorrelimit; dhe mbështjellja me rrjete ose teknikat harmonike. Në sistemin e ASR teknika më e përdorur është metoda e autokorrelacionit si arsye e efikasitetit në llogaritje dhe në ndërtimin e qëndrueshmërisë dhe lidhjes logjike.

Metoda e autokorrelimit : Kur kemi të bëjmë me sinjale zëri të ‘dritarizuara’ ne duhet të marrim parasysh efektet kufitare në mënyrë për të shmangur gabimet e mëdha të parashikimit në skaje. Kjo metodë mund të rafinojë zonën në të cilën zbatohet metodën e katrorëve më të vegjël duke pasur parasysh se kampionët janë 0 jashtë ‘dritares’ për ta rishikuar $\phi_{i,j}$ si:

METODAT E ANALIZËS SPEKTRALE PËR ASR

$$\phi_{i,j} = \sum_{n=0}^{N-1-(i-j)} S_n S_{n+(i-j)} \quad (2.3)$$

Tani $\phi_{i,j}$ varet vetëm nga $(i-j)$ dhe mund të shkruhet në termat e funksionit autokorrelues $\phi_{i,j} = r_i - j$

$$r_k = \sum_{n=0}^{N-1-k} S_n S_{n+k} \quad (2.4)$$

$$R_n(k) \quad (2.5)$$

Metoda e autokorrelimit në mënyrë të pavarur zhvillon një filtër parashikimi me vlera zero që bazohen në rrethin e brendshëm të planit Z. Autokorrelimi bazohet në supozimin se të gjitha përfitimet jashtë boshtit të kohës së ndërhyrjes së vlerësimit janë konsideruar 0. Për më tepër, një dritare aplikohet me kufizime rreth vlerësimit të kohëzgjatjes specifike. Zakonisht arrihet me një dritare ‘Haming’.

Në procesimin e sinjalit një funksion dritare është një funksion i cili vlerësohet si ‘0’ jashtë disa intervaleve të zgjedhura. Për shembull : Një funksion që është konstant brenda intervalit dhe 0 jashtë tij quhet një ‘dritare’ katrore e cila përshkruan formën e paraqitjes së saj grafike.

Trajta e përgjithshme e formës së dritares Haming është :

$$\omega(n) = \alpha - \beta \cos\left(\frac{2n\pi}{N-1}\right) \quad (2.6)$$

Dritarja Haming u propozua nga Richard W. Hamming. Dritarja është e optimizuar për të minimizuar lobin maksimal anësor më të afërt duke i dhënë një gjatësi prej 1/5 së dritares Han.

Avantazhi i dritares është në sigurimin e një segmenti tranzitor të zbutur brenda modelit parametrik. Teknika e dritares zgjidh problemin e luhatjes së ndryshimeve të shpejta në

METODAT E ANALIZËS SPEKTRALE PËR ASR

audio që janë në ekstremet e dritares gjatë një analize që mbivendoset. Sipas Piccone, autokorrelimi na krijon një mundësi për përsëritjen Levinson-Durbin e cila na lejon llogaritjen e koeficientëve të parashikimit. Ekuacionet e përsëritjes Levinson-Durbin na lejojnë llogaritjen e koeficienteve të parashikimit me proporcionalitet kompleks; dhe për më tepër, LP mund të llogaritet në një përfaqësi komplekse. Autokorrelacioni na siguron një thjeshtëzim për të krahasuar spektrin e sinjalit audio me spektrin e modelit LP. Një model i normalizuar i spektrit të sinjalit të zërit krijohet aty ku koeficientët e parashikimit dhe vlerat e fuqisë së sinjalit janë të pavarura.

Ekzistojnë disa zona të rëndësishme sipas teorisë së LP ose të teorisë së gabimit të katrorëve më të vegjël, e cilat llogariten duke përdorur koeficientët e reflektimit të cilat janë vlera të ndërmjetme. Të cilat rezultojnë të jenë tepër të nevojshme për programet e zonave të ruajtjes dhe kompresimit pasi koeficientët e reflektimit janë të lidhura. Për shembull, kompresimi i koeficientëve të LP në programet e procesimit të zërit mund të ruhen dhe kompresohen në një mënyrë të tillë sipas të cilën maksimumi i kompresimit të të dhënave të sinjalit origjinal të arrihet pa e degraduar cilësinë.

Teknika e LP përfshin iteracione dhe një llogaritje të zgjidhjes për të gjitha rendet e këtij modeli. Në procesimin e sinjalit kjo është e dobishme pasi çdo rend modeli ka një parametër sistemi që është i fiksuar dhe një vlerësim i rendit të modelit është pjesë e punës. Për më tepër ky funksionalitet është një përfitim për programet e sistemeve automatike të njohjes së zërit.

Teknika e LP me rritjen e rendit përmirëson përshtatjen e modelit. Për më tepër, përshtatja e modelit vazhdon të përmirësohet edhe me kohën me rritjen relative të rendit. Siç mund të prezumohet qartësisht, gabimi ekspozon një tendencë për të ulur në mënyrë monotone me rritjen e rendit. Modeli LP ka si qëllim krijimin e një lidhje sa më të mirë midis spektrit të sinjalit të zërit të modelit në lidhje me rendin

METODAT E ANALIZËS SPEKTRALE PËR ASR

Ky funksionalitet tregohet qartësisht në figurën më poshtë. Figura 6 na paraqet një spektër zëri me dy modele të lidhura me LP. Lidhja midis modelit dhe e spektrit të sinjalit të zërit, përmiresohet qartësisht me rritjen korresponduese në rend.

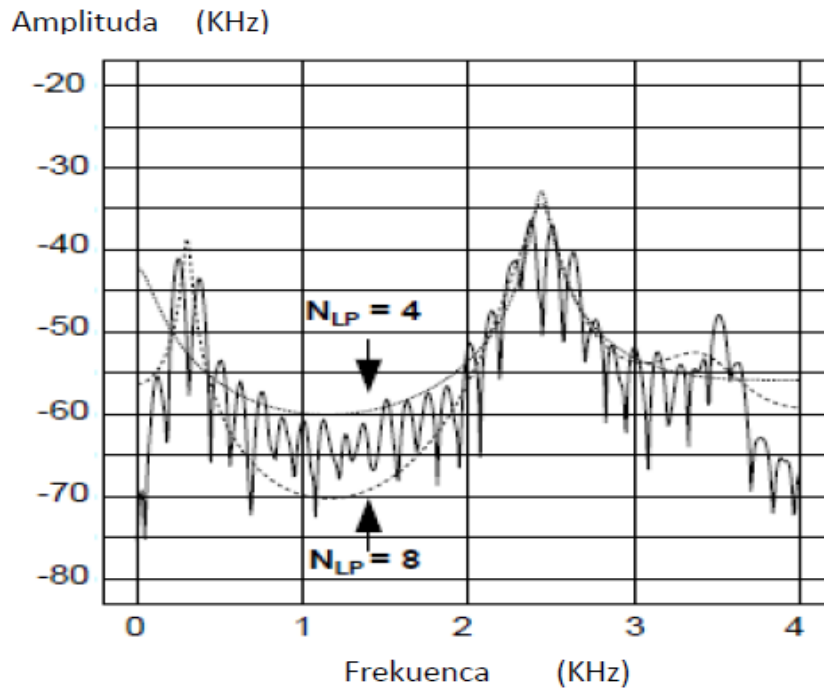


Fig. 6 : Rritja e rendit për përshtatjen e modelit (Piccone,1993)

Për këtë arsye, nevojitet një rend më i lartë për të maksimizuar përfitimet e teknikës LP. Një rend i vogël i reduktuar, thjesht kap formën e spektrit. Një rend i lartë na jep mundësinë të zbulojmë deri detajet më të vogla të spektrit të kapura nga modeli LP. Ashtu sic u diskutua, në pjesët e mësipërme të këtij punimi, modelet e spektrit të bazuara në zonat me energji të ulët janë zakonisht jo të sakta.

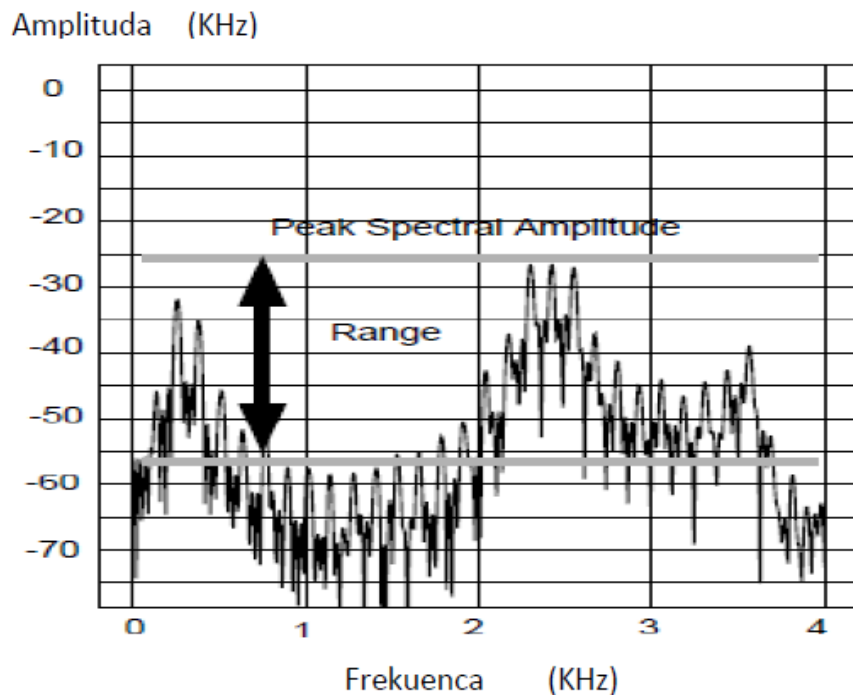


Fig. 7 : Rritja e rendit për përshtatjen e modelit (Piccone, 1993)

Në figurën 7 paraqitet kufiri i një rangu dinamik. Kur përdorim teknikën LP, kufiri i rangut dinamik përdoret në tre mënyra. Mund të realizohet duke reduktuar rangun spectral nëpërmjet metodës së kovariancës e cila e stabilizon. Për më tepër, nëpërmjet modelit LP mund të zgjerohen bandatme anë të një metode peshash perceptimi. Është e mundur që të shtojmë një pjese shumë të vogël zhurmash duke përdorur metodën e autokorrelacionit të stabilizuar. Nga të tre metodat e përshkruara, më efiçentja dhe e përdorura është metoda e autokorrelacionit të stabilizuar.

METODAT E ANALIZËS SPEKTRALE PËR ASR

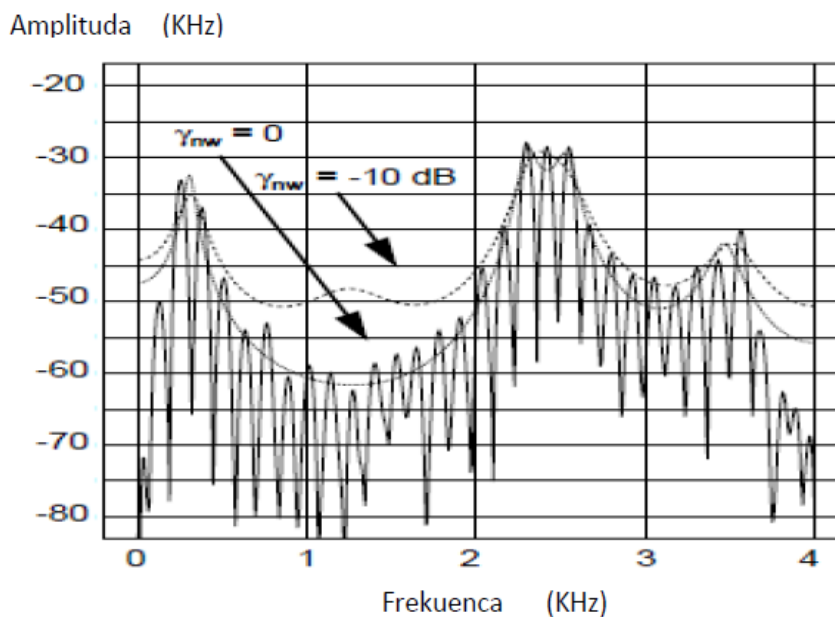


Fig. 8 : Metoda e Autokorrelacionit e stabilizuar (Piccone 1993)

Ashtu si paraqitet në figurën 8, kjo metodë e fundit është e thjeshtë dhe efektive. Ky sistem është i ngjashëm me kombinimit e zhurmës së bardhë të pakorreluar me sinjalin e zërit, para analizës LP. Impakti i shtimit të zhurmës ka të bëjë me sigurinë që modeli nuk do të përmbajë zero të mprehta në spektër. Në zonat spektrale me energji të lartë, ndonjëherë nga kjo metodë krijohet një shtrëmbërim i nivelimit të formës spektrale. Duke qënë se gjerësia e bandave të rezonancës zgjerohet, bashkangjitja e rezonancave spektrale ndonjëherë bie në kolaps duke rezultuar në një rezonancë të gjerë. Një tipar tjetër i modelit LP është se duke shtuar të dhëna mbi fuqinë dhe frekuencën në koeficientët LP, krijohet mundësia e rindërtimit të versionit audio të sinjalit.

Gjenerimi i një modeli të vlefshëm LP është hapi fillestar dhe më i rëndësishëm në analizën spektrale. Përshkrimi parametrik i sinjalit, është shumë i rëndësishëm, veçanërisht në modelet e njohjes së zërit. Duke i dëgjuar ato është e lehtë të vrojtohen problemet dhe difektet. Disa transformime parametrike nuk hartohen tërësisht me të dhënat fillestare të LP. Një shëmbull janë koeficientët Kepstralë. Për këtë arsye, vlerësimi i vlefshmërisë së bashkësisë së parametrave është i vështirë. Sistemet ASR

METODAT E ANALIZËS SPEKTRALE PËR ASR

fillimisht përdornin parametrat LP direct në procesin e njohjes. Përpara aplikimit filestar kemi aplikuar transformime të avancuara në parametrat e dizenuar LP.

Teknika LP nuk është një operacion linear. Për këtë arsye, implementimi në ambiente me shumë zhurmë ndonjëherë përbën problem. Për shkak të performancës së dobët shumë sisteme përdorin analiza me Banka filtrash që në thelb kanë Transformimin Furie.

Për fotografimin apo incizimin e spektrit përdorim një spektrograf, i cili i ndan në gjatësi valësh sipas rrezatimit elektromagnetik dhe mat secilën prej tyre. Në figurën 17 sistemi i modelimit LP ilustron duke parë disa modele LP dhe shfaqjen spektrale në varësi të ambienteve të ndryshme.

Në analizën spektrale koeficienti i parashikimit linear, (LPC) parashikon polet e lidhura me rezonancat. Kur këto kombinohen me burimin e spektrit zanor, ose “residual” në analizën spektrale rezultojnë në formën e valës origjinë. Analiza LPC bazohet në konceptin e ndarjes; ajo ndan karakteristikat rezonante nga ato të burimit të një sinjali. Spektri rezultat është një spektër i zbutur, ku kulmet përfaqësojnë formantë¹¹ të cilat vijnë si rezultat i rezonancës në traktin vokal kur flasim një zanore, ose një bashkëtingëllore të fortë.

Në kimi, një trup i ngurtë dallon nga lëngu duke kaluar një filtër i cili ndan veç lëngun nga copëzat e ngurta në të. Lëngu kalon nga letra dhe copëzat e ngurta mbeten te ajo. Filtrimi është ndarja selektive e një njësie fizike nga një tjetër. Filtrimi përdoret gjithashtu dhe në akustikë. Një tingull kompleks filtrohet kur disa frekuenca lejohen të kalojnë nga filtri kurse të tjera jo.

Një filtër me frekuencë të ulët LP, lejon që disa frekuenca nën një kufi të caktuar të kalojnë pa u sforcuar ndërsa ato mbi kufi të bllokohen.

¹¹ Secila nga bandat mbizotëruese të frekuencës që përcaktojnë cilësinë fonetike të një zanoreje

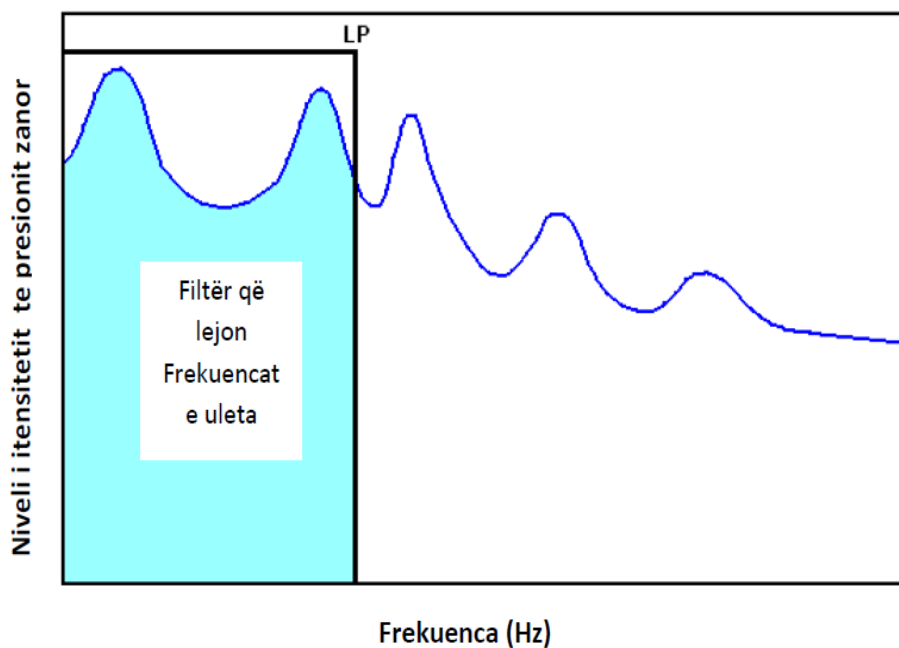


Fig. 9 : Paraqitja e prerjes së frekuencës së LP

Në shumicën e filtrave ekziston një zonë rreth frekuencës kufi ku frekuencat lejohen të kalojnë pjesërisht. Kjo lejon një tranzicion të butë ndërmjet bandës kaluese të cilat përfshijnë frekuenca të lejuara dhe bandës ndaluese që përfshin frekuenca të ndaluara.

METODAT E ANALIZËS SPEKTRALE PËR ASR

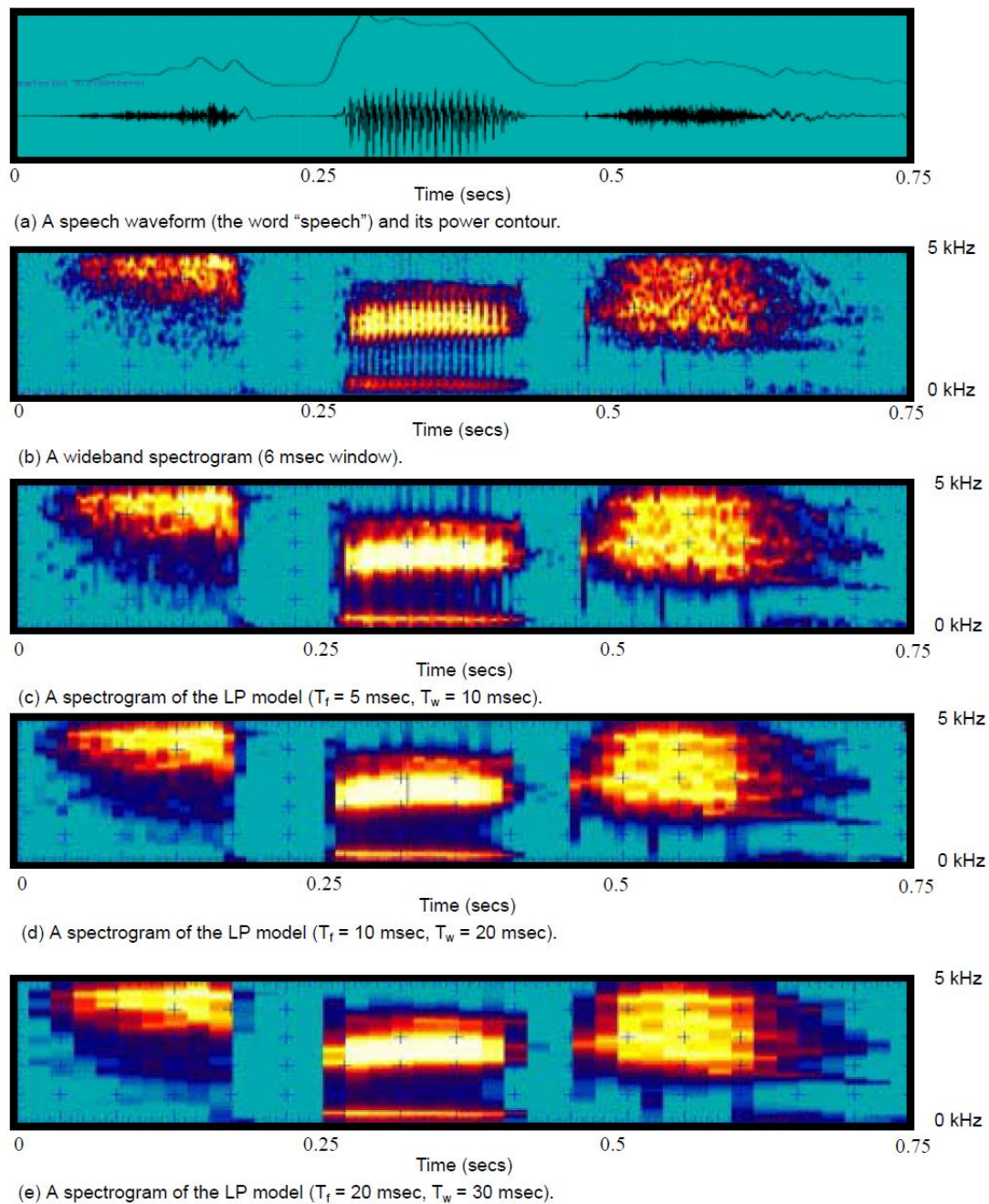


Figura 10 : Shembuj të procesit të LP (Piccone, 1993)

Ilustrimi spektrografik tregon që sekuenca dhe zgjatja kohore zvogëlohen në përpjestim me rezolucionin kohor në spektrogram. Frekuencat me gjatësi 20ms ishin më të shpeshtat në sistemet ASR. Në kohët e fundit, kërkimet mbi njohjen e zërit e kanë ndryshuar fokusin drejt njohjes fonetike, dhe gjatësitë kohore të sekuencave janë

METODAT E ANALIZËS SPEKTRALE PËR ASR

përdorur më shpesh. Me maturimin e teknologjisë mbi njohjen e zërit; përmirësimi i rezolucionit kohor vazhdon.

2.3 Amplitudat e bankave të filtrave të derivuara nga LP

Koncepti i bankës së filtrave mund të kombinohet me modelin LP. Amplitudat që vijnë nga parashikimi linear konsiderohen si amplitudat e bankës që rezultojnë nga kampionimi i modelit LP të spektrit në vend të spektrit të sinjalit audio.

Për këtë arsye, spektri LP kampionohet në frekuencat e bankës së filtrave. Avantazhi kryesor i këtij kombinimi është modelet me rezolucion të lartë japin vlerësime spektrale që janë më të qëndrueshme. Për shkak se spektri kalon në një funksionit stabilizues, parametrat e derivuara për fazat e tjera janë më të qëndrueshëm. Sidoqoftë, me inovacionet e reja dhe teknologjitë DPS, ky kombinim nuk shfaqet më efektiv sa më parë. Për të kampionuar sic duhet spektrin sipas modelit LP vlerat spektrale duhet të aplikohen direkt duke llogaritur amplitudat e bankave të filtrave:

$$S_{LP}(f) = \frac{G_{LP}}{\sum_{i=0}^{N_{LP}} a_{LP}(i) e^{-j2\pi(\frac{f}{f_s})}} \quad (2.7)$$

Vlerat e f_s përfaqësojnë frekuencat dhe kjo metodë ka nevojë për një rend $4p+3$ që të shumëzohet/akumulohet në operacione për kampionin frekuencë. Në këtë metodë kombinimi; spektri zakonisht mbikampionohet dhe vlerësimet dalin duke gjetur mestaret dhe gjenerohet për amplitudat aktuale të bankave.

Në vite 90' mund të llogaritej spektri i fuqisë nisur nga autokorrelacioni i impulsit. Impulsi përgjigje përdor koeficientët LP për llogaritje direkte. Me të dy teknikat, spektri i përkulur nuk është linear dhe mund të përpunohen thjesht nëpërmjet zgjedhjes së frekuencave kampion të bankës së filtrave. Bankat kanë tendencën të paraqesin spektrin nga ana sasiore shumë të rëndomtë ose pa cilësi. Për këtë arsye, edhe pse modeli LPC mund të prodhojë një formë spektrale të lëmuar, është me vend ta mbikampionojmë atë

METODAT E ANALIZËS SPEKTRALE PËR ASR

kështu që majat e frekuencave të mund të karakterizohen sa më mirë. Kjo teknikë kombinimi tregon që ASR po zhvillohen vazhdimisht. Kombinimi i FFT me frekuencat kampion të bankave, është një shembull i këtij zhvillimi dinamik.

2.4 Koeficientët Kepstrale LP

Për të gjeneruar amplitudat e bankave u përdor modeli LP. Për të përmirësuar performancën, një hap llogjik në kombinimin e elementëve sistematikë ASR do të ishte përdorimi i LP për llogaritjen e koeficientëve kepralë. Këto koeficientë mund të përlllogariten direkt nga LP nëse filtri i parashikimit linear është stabil dhe mund të vazhdojë të jetë i tillë dhe gjatë analizës së autokorrelacionit. Ky logaritëm i Filtrit Invers mund të shprehet:

$$C_{LP}(z) = \sum_{i=0}^{N_C} C_{LP}(i)z^{-i} = \log H(z) = \log \left(\frac{G_{LP}}{\sum_{j=0}^{N_{LP}} a_{LP}(j)z^{-j}} \right) \quad (2.8)$$

Koeficientët e mësipërm u referohemi si Koeficientët Kepstralë Lp. Sipas Markel(1980), ato janë përkufizuar si logaritmi i gabimit të LP në fuqi. Vlera e fuqisë merret si parametër më vete, kështu që nuk ka nevojë ta kombinojmë në barazimin e mësipërm. Në këtë këndvështrim, modeli Kepstral shikohet si një model i normalizuar, më shumë si formë e LP.

Ashtu si e thamë më sipër, CC përfshin një transformimi Fourier. Rekursi i koeficientëve kepralë krijon një problem pasi numri i koeficientëve kepralë që duhet të llogaritet nuk specifikohet. CC në fakt është një transformim Invers Fourier, i aplikuar si përgjigje impuls i modelit LP. Kështu, të paktën teorikisht është e mundur të llogarisim një numër të pafundmë koeficientësh kepralë.

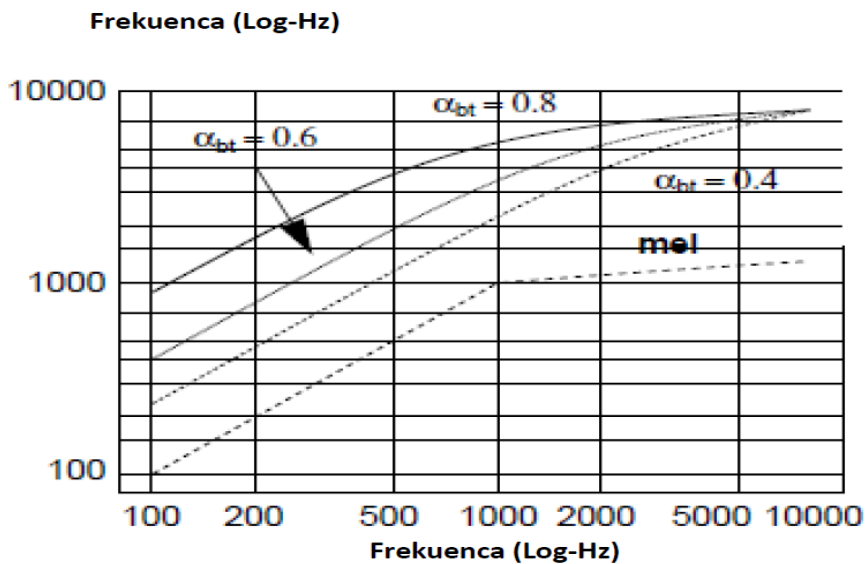
Zakonisht, numri i koeficientëve kepralë të llogaritur është sa numri i koeficientëve LP. Duke llogaritur koeficientët Kepstralë me modelin e mësipërm LP kemi reflektimin e një shkalle frekuence lineare. Një disavantazh i koeficientëve Kepstralë të derivuar

METODAT E ANALIZËS SPEKTRALE PËR ASR

nga modeli LP është se nevojiten hapa të tjerë në mënyrë që të përshtasim një shkallë frekuence jolineare. Për këtë arsye, kombinimi i koeficientëve LP me amplitudat e bankave është më i përshtatshëm pasi vetë ai paraqitet si jolinear. Metodika më e pëlqyer në këto raste bazohet në një metodë e cila përkul frekuencat në një filtër dixhital. Kjo metodë përdor një transformim në përpunimin e sinjalit dixhital: transformimin bilinear ku α_{bt} konsiderohet si parametri i përthyerjes së frekuencave.

$$f_{new} = 2\pi \frac{f}{f_s} + 2 \tan^{-1} \left(\frac{\alpha_{bt} \sin \left(2\pi \frac{f}{f_s} \right)}{1 - \alpha_{bt} \cos \left(2\pi \frac{f}{f_s} \right)} \right) \quad (2.9)$$

Përkulja e këtyrë frekuencave sipas transformimit bilinear është i ngjashëm me shkallën Mel. Kjo përfaqësohet në figurën e mëposhtme



**Fig. 11 : Transformimi i përkuljes së frekuencës në ngjashmëri me shkallën Mel
(Piccone, 1993)**

METODAT E ANALIZËS SPEKTRALE PËR ASR

Një vlerë e zakonshme e α_{bt} është 0.6. Ky transformim mund të përdoret si një nga gjendjet e shumta të llogaritjes së koeficientëve kepstralë LP. Për shembull, ashtu si përmendet dhe nga Piccone, mund të përdoret në funksionin e autokorrelacionit, në parametrat parashikues dhe në parametrat kepstralë të derivuar LP. Ashtu sic pohohet dhe nga Shikano(1986), metoda e post përpunimit të koeficientëve kepstralë ishte më efektive gjithashtu dhe më e thjeshta.

Nëse një proces hapësire frekuence aplikohet duke përdorur një metodë transformimi kampionësh, për të llogaritur atë duhet përdorur një transformim Invers Furie i Koeficientëve Kepstralë. Një llogaritje rekursive e drejtpërdrejtë e koeficientëve rezulton shumë më e thjeshtë. Ky rekurs ekziston. Sipas Piccone, ky rekurs mund të shihet si një sekuenca kaskadë procesesh lineare filtrash të fiksuar, dhe mund të aplikohet si më poshtë me kushtin që të gjitha kushtet fillestare janë zero.

$$Pwr \ 0 \leq n \leq N_c \begin{cases} c_{bt}^{(n)} = \alpha_{bt}[c_{bt}^{(n-1)}(\mathbf{0}) - \mathbf{0}] + c_{LP}(N_c - n) \\ c_{bt}^{(n)}(\mathbf{1}) = \alpha_{bt}[c_{bt}^{(n-1)}(\mathbf{1}) - \mathbf{0}] + (1 - \alpha^2)c_{bt}^{(n-1)}(\mathbf{0}) \end{cases} \quad (2.10)$$

$$Pwr \ 2 \leq k \leq N_{bt} \{c_{bt}^{(n)}(k) = \alpha_{bt}[c_{bt}^{(n-1)}(k) - c_{bt}^{(n)}(k-1)] + c_{bt}^{(n-1)}(k-1)\} \quad (2.11)$$

Ky rekursion është iterativ dhe rezultati në fund janë koeficientët e transformuar. Ky rekurs kërkon rendin e mëposhtëm në mënyrë që të përdoret për operacione shumëzimi/mbledhje:

$$N_c \times N_{bt} \quad (2.12)$$

Kompleksiteti i tij është pak a shumë i njëjtë me LP. Është shumë e rëndësishme të marrim në konsideratë dhe thjeshtimin. Sekuenca e ndryshuar bilineare, që vjen si rezultat i përpunimit jo linear të keprstrimit të thjeshtuar, në kohë është e pafundme. Por, sipas Piccone, në terma praktikë, nëse sekuenca nga një zgjatje të fundme, sekuenca e transformuar do të zvogëlohej në mënyrë eksponenciale. Për këtë arsye është e mundur të thjeshtojmë sekuençën me shtrembërime minimale të frekuencës (1993).

METODAT E ANALIZËS SPEKTRALE PËR ASR

Figura 12 përfaqëson efektet e kombinuara të analizës cepstrale dhe koeficientëve kepstrale të transformuar bilinearë.

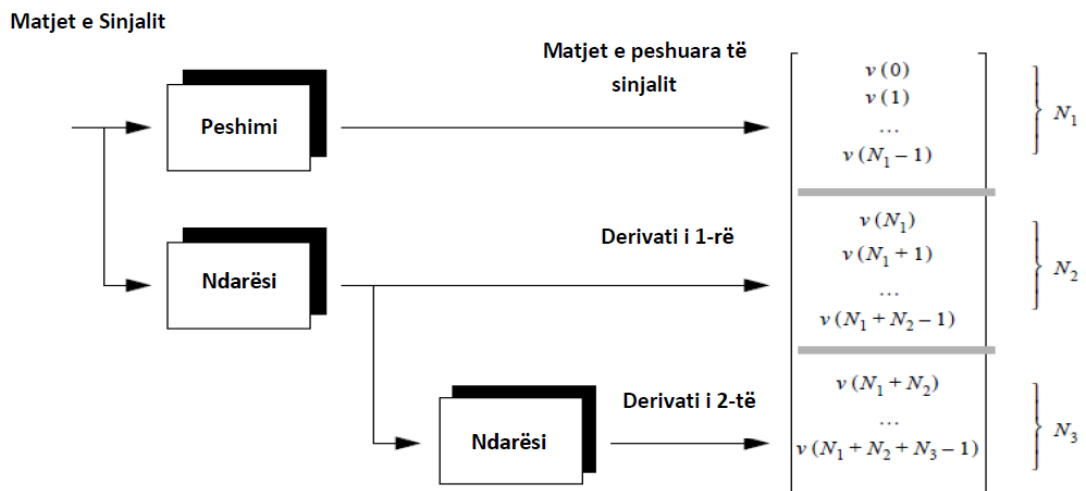


Fig. 12 Transformimi i përkuljes së frekuencave

Kjo figurë tregon sinjalin kampion të zërit me frekuencë 8kHz sipas modelit LP. Spektri sipas madhësisë log i koeficientëve kepstrale ilustron për analizë kepstrale. Në të njëjtën mënyrë, spektri i madhësisë log ilustron dhe për kepstrumin e transformuar në mënyrë bilineare. Në këtë mënyrë koeficientët kepstrale u transformuan në koeficientë bilinearë.

Shuma e duhur e kompresimit bilinear është pak a shumë sa frekuenca provë. Piccone(1993) argumenton se nëse faktori i kompresimit është 0.6 atëherë gjerësia e bandës del e pafundme. Shumica e energjisë duket në çerekun më të ulët të shkallës së frekuencës për sonorantët.

Në bazë të studimeve të deritanishme, për arsye të shumta, transformimi bilinear nuk është shumë produktiv në frekuencën kampion. P.sh., sinjali zanor zë pjesën madhore të bandës së disponueshme, duke pasur kështu pak vënd për të zgjatuar spektrin. Ekziston rrisku që të humbim informacion të rëndësishëm nëse aplikohet një shkallë ekstreme. Së dyti, faktori i kompresimit duhet të reduktohet në mënyrë që të ruhet përafrimi i shkallës

METODAT E ANALIZËS SPEKTRALE PËR ASR

Mel. Megjithatë kur ulet faktori i kompresimit ulet dhe kompresimi përfundimtar. Kjo bën që të spekulojmë që ky mund të mos jetë proçesi më i mirë i përafrimit të shkallës Mel.

Kapitulli 3

3.1 Bankat e filtrave dixhitalë

Bankat e filtrave dixhitalë, janë një nga konceptet themeltare në procesin e përpunimit të zërit. Bankat e filtrave dixhitalë janë evolucion i filtrave analog që përdorreshin më parë, kur kapaciteti i hardwareve kompjuterik për të realizuar procese matematikore komplekse ishte i kufizuar. Bankat e filtrave analoge kishin një të mirë që kosto e përpunimit të tyre ishte shume e ulët. Filtrat kryesisht ndërtoheshin nga komponentë diskretë, dhe sinkronizoheshin me kujdes, duke rregulluar tranzistorët dhe kapacitorët para përdorimit.

Një bankë filtrash dixhitalë mund të shikohet si një model bazik i fazave të para të imitimit të sistemit dëgjimor njerëzor. Sipas Piccone, ka dy arsye për përdorimin e tyre. Së pari, pozicioni i distancës maksimale gjatë membranës basilare për stimuj si tone të pastra ka lidhje logaritmike me frekuencën e tonit. Kjo hipotezë është pjesë e një koncepti të të dëgjuarit që quhet “Teoria e Vendit”. Së dyti, eksperimentet në perceptimin e të degjuarit njerëzor, kanë treguar që frekuencat e tingujve komplekse në një gjerësi bande specifike nuk mund të identifikohen individualisht. Kur njëri nga komponentet bie jashtë bandës atëherë ai mund të identifikohet individualisht. Banda kritike konsiderohet si 10-20% e frekuences qendrore të zërit. Njësitë e kësaj shkalle perceptimi frekuencash njihen si banda kritike ose shkalla Bark.

Ashtu sic thuhet dhe nga Rossel, informacioni në çdo sekuencë është shpjegim i shpërndarjes së frekuencës. Për këtë arsye informacioni në pjesëza tregon si shpërndahet fuqia e sinjalit në frekuenca të ndryshme. Një bankë filtrash, në formën e saj më të papërpunuar, është një bashkësi filtrash me frekuenca të ndryshueshme që mbulojnë pjesët kyç të spektrit që janë, frekuenca bazë dhe formantet (zonat me koncentrim më të lartë energjitik). Daljet nga filtrat gjatë një sekuence mund të përdoren si tipar. Frekuencat qendrore të filtrit sigurohen në disa mënyra. Zakonisht, ato ndërtohen në varësi të një shkalle në bazë perceptimi. Tingulli i perceptuar i një sinjali është i ndryshëm nga frekuenca aktuale e tij.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Një përafrim i përdorur gjerësisht i hartimit real të zërit dhe njohjes së tij është shkalla Mel, ose frekuenca Mel (f_{mel}).

$$frekuenca\ mel = 2595 \log_{10} \left(1 + \frac{f}{700.0} \right) \quad (3.1)$$

ose

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (3.2)$$

Shkalla Mel fokusohet në krijimin e hartës për frekuencën e perceptuar të një tingulli, thirrjeje në një shkallë lineare. Zakonisht përafrohet me një shkallë lineare nga 0-1000 Hz dhe pastaj në një shkallë logaritmike përtej 1000Hz. Sistemi dëgjimor tek njeriu nuk mundet të dallojë tinguj që janë afër njëri tjetrit. Frekuencat e larta rezultojnë në banda kritike më të larta. Bandat kritike i referohen intervaleve ndërmjet të cilave frekuencat nuk mund të ndahen nga frekuenca qendrore. Në filtrat e bandave kritike, filtrat janë të pozicionuar në mënyrë lineare në një shkallë të motivuar nga perceptimi; shkalla mel është një shembull i kësaj që thamë.

$$BW_{kritik} = 25 + 75 \left[1 + 1.4 \frac{f_{mel}}{1000^2} \right]^{0.69} \quad (3.3)$$

Gjerësia e bandës së filtrit përshtatet me gjerësinë e bandës kritike për frekuencat qendrore. Një përafrim i rëndësishëm i bankave të filtrave kritike bazohen në shkallën Mel. Në hapësirën 100hz-1kHz filtrat pozicionohen në mënyrë lineare. Mbi 1kHz caktohen pesë filtra për çdo dublim të frekuencës. Me fjalë të tjera ato pozicionohen në formë logaritmike. Gjerësia e bandës përcaktohet në atë formë që pika 3 db (gjysma e fuqisë) të ndodhet në gjysmën e qendrave të dy filtrave të njëpasnjëshëm. Zakonisht përdoren kampionë nga 20 filtrat e parë.

Figura 13 tregon bankën e filtrave kritike dhe bankën e filtrave Mel, duke përfshirë dhe shfazimin e lidhur me to. Një tjetër mënyrë për të arritur një vektor korrespondues për tiparet është të përdoret transformimi Furie për të kampionuar sinjalin në frekuencat e dëshiruara, p.sh. si qendrat e filtrave në bankën e filtrave Mel. Kjo rezulton në një bankë

METODAT E ANALIZËS SPEKTRALE PËR ASR

filtrash Furie dhe arrihet nëpërmjet Transformimin Diskret Furie që përdor numrat e kampionëve në dritaren në shqyrtim. Shpesh sinjali kampionohet në rezolucione të larta në frekuencë dhe pastaj për cdo “filtër” llogaritet si një mesatare e disa frekuencave. Dikush mund të përdorë Transformimin e Shpejtë Furie (FFT) për të arritur objektivat e tij.

Ky transformim mund të përdoret për të llogaritur gjerësinë e bandës në një shkallë perceptual për filtrat në një frekuencë të caktuar të bankave Bark dhe Mel. Funkcioni i bandës kritike jepet në ekran. Si shkalla Bark ashtu dhe Mel mund të shikohen si një transformim i frekuencave në një shkallë perceptuale lineare që ka një kuptim. Kombinimi i këtyre dy teorive i dha hov një teknike të njohur si Banka e Filtrave të Bandave Kritike.

Një bankë Bandash Kritike është një bankë fazash lineare FIR që sistemohen në një vijë të drejtë përgjatë shkallës Mel dhe Bark. Gjerësitë e bandave zgjidhen të barabarta me një bandë kritike që i korrespondon frekuencës qendrore. Në shumë aplikacione të përpunimit të zërit, filtri i parë anashkalohet pasi rangu i tij është më i madh se qëllimi i konvertuesit A/D.

Shpesh, të dhënat e kampionuara për rangun e frekuencave korresponduese janë shumë të zhurmshme. Bisedat telefonike përpunohen duke përdorur një filtër me 16 banda. Output i filtrave zakonisht përpunohet nëpërmjet një metode të vlefshme që vlerëson fuqinë. Zakonisht, në sisteme që mundohen të imitojnë fizionominë dëgjimore dhe folëse të njeriut, përdoret një bankë filtrash dixhitalë . Llogaritjet përseritëse janë tëpër të dobishme për post përpunimin në këto aplikime.

Kjo analizë zakonisht përdor një vektor me vlera të fuqive/frekuencave për çdo segment të dhënash. Normalisht këto të dhëna çifte, kombinohen me parametra të tjerë si fuqia totale, për të formuar një vektor për matjen e sinjalit. Banka e filtrave mundohet të shpërthejë sinjalin në bashkësi të ndara kampionësh spektralë që përfshijnë të dhëna që imitojnë mënyrën si ato përpunohen nga sistemi dëgjimor njerëzor.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Analiza e tingullit zanor zakonisht bazohet në përpunimin sekuencial, për këtë arsye kjo teknike rezulton e efektive në rastin kur përballemi me zhurmat.

3.2 Banka e filtrave MEL

Banka e filtrave MEL është një bankë filtrash për njohjen e zërit e cila bazohet në shkallën MEL. Banka e filtrave MEL përdoret gjerësisht në këto procese. Frekuenca/banda për këtë bankë filtrash jepen në kolonën e katër dhe të pestë të tabelës në figurën 13.

Index	Shkalla Bark		Shkalla Mel	
	Frek. Qënd	BW (Hz)	Frek. Qënd	BW (Hz)
1	50	100	100	100
2	150	100	200	100
3	250	100	300	100
4	350	100	400	100
5	450	110	500	100
6	570	120	600	100
7	700	140	700	100
8	840	150	800	100
9	1000	160	900	100
10	1170	190	1000	124
11	1370	210	1149	160
12	1600	240	1320	184
13	1850	280	1516	211
14	2150	320	1741	242
15	2500	380	2000	278
16	2900	450	2297	320
17	3400	550	2639	367
18	4000	700	3031	422
19	4800	900	3482	484
20	5800	1100	4000	556
21	7000	1300	4595	639
22	8500	1800	5278	734
23	10500	2500	6063	843
24	13500	3500	6964	969

Fig. 13 : Gjerësia e bandës së bankës së filtrave Mel (Piccone, 1993)

METODAT E ANALIZËS SPEKTRALE PËR ASR

Në filtrat Mel, përcaktohen dhjetë filtra sekuencialë nga 100 Hz në 1000 Hz (Fig 6). Mbi 1000 Hz, pesë filtra aplikohen për çdo dublim të shkallës së frekuencës (oktavë). Këto filtra janë të pozicionuar në mënyrë të barabartë në një shkallë logaritmike. Gjerësitë e bandave caktohen në atë mënyrë që pika 3dB është qendra ndërmjet serisë aktuale dhe asaj paraardhëse/pasuese. Zonat e hijëzuara në tabelë tregohen për efekt kontrasti dhe krahasimi. Zakonisht, përdoren vetëm 20 kampionët e parë nga bankat e filtrave. Çdo filtër në bankën dixhitale përdoret normalisht si një filtër linear faze në mënyrë që vonesa për çdo grup filtrash të jetë zero dhe sinjalet në dalje të sinkronizohen në kohë. Ekuacionet e filtrave janë për një implementin të filtrit fazë sekuencial. Përpunimi i zërit nëpërmjet dy filtrave në këtë lloj banke ilustron në figurën 14. Figura 14 tregon një sinjal zanor së bashku me daljen nga dy filtra kalimtarë bande. Njëri filtër bande qëndërzohet në 250Hz dhe tjetri në 2500Hz. Figura tregon qartë që fuqia e outputeve varion në lidhje me llojin e sinjalit të folur.

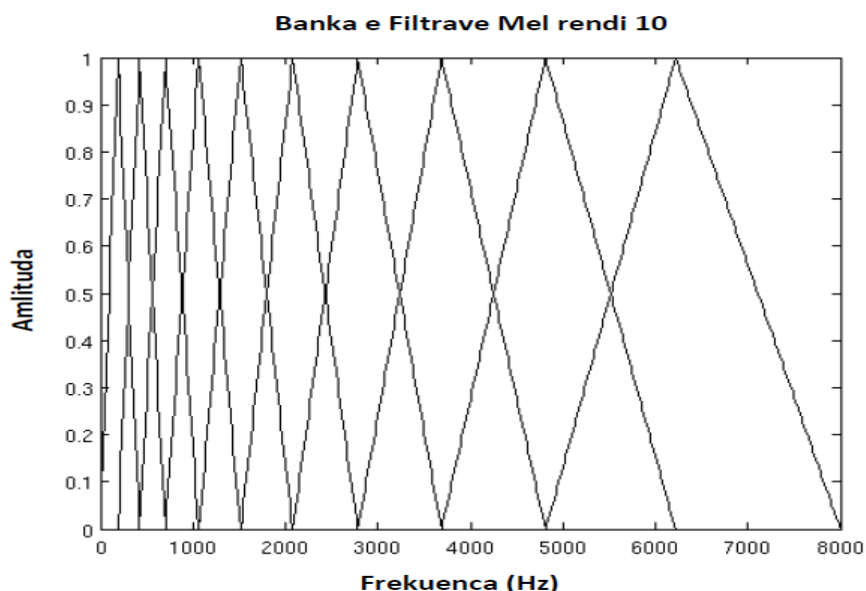


Figura 14 : Një bankë filtrash Mel me 10 filtra

METODAT E ANALIZËS SPEKTRALE PËR ASR

Avantazhi themelor i një banke filtrash është që disa outpute mund të bëhen të përshtatshëm me disa klasa të tingujve zanorë.

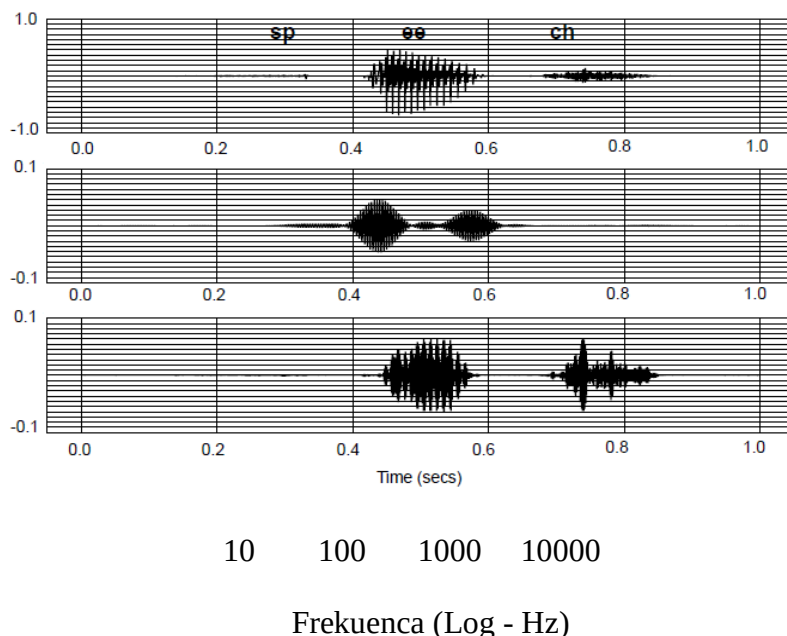


Fig. 15: Frekueca Mel

Ndërtimi i përfaqësimeve të përshtatshme për njohjen e zërit ka qënë një fushë e rëndësishme kërkimore për shumë vite. Popullariteti i filtrave Mel vjen si pasojë e teorive motivuese për prodhimin dhe perceptimin e zërit. Sidoqoftë, ashtu siç shkruan dhe Sainath etj(2013)., edhe pse tiparet e derivuara të Mel janë të përdorura gjerësisht, bankat e filtrave Mel nuk janë zgjedhja korrekte për njohjen e zërit pasi ato nuk janë ndërtuar për këtë qëllim.

Në sistemet e vazhdueshme të njohjes së zërit me fjalor të gjerë, përdorimi i sistemeve folës të adaptuar dhe diskriminative ka treguar një përmirësim të performancës. Këto sisteme trajnohen nëpërmjet funksionit të objektivave si regresioni linear i ndodhjes maksimale tipar-kohë (fMLLR), si dhe funksionit të informacionit maksimal tipar-hapësirë (fBMMI) (2013).

METODAT E ANALIZËS SPEKTRALE PËR ASR

Është traditë që dizenjimi i tipareve të bëhet vecmas nga dizenjoja e klasifikuesit. Kërkuesit kanë përdorur për shumë vite module të ndarë për përshtatjen e folësit (dmth fMLLR) dhe trajnimin diskriminues (dmth FBMMI) për trajnimin e GMM. Rrjetat e thella Neurale janë shumë më të forta se Modelet e Përzierjes Gausiane për faktin se kur bëhet ekstraktimi i tipareve bashkarisht bëhet dhe klasifikimi i tyre. Për këtë arsye, tiparet sinkronizohen me procesin klasifikues. Sidoqoftë, bashke me DNN ato nuk janë të ndërtuar në mënyrë specifike për të akomoduar stilet e ndryshme të se folurës dhe reduktojnë variancën e frekuencave përkthyes në sinjalet audio. Sanaith etj, pohojnë që fMLLR kur përdoret së bashku me DNN kanë tendencën të japin një performacë të optimizuar. Ky është një tjetër shëmbull i kombinimit të sistemeve ASR për maksimizimin e performancës.

Banka e filtrave Mel, dizenjohej si imitues i sistemit dëgjimor njerëzor dhe i mënyrës se si njerëzit i perceptojnë sinjalet audio. Për këtë arsye, ai dizenjohej për evidencë perceptual. Në një strukturë modelimi statistikor, qëllimi përfundimtar është të reduktojë shkallën e gabimeve në fjalë pasi ajo prek direkt cilësinë në dalje të sistemit ASR. Për të arritur këtë, bankat e filtrave Mel dhe Gamatone(për të cilët do flitet më poshtë) që janë të ndërtuara si imitues të dëgjimit njerëzor janë në disavantazh.

Një shëmbull i kësaj është përdorimi i madh i Modeleve të fshehura Markov (HMM) të modelimit akustik. HMM kanë pavarësi në strukturën e tyre që është një formë e ndryshme e modeleve dëgjimore njerëzore. Deri më sot HMM janë teknika dominante e modelimit akustik.

Tiparet logaritmike Mel krijohen duke kaluar një spektër fuqie nëpërmjet një banke filtrash Mel, dhe më pas duke kaluar në një operacion log jo-linear. Sanaith etj. (2013), argumentojnë që ky sistem mund të modelohet nga një shtresë rrjetash neuronesh, me një “shumëzim peshash lineare (dmth një shtresë banke filtrash), të ndjekura nga një shtresë jo lineare (dmth Log)”. Kjo mundëson që Banka e Filtrave të përshtatet me detyrën e kërkuar. Ky është një sugjerim për të mundësuar trajnimin e bankave të filtrave në varësi të të dhënave. Kjo fushë është studiuar në shumë kontekste të ndryshme. P.sh., bankat e filtrave të derivuara nga të dhëna të etiketuara fonetike duke

METODAT E ANALIZËS SPEKTRALE PËR ASR

përdorur Analizën Lineare Të Diskriminimit. Sidoqoftë, autorët arritën në përfundimin që bankat e filtrave të përfutuara nuk ishin optimal për shkak të LDA pret që të dhënat të kenë shpërndarje Gaussiane, gjë që nuk mund të thuhet për spektrat e fuqisë. Për më tepër, duke parë distancën Kullback-Leibler (KL) si masë në dizenjimin e bankës së filtrave, filtri ishte dizenuar në mënyrë të pavarur nga modeli akustik. Sanaith etj, prezaton të mësuarit në bankat e filtrave, që është një model i fuqishëm dhe i përparuar akustik CNN, ku në parametrat CNN janë të trupëzuar dhe elementët trajnues të CNN.

Avantazhi kryesor i përdorimit të rrjetave neural është që trajnimi i bankave të filtrave mund të shikohen si një shtesë e tyre, ku parametrat e rrjetave neurale përditësohen bashkë me parametrat e shtresave të tjera. Kjo teknikë ka të njëjtat disavantazhe si të bankave të filtrave Mel dhe Gammatone. Experimentet e Sanaith etj u bazuan në krijimin e bankave të filtrave me rrjeta neurale të thella dhe u testuan mbi një transmetim 50h të një stacioni televiziv anglez. Ata gjetën që duke aplikuar elementët trajnues të bankave të filtrave direkt në CNN, ku përmirësimi modest është 22.0% WER (raporti i gabimit në fjalë). Eksperimenti u ndërtua me premisën që si shumëzimi i peshave dhe jolineariteti janë funksione diferenciale dhe në këtë mënyrë ishte e mundur të “mësonin” bashkarisht si filtrat ashtu dhe pjesa tjetër e rrjetave neurale. Avantazhi i kësaj metode është që filtrat mësuan nëpërmjet një funksioni objektiv me entropi të tërthortë dhe në lidhje me subjektin e testuar në vend që të dizenohej më përpara.

Në mënyrë specifike, tiparet Mel gjenerohen duke kaluar spektrin e fuqisë nëpërmjet një bashkësie filtrash Mel. Numri standart i filtrave Mel të përdorur në procesin e njohjes së zërit është 40. Kjo pasohet nga një aplikim kompresimi log për të gjeneruar tiparet log-Mel të bakës së filtrave.

Për këtë arsye studimi i Sanaith tregoi që futja e trajnimit të bankës së filtrave, si një hap shtesë në rrjetat neurale, i jep mundësi peshave specifike të përditësohen në ndryshim nga filtrat e paracaktuar Mel dhe Gamatone të cilat nuk janë të akorduara ose të ndërtuara posaçërisht për sistemet ASR.

3.3 Koeficientët Kepstral të filtrave Mel

Koeficientët kepstrale të frekuencave Mel (MFCC) shpesh përdoren në njohjen e zërit dhe në mundësimin e të folurës automatike. MFCC u përdorën për herë të parë në 1980 nga Mermelstein dhe Davis dhe popullariteti i tyre vazhdon akoma. Para MFSS funksioni i të folurës automatike realizohej duke përdorur LPC dhe LPCC. Në këtë pjesë të studimit, MFCC do të vëzhgohen dhe diskutohen në mënyrë kritike në kontekstin e njohjes automatike të zërit.

MFCC përdoren në njohjen automatike të zërit si më poshtë. Së pari, sinjalet e zërit transformohen në segmente dhe më pas i nënshtrohen vlerësimit të periodogramës që bazohet në spektrin e fuqisë, këto janë vlerësime që matin energjinë. Më pas merret një logaritëm nga energjitë e bankave të filtrave. Më pas merren DCT e logaritmave, ndërtohen koeficientët e duhur dhe pjesa tjetër skartohet. Si hap final zbatohet Liftering

Më vonë, zbatohet transformimi parametrik i sinjalit zanor. Është e mundur që energjisë në secilin segment t'i bashkangjet çdo vektori tipar dhe Delta, d.m.th bëjmë bashkimin e tiparve Delta-Delta. Parametri Delta përdoret për të përfaqësuar sinjalin në dalje që merret nga procesi i diferencimit. Një parametër delta-delta arrihet duke përdorur edhe njëherë outputin e diferencimit të rangut të parë, duke dhënë dhe derivatin kohor të rendit të dytë, (Piccone 1993) theksoi që një proces i tillë aplikohet edhe për derivate të një rendi më të lartë.

Secili nga hapat e diskutuar më sipër është jetësor në procesin MFCC. Sinjalet zanore dhe audio janë në fluks konstant; supozimi bazë është që nëse shkalla kohore shkurtohet, fluksi i sinjalit nuk është i mirë. Problem atëherë është të gjejmë shkallën e duhur kohore, e cila shtrihet ndërmjet 20-40ms. Shkurtimi i shkallës kohore përtej parametrave minimale mund të rezultojnë në pamjaftueshmëri kampionësh për llogaritjen e besueshme të vlerësimeve spektrale. Zgjatimi i hapësirave kohore pastaj sjell problemin e fluksit të tepërt.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Llogaritja e spektrit të fuqisë për secilën pjesë varet nga karakteristikat e kërmilit të veshit njerezor. Kërmilli i veshit është organ i veshit të brendshëm, i cili dridhet në sinkron me tingujt që hyjnë në vesh dhe merret me kodimin e sinjalit zanor që çon në tru. Vlerësimet e periodogramit dhe bankat e filtrave kryejnë një funksion të ngjashëm duke kopjuar aparatit e dëgjimit njerezor dhe më pas i kalojnë këto sekuenca për të identifikuar frekuencat e tyre. Vlerësimi i periodogramit kodon një masë shumë të madhe informacioni, një pjesë e të cilit është thelbësor për njohjen automatike të zërit për shkat të kufizimeve të diagramës së kërmilit të veshit për të diferencuar dy frekuenca afër njëra tjetrës. Kur frekuencat rriten, fuqia diskriminuese e kërmilit të veshit bie. Për shkak të karakteristikave të kërmilit të veshit, grupe të njësive të periodogramës duhet të mesatarizohen për të arritur një përafrim të energjisë për secilin nivel 0 Hz. Ashtu siç theksuan Huang dhe Acero në (2001), shkalla Mel siguron informacion për vendosjen optimale dhe gjerësinë e Bankave të Filtrave.

Matja e energjive të bankave të filtrave vazhdohet nga marrja e logaritmit. Qëllimi i kësaj metode gjithashtu ka bazë në procesin dëgjimor. Tek njerezit, lidhja ndërmjet volumit të perceptuar dhe energjisë nuk kanë lidhje lineare; kur tingujt janë të lartë, luhatjet e mëdha në energji nuk kodohen si shumë të ndryshme. Sigurisht, rrënja kubike mund të përdoret në vend të një logaritmi, por logaritmat suportojnë përdorimin e teknikës së normalizimit të kanalit nëpërmjet zbritjes së mesatares kepstrale.

Kemi dy motive kryesore për llogaritjen e DCT për energjitë e bankave të filtrave log. Mbivendosja e tyre nënkupton një shkallë të madhe korrelacioni të brendshëm ndërmjet energjive të bankave të filtrave. DCT është një mjet dekorrelacioni që mbështetet në përdorimin e matricave diagonal të kovariancës për modelimin e tipareve. P.sh., siç ndodh në një klasifikues HMM. Ka një arsye pse ngelemi vetëm tek një nënbashëksi e koeficientëve DCT; koeficientët e mëdhenj DCT përfaqësojnë ndryshimet më të shpeshta në energjitë e bankave të filtrave dhe ndryshimet me një shpejtësi të tillë dëmtojnë performancën e njohjes automatike të zërit. Për këtë arsye, eliminimi i tyre mund të sjellë përmirësim.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Përdorimi i shkallës Mel shkon në të njëjtën vijë me aftësinë e njeriut për të ndjerë ndryshimet e lartësive të zërit më mirë në frekuenca të ulëta se të larta.

Formula e konvertimit nga frekuencë në shkallë Mel është si më poshtë:

$$M(f) = 1125 \ln \left(1 + \frac{f}{700} \right) \quad (1) \quad (3.4)$$

Për të bërë të anasjelltën:

$$M^{-1}(m) = 700 \left(\exp \left(\frac{m}{1125} \right) - 1 \right) \quad (2) \quad (3.5)$$

Për arsye ilustrimi të bankave të filtrave në një aplikimi automatik të njohjes së zërit marim një shëmbull për një sinjal zëri të kampionuar në 16kHz.

1. Ndajmë sinjalin në sekuenca 20-40ms, me preferencë 25ms. Kështu për një sinjal 16kHz marrim $0.025 \cdot 16000 = 400$ kampionë. Hapi sekuencave vendoset 10ms, ose 160 kampionë duke përfshirë këtu dhe mbivendosje ndërmjet tyre. Sekuenca e parë me 400 kampionë fillon në kampionin 0, frekuenca pasuese fillon në 160 e kështu me rradhë. File i zërit duhet të zgjerohet me 0 në mënyrë që të plotpjestohen në një numër çift sekuencash.

Hapat e tjerë përfshijnë të gjitha sekuencat. Si fillim, një bashkësi me 12 koeficientë MFCC merret nga secila sekuencë. Domain kohor i sinjalit dizenjohet $s(n)$. Pasi ndahet sekuenca, $s_i(n)$, ku n përfshin 1-400 me supozimin që një sekuencë ka 400 kampionë dhe 1 varion sipas numrit të sekuencave. Llogaritjet komplekse DFT japin $S_i(k)$, ku i është numri i sekuencave që masin hapësirën sekuenciale kohore. $P_i(k)$ në këtë formë bëhet dhe spektri i fuqisë së sekuencës i .

2. Për të arritur transformimin diskret Furie shfrytëzojmë formulën e mëposhtme

$$S_i(k) = \sum_{n=1}^N s_i(n) h(n) e^{-\frac{j2\pi kn}{N}} \quad 1 \leq k \leq K \quad (3.6)$$

METODAT E ANALIZËS SPEKTRALE PËR ASR

Në këtë ekuacion $h(n)$ është një dritare Hann me N kampionë e njohur ndryshe si një dritare me një analizë të gjatë, dhe K është gjatësia e DFT. Vlerësimi i bazuar në periodogramë për sekuencën zanore $s_i(n)$ jepet nga:

$$P_i(k) = \frac{1}{N} |S_i(k)|^2 \quad (3.7)$$

Sipas vlerësimit të periodogramës për spektrin e fuqisë, vlera absolute e transformimit kompleks Furie, duhet të ngrihet në katror. Metodologjia e duhur është të bëhet një FFT me 512 pika dhe të mbahen 257 koeficientët e parë që derivohen.

3. Llogarisim Bankën e Filtrave me hapësirën Mel. Banka e filtrave në hapësirën Mel përbëhet nga 20-40 filtra trekëndorë (ku 26 janë standard) që më pas aplikohen në hapin e dytë të vlerësimit në periodogramë të spektrit. Secili vektor përbëhet kryesisht nga zero, me disa vlera jo zero në pjesë të caktuara të spektrit. Llogaritja e energjisë së bankës së filtrave arrihet nëpërmjet shumëzimit të secilës bankë me spektrin e fuqisë, pasuar kjo me shtimin e koeficientëve. 26 numrat që rezultuan japin një tregues të niveleve të energjisë në secilën bankë filtrash.

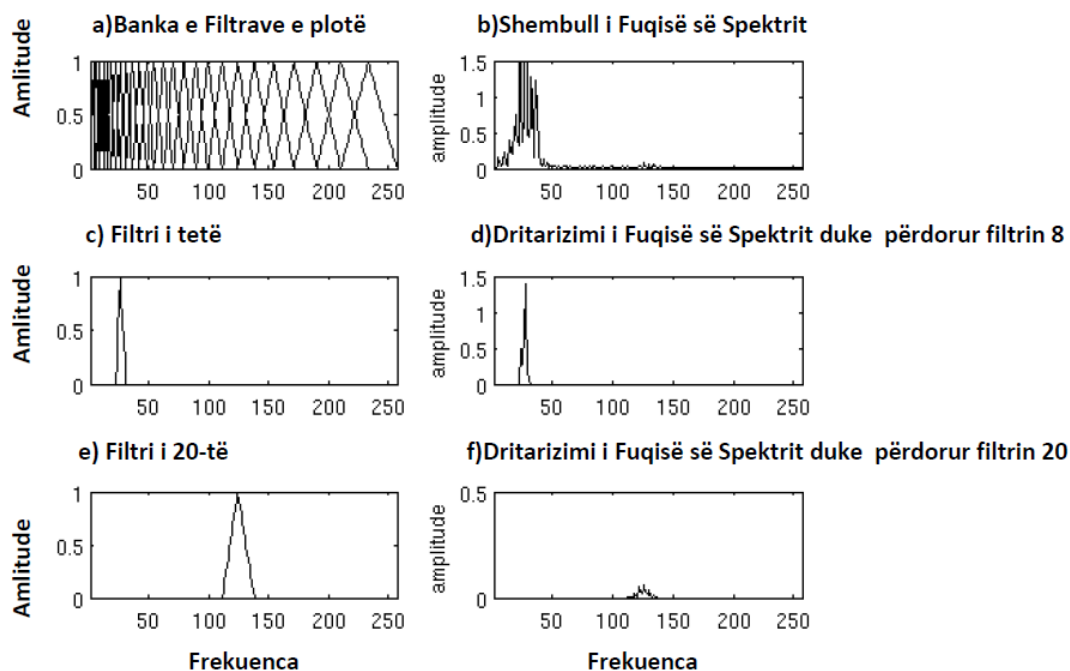


Fig. 16: Banka e filtrave Mel dhe fuqia spektrale e ‘Dritarizuar’

4. Merr log e secilit prej 26 energjive nga hapi i mësipërm, dhe gjenerojme rezultatet e energjisë për 26 bankat e filtrave.
5. Kryejmë Transformimin Direkt Cosin nga energjitë e 26 bankave për të përfutur 26 koeficientë cepstralë. Për njohjen automatike të zërit, gjysma më e vogël (12-13) nga 26 koeficientët e përfutur, mbahet. 12-13 numrat e mbajtur njësohen si koeficientët cepstralë të Frekuencës Mel. Për këtë arsye, është e qartë që ka shumë hapa të pëfshirë në ASR. Në seksionet e mëposhtme, do të analizojmë mundësinë e reduktimit të hapave të procesit.

Përdorimi i koeficientëve cepstralë të koeficientëve Mel mund të konsiderohet si një nga metodat standard për nxjerrjen e tipareve (Motlíček, 2002). Përdorimi i rreth 20 koeficientëve MFCC është i zakonshëm në ASR, edhe pse 10-12 koeficientë konsiderohen të mjaftueshëm për kodimin e të folurit (Hagen et al 2003). Ana më e dukshme negative e përdorimit të MFCC është ndjeshmëria që kanë ndaj zhurmave për

METODAT E ANALIZËS SPEKTRALE PËR ASR

shkak të varësisë që kanë nga forma spektrale. Metodatat që përdorin informacion në periodicitetin e sinjalit të të folurit mund të përdoren për të anashkaluar këtë problem, edhe pse në të ka dhe element jo periodikë (Ishizuka&Nakatani 2006).

Shkalla jolineare e përdorur është një përafrim i shkallës së frekuencës Mel, ose është pothuajse lineare për frekuencën nën 1kHz dhe logaritmike për frekuencën mbi 1kHz. Kjo motivohet nga fakti që sistemi dëgjimor njerëzor bëhet më pak selektiv në lidhje me frekuencat kur tejkalohet frekuenca 1kHz.

Tiparet MFCC i korrespondojnë kepstrumeve të log të energjive në bankat filtrave. Përderisa sistemi dëgjimor është i ndjeshëm ndaj evolimit kohor të përbërjes spektrale të sinjalit, një përpjekje zakonisht bëhet për të pëfshirë nxjerrjen e këtij informacioni, si pjesë e analizës së tipareve. Në mënyrë që të kapim ndryshimet në koeficientë me kalimin e kohës, koeficientët diferencial të rendit të parë dhe të dytë llogariten respektivisht, me të vetmin ndryshim që hapi i përkuljes së Frekuencës Mel anashkalohet. Për këtë arsye, rangi i dëshiruar i frekuencave implementohet nga një bankë filtrash me 40 filtra të së njëjtës gjerësi dhe të pozicionuar në të njëjtën gjatësi. Gjerësia e bandës për secilin filtër është 164 Hz dhe gjithë banka e filtrave mbulon zonën [133,6857] Hz. Sigurisht, fakti që të gjithë filtrat kanë të njëjtën gjerësi bande e bën të panevojshme normalizimin e zonës poshtë secilit prej tyre.

3.4 Filtrat Gamatone

Sistemet ASR përfshijnë një filtër të ngjashëm me Mel në kontekstin e imitimit të sistemit dëgjimor/auditor të njeriut, që quhet filtri Gamatone. Ky filtër është një përafrim i përpunimit të sinjalit ashtu siç ndodh në sistemin dëgjimor njerëzor që përdoret, nëpërmjet një dizenjimi të dedikuar formash matematikore dhe përgjigjesh frekuencash në mënyrë që t'i përshtaten sistemit auditor njerëzor. Ashtu si MFCC që u diskutuan në seksionin e mësipërm, ky tipar i filtrave Gamatone emërtohet Koeficienti Kepstral të Frekuencës Gamatone (GFCC). Banka e filtrave Gammatone u prezantua për herë të parë nga Roy Patterson dhe kolegët në 1992. Filtrat Gammatone u konceptuan për të imituar thjesht dhe saktë vërtetimet eksperimentale të kërmillit të

METODAT E ANALIZËS SPEKTRALE PËR ASR

veshit tek gjitarët. Filtrat Gamatone kane një strukturë polare repetitive që çojnë në një përgjigje impulsive që është produkt i një ‘mbështjellëse’ Gama $g(t) = t^n e^{-t}$ dhe një sinusoide.

Një filtër Gamatone është një filtër linear ose sekuencial që përfaqësohet nga një përgjigje impuls. Ai përdoret gjërësisht në sistemet ASR. Formalisht paraqitet nga ekuacioni i mëposhtëm:

$$g(t) = at^{n-1}e^{-2\pi bt} \cos(2\pi f_c t + \Phi) \quad (3.8)$$

Ne kete ekuacion, f_c eshte frekuenca qendrore e filtrit dhe Φ eshte faza e cila zakonisht vendoset 0. Konstantja a kontrollon rritjen, dhe n është rendi i filtrit i cili zakonisht vendoset 4 ose më i vogël se 4 (Patterson & Moore, 1986). Ndersa b është faktori i shpërbërjes ose dobësimit i cili i referohet frekuencës f_c dhe jepet nga:

$$b = 1.019 * 24.7 * \left(4.37 * \frac{f_c}{1000} + 1 \right) \quad (3.9)$$

Një set filtrash gammatone me frekuencë të ndryshme qendrore f_c përbëhet nga një bankë filtrash Gamatone. Kjo mund të përdoret dhe të aplikohet për të marrë karakteristikat e sinjalit në frekuenca të ndryshme të cilat na sjellin një paraqitje temporale frekuencash të cilat janë të krahasueshme me FFT-në të bazuar në analizën spektrale me kohë të shkurtër. Në një përpjekje për të imituar sistemin auditor te njeriut, frekuencat qendrore të bankës së filtrave zakonisht shpërndahen në mënyrë të barabartë në nje shkallë fiziko-akustike (bark scale).

Një sfidë madhore për sistemin automatik të njohjes së zërit (ASR) lidhet me reduktimin në mënyre domethënëse të performancës në një ambient me zhurma. Një mori kërkimesh janë kryer në njohjen e zërit për të zbutur keto probleme. Kërkimet egzistuese për të luftuar këtë fenomen përfshijnë hapat e mëposhtëm: normalizimin e kanaleve, përmirësimin e sinjalit, përshtatjen e modelit dhe njohja audio-vizuale. Një studim i kryer ka analizuar nivelet e zbërthimit të parametrave të ‘zhurmës’, kryesisht rreth parametrave të auditorit (Rahali et al, 2014). Kryesisht kjo ndodh pasi karakteristikat e sistemit auditor ngjasojnë me sistemin auditor të njeriut. Kjo

METODAT E ANALIZËS SPEKTRALE PËR ASR

karakteristikë mund të llogaritet për banka filtrash të veçantë të cilat na drejtojnë drejt koeficientëve spektral të frekuencave Mel (MFCC) dhe perceptimi i parashikimit linear (PLP). Një mënyrë më e përgjithshme për të imituar sistemin auditor te njeriut paraqitet në filtrat gammatone (GF), të cilat përfshijnë një strukturë të dedikuar matematikore të frekuencës së përgjigjes që perputhet me rezultatet eksperimentale fizologjike. Në ngjashmëri me MFCC, ky tipar zakonisht i referohet frekuencës së koeficientëve spektral të filtrave Gammatone. Pre-emphasis zakonisht shquhet për ndihmesën në reduktimin e gamës dinamike të spektrit dhe në intensifikimin e komponentët e rëndësishëm të frekuencave të ulëta të cilat zakonisht na japin më shumë material për sinjalet e zërit. Në studimin e tyre Qi et al e implementoi pre-emphasis si një filtër të frekuencave të ulëta të rendit të 2-të që jepet më poshtë, ku b dhe f_s përcaktohen tek (2) dhe (4) përkatësisht (Qi et al, 2013)

$$H(Z) = 1 + 4e^{-\frac{2\pi b}{f_s}} Z^{-1} + e^{-\frac{2\pi b}{f_s}} Z^{-2} \quad (3.10)$$

Me ‘diagramen e kermillit’(Një diagramë me paraqitje kohë-frekuencë të sinjalit në hyrje, duke imituar komponentët e veshit të jashtëm dhe të mesëm. Sinjali ndahet ose coptohet në frekuenca të ndryshme që përzgjidhen nga kërmilli i veshit.) të veshit dhe Transformimit Diskrete Furie (DCT) kërkohet të arrihen komponente të palidhura me tiparet spektrale. Ndërsa DCT-ja mund të aplikohet në ‘diagramën e kermillit’ direkte ashtu si logaritmi mund të aplikohet në diagramën e kermillit në mënyrë të drejtëpërdrejtë; zakonisht zbatohet në procesimin e MFCC. Pavarësisht se nuk ka shumë avantazh teorik me logaritmin, është treguar se kjo na udhëheq në një procesim numerik me stabel (Qi et al, 2013). Ekuacioni i mëposhtëm paraqet formën ekzakte spektrale :

$$F(n, u) = \left(\frac{2}{M}\right)^{0.5} \sum_{i=1}^M \left\{ \frac{1}{3} \log(v(n, i)) \cos \left[\frac{\pi u}{2N} (2i - 1) \right] \right\} \quad (3.11)$$

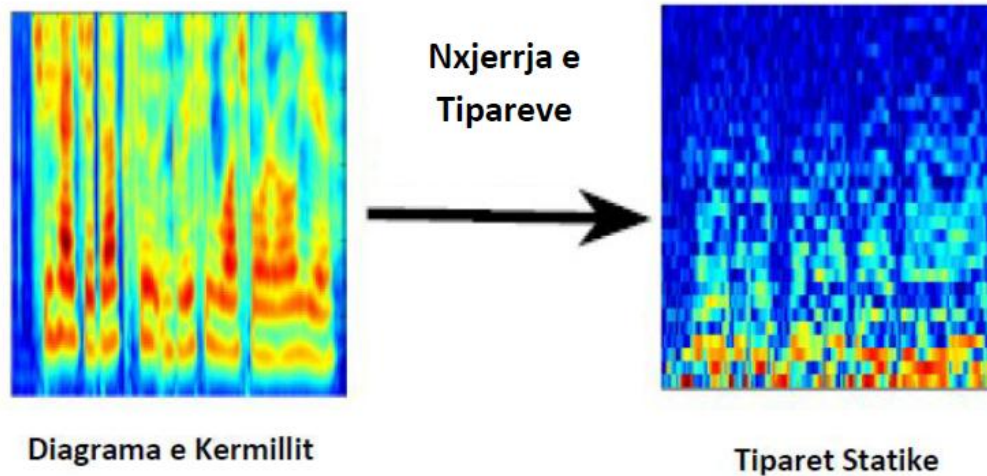


Fig. 18: Diagrama e kërmillit të veshit dhe GFCC (Qi et al,2013)

Me ekuacionin e filtrit gamatone të ndërtuar në boshtin e kohës në seksionin e mëparshëm, parametrat GFCC mund të derivohen nga diagram e kërmillit të filtrit Gamaton. Gjithsesi është treguar se modifikimi i hapave të implementimit mund të na udhëheq në parametra numerik më të mira.

Një testim i njohjes u krye nga Qi et al, që përfshin një studim krahasimor me katër parametra. Rezultatet treguan se GFCC në boshtin kohor i tejkalon në performancë PLP dhe MFCC në çdo ambient me zhurma pothuajse në çdo rast. Kjo gjë tregon se performanca e GFCC-së varet shumë në detajet e implementimit, dhe perfitimet e marra të raportuara nuk janë të thjeshta për tu përshtatur në database të tjera (Qi et al 2013). Figura 19 tregon mesataren kohore të procesimit për një sekondë të sinjalit të zërit. Një kompjuter desktop dual core 2.4GHz dhe një memorie 2GB ram u përdorën për të kryer krahasimin. Kjo figurë jep një ide se sa fuqi procesuese duhet për ASR-në.

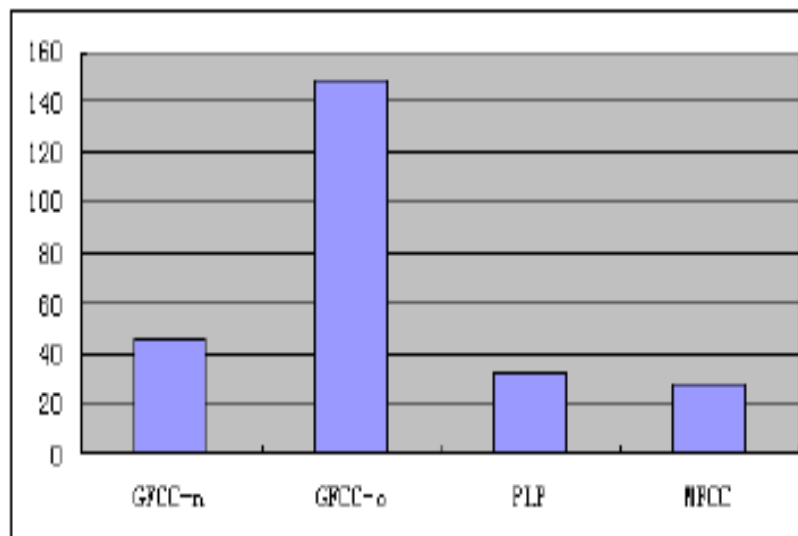


Fig. 19: Koha mesatare e procesimit (në një million sekonda) për një second të zërit

Gjetjet e Qi et. al. treguan se GFCC e boshtit kohor është goxha më e shpejtë se implementimi i boshtit të frekuencës, megjithëse është relativisht akoma më e ngadaltë se MFCC dhe PLP. Një variant i filtrit gamatone, filtri gammachirp është një ASR domethënëse i sistemit të bankës së filtrave, dhe analizohet në seksionin e mëposhtëm.

3.5 Banka e filtrave gamachirp

Filtri gamachirp është një variant i filtrit gamaton. Sistemet e ASR-së përdoren akoma në shumë aplikacione të ditëve të sotme. Si pasojë e zhvillimit të shpejtë në këtë fushë, në të gjithë botën ekzistojnë një shumëllojshmëri sistemesh dhe paisjesh me futjen dhe nxjerrjen e zërit. Një fushë kaq e gjerë kërkon përdorim të shpeshtë të sistemeve të tilla në ambiente me zhurmë. Kryesisht fortësia e zhurmave paraqet një fushë të rëndësishme në aktivitetin kërkimor. Një pengesë kryesore që ndikon në procesimin e sistemeve të zërit është zhurma ambientale dhe efektet e saj të dëmshme në performancën e sistemit. Prezenca e zhurmes zakonisht ul performancën e njohjes së zërit. Një kohë e

METODAT E ANALIZËS SPEKTRALE PËR ASR

konsiderueshme është shpenzuar në këtë zonë dhe ekzistojnë plot teknika që përmisojnë qëndrueshmërinë e sistemit në kushte me zhurma. Në punën e Rehali et al, një filtër auditor që stimulon funksionin e organeve auditor të njeriut është aplikuar për të procesuar përgjigjen e impulsit dhe sinjalet impulsive. Modelet auditore janë kryesisht një bankë filtrash, dhe nuk janë të shpërndara uniformisht në frekuencë; të paisura me gjërësi bandash jo uniforme; të kufizuara në frekuenca të ulëta; dhe tolerante në frekuenca të larta; dhe konverton sinjalin në hyrje në një set nënbandash. Filtri auditor gamachirp është një zgjerim i filtrit Gama; dhe ka një term modulim-frekuence për të prodhuar një spektër amplitude jo simetrike. Filtri Gamachirp ka një përgjigje impulsi më të thjeshtë se modelet fizologjike në mekanikën e kërmillit.

Për më tepër termi ‘chirp’ në fjalën Gamachirp konsiston me observacionet fiziologjike në matjet e modulimeve të frekuencave të përgjigjeve mekanike të membranës kryesore. Irino dhe Patterson, kanë zhvilluar një filtër teorik auditori , gamachirp, parametrat e së cilës mund të zgjidhen për tu përshtatur të dhënave fiziologjike dhe psiko-fizike. Parametrat kryesore në dizajnin e bankës së filtrave janë përgjigja frekuencave të cilat përcaktojnë formën e filtrave; frekuencën qendrore dhe gjerësinë e bandës (Irino & Patterson, 1987)

Këto parametrat e selektuar janë të bazuar në sistemin auditor të njeriut. Filtrat Gamachirp përdorin një përafrim të përgjithshëm të cilat përfshijnë një dizajn të dedikuar të formave matematikore të përgjigjes së frekuencës që përputhen me përgjigjet eksperimentale dhe fiziologjike. Në ngjashmëri me MFCC-në ky tipar zakonisht i referohet si frekuenca e koeficientëve spektral gamachirp. Implementimi i bankës së filtrave gamachirp tregon qëndrueshmëri dhe rritje të performancës në lloje të ndryshme zhurmash dhe nivelesh.

Filtri gamachirp është një përafrim i mirë ndaj sjelljes selektive të frekuencës së kërmillit të veshit. Në një filtër auditor i cili paraqet një asimetri dhe një karakteristik të varur nga niveli i filtrave të kërmillit dhe mund të konsiderohet që mund të jetë një përforsim dhe përmisim i filtrit Gamatone.

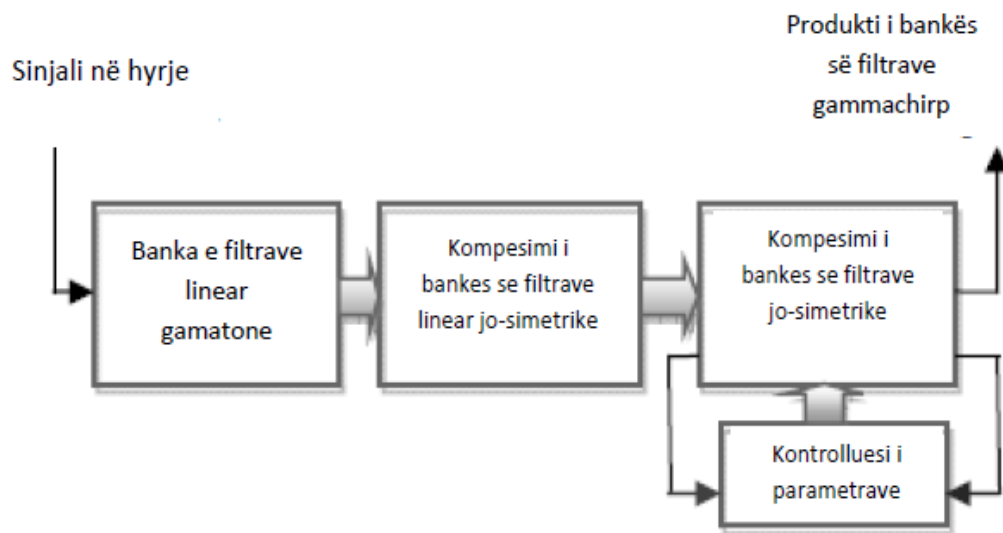


Fig. 20: Ndryshimi ndërmjet filtrave Gamaton dhe Gamachirp (Rahali et al, 2014)

Figura 20 tregon një diagramë të bankës së filtrave Gamatone. Me sa vëmë re, diagrama është një kaskade prej tre banka filtrash: një filter Gamatone, një filter AC me frekuencë të ulët, dhe një filtër AC me frekuencë të lartë. Një studim i kryer nga Rahali et al në parametrizimin e sinjaleve të zërit të bazuar në një bankë filtrash Gamachirp ka ndjekur një metodë të përdorur në teknikën e MFCC dhe PLP. Një sistem për njohjen automatike të fjalëve të izoluar me zhurmë impulsive të bazuar në HMM/GMM u zhvillua. Studimi u përqëndrua kryesisht në analizën dhe dizenjimin e filtrit Gamachirp dhe u përdor me sukses për ‘fuqinë e zhurmës’ në njohjen e zërit. Tingujt u kombinuan me fjalë duke përdorur nivele të ndryshme të sinjalit të zhurmës (20dB, 15dB, 10dB, 5dB, 0dB, dhe -5dB). Pas nxjerrjes së parametrave ato u përdorën në mënyrë krahasimore për të matur performancën me MFCC standarte dhe PLP (Rahali et al, 2014).

Kapitulli 4

4.1 Baza e të Dhënave Aurora

Aurora është një database me ambjent me zhurma të larta të zërit dhe është dizenuar për të vlerësuar performancën e sistemit të njohjes së zërit në ambjente të zhurmshme. Detyrat e Aurorës janë përcaktuar nga Instituti i Standarteve të Telekomunikacionit Europian (ETSI) si një industri celulare me qëllim për të standartizuar një metode për procesimin e zërit dhe strukturën e njohjes së zërit.

Filtri gamachirp u krahasua me bankën e filtrave trekendore Mel dhe u vu re se filtrat Gamachirp janë lehtësisht superior ndaj bankës së filtrave trekendore Mel. Rezultatet më të dobëta u gjeten aty ku është përdorur modeli bazik dhe rezultatet më të mirat u gjetën aty ku kemi përdorur modelin e bankës së filtrave Gamachirp. Kërkimi që u krye në implementimin e modelit Gamachirp të filtrit të kërmillit të veshit u analizua me kujdes. Implementimi u konfirmua nga përdorimi i tij në disa fjalë me zhurma impulsive. Rezultatet e marra pas aplikimit të këtij filtri në fjalët që përmendëm më sipër na tregojnë që ky filter na jep rezultate të pranueshme ose më të mira nga krahasimi me ato metoda parametrizimi, siç janë MFCC dhe PLP (Rahali et al, 2014)

4.2 Transformimi Furie i bankës së Filtrave

Seritë Furie

Në matematikë, një seri Fourier, është një mënyrë për të përshkruar një funksion vale, si shumë e disa valëve të thjeshta sinusoidale. E thënë ndryshe, ato shpërbëjnë çdo funksion, ose sinjal periodik në shumën (mbase infinite) e disa funksioneve të thjeshta luhatëse, sinuso ose kosinusoidale. Transformimi Fourier me kohë diskrete është një funksion periodik, të cilit shpesh i referohen si seri Fourier. Një tjetër rast i serive Fourier është transformimi Z, i cili reduktohet në seri për $|z|=1$. Seritë Fourier luajnë një rol të rëndësishëm në provën e teoremës së grumbullimit Nyquist-Shannon dhe janë degë e analizës Fourier.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Përkufizimi

Funksioni $s(x)$, përfaqëson një funksion të variablilit real x , dhe s është i integrueshëm në një interval $[x_0, x_0+P]$. Ne do të përqijemi të shpërthejmë s në këtë interval si një shumë infinit funksionesh sinusoidale të lidhura në mënyrë harmonike. Jashtë intervalit seritë kanë periodë P (frekuencë $1/P$). Nëse s plotëson një kusht të tillë, atëherë përafrimi është i vlefshëm në të gjithë R . Le të fillojmë me shumën e pjesshme:

$$s_N(x) = \frac{A_0}{2} + \sum_{n=1}^N A_n \sin\left(\frac{2\pi nx}{P} + \theta_n\right), \quad \text{për } N \geq 1 \quad (4.1)$$

$s_N(x)$ është një funksion periodik me periodë P . Duke përdorur shprehjet:

$$\sin\left(\frac{2\pi nx}{P} + \theta_n\right) \equiv \sin(\theta_n) \cos \frac{2\pi nx}{P} + \cos(\theta_n) \sin \frac{2\pi nx}{P} \quad (4.2)$$

$$\sin\left(\frac{2\pi nx}{P} + \theta_n\right) \equiv \operatorname{Re} \left\{ \frac{1}{i} e^{i\left(\frac{2\pi nx}{P} + \theta_n\right)} \right\} = \frac{1}{2i} e^{i\left(\frac{2\pi nx}{P} + \theta_n\right)} + \left(\frac{1}{2i} e^{i\left(\frac{2\pi nx}{P} + \theta_n\right)} \right)^* \quad (4.3)$$

Ky funksion mund të shkruhet dhe në format ekuivalente të mëposhtme:

$$s_N(x) = \frac{A_0}{2} + \sum_{n=1}^N \left(\underbrace{A_n \sin(\theta_n)}_{\widetilde{a}_n} \cos \frac{2\pi nx}{P} + \underbrace{A_n \cos(\theta_n)}_{\widetilde{b}_n} \sin \frac{2\pi nx}{P} \right) =$$

$$\sum_{n=-N}^N c_n e^{i\left(\frac{2\pi nx}{P}\right)} \quad (4.4)$$

Ku:

$$c_n \stackrel{\text{def}}{=} \begin{cases} \frac{A_n}{2i} e^{i\theta_n} = \frac{1}{2}(a_n - ib_n) & \text{per } n > 0 \\ \frac{1}{2} A_0 = \frac{1}{2} a_0 & \text{per } n = 0 \\ c_{|n|}^* & \text{per } n < 0 \end{cases} \quad (4.5)$$

Lidhja inverse ndërmjet koeficientëve është :

$$A_n = \sqrt{a_n^2 + b_n^2} \quad \theta_n = \operatorname{atan2}(a_n, b_n) \quad (4.6)$$

Dhe koeficientët Fourier llogariten si më poshtë:

METODAT E ANALIZËS SPEKTRALE PËR ASR

$$a_n = \frac{2}{P} \int_{x_0}^{x_0+P} s(x) \cos\left(\frac{2\pi nx}{P}\right) dx \quad (4.7)$$

$$b_n = \frac{2}{P} \int_{x_0}^{x_0+P} s(x) \sin\left(\frac{2\pi nx}{P}\right) dx \quad (4.8)$$

$$c_n = \frac{1}{P} \int_{x_0}^{x_0+P} s(x) e^{\frac{-i2\pi nx}{P}} dx \quad (4.9)$$

$S_n(x)$ është përafrim i $s(x)$ për $[x_0, x_0+P]$, dhe ky përafrim përmirësohet kur $N \rightarrow \infty$. Shuma e pafund $s_\infty(x)$, quhet seri Fourier përfaqësuese e s . Në aplikimet inxhinierike, seritë Fourier presupozohet se konvergjojnë kudo përveç se në pavazhdueshmëri, duke qenë se funksionet e hasura në inxhinieri kanë një sjellje më të qëndrueshme se ato që mund të sjellë matematika si kundërsimbuj. Në veçanti seritë Fourier konvergjojnë absolutisht dhe në mënyrë uniforme me $s(x)$, sa herë që derivati i dytë $s''(x)$, i cili mund të mos ekzistojë në çdo pikë është i integrueshëm. Nëse katrori i një funksioni është i integrueshëm në intervalin $[x_0, x_0+P]$, seritë Fourier konvergjojnë drejt funksionit pothuaj në çdo pikë. Konvergjenca e serive Fourier, varet gjithashtu edhe nga numri i fundëm i maximumeve dhe minimumeve në një funksion, icili njihet masivisht si kushti Dirichlet për seritë Fourier. Për funksione më gjenerikë, ose shpërndarje gjenerike mund të përcaktojmë koeficientët Fourier, dhe në këto raste konvergjenca në normë ose e dobët është me interes.

Teorema e integralit të anasjelltë Furie

Teorema e Fourier të anasjelltë thotë që nëse dimë transformimin Fourier të disa funksioneve, mund të gjenerojmë funksionin bazë. E thënë ndryshe, nëse dimë të gjitha frekuencat dhe fazat, mund të rindërtojmë valën fillestare në mënyrë identike.

Teorema thotë që nëse kemi një funksion $f: \mathbb{R} \rightarrow \mathbb{C}$, i cili përmbush disa kushte të caktuara, të cilat i përdorim në transformimin Fourier si:

$$(Ff)(\xi) := \int_{\mathbb{R}^n} e^{-2\pi i y \cdot \xi} f(y) d(y), \quad (4.10)$$

Atëherë:

$$f(x) = \int_{\mathbb{R}^n} e^{2\pi i y \cdot \xi} F(f)(\xi) d(\xi) \quad (4.11)$$

Me fjalë të tjera, Teorema thotë që:

$$f(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{2\pi i (x-y) \cdot \xi} f(y) d(y) d\xi \quad (4.12)$$

Ekuacioni i fundit përfaqëson teoremën e Fourier e integralit të anasjelltë.

Lidhja me seritë Furie

Kur marrim në konsideratë seritë Fourier të një funksioni, zakonisht përcaktohet që ato veprojnë në shkallën $[0, 2\pi]$, ose e thënë ndryshe kanë një periodë 2π . Në rastin më poshtë supozojmë që f vepron në segmentin $[0, 1]$.

Teorema e integralit të anasjelltë Fourier është analoge me konvergjencën e serive Fourier. Në rastin e transformimit Fourier kemi:

$$f: \mathbb{R}^n \rightarrow \mathbb{C}, \quad \hat{f}: \mathbb{R}^n \rightarrow \mathbb{C} \quad (4.13)$$

$$\hat{f}(\xi) := \int_{\mathbb{R}^n} e^{-2\pi i y \cdot \xi} f(y) d(y), \quad (4.14)$$

$$f(x) = \int_{\mathbb{R}^n} e^{2\pi i y \cdot \xi} \hat{f}(\xi) d(\xi) \quad (4.15)$$

në rastin e serive Fourier kemi:

$$f: [0, 1]^n \rightarrow \mathbb{C}, \quad \hat{f}: \mathbb{Z}^n \rightarrow \mathbb{C} \quad (4.16)$$

$$\hat{f}(k) := \int_{\mathbb{R}^n} e^{-2\pi i y \cdot k} f(y) d(y), \quad (4.17)$$

$$f(x) = \sum_{k \in \mathbb{Z}^n} e^{-2\pi i y \cdot k} \hat{f}(k) \quad (4.18)$$

ku k është numër i plotë dhe shuma e funksionit varion nga $-\infty$ në ∞

Një nga metodat më të përshtatshme dhe më efikente për llogaritjen e një modeli pa hapësirë dhe jolinear të një banke filtrash të sinjalit të zërit është aplikimi i një transformimi Furie në sinjal. Një transformim Furie na krijon mundësinë për të llogaritur daljen e transformuar në frekuencën e dëshiruar.

4.3 Transformimi Furie Diskrete

Transformimi Diskrete Furie (DFT) është një nga makineritë matematikore më të besueshme dhe më të rëndësishme në procesimin e sinjalit. DFT-ja e një sinjali përcaktohet si më poshtë:

$$S(f) = \sum_{n=0}^{N_s-1} s(n) e^{-j\left(\frac{2\pi f}{f_s}\right)n} \quad (4.19)$$

DFT-ja përdoret për të kampjonuar spektrin në një rreze të caktuar frekuencash. Fatkeqësisht spektri është i tejkampjonuar në një cilësi më të mirë dhe cdo output i bankës së filtrave në të cilën madhësia e fuqisë spektrale procesohet si një shumë e peshuar e vlerave të përafërta. Theksojmë fjalën ‘fatkeqësisht’ pasi përveç avantazheve që na sjell tejkampionimi siç janë : ulje të nivelit të zhurmave, dhe shtrëmbërimet fazore. Ndërsa ndër disavantazhet mund të përmendim rritjen e kohës së përpunimit të sinjalit gjatë kampionimit (e cila mund të na çojë dhe në rritje kostosh) dhe rritjen e nivelit të zhurmave në rastin kur i rikthehemi sinjalit origjinal nëpërmjet interpolimit.

DFT-ja përdor mesatarizimin për të implementuar zbutjen spektrale të një funksioni. Mesatarizimi është një teknikë e përdorur gjërësisht për zbutjen spektrale. Teknika e mesatarizimit përdoret shpesh në boshtin e frekuencave Mel nëse një DFT përdoret pasi

METODAT E ANALIZËS SPEKTRALE PËR ASR

ngarkesa kompjuterike e shtuar është minimale. Duke kryer mesatarizimin në boshtin e logaritmeve ose në boshtin e vlerave të fuqisë logaritmike të propozuar në amplitudat spektrale, është në dobi për analizën spektrale të treguar në figuren 21.

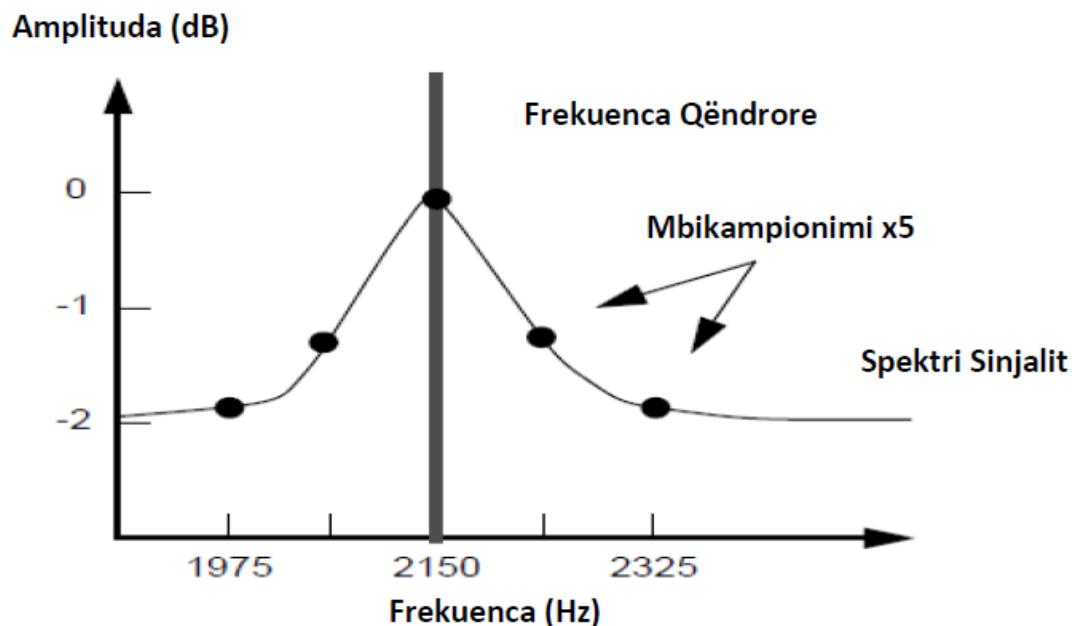


Fig. 21: Impakti i mesatarizimit të boshtit log në analizën spektrale (piccone,1993)

Kombinimi i tingujve të pastër ose i valëve sinusoidale rezulton në një kompleksitet tingujsh. Një analizë frekuencash i një tingulli të tille shpesh përpqet që të identifikojë tingujt e pastër. Transformimi Furie u dizenjua nga matematikani francez Fourier në vitin 1820 dhe qëndron akoma edhe sot si një nga metodat primare për të zhvilluar analizën e frekuencave të tingullit si dhe të fenomeneve të tjera. Një numër i ndryshëm rrugësh për implementimin e transformimit Furie janë zhvilluar duke përfshirë dhe transformimin diskret Furie (DFT) dhe transformimin e shpejtë Furie (FFT) dhe transformimin e Rrallë Furie (SFT). Këto janë projektuar për të punuar me sinjale dixhitale, siç janë sinjalet e zërit.

4.4 Transformimi i Shpejtë Furie

Ndër teknikat e transformimit Furie, algoritmi i transformimit të Shpejtë Furie (FFT) është një ndër metodat më të përdorura. FFT u shpik gjate viteve 1960, dhe llogarit frekuencën e sinjalit të madhësisë N në kohën $O(N \log N)$. Një transformim i Shpejtë Furie (FFT) mund të përdoret si një metodë alternative për procesimin e spektrit të sinjalit. FFT është një vendosje efektive e Transformimit Diskret Furie nën kufizimet që spektri të vlerësohet në një platformë frekuencash diskrete që janë shumfishe të :

$$\frac{f_s}{N} \quad (4.20)$$

Këto frekuenca konsiderohen si frekuenca ortogonale. Avantazhet kryesore të FFT është që është shumë e shpejte. Është normale që të shtojmë një hap shtesë procesimi. Supozohet që zonat me vibrim maksimal ose amplitudë maksimale u jepet më shumë peshë në sistemin e dëgjimit njerzor që janë zona me amplituda të ulëta. Piccone,1993). Për më tepër në ambjente të zhurmshme, zhurma në sfond përpiket të ndikojë negativisht në vlerësimet tona në zonat spektrale me amplitudë të ulët. Me fjalë të tjera, ne përpiqemi të jemi më të vetëdijshëm në besueshmërinë e vlerësimeve tona për zonat me amplitudë të lartë të spektrit. Për shkak të kësaj, zakonisht vendoset një limit në rrezet dinamike të spektrit. Ky limitim i poshtëm zakonisht quhet pragu i gamës dinamike. Duke zgjedhur një prag specifik nga maksimumi i spektrit, konsiston në atë që ne thjesht heqim vlerësimet poshtë një pragu nga pika më e lartë e një spektri. Për të implementuar algoritmin e pragut, është e nevojshme që vlerat spektrale të jenë të rrumbullakuara në lidhje me shpërndarjen. Nëse ky hap nuk ndërmeret zonat me energji të ulët që janë produktive mund të shkatërrohen aksidentalisht. Në lidhje me faktin se spektri i sinjalit audio i tingullit të njeriut bie 20 dB cdo 10 vjet, vendosja e një pragu të instaluar në energjinë e frekuencave të ulëta është realiste. Kjo ndodh për faktin se kur nga amplituda spektrale nga më e vogla tek më e madhja diferenca e rrezes është

METODAT E ANALIZËS SPEKTRALE PËR ASR

mjaftueshmërisht e madhe, energjia e audios në frekuenca më të larta mund të shkatërrohet. Për më tepër FFT përdor një teknikë efikase për vendosjen e një pragu dhe mbrojtjen e të dhënave më të rëndësishme të sinjalit të zërit. Ekzistojnë metoda të tjera për të përcaktuar prapun e spektrit duke përdorur modele parametrike.

4.4.1 Shpejtësia e FFT

FFT përdoret gjerësisht dhe u konsiderua që të jetë një nga algoritmet më të rëndësishme dhe më të fuqishme të shekullit të 20. Aktualisht të dhëna të ngarkuara përdoren për sistemet ASR. Kjo na udhëheq drejt përmasave të procesimit të cilat mund ti kalojnë terrabajtet. Llogaritja e GPU ka kapacitetin për të bërë aplikacionet që të punojnë më shpejtë duke kombinuar një procesues grafik (GPU) me një CPU. Edhe duke përdorur këto metoda mund të ndodh që për shkak të madhësive të informacionit metoda e transformimit të shpejtë Furie mund të mos jetë mjaftueshëm e shpejtë. Një problem tjetër potencial lind nëse platforma e të dhënave është shumë e vogël. Për shembull në një ASR është e rëndësishme modernizimi i procesit të maksimizimit të cilësisë dhe minimizimi i kohës së peshuar. Për më tepër, ka nevojë për metoda dhe algoritme që mund të llogarisin transformimin Furie në një kohë minimale ose në një afat kohor që është goxha më e vogël se madhësia e të dhënave. Këto metoda kane potencialin për të përdorur vetëm një nën-platformë të dhënash të inputit që shërben si bazë për outputin. SFT-ja na siguron pikërisht këtë funksionalitet.

SFT(Transformimi Furie në Të Rrallë) është një algoritëm i ri i cili mund të procesojë një mori të dhënash nga 10 deri në 100 herë më shpejt se FFT. Ky transformim nuk mund të përdoret në çdo lloj të dhënash por vetëm në ato grupe të dhënash që kanë një strukturë të veçantë. Për shembull të dhënat e sinjaleve të zërit, muzikës apo të sinjaleve video mund të marrin vetëm disa vlera të veçanta. Pikërisht duke u bazuar tek këto vlera të veçanta të sinjaleve që sapa përmendëm më lart konsiston funksionaliteti dhe shpejtësia shumë e lartë e kësaj metode. Theksojmë që përdorimi i kësaj metode është më i limituar se i FFT.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Zhvillimet në teknologjinë kompjuterike do të vazhdojnë të rrisin standartet e asaj që është e mundur. Një FFT zakonisht na siguron në një masë të madhe të detajeve spektrale. Kur një algoritëm FFT është duke punuar, shumica e maksimumeve spektrale të spektrit të zërit të cilat lidhen me frekuencat tingëluese të traktit të fytit dhe janë të mbivendosura në harmonikat e mirë detajuara të cilat janë shumfishe të frekuencave themelore. Kjo na bën të mundur të ekzaminojmë modelin spektral të një spektri pa ndërhyrjen e një modeli harmonik të tingullit të folur ose një model çfardo jo i folur nga njeriu. Në këtë seksion komponentet bazike dhe teknikat e përdorura në algoritmet SFFT janë të përshkruara. Fillimisht spektri i sinjalit konsiston ose është i dominuar, nga vetëm një frekuencë jo-zero. Në këtë rast, mund të tregohet se pozicioni dhe vlerat e frekuencës jo-zero gjenden duke përdorur vetëm dy kampjone të sinjalit nëse sinjali është një tingull i pastër pa asnjë zhurmë ose një selektim kampjonesh logaritmike nëse sinjali degradohet nga zhurmat. Së dyti, është e mundur që t'i adresohet rasti i përgjithshëm duke reduktuar në disa nën-probleme dhe duke përfshirë një frekuencë të vetme jo-zero. Kjo arrihet duke grupuar nëngrupet e hapsires Furie në një numër të vogël 'vatrash'.

Analiza e frekuencës së sinjalit kryhet në një hapsirë diskrete që përfshin FFT. FFT përcaktohet nga një numër pikash (p.sh 1024) gjë që na jep shumë frekuenca diskrete me rezultate përkatëse për çdo pikë. Këtyre frekuencave diskrete ne do tu referohemi si 'vatra' ose kosha thënë më ndryshe. Për shembull nëse ne kemi pika frekuencash të përcaktuara në interval prej 1 Hz, atëherë këto mund të konsiderohen si qendrat e 'vatrave'. Vatrata mund të jenë : 0.5-1.5, 1.5-2.5, etj.

Në rastin më të thjeshtë çdo vatër korespondon me një bandë frekuence. Ndërkohë mund të krijohen metoda të tjera të etiketimit. Nëse çdo vatër përmban vetëm një frekuencë të vetme dhe kjo frekuencë është e izoluar; atëherë problemi mund të zgjidhet ekskluzivisht për çdo vatër duke përdorur metodat e përmendura më sipër. Kjo na drejton drejt kompleksitetit të kampionit dhe të kohës së punës përpjestimore me numrin e vatrave, të cilat kufizohen nga poshtë nga sparsiteti i parametrin k . (k -na paraqet rendin e modelit)

METODAT E ANALIZËS SPEKTRALE PËR ASR

Ku $k = n/\log n$ (4. 21)

Është e mundur që të ndikojmë te spektri i sinjalit duke ndryshuar rradhën e boshtit të kohës së sinjalit. Meqënëse vendodhjet e frekuencave jo-zero në sinjalet e riorganizuar janë të çfardoshme, është e mundur që fillimisht të izolohet frekuencat e sinjalit dhe më pas të rikuperohen duke përdorur procedurën e ‘vatrave’. Për tu siguruar që të gjithë koeficientët jo-zero janë rikuperuar, procedura e riorganizimit dhe e mbledhjes (tek vatrat) të përsëritet disa herë, duke përdorur rastësinë çdo herë në organizimit e të dhënave.

Të gjitha algoritmet SFTT mbështeten tek kjo teknikë e përgjithshme. Në këtë seksion ne japim një paraqitje të përdorimit të të dhënave të algoritmit të FFT-së dhe të teknikave që kanë dalë në dritë këto kohët e fundit. Këto aplikacione përfshijnë për shembull: GPS, dhe dixhitalizimin e çdo sinjali analog.

Në një problem sinkronizimi me GPS jepet një kod i cili korrespondon me një satelit të veçantë. Sateliti në mënyrë të vazhdueshme transmeton kodin. Për më tepër neve na jepet një sinjal i cili regjistrohet nga një marrës GPS që konsiston me një pjesë të sinjalit i cili gjenerohet nga sateliti, i korruptuar me zhurmë. Qëllimi është që të lidhim kodin me sinjalin e regjistruar. Kështu që është e nevojshme të indentifikojmë se ku fillon dhe ku mbaron kodi. Në fund të kodit, marrësi llogarit konvolucionin e kodit dhe të sinjalit të marrë, dhe raporton tendencën që maksimizon lidhjen ndërmjet kodit dhe sinjalit. Ky veprim kryehet duke përdorur FFT. Ne përdorim FFT tek kodi dhe sinjali dhe më pas llogarisim produktet e daljes/output, dhe aplikojmë inversin e FFT tek produkti (sinjali në dalje).

Teknikat SFFT mund të përdoren për të përshpejtuar proceset. Ky përmirësim bazohet në vëzhgimet e mëposhtme: meqënëse outputi i inversit të FFT përmban një maksimum që korrespondon me zhvendosjen ekzakte, hapi Invers mund të implementohet duke përdorur SFTT. Faktikisht meqënëse $k=1$, algoritmi është i thjeshtë, dhe bazohet në një prezantim të deformimit ose gabimit gjatë pastrimit të një sinjali nga zhurmat. Për më tepër meqënëse algoritmi Invers i SFFT përdor vetëm disa nga kampionet e produktit, struktura e tij llogarit vetëm ato kampione. Avantazhet e kësaj metode janë reduktimi i

METODAT E ANALIZËS SPEKTRALE PËR ASR

kostove të hapave të mëtejshëm. Të gjitha eksperimentet me sinjale reale na tregojnë se nëse dimë vlerat e zhvendosjes Doppler algoritmi i ri redukton medianen me një faktor 2.2 ose më tepër.

4.5 Transformimi Furie në të rrallë

SFT është gjetja më e fundit në familjen e FFT. Është zhvilluar në dekadën e fundit dhe algoritmet SFT llogarisin një përafrim ose një version të kompresuar të transformimit Diskret Fourier (DFT) në kohe proporcionalisht me shpërhapjen e spektrit të sinjalit të zërit. Frekuencat dominante të sinjalit të zërit janë ato që përcaktojnë spektrin më shumë sesa gjatësia e sinjalit. Algoritmet e përdorura nga SFT aplikohen vetëm në një numër të reduktuar të nëngrupeve të të dhënave pasi siç e shpjeguam në fillim SFT ajo zbatoheshte vetëm për të dhëna me karakteristika të veçanta duke i mënjeluar të dhënat e tjera.

Kjo teknikë bën të mundur që SFT të jetë goxha më e shpejtë, në vënd që të merremi duke përcaktuar spektrin duke u bazuar në gjatësinë e sinjalit. SFT realizohet duke kërkuar algoritmat që raportojnë vetëm frekuencat jo-zero ose frekuencat e mëdha brenda amplitudave të përbëra ose frekuencave të mëdha. Ajo që bie në sy në metodën e bazuar në algoritëm është që një vektor të përmbajë të njëjtin informacion duke raportuar në çdo frekuencë. Duke u bazuar në faktin se sinjalet audio të zërit dhe matjet spektroskopike të cilat prodhojnë të dhëna masive janë në gjëndje të kompresohen ose të rrallohen. Këto rezultate të arrira nëpërmjet SFT na garantojnë një impakt praktik në shumë boshte më të dhëna të shumta. Qëllimi i një studimi analitik të spektrit është që të skanojë spektrin në dispozicion dhe të indentifikojë hapësirat ndërmjet frekuencave që janë në përdorim. Në shumë aplikime kjo detyrë ka nevojë që të kryhet në mënyrë efçente, meqënëse spektri ka tendencë që të ndyshojë në mënyrë dinamike. Fatkeqsisht skanimi i një spektri të gjerë në GHz është një operacion i cili kërkon konsumimin e shumë energjie. Për të reduktuar fuqinë dhe kohën e marrjes , është e mundur që të

METODAT E ANALIZËS SPEKTRALE PËR ASR

përdoret një SFT për të llogaritur frekuencën përfaqësuese të një sinjali të rrallë dhe ta neglizhojmë atë që të kampjonohet në tërë gjerësinë e bandës.

4.6 Transformimi Wavelet (WD)

Transformimi Wavelet konsiderohet që të jetë i përshtatshëm për përpunimin e zërit për shkak të ngjashmërisë së tij me mënyrën se si veshi i njeriut e proçeson zërin. Tre metodat e transformimit WD të përdorura janë: Transformimi Diskret Wavelet (DWT), modeli i dekompozimit i transformimit Wavelet (WPD) dhe modeli Diskret i Dekompozimit të Transformimit Wavelet (DWPD) të cilat jepen më poshtë. Informacionet rreth sinjaleve jo-stacionare siç janë sinjalet audio mund të ekstrahohen duke përdorur DWT pasi është një metodë e kohëve të fundit dhe një teknikë efiçente për nxjerrjen e parametrave. WPD është thjesht një përgjithësim i DWT dhe është një metodë më fleksibël dhe më e detajuar se DWT. Në DWT, sinjali i zërit dekompozohet në komponentë me frekuenca të ulëta dhe në komponentë me frekuenca të larta në çdo nivel si në DWT, por diferenca kryesore ndërmjet DWT dhe WPD është se Transformimi Diskret Wavelet është paraqitur në rezultatet me frekuenca të ulëta. WPD ndryshon pasi aplikon hapin e transformimit si në rezultatet me frekuenca të ulëta ashtu dhe në rezultatet me frekuenca të larta. Avantazhi kryesor i DPWD është se ai dekomponon edhe bandën e frekuencave të larta në më shumë kampione por për më tepër ruan kompleksitetin në llogaritje. Teknikat e WD kanë treguar që përmirësojnë efiçencën e sistemeve ASR (Hibare&Vibhute 2014).

Kapitulli 5

5.1 Përmirësimet në sistemin ekzistues të ASR

Është e nevojshme zhvillimi i teknikave hibride që do të na udhëheqin në aplikacione me performancë të lartë të ASR. Në mënyrë për të pasur një saktësi më të madhe, modeli gjuhësor, proçesimi i tekstit dhe i shqiptimit ka nevojë akoma për shumë

METODAT E ANALIZËS SPEKTRALE PËR ASR

kërkime dhe inovacione që janë të nevojshme. Filtrat Mel; parashikimi linear (LP) dhe filtrate Gammatone kanë treguar që janë efikase për problemet e zërit. Siç u diskutua më parë, për MFCC-të u tregua që zakonisht llogariten duke integruar termat e shkurtër (të fuqisë spektrale duke përdorur një bankë filtrash Mel (MelFB) që zakonisht konsiston në mbivendosjen e filtrave trekëndor me ata të tipit GFCC që është filtri ekuivalent Gamatone. Me terma të shkurtër të fuqisë spektrale do të kuptojmë ato sinjale të cilat merren si rezultat i analizës që i bëhet një sinjali të vetëm nga një bashkim filtrash të cilët janë selektiv te disa frekuenca të caktuara. Të dyja MFCC dhe PLP performojnë kënaqshëm në kushte testimi por diferenca e performancës midis njohësve automatik të zërit dhe dëgjuesve njerzor në botën reale është shumë e madhe (Huang & Acero,2001). Kushtet e ndryshme të operimit gjatë marrjes së sinjalit siç janë përgjigja e kanaleve, receptori i telefonave, zhurma e ambjenteve, jehonat e të tjera ... na çon tek mosnjohja korrekte e karakteristikave.

Transformimi i një frekuence lineare në një frekuence jo-lineare quhet përkulje (warping). Tregohet që përkulja e drejtpërdrejtë e DFT-së në vend që të përdorim mesatarizim të bankës së filtrave na jep një përafrim më preçiz të shkallës së perceptimit (QI et al 2013). Ky ishte një përfundim i studimit në lidhje me shtimin e zhurmës degraduese në sistemet ASR. Ekzistojnë kërkime të shumta në përmirësimin e ASR-së në ambjente akustike të ndryshme. Transformimi i Shpejtë Furie (FFT) dhe Transformimi Diskrete Furie (DFT) janë në thelb njësoj; me diferencën e vetme që FFT-ja është më i shpejtë. Përdorimi i FFT-së ose DFT-së kanë treguar që japin gabime shumë më pak se koeficientët spektral të DFT-së.

Në metodën tradicionale të MFCC-së, procesimi i një sinjali zëri fillon me fazën para-përpunuese. Kjo përfshin heqjen e DC-së dhe pre-emphasis duke përdorur një filtër të rendit të parë me frekuenca të larta me një funksion transferimi të ndjekur nga një transformim Furie të aplikuar siç u diskutua më lartë. Një menyrë për të arritur përkulje të frekuencës realizohet duke aplikuar një bank filtrash jolineare, siç janë bankat e filtrave Mel, në paraqitjen e frekuencës lineare. Një mënyrë tjetër është të përdorësh një ‘conformal mapping’ siç janë transformimet bilineare të cilat ruajnë njësinë e rrethit.

$$H(z) = \frac{z^{-1} - \alpha'}{1 - \alpha' z^{-1}} \quad -1 < \alpha' < 1 \quad (5.1)$$

Ekuacioni mësipërm është një shëmbull i përkuljes së DFT. DFT-ja arrihet duke aplikuar algoritmin e FFT-së. Në përkuljen e DFT ose FFT maksimumet e pozicionit të frekuencave modifikohen duke përdorur një transformim që lejon çdo lloj frekuence për të ‘përkulur’ boshtin e frekuencave. Më pas pikat e ndara uniformisht në boshtin e frekuencave të përkulura janë të ngjashme me maksimumet e hapsirave jo-uniforme të boshtit të frekuencës origjinale. Duke zgjedhur parametrat e përkuljes me vëmendje, ne mund të vendosim disa nga kampionet e frekuencave në distanca të afërta me njëra tjetrën për të marrë një zgjidhje më të mirë në rrezen e frekuencave të interes (pa arritur gjatësinë e DFT-së, duke përdorur përkuljen e frekuencave ne mund të përmisojmë paraqitjen spektrale të sinjalit të zërit në rajonin e frekuencave të ulëta.

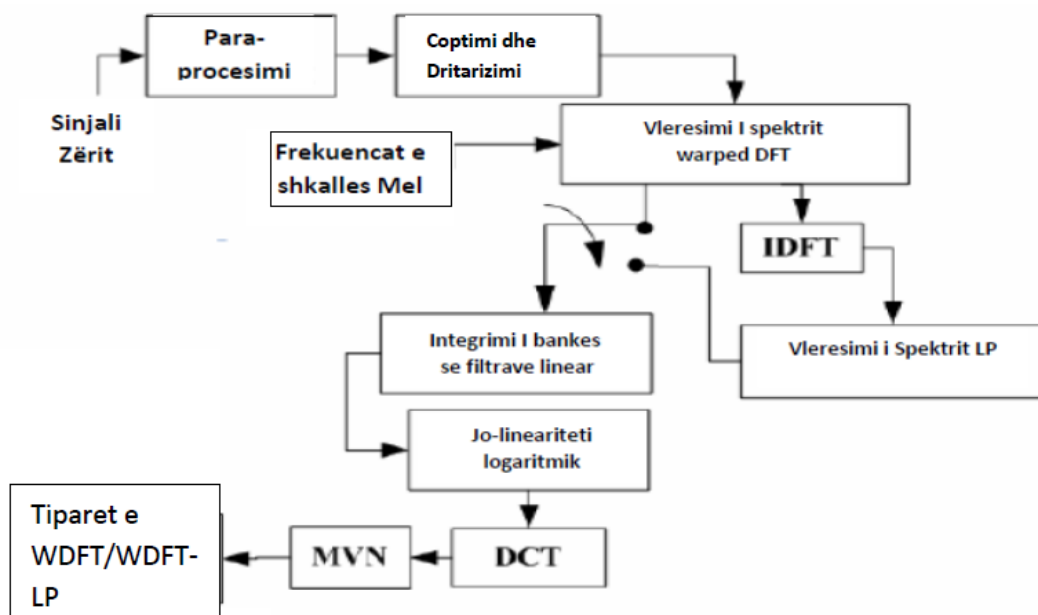


Fig. 22: Teknika Hibride e nxjerrjes së Tipareve

Siç përmendet nga (Yapanel, 2008) duke përkulur spektrin e DFT-së direkte, pa përdorur mesatarizimin e bankës së filtrave, na siguron një përafrim më të saktë të shkallës së perceptimit. Meqënëse spektri afërsisht i ‘para-përkulur’ duke përdorur

METODAT E ANALIZËS SPEKTRALE PËR ASR

përkuljen e frekuencës Mel banka e filtrave Mel në formë trekëndore dhe duke e ndarë në mënyrë jolineare, zëvendësohet nga një bankë filtrash e shpërndarë në mënyrë uniforme. Teknika ‘para-përkulur’ (pree warping) është një teknik e cila zbatohet në frekuencat e sinjalit para se sinjali ti nënshtrohet analizës spektrale në mënyrë që të kemi një përshtatje dhe konvertim më të shpejtë dhe më të saktë të frekuencës nga lineare në të përkulur. Këta filtra trekëndor janë gjysëm të mbivendosur, për të siguruar zbutjen spektrale. Figura më poshtë na tregon spektrin e zërit aktiv të (a) pastër dhe (b) sinjalit të zërit me zhurma që janë kontaminuar nga zhurmat e flluskave me një sinjal me nivel zhurme deri në 6dB, që merret duke përdorur vlerësuesit e spektrit DFT, WDFT (Warped DFT) dhe WDFT-LP. Bazuar në këtë ekzaminim visual, WDFT dhe WDFT-LP na sigurojnë vlerësime spektrale më të mira krahasuar me metodat DFT dhe LP. Si shkallë e reduktimit të shkallëve të lirisë në modelimin me një numër të çfarëdoshëm polësh (modeli i rendit me koeficient $p=24$ përkundrejt $N=256$ vatra), spektrat WDFT-LP janë përgjithësisht më të zbutur se WDFT. Kjo gjë na çon në përmisimin e qëndrueshmërisë së zhurmës mbi WDFT.

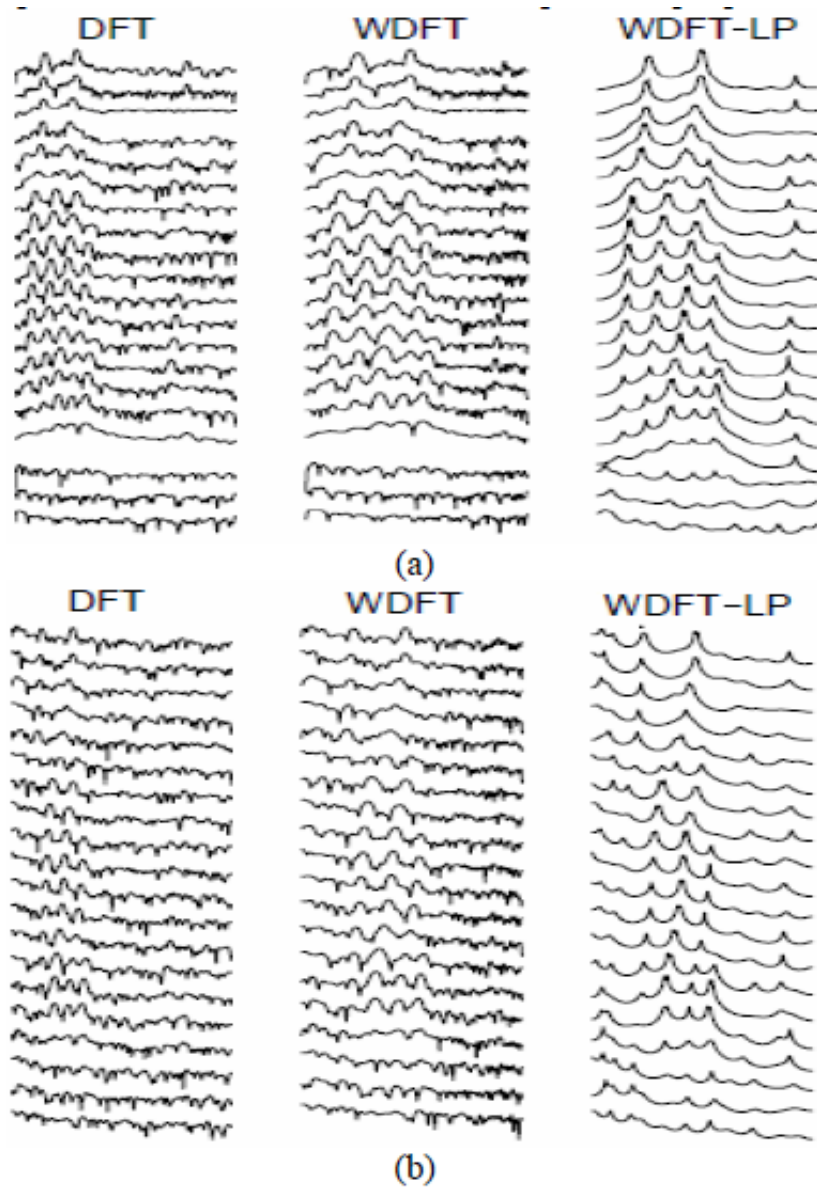


Fig. 23: Krahasimi ndërmjet dy spektrave të sinjalit të zërit që janë kontaminuar me zhurma fluskash (a) të pastër (b) me zhurma

Përkulja e DFT (WDFT) dhe WDFT –parashikimi linear (WDFT-LP) të bazuar në ekstraktorët spektral të parametrave, të paraqitura si në figurën më lartë, u vlerësuan dhe u krahasuan me MFCC tradicionale, PLP front-ends në korpusin Aurora-4 në kontekstin e njohjes së zërit. Të dyja modelet e trajtimit të pastër dhe shumë stilësh u analizuan në

METODAT E ANALIZËS SPEKTRALE PËR ASR

kërkimin primar. Norma e gabimit (WER) është përdorur si një vlerësim metrik. WER është një algoritem i cili na bën një vlerësim metrik të fjaleve të bazës së të dhënave të cilat makina e përkthimit nuk i përkthen në mënyrë të saktë. Pra shkurtimisht themi që WER na jep një vlerësim të performancës të një sistemi ASR. WER mund të llogaritet si më poshtë :

$$WER = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}$$

S- numri i sinonimeve të fjalës

D- numri i fjalëve të fshira nga sistemi

I- numri i fjalëve të vendosura në sistem

C- numri i fjalëve të sakta të përdorura

Variante të frekuencës Mel të përkulja e DFT(FFT) , u gjet që janë me frekuencë më të qëndrueshme ‘përkulje’ të bazuar në tiparet spektrale të paraqitura si në tabelën 1 dhe 2 më poshtë. Nga të dhënat në Tabelën 2 mund të shikohen parametrat spektral të bazuar në WDFT që në përgjithësi performuan më mirë se DFT e bazuar në MFCC dhe PLP. Duke krahasur rezultatet të Tabelës 1 dhe 2 mund të thuhet qartësisht se parametrat spektral të bazuar në WDFT kanë performuar më mirë se MFCC dhe PLP të dyja në modelet e pastra dhe disa-llojesh. Parametrat MFCC të llogaritura duke u bazuar nga spektri ‘front end’ dhe Mel WDFT na siguruan gabime më të vogla se MFCC-ja konvencionale dhe PLP-ja në korpusin Aurora-4.

‘Front end’ është një metodë efikase që përdoret tek sistemet ASR dhe që shërben për të performuar një analizë linguisitike dhe fonetike të sistemit

Detyra e LVCSR na tregoi qëndrueshmërinë e WDFT dhe WDFT-LP të bazuar në parametrat spektral.

LVCSR është një fjalor i madh që shërben për njohjen e zërit të vazhdueshëm në sistemet ASR.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Tabela 1: Norma e gabimit të fjalëve (WER ne %) të marra nga tipare të ndryshme të nxjerra nga ky punim, në korpusin Aurora-4 LVCSR në kushte të një formimi të pastër. Rendi i modelit të selektuar në këtë punim është : p=24 për WDFT-LP dhe p=14 për PLP. Sa më e vogël të jetë WER aq më e mirë është performanca e ‘nxjerrësit të tipareve’.

	A	B	C	D	Mesatarja
MFCC	9.98	50.81	28.88	64.55	38.56
PLP	10.28	49.59	25.56	60.36	36.45
WDFT-MFCC	10.90	49.07	24.27	60.67	36.23
WDFT-LP	11.01	43.08	23.65	54.68	33.10

Tabela 2: Norma e gabimit të fjalëve (WER ne %) të marra nga tipare të ndryshme të nxjerra nga ky punim, në korpusin Aurora-4 LVCSR në kushte të një formimi shumë stilësh. Rendi i modelit të selektuar në këtë punim është : p=24 për WDFT-LP dhe p=14 për PLP. Sa më i vogël të jetë WER aq më e mirë është performanca e ‘nxjerrësit të tipareve’.

	A	B	C	D	Mesatarja
MFCC	14.62	23.84	19.19	31.47	22.28
PLP	16.10	24.98	18.27	30.23	22.40
WDFT-MFCC	15.43	23.65	17.97	30.41	21.87
WDFT-LP	15.46	23.98	17.50	30.66	21.90

Përmbledhje e shkurtër për pjesën eksperimentale

Baza e të dhënave Aurora-4 na siguron një bashkësi të dhënash standarte për krahasimin e qëndrueshmërisë së teknikave që merren me qëndrueshmërinë e sistemeve ASR. Kjo bashkësi përbëhet nga 7,180 fjali trajnuese që përbëjnë afërsisht 15 orë regjistrim dhe 330 fjali testuese me një kohëzgjatje mesatare prej 7 sek.

Të dhënat akustike (trajnim & testim) janë të aksesueshme në 2-lloje frekuencash kampionimi (8 dhe 16 KHz), të kompresuara ose të pakompresuara.

Bashkësia e të dhënave për trajnim janë të ndara në 2-platforma. Platforma e parë përbëhet nga të dhëna të cilat janë regjistruar me anë të një kanali komunikimi dhe të pandikuar nga zhurmat. Ndërsa në platformën e 2-të gjysma e të dhënave është regjistruar me kanalin e parë të komunikimit dhe gjysma e dytë me kanalin e dytë të komunikimit në kushte të shumllojshme.

Baza e të dhënave Aurora -4 përbëhet nga 14 grupe testimi të regjistruara nga 2-kanale regjistrimi dhe 6-lloje zhurmash të shtuara. Niveli i amplitudës së zhurmave të shtuara varion nga 0-15 dB. Setet e testimit i kemi ndarë si më poshtë:

1. Set (01) - Pa zhurma
2. Set (02) - Makina
3. Set (03) - Flluska sapuni
4. Set (04) - Restorant
5. Set (05) - Rruga
6. Set (06) - Aeroport
7. Set (07) - Stacion treni
8. Set (08) - Pa zhurma
9. Set (09) - Makina
10. Set (10) - Flluska sapuni
11. Set (11) - Restorant
12. Set (12) - Rruga
13. Set (13) - Aeroport
14. Set (14) - Stacion Treni

Setet nga 1 tek 7 janë kryer nëpërmjet kanalit të parë të komunikimit. Ndërsa seti nga 8-14 nëpërmjet kanalit të 2-të të komunikimit.

Kanali i parë i komunikimit është realizuar nëpërmjet një mikrofoni 'Sennheiser' dhe

METODAT E ANALIZËS SPEKTRALE PËR ASR

kanali i dytë i komunikimit është një mikrofon i çfardoshëm i zgjedhur në mënyrë të rastësishme nga një set prej 18 mikrofonësh çfaredo.

Fillimisht sinjali i nënshtrohet procedurës së ‘Pree-emphase’ me një koeficient $k=0.97$. Llogaritjet janë kryer nëpërmjet algoritmave të caktuar nëpërmjet programit ‘LabView’. Energjitë për frekuencat ndërmjet 2 dhe 16 Hz të cilat paraqesin pjesët kryesore të sinjalit të zërit llogariten të gjitha nëpërmjet algoritmeve përkatëse. Për kryerjen e këtyre proceseve është përdorur dritarja ‘Hamming’.

Kërkimi i marketingut

Në “Marketingun e Kellogg” thuhet që të dhënat të marra nga vetë menaxheri nuk i ndihmojnë manaxherët në vendimmarrje. “E vërteta është që të dhënat mund të jenë frymëzuese, por frymëzuese janë dhe eksperiencia dhe intuïta. Në cdo rast duhet të mendosh dicka dhe kjo s’mund të jetë aq e lehtë si të ngresh hipoteza mbi të dhënat, pasi kjo do të thotë parashikim i thjeshtë por pa shpjegime” (Marketingu i Kellogg, 2000). Kërkimi në marketing është një proces sistematik që përfshin mbledhjen dhe analizimin e të dhënave për të pajisur vendimtarët me informacionin e nevojshëm për të marrë vendime. Shoqata Amerikane e Marketingut e ka përkufizuar kërkimin në marketing në këtë mënyrë: “proçesi ose numri i proçeseve që lidhin konsumatorin, klientin dhe përdoruesit e fundit me tregtarin nëpërmjet informacionit. Më në detaje, ai përfshin informacionin e përdorur për të identifikuar dhe përcaktuar rastet dhe problemet e marketingut; gjenerimi, saktësimi dhe vlerësimi i veprimeve të marketingut; monitorimi i performancës së marketingut; dhe përmirësimi i të kuptuarit të proçesit të marketingut. Kërkimi marketingut specifikon informacionin e nevojshëm për të zgjidhur këto problem, dizenjon metodat për të mbledhur informacionin menaxhon dhe implementon procesin e mbledhjes së të dhënave, analizon rezultatet si dhe komunikon gjetjet dhe implikimet e tyre të mundshme (2007).

Kërkimi i marketingut është thelbësor për çështjet e marketingut të organizatave pasi është faktori kyç për një proçes vendimmarrje të mirëmenduar. Çdo pjesë e proçesit të manaxhimit të informacionit përfshin kërkimin e marketingut. Kjo përfshinë:

METODAT E ANALIZËS SPEKTRALE PËR ASR

përcaktimin e informacionit të nevojshëm; mbledhjen dhe analizimin e këtij informacioni; dhe interpretimin e informacionit të mbledhur nën dritën e objektivave që e shtynë në rradhë të parë studimin. (Iacobucci & Churchill, 2009). Kjo gjë është njësoj e rëndësishme dhe në fushën e kërkimit akademik. Si kërkimi akademik ashtu dhe kërkimi i marketingut janë të nevojshme në fushën e sistemeve ASR.

Është shumë e rëndësishme të mos e nënvlerësohë rëndësinë e kërkimit. Duke marrë parasysh natyrën dinamike të kësaj fushe, është shumë e rëndësishme të merret në konsideratë një sistem i monitorimit të vazhdueshëm të fushës së ASR.

Kërkimi i Marketingut është një proces i vazhdueshëm i mbledhjes së informacioneve që mund të ndihmojnë manaxherët e lartë në fushën e ASR për të qenë në kontakt të vazhdueshëm me risitë dhe dhe ndryshimet në mjedis. Për qëllimet e këtij punimi janë përdorur një larmishmëri burimesh dytësore nga shumë burime.

Të dhënat dytësore u morën nga rishikime literaturash, revista online, raporte, libra, publike, gazeta dhe forma të tjera të medias. Pasi u mbodh literature për këtë fushë studimi, hapi tjetër ishte analizimi i të dhënave. Si të dhënat cilësore ashtu dhe ato sasore u organizuan, analizuan dhe interpretuan në mënyrë të tillë që të gjeneronin informacionin e nevojshëm për të kënaqur studimin brenda kornizave të hipotezës së vendosur dhe kuadrit teorik. Ka shumë autorë që e kanë mbuluar këtë fushë, gjithsesi disa burime dytësore të kërkimit ofrojnë një panoramë të njëanëshme ose të pasaktë. Për këtë arsye, ishte shumë e rëndësishme të fitohej një kuptim i qartë i pikave të forta dhe të dobta të metodave ekzistuese të analizës spektrale për aplikimet ASR, nëpërmjet një pune kërkimore të plotë e serioze.

Identifikimi i folësit

Pjesa më e madhe e aplikacioneve audio zakonisht duhet të përballen me mjediset kumbuese. Për shembull në fushën e kriminologjisë teknikat e identifikimit të folësit luajnë një rol thelbësorë në aplikimet e kësaj fushe. Gjithsesi, në kontestin e regjistrimeve ambientale identifikimi i një folësi të caktuar është i vështirë kur ai ose ajo

METODAT E ANALIZËS SPEKTRALE PËR ASR

po regjistrohet në një ambient bashkë me folës të tjerë. Një studim i bërë nga Silveira et al (2013), sugjeron Analizën Komplekse (Konvolute) të Komponentëve të Pavarur (ICA) si një teknikë për të tejkaluar këtë dobësi në aplikimin tradicional të ASR.

Është shumë e rëndësishme që mjetet e përdorura për të tejkaluar këtë problem të jenë të duhurat. Kërkimet kanë demonstruar që nuk duhet aplikuar Analiza lineare e Komponentëve të Pavarur. Për këtë arsye për të trajtuar mostra me vonesa kohore preferohet të aplikohet teknika e Analizës Komplekse (Konvolute) të Komponentëve të Pavarur.



Fig. 24: Analiza Komplekse (Konvolute) e Komponentëve të Pavarur (Silveira et al, 2013)

Kjo metodë mund të përdoret për të ndarë përzierjet komplekse (konvolute) të sinjaleve me bandë të ngushtë, pasi korigjon ndryshimet frazore të shkaktuara nga mostrat e shumta të përsëritura në ambient. Figura e mëposhtme tregon një diagram të cilat paraqet një pamje të përgjithshme të mënyrës se si punon ICA konvolute. Rreshtimi i mikrofonave regjistruar prodhon sinjalin N , një për secilin mikrofon. Sinjalet N përfaqësojnë përzierjen N që do të aplikohet në algoritmin e ICA konvolute. STDFT përdoret për analizimin e këtyre sinjaleve në fushën e kohë-frekuencës me qëllim trekëndëshëzimin e mostrës origjinale dhe me kohë të vonuar. Për më tepër, secilit component me bandë të shkurtër të përzierjes së sinjaleve i aplikohet një metodë e përgjithshme e vlerësuar e ICA.

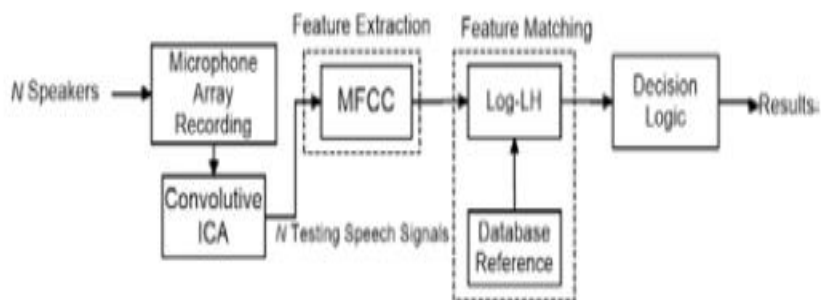


Fig. 25: Sistemi i identifikimit të folësit bazuar në sistemin e ICA konvoluteve
(Silveira et al, 2013)

Me qëllim zgjidhjen e anomalive të ndërrimit për secilin vatrë/kosh frekuence mund të zbatohet një matricë ndërrimi e cila mund të llogaritet mbi bazën e amplitudës së komponentëve të frekuncës së përafërt, gjithashtu të njohur dhe si Mbështjella spektrale. Së fundmi, aplikohet STDFT inverse, e cila rezulton në N testime sinjalesh të së folurës.

Sistemi i propozuar nga Silveira et al jep një vlerësim të mirë e të menjëhershëm rreth identitetit të folësve N . Për të arritur këtë qëllim përdoret një matricë M nga N *log-likelihood*. Matrica zëvendëson vektorin e kolonës të përdorur tradicionalisht në sistemin e prezantuar. Rreshtat e matricës N *log-likelihood* përfaqësojnë individët në sinjalin e regjistrimit të së folurës, në database.

Po ashtu në logjikën e vendimmarrjes u bë dhe një modifikim. Në sistemin e propozuar të ICA konvoluteve, logjika e vendimmarrjes zgjedh për secilin nga testimet e sinjaleve të të folurit, individin që ka më shumë gjasa të jetë folësi. Kjo arrihet duke zgjedhur për cdo kolonë j të matricës *log-likelihood*, elementin ose personin me vlerën më të madhe të *log-likelihood* dhe duke e zëvendësuar këtë element ose person si folësi i identifikuar i sinjalit j të testuar. Sinjalet origjinale të të folurit u ndanë duke përdorur Teknikën e Ndarjes së Verbër të Burimit (Blind Source Separation Technique) dhe sinjalet e ndara u aplikuan në një sistem ASR të bazuar në Koeficientët Cepstral të Frekuencës Mel dhe në Modelet Mikse Gaussiane.

Në këtë studim të bërë nga Silveira et al, përqasja e bazuar në sistemin e ICA konvoluteve u krahasua me performancën e teknikave tradicionale të identifikimit të folësve MFCC-GMM. Ky studim vlerësoi normën e suksesit të parametrave të performancës, si dhe bëri një krahasim të rasteve ku fajlet e të folurit përmbanin një përzierje konvoluteve zërash me situata ku zërat ndahen sipas algoritmit ICA. Tekstet (testet) e bëra në këtë mjedis eksperimental treguan rezultate të qarta që qasja e propozuar ICA funksiononte më mirë mbi mostrat e tingullit me folës të shumëfishtë në sfonde ambientale se sa teknikat tradicionale ASR të përdorura pa qasjen e propozuar ICA (2013).

Rrjetet neutrale

Rrjetet neutrale kanë treguar një potencial të madh në përmirësimin e performancës së ASR. Për shembull, Perceptronës Shumë-Shtresësh (Multi-Layer Perceptron, MLP) u futën në ekstraktimin e tipareve, ku vlerat e shtresë së outputit “tandem” ose të shtresës së fshehur të “Bottleneck” përdoren në hapin e pre-procesit në vend të tipareve tradicionale MFCC. Në shumë kontekste dhe eksperimente tiparet MLP kanë treguar që janë të një fuqie të lartë diskriminuese dhe shumë të qëndrueshme ndaj variacioneve të folësve dhe ambienteve.

‘Tandem’ është një metode e perzgjedhjes artificiale e cila përdor kritere të vecanta për ‘Bottleneck’ është një metodë e cila na ndihmon të stabilizojmë punën tonë gjatë procesimit të sinjalit. Në rast se informacioni në hyrje është shumë më i shpejtë se në dalje gjë e cila mund të na interferojë me punën tonë, atëherë përdorim algoritmin e ‘Bottleneck’ e cila gjen një ekuilibër ndërmjet inputit dhe outputit për të mos penguar procesin e përpunimit të sinjalit të zërit.

Për më tepër MLP kanë treguar dhe që kanë një nivel pavarësie gjuhësore. Në një studim të Bresacier et al në kontekstin e ASR për gjuhën me pak burime (nder-resourced); MLP ka treguar që i lejon zhvilluesit të ndërtojnë sisteme të përpunimit të të folurit me një sasi shumë të vogël të dhënash dhe ti ndajë këto të dhëna të të folurit në disa gjuhë për të rritur efikasitetin e sistemeve ‘**bootstrap**’ në gjuhë akoma të papara.

‘Bootstrap’ i referohet një procesi vetë-startues i cili nuk ka nevojë për të dhëna të jashtme në hyrje. Nëse ne nuk mund ta ristartojmë përsëri eksperimentin tonë atëherë kjo metodë krijon të dhëna të reja nga kampionimi i rastësishëm i sinjalit fillestar. Meqë rikampionimi bëhet në sinjalin origjinal ne mund të gjenerojmë të dhëna sa të duam dhe ta rivlerësojmë eksperimentin tonë në të njëjtën mënyrë që e kryem herën e parë.

Studime të shumta kanë treguar se tiparet e nxjerra nga një MLP të trainuar me një ose shumë gjuhë mund të aplikohen edhe ndaj gjuhëve të tjera. Këto kërkime kanë treguar

METODAT E ANALIZËS SPEKTRALE PËR ASR

se si data set-et nga shumë gjuhë mund të përdoren për të nxjerrë tipare për një gjuhë me pak burime që cojnë në një performancë të përmirësuar të ASR.

Ky punim mbi nxjerrjen e tipareve bazuar në MLP vjen si rezultat i një qasje të orientuar nga të dhënat që nuk përfshin asnjë njohuri të mëparshme mbi phone set-in e gjuhëve të targetuara. Autorët futën eksperimente dhe gjetje mbi përdorimin e MLP shumëgjuhëshe për iniciimin e një MLP-je për gjuhët me pak burime bazuar mbi hartën telefonike IPA. Kjo teknikë tregoi një përmirësim të rëndësishëm porsa i përket performancës së ASR-së si dhe tregoi që është shumë e qëndrueshme ndaj gabimeve të transkriptimit në trainimin e të dhënave. (Besacier et al, 2014).

Modelimi akustik

Nuk është e thjeshtë të marrësh transkriptime gjuhësore në gjuhët me pak burime. Për këtë arsye, metodat e pa mbikëqyrje ose me mbikëqyrje të vogël janë shumë atraktive në këtë kontekst. Një kërkim nga Cetin propozonte metoda të pa mbykqyrura të përshtatjes gjuhësore për të zhvilluar një njohës fjalësh të izoluar për gjuhën Tamile. Qasje të tilla madje dhe më të zhvilluara janë propozuar për gjuhën polake dhe vietnameze (2008). Prandaj në një situatë kur kemi një farë informacioni mbi gjuhën e targetuar këto metoda janë shumë të mira për të kursyer kohën dhe ulur kostot duke ndërtuar një sistem ASR për një gjuhë ende të pasuportuar. Informacioni ekzistues mund të përfshijë fjalorin e shqiptimit; modelin e gjuhës; ose identifikimin e gjuhës së të dhënave të patranskriptuara. Kërkimet e bëra nga Vu në 2010 dhe 2011 treguan se kjo gjë është e mundur edhe nëse gjuha burimore dhe gjuha e targetuar nuk kanë lidhje me njëra tjetrën. Ata përdorën disa sisteme ASR për një varietete të madhe gjuhësh të palidhura me njëra tjetrën për të koduar të dhënat audio të gjuhës së targetuar dhe paralelisht për të llogaritur një rezultat besimi (confidence score) të quajtur “Multilingual A-stabil”, (Besacier et al, 2014).

METODAT E ANALIZËS SPEKTRALE PËR ASR

Kjo llogaritje bën që në të gjitha fjalët që janë zgjedhur nga të paktën dy gjuhë të ndryshme të përzgjidhen për adaptimin e modelit akustik të gjuhës së targetuar. Në strukturën e tyre adaptimi i MAP u implementua në mënyrë iterative për të maksimizuar sasinë e informacionit për trainim dhe për të përmirësuar cilësinë e transkriptimit automatik. Në të gjitha këto zhvillime, të dhënat e transkriptuara nga gjuhë të studiuar mirë përdoren për të zhvilluar sisteme themeltare dhe për të transkriptuar të dhëna mbi të folurën nga gjuha në target mundësisht kur i përdorur bashkë me një sasi të vogël të foluri të transkriptuar ka treguar se është i përshtatshëm të trajnojë sisteme të përdorshme ASR. Është interesant fakti që lidhja mes gjuhës burimore dhe asaj target si rregull nuk është një variabël thelbësorë: edhe gjuhët shumë të ndryshme nga njëra tjetra performojnë mirë në lidhje me këtë gjë. (Besacier et al, 2014).

Sistemet ASR të avancuara në gjuhët me burime të mira zakonisht përdorin Modele “Hidden Markov”, të varura nga konteksti për të modeluar fonemat e një gjuhe dhe e njëjta metodë përdoret në përgjithësisht për gjuhët me pak burime. Në këtë rast konteksti me pak burime sjell një numër sfidash dhe mundësishë të reja. Për shembull, përkufizimi i një foneme të përshtatshme e marrë për model, si rregull është një hap i rëndësishëm. Edhe pse këto bashkësi janë përkufizuar në një gjuhë, shpesh ato nuk kanë themele të forta empirike. Po ashtu, fonemat e supozuara si për shembull afrikatet, diftongjet the tingujt e klikimit mund të modelohen shumë mirë si njësi të vetme por edhe si sekuenca, edhe alofonet të cilat nga ana akustike janë goxha të dallueshme mund të modelohen të ndara.

Për të gjitha këto probleme disa zgjidhje mund të na vijnë nga zgjedhjet që janë bërë për gjuhë të ngjashme, gjithsesi një sasi e caktuar kërkimi dhe vigjilence është gjithmonë e nevojshme. Kur kemi në dispozicion një gjuhë burimore të ngjashme me shumë burime atëherë është e mundur të përdorim të dhëna nga kjo gjuhë për të zhvilluar modele akustike për një gjuhë target me burime të pakta. (Besacier et al, 2014).

Për këtë qëllim janë përdorur qasje të ndryshme duke filluar nga grumbullimi i të dhënave mes gjuhëve (Van Heerden, 2010) tek bootstrapping nga vijëzime modelesh

METODAT E ANALIZËS SPEKTRALE PËR ASR

burimore, tek gjurmimi telefonik (phone mapping) për identifikim me modelin burimor e deri tek adaptimi Maximum A-Posteriori (MAP) me të dhënat e gjuhës target.

Shumë autorë në fushën e ASR kanë sugjeruar që përdorimi i modeleve në vend të Modelit standard të fonemave, të varura nga konteksti Hidden Markov është një teknikë ASR e përshtatshme për gjuhët me burime të pakta. Për shembull, në egzemplarët e bazuar mbi njohjen e të folurit (Gemmeke, 2011) përfaqësimet e njësive akustike paraqiten si vektorë të shembujve të peshuar (matur). Kjo teknikë e cila përdor një numër të vogël parametrash duket në mënyrë të vecantë interesante në rastet kur kemi pak të dhëna për training. Një teknikë tjetër e cila ndahet në mënyrë më pak radikale nga modelet konvencionale ASR përdor Modelet Hidden Markov për të modeluar rrokjet në vend të fonemave (Tachbelie, 2013). Ky studim arriti në përfundimin se ulja e parametrave të modeleve vjen nga fakti që varësitë nga konteksti janë si rregull më pak të rëndësishme për modelet e rrokjeve (Sinischalchi, 2013). Po ashtu propozon aftësinë për të përshkruar çdo gjuhë të folur me një set të përgjithshëm njësishë bazë që mund të përkufizohen si “universale” rreth spektrit të gjuhëve të folura.

Toni dhe struktura e çdo folësi të vecantë duke përfshirë theksin dhe artikulinin përzgjidhen për të formuar këtë njësi inventari si dhe përdoren për të ndërtuar një model bashkësish atributësh për gjuhë universale të derivuara nga IPA (Stuker, 2003) ose me metoda modelimi të orientura nga të dhënat. Punimi i fundit i propozuar nga (Sinischalchi, 2013) është i përshtatshëm për rrjete të thella arkitekturash neutrale për ASR (Yu, 2012).

Modelimi leksikor

Për të krijuar shqiptimin e fjalorëve në gjuhë si Tajlandeze, Amharike, Vietnameze u përdorën metoda të bazuara në grafema. Në modelet e bazuara në grafema, çdo fjalë në fjalor shpërbëhet në grafema; të cilat shërbejnë si njësia bazë e modelit akustik. Këto procese japin rezultate mjaft të mira sidomos për ato gjuhë të cilat kanë një marrëdhënie të afërt grafema-fonema (Besacier et al 2014).

Përkthimi me makina

Ka pasur një zhvillim të konsiderueshëm në metodat të cilat përdorin principet e përkthimit statistikor mekanik për të përkthyer grafemat në fonema (Laureat,2009; Karansowu,2010). Në këto metoda grafemat shikohen si “fjalë” në gjuhën bazë dhe fonemat si “fjalë” në gjuhën objektive. Një sistem “përkthimi mekanik” trajnohet nëpërmjet një fjalori fillestar fonetik dhe më pas sistemi aplikohet për të konvertuar çdo fjalë në formën e saj fonetike. Kjo metodë u propozua në fillim për gjuhën Rumune (Cucu 2011). Zhvillimet në këto fusha do të kenë një impakt të paevitueshëm në metodat në metodat e analizës spektrale si dhe në standardet e përgjithshme të ASR.

Një sasi zhurme ambienti është e paevitueshme. Fatkeqësisht, edhe pse rrallë shkakton probleme në dialogje, ajo degradon performancën e ASR. Kur nxjerrim tiparet e zërit, ndonjëherë është e vështirë të kuptohen fjalët e sakta për shkak të zhurmës së ambientit. Përmirësimet në sistemet ASR mund të përmirësojnë dukshëm eksperiencën e përdoruesit sidomos në bisedat online. Për shembull, njohja e zërit të sistemit Windows është i mirë por gjithsesi konsiderohet si një sistem ASR me një rrugë. Kur flasim, bëhet procesimi i ambientit dhe në përgjigje hapen programe ose kryhen detyra nga sistemi. Por ai nuk përfshin një përgjigje me zë; thjesht një përgjigje hardware apo nëpërmjet programit. Për t'u bërë një sistem miqësor me përdoruesin, përgjigja me zë për komandat e dhëna është e rëndësishme. Sidoqoftë, në sistemin e zërit Windows API ekzekutohen vetëm komandat e lidhura me OS ekzekutohen. Këto komanda janë të rëndësishme por nuk kanë si objekt ta lehtësojnë punën e përdoruesit, dhe nuk krijojnë as një ndërveprim të specializuar. Kjo ishte e vërtetë deri pak më parë, kur Google mund të hapej me anë të zërit por nuk mund të realizonim kërkimin nëpërmjet përpunimit të zërit.

Kërkimi është diçka thelbësore për zhvillimin e inovacioneve në këtë fushë. Ghoshal 2009 dhe Schlippe 2013 në kërkimet e tyre i përshkruajnë metodat automatike për të

METODAT E ANALIZËS SPEKTRALE PËR ASR

prodhuar fjalor nisur nga çiftet fjalë-shqiptim të gjetura në WWW. Duke qene se Wikitionary është një fjalor i hapur i bazuar në wiki, përfshin shënime fonetike të shkruara në alfabetin ndërkombëtar fonetik (IPA). Schlippe dizenjoji dhe një sistem i cili nxjerr automatikisht shënimet fonetike nga IPA. Autorët kanë rezultate për katër gjuhë, Anglisht, Frëngjisht, Gjermanisht dhe Spanjisht, për sa i përket cilësisë dhe sasisë së kërkimeve fonetike. Cilësimet sasiore tregojnë që edhe nëse emrat e vendeve dhe qyteteve nuk mund t'i gjejmë nga ana fonetike mund ti ekstraktojmë nga Wikitionary bashkë me shënimet fonetike. Funksionaliteti i tyre lidhej fort me sasinë dhe cilësinë e informacionit të gjetur në Wiki. Është për të ardhur keq që shumica e gjuhëve në botë akoma nuk janë pjesë e këtij fjalori. Ashtu siç theksohet dhe nga Schlippe, gjenerata e modeleve G2P për gjuhët evropiane u investigua nëpërmjet çifteve fjalë shqiptim nga 6 edicione të Wikitionary dhe 10 Global Phone fjalorë. Duke përdorur shqiptimin e gjeneruar nga Wikitionary, modelet G2P për trajnimin dhe dekodimin e ASR rezultuan me një degradim të pranueshëm të performancës, duke pasur në konsideratë dhe faktorë koha dhe kosto e sistemit. Ai ka propozuar metoda automatike për të gjurmuar, hequr dhe zëvendësuar fjalë jokonsistente dhe difektoze për sa i përket shqiptimit nga WWW dhe ka trguar që ka përmirësime të cilësisë. (Schlippe 2012)

Modelimi gjuhësorë

Modelet gjuhësore statistikore paraqesin një vlerësim të probabilitetit të një sekuence fjalësh. Një nga skemat më eficiente të modelimit gjuhësorë statistikor bazohet mbi fjalë n-gramet (bigrame, trigrame e më shumë) të cilat masin përsëritjen e çdo sekuencë fjalësh në një tekst. Probabilitetet në modelet gjuhësore n-gram përcaktohen si rregull nëpërmjet vlerësimit të ngjashmërisë maksimale. Kjo gjë e bën shpërndarjen e probabilitetit të varur nga të dhënat e dispoznueshme të trajnimit. Prandaj për të siguruar një domethënie statistikore në modelimet gjuhësore statistikore, nevojiten sasi të mëdha të dhënash trajnimi.

Web teksti

Grumbullimi i të dhënave tekstuale në një gjuhë të caktuar dhe për një domain të caktuar është gjithashtu një çështje e rëndësishme që mund të adresohet duke përdorur ueb-in si një korpus ose duke përdorur sisteme të përkthimit automatik për të përkthyer korpuset tekstesh nga një gjuhë në tjetrën (Cocu, 2012). Duhet theksuar se ekzistojnë disa probleme specifike në lidhje me zhvillimin e modeleve gjuhësore për sa i përket disa gjuhëve me pak burime. Për shembull, gjuhë të tilla si rumanishtja ose turqishtja përdorin shumë shenja diakritike. Edhe pse për një lexues njerëzor kuptimi i një teksti pa diakritikë përgjithësisht është i kuptueshëm duke parë kontekstin, restaurimi diakritik automatik nuk është një çështje pa rëndësi, madje në disa raste ka shumë rëndësi. Për shembull, për një numër të caktuar gjuhësh që përdorin diakritikët edhe pse korpuset e tekstit mund të merren nga ueb-i, ato vijjnë pa diakritikë. Output-i i një sistemi ASR të cilit i mungojnë diakritikët mund të jetë i vështirë për tu kuptuar madje dhe i pakuptueshëm. Prandaj, për këto gjuhë sistemi i restaurimi automatik të diakritikëve është i domosdoshëm (Cucu, 2013). Problemet teknike janë nevoja për normalizim (standardizim) në numra, akronime, shkurtime etj., po ashtu përdorimi i identifikimit gjuhësor ashtu si dhe sistemi i shkrimit janë gjithashtu probleme të rëndësishme që duhen trajtuar në kontekstin e gjuhëve me pak burime në ASR (Besacier et al, 2014).

Problemet e copëzimit të fjalëve

Sistemeve të shkrimit të disa gjuhëve si psh. të gjuhëve kineze, vietnameze, kmere, tajlandeze etj., u mungojnë ndarësit e fjalëve totalisht ose edhe kur i kanë i përdorin në mynyrë jo konsistente. Ky përkufizim i njësive të fjalëve është thelbësore për ASR-në, pasi fjalori dhe modeli gjuhësor varen prej tij. Copëzimi në fjalë njësi ose ‘fjalë’

METODAT E ANALIZËS SPEKTRALE PËR ASR

identifikimi”, nuk është një çështje e thjeshtë edhe për gjuhët të cilat i ndajnë fjalët me karaktere të veçanta. Për gjuhët që kanë një sistem shkrimi pa një ndarje të qartë mes fjalëve, n-gramet e fjalëve përgjithësisht vlerësohen nga një korpus teksti të copëzuar në fjalë duke përdorur metoda automatike. Copëzimi automatik i tekstit nuk është një punë e lehtë dhe gjeneron gabime për shkak të ambiguitetit në gjuhën normale dhe pranisë së fjalëve jashtë fjalorit në tekst. Kërkimet kanë treguar se një alternativë e mundshme është të llogarisësh probabilitetet nga karakteret logografike si p.sh Kanji në Japonisht ose Hanja në Koreane (Denoual, 2006) (besacier et al, 2014).

Norma e fjalëve gabim

Norma e fjalëve gabim (Word Error Rate, WER) është një mjet matës intuitiv dhe i përshtatshëm për gjuhët analitike të orientuara nga fjala dhe me një morfologji të thjeshtë. Është e rëndësishme të theksojmë se gjuhët janë të bazuara në morfema dhe disa të tjera janë të bazuara në rrokje. Po ashtu është e rëndësishme të mbahet parasysh se disa gjuhë të tilla si gjuha tajlandeze ose vietnameze nuk kanë ndarës të qartë mes fjalëve ortografike. Prandaj, gjuhë të tilla mund të sintetizojnë forma koherente fjalësh goxha të gjata nga një numër njësi nën-fjalësh. Për shembull, në shumë gjuhë aglutinative si estonishtja ose finlandishtja word-forms fjalë-format

Një sistem “përkthimi automatik” është i bazuar (stërvitur) mbi një fjalor fonetik fillestar dhe më pas ky sistem mund të aplikohet për të konvertuar çdo fjalë në formën e saj fonetike. Një qasje e tillë u propozua, për shembull, për gjuhën rumune (Cocu, 2011). Zhvillimet në këtë fushë do të kenë patjetër impakt mbi metodat ‘Front End ‘të analizës spektrale si dhe standardin e përgjithshëm të ASR-së.

Një nivel i caktuar zhurmash ambienti është i pashmangshëm. Fatkeqësisht, edhe pse rrallë shkakton vështirësi për njerëzit për të kuptuar sinjalet e të folurit gjatë një dialogu, këto zhurma ambienti e ulin në mënyrë të konsiderueshme performancën e sistemeve automatike të njohjes së të folurit (ASR). Duke nxjerrë karakteristikat nga një e folur ndonjëherë është e vështirë njohja e një fjale të caktuar për shkak të zhurmave dhe

METODAT E ANALIZËS SPEKTRALE PËR ASR

kushteve të tjera ambientale. Përmirësimet në sistemet ASR mund ta bëjnë eksperiencën e përdoruesit shumë më të mirë, sidomos në kontekstin e eksperiencave online. Për shembull, njohja e të folurit e Windows-it është goxha efiçente por është akoma një sistem ASR një drejtimsh.

5.2 Aplikimet e ASR dhe mjetet

Punët kërkimore të zërit në Afriken e Jugut

Afrika e Jugut është një shtet me shumë diversivitet, me diferenca të mëdha midis klasave sociale dhe me 11 gjuhë zyrtare. Projektet teknologjike të cilat adresojnë problemet sociale ndërsa janë në gjendje për të krijuar ura komunikimi midis barrierave të gjuhëve gjë që ka tërhequr vëmendjen në Afriken e Jugut këto vitet e fundit. Ka pasur një progres të madh në zhvillimin e burimeve dhe sistemeve të zërit që grupojnë/rrethojnë 11 gjuhët. Si pasojë e zhvillimit dhe komercializimit të këtij aktiviteti ishte zhvillimi i programeve që kryejnë kërkime në web të bazuar në gjuhën e të folurit në tre gjuhë të Afrikës së Jugut të quajtura isizulu, anglishtja Afrikano Jugore, dhe Afrikanshja. Përdorimi i disa teknikave të përmendura më lart, të dhënat e këtyre gjuhëve u përmbushën dhe sistemet ASR u zhvilluan duke përdorur mjetet dhe infrastrukturën të siguruar nga Google; për këto sisteme u pa që janë më pak të sakta sesa sistemet state of –the-art në gjuhën Amerikane, por ishin të një standarti të mjaftueshëm për tu nxjerrë në përdorim komercialisht. Edhe gjuha Afrikane dhe gjuha Anglo-Afrikano Jugore kanë tërhequr përdoruesit aktiv në popullsi; në isizulu, gjithsesi sasia e informacionit e aksesueshme në Web është shumë e limituar për të suportuar bazën e një përdoruesi aktiv (Besacier et al, 2014).

Forumet interaktive të zërit të fermerëve Indian

Kjo përpjekje fantastike u quajt Avaaj Otalo dhe u dizenjua në verën e 2008 si një projekt i përbashkët midis një organizate jo-qeveritare në Indi dhe IBM India, laboratoritë kërkimor. Një forum mesazhezh zanor u sugjerua për fermerët në Indi. Fermerët indian shpesh herë kanë një edukim shumë të limituar. Forumi i sugjeruar ishte për të siguruar

METODAT E ANALIZËS SPEKTRALE PËR ASR

akses në njohuritë Agro-Kulturore. Përmbajta e zërit u aksesua duke përdorur telefone celularë me kosto të ulët, të cilat janë përshtatur shumë nga komunitetet rurale anembanë botës. Elementi më popullor i këtij projekti ishte një forum për bërjen e pyetjeve dhe kërkimi i pyetjeve të tyre dhe i përgjigjeve në një varietet temash rreth agrokulturës. Për shëmbull, fermerët mund të kontrollojnë parashikimin e motit, të merrnin ndihmë për të vendosur se kur do të fekondonin të korrat; të mësonin se kur doktorët ishin të disponueshëm për të ardhur në fshat ; dhe të gjejnë çmimet më të mira për mallrat e tyre etj. Në kontekstin e sistemeve të ASR hyrjet e përdoruesve u dërguan drejt makinës së njohjes së zërit, IBM websphere Voice Server (WVS). WVS ka një fjalor shumë të gjërë me një njohës zëri të vazhdueshëm të trainuar me anglishten Amerikane. Kjo u përshtat me gjuhën Gujarati e cila është gjuha e forumit dhe flitet nga afërsisht 50M persona në Indi. Për këtë qëllim, komandat e zërit Gujarati u konvertuan duke përdorur Anglishten Amerikane. Me këtë afirim një saktësi prej 94% të njohjes së zërit u arrit në një ambient të pastër shtëpie. Gjithsesi, në termat e përdorshmërisë, gjatë lëshimit të projektit kërkimor u tregua se shumica e përdoruesve parapëlqyen hyrjet numerike në vënd të zërit (Besacier et al, 2014).

Projekti PI

Projekti PI u themelua nga franca ANR-Agence Nationale de la Recherche dhe iu dedikua kryesisht njohjes automatike të zërit për nën-resurset e gjuhës. Kjo ju dedikua veçanërisht për gjuhët që janë në shtete si Vietnami, Laos dhe Kamboxhia. Nga një vështrim operacional, ky projekt synon në gjetjen e teknikave dhe mjeteve për zhvillimin e ASR në resurset e nën-gjuhëve. Një zhvillim tjetër pozitiv i projektit PI është kontributi i fuqishëm për strukturimin e komunitetit të shkencës rreth temës së ‘processing under-resources languages’ (Besacier et al, 2014).

Normalizimi i gjatësisë së traktit vokal (VTLN)

METODAT E ANALIZËS SPEKTRALE PËR ASR

VTLN është një teknikë shumë e përdorur për të reduktuar transformimin e frekuencës të sinjalit në hyrje. VTLN përpiket të lokalizojë faktorin e mbështjelljes së frekuencës optimale për çdo folës gjatë procesimit të zërit, dhe e skicon zërin në një hapsire konike. Cdo faktor përkulje prodhon një version të zhvendosur dhe të shkallëzuar për bankën e filtrave Mel në çdo rajon frekuencash (i.e., 1-40). Faktori i përkuljes për çdo folës zakonisht selektohet nëpërmjet një procesi përafrimi. VTLN mund të përdoret si një strategji bashkuese. Proçesi i bashkimit të frekuencave implementohet për të reduktuar përkthimet e fjalëve si pasojë e mënyrave të stileve të ndryshme të folurit, gjatësia e kordave zanore, gjinia, thekset, etj. Siç na sygjerohet nga Sainath et al, një strategji bashkuese e VTLN mund të përdoret në një mënyrë jo të supervizuar, duke përdorur një filtër për çdo rajon frekuence.

Në mënyrë rajonale vendodhje unike të frekuencës qëndrore të filtrave të VTLN llogariten, dhe ndryshimet në format e filtrave të VTLN nuk merren parasysh. Nuk selektohet asnjë faktor përkulje ‘optimal’ dhe bankë filtrash për çdo folës pasi kryhet në VTLN normale, vetëm një filtër përdoret për çdo rajon (2013).

Ekzistojnë një numër i madh kombinimesh dhe ndryshimesh të modelit të sinjalit që ne kemi analizuar në përdorim sot. Le të fillojmë diskutimin tone duke numëruar disa gjëndje të sistemit të njohjes së zërit së bashku me teknikat e modelimit të sinjalit të përdorur. Vëzhgimi i sistemeve bashkëkohor të Piccone në vitet 90 tregoi se koeficientët e derivuara kepsral të FFT duket që dominojnë fushën. Analiza e LP e cila ishte dikur guri i themelit të njohjes së zërit u spostua në nivel dytësor. Një parameter i vektorit të sinjalit që konsiston në koeficientë spektral, derivati i parë i koeficientëve kepsral, fuqia, dhe derivati i fuqisë është bërë de facto një standart. Peshimi i variaciones i këtij vektori parametrik ishte më e kërkuara në teknikën e normalizimit.

është pozitive të vësh re se pavarësisht diferencave të mëdhaja në afrim, shumë këto afrime kanë pasur një sukses në rritje. Shpesh, diferencat e rëndësishme në sistemet e njohjes shtrihen në detaje përtej modelit të sinjalit. Kapaciteti dhe potenciali aktual i sistemeve aktuale të njohjes së zërit janë akoma shumë larg performancës njerzore. Në kundërshtim mjediset ambjentale, siç janë zonat e aplikimeve të mjeksisë-ligjore siç u

METODAT E ANALIZËS SPEKTRALE PËR ASR

diskutuan më parë; diferenca në performancë midis njerëzve dhe makinerive është akoma më e madhe. Siç u përmend më parë, koeficientët cepstral janë matëset dominante akustike. Përshembull, 21 nga 31 sistemet përdorin disa forma të koeficientëve cepstral. Ky raport është akoma më i lartë (21 nga 26) nëse ne do të përjashtojmë sistemet e bazuara në rrjetin nervor. Për më tepër, koeficientët cepstral të derivuara nga FFT janë një formë e analizës cepstrale të përdorur. Koeficientët cepstral të derivuara të LP përdoren vetëm nga 1/3 e sistemeve duke përdorur analizën cepstrale. Ekziston një preferencë definitive përkundrejt përdorimit të shkallëzimit Mel. Inovacionet në fushën e ASR siç janë transformimi Wavelet; po i rritet popullariteti dhe ka potencialin për të rritur standartin në sistemet që imitojnë sistemin auditor të njeriut.

Sipas Piccone, dy institucione që janë avokatët e dukshëm të koeficientëve cepstral janë institucionet e vetme që përdorin teknikat e tyre respektive; koeficientët e hequr cepstral dhe koeficientët spektral bilinear të transformuar. Ai e bëri studimin gjatë viteve 90 se sistemet që performojnë me fjalor të gjërë të njohjes së zërit kanë tendencën të përdorin përafrimin e densitetit diskret të bazuar në kuantizimin e vektorëve, ndërsa sistemet që performojnë me fjalore të vogla në ambiente të vështira kanë tendencën për të përdorur disa lloj filtrash pre-whitening. Gjatë viteve 90 sistemet e reja përdornin një matricë kovariance (1993).

Puna e Piccones në vitet 90 na siguroi bazat shumë të mira të programeve të ASR. Ai gjeti që shumica e sistemeve të ASR përdorin një diferencë të rendit të parë. Sisteme të ndryshme, së shumti përdorin një analizë regresi prej 5 segmentesh për procesimin e zërit. Dhe me mundësi shumë të lartë kjo na siguron një paraqitje të zbutur dhe më të qëndrueshme të parametrut (1993).

Nuk ka nevojë për të dhëna gjithëpërfshirëse për analizat krahasuese të modelimit të sinjalit në njohjen e zërit. Me të dhëna të mëdha dhe llogaritjen e GPU, eksperimentet në shkallë të gjërë të njohjes së zërit janë bërë të realizueshme. Bazat e të dhënave janë tashmë mjaftueshëm të mëdha për të suportuar statistikisht krahasimet e rëndësishme dhe kompjuterat janë tashmë mjaftueshëm të shpejtë për të bërë vlerësime parametrike. është esenciale të përdorim teknikat akademike dhe të tregut për të simuluar

METODAT E ANALIZËS SPEKTRALE PËR ASR

nënbashkësitë e shumë qasjeve konkurruese dhe zhvillimi i kërkimeve të gjëra në çdo zhvillim të ri. Në këtë tezë kemi bërë një analizë të gjerë të teknikave të analizës spektrale në programet e ASR. Shumica e teknikave funksionojnë brenda një kuadri që theksojnë rëndësinë e saktësisë së analizës spektrale dhe normalizimin statistikor. E parë brenda këtij kuadri, diferenca brenda këtyre teknikave konkurruese ngjason si e pafundme në krahasim me sfidat e mëdha me të cilat ndeshemi akoma në fushën e ASR.

Pavaresisht popullaritetit të MFCCs, LPCs dhe GFCCs; përdorimi i drejtpërdrejtë i vektoreve që përmbajnë koeficientët e FFT të fuqisë spektrale janë gjithashtu të mundshme për nxjerrjen e karakteristikave. Në krahasim me metodat e bazuara në njohuritë rreth sistemit auditor të njeriut sisteme si bankat e filtrave Mel ose Gammatone, spektri i pastër i FFT përmban shumë informacione të sinjalit të zërit. Gjithsesi, shumë nga ky informacion ekstra ndodhet në bandat e frekuencave të larta kur përdorim shkallë të larta kampionimi (p.sh 44.1 KHz etj). Këto frekuenca të larta zakonisht nuk konsiderohen produktive për procesimin e zërit. Logaritmi i spektrit të FFT përdoret shpesh për të modeluar perceptimin e volumit të lartë.

Një nga metodat më të përdorura për studimin e sinjalit të zërit është nëpërmjet spektrit të fuqisë. Fuqia spektrale e sinjalit të zërit na përshkruan përmbajtjen e frekuencës të sinjalit në varësi të kohës. Hapi i parë drejt llogaritjes së fuqisë spektrale të sinjalit të zërit është kryerja e një transformimi diskrete furie (DFT). Një DFT llogarit frekuencën e informacionit të sinjalit ekuivalent të varur nga koha. Meqënëse një sinjal zëri përmban vetëm vlerat e pikave reale, një pikë reale e transformimit të shpejte furie (FFT) mund të aplikohet për rritjen e efikasitetit. Rezultati dalës përmban edhe magnituden dhe informacionin e fazës së sinjalit origjinal të varur nga koha. Analiza cepstrale e shkallëzimit Mel është shumë e ngjashme me analizën e zërit të Parashikimit Linear Perceptiv.

Në këtë metodë, gjithsesi spektri është ‘përkulur’ sipas shkallëzimit Mel, duke pasur parasysh që në PLP spektri është ‘përkulur’ sipas shkallës bark. Diferenca kryesore midis analizës cepstrale të shkallës Mel dhe PLP lidhet me outputin e koeficientëve

METODAT E ANALIZËS SPEKTRALE PËR ASR

cepstral. Modeli PLP përdor një model shumë polësh për të zbutur fuqinë spektrale të modifikuar. Kjo bëhet duke kryer transformimin direkt të fuqisë spektrale të logaritmit në boshtin cepstral duke përdorur transformimin diskrete furie (DFT).

Përdorshmëria e shpeshtë e analizës spektrale të bazuar në FFT vazhdon të bazohet në ‘resilient noise’ të FFT-së. Analiza e LP u përdor gjërësisht në 1970 dhe 1980 dhe përdoret akoma sot por në një formë hibride me FFT. Trendi i kohëve të sotme është drejt analizës që bazohet në FFT ose formës hibride që përfshin analizën e bazuar në FFT. Është logjike të spekulojmë që kjo ndodh si arsye e lehtësisë me të cilën shkalla Mel mund ti imponohet sistemit. Për më tepër është e kuptueshme që aplikimi i vlerave spektrale të FFT në mënyrë të drejtpërdrejtë duke përdorur frekuencën Mel është mjaftueshëm efiçente për të hequr hapin e bankës së filtrave komplet siç u gjet në punën e Parinam (2013).

Një studim në vlerat spektrale të FFT që aplikohen në mënyrë të drejtpërdrejtë në sinjalin e zërit duke përdorur frekuencat Mel ose të gjashme me të u krye nga Parinam. Ky kërkim ka treguar në mënyrë të përvokueshme se përdorimi i drejtpërdrejtë i vlerave spektrale të FFT-së dhe e tipareve që inkorporojnë një shkallë frekuence Mel ose të ngjashme me të mund të anashkalojë procesimin e bankës së filtrave. Megjithatë gjetjet e Parinamit nuk kan gjetur ndonjë disavantazh në përdorimin e bankës së filtrave por kanë paraqitur këtë sistem në mënyrë që të thjeshtojë procesin e ASR-së. Një tjetër paraqitje e tipareve të zërit njihet si RASTA-PLP një akronim për ‘**Relative Spectral Transform-Perceptual Linear Prediction**’. Teknika e PLP u prezantua zyrtarisht nga Hynek Hermansky si një mënyrë për të ‘përkulur’ spektrin për të minimizuar diferencën midis folësve duke ruajtur informacionet e rëndësishme të zërit. RASTA është një teknikë e ndarë që aplikon një filtër me bandë lejuese për energjinë në çdo frekuencë të nën-bandave në mënyrë që të zbusë në kohë me terma afat-shkurtër vibrimet e zhurmës

METODAT E ANALIZËS SPEKTRALE PËR ASR

dhe të largoje çdo konstante të mënjeluar që rezulton nga ngjyrosja spektrale statike në kanalet e zërit aq statike si një linje telefoni.

Tipet e dëshiruara të ASR duhet të jenë resilient, të qëndrueshme dhe të hapura me paraqitjen e të qenit të përhapura dhe fizikisht të interpretueshme. Autori shpjegoi se përdorimi i konceptit të MP(Matching Pursuit) na e bën të mundur këtë gjë. Avantazhet e kësaj janë aftësia për të kapur strukturën e pandare të çdo tipi sinjali, dhe të skicosh nga një sinjal të madh në një sinjal të vogël. Më e rëndësishmja është me përshtatshme me zhurmën në sfond dhe mund të kapë karakteristika në sinjale ku MFCC-ja mund të dështojë. MP është një metodë e pëlqyeshme për të bërë një prezantim të përgjithshëm dhe për të reduktuar energjinë e mbetur në sa më pak atome. Meqënëse MP selekton atomet në mënyrë lineare midis platformës së valëve në fjalorin që lidh më mirë strukturat e sinjalit të zërit. Iteracioni mund të ndalet nëse koeficientët e lidhura me selektimin e atomeve bien poshtë një limiti ose kur një numër i caktuar i atomeve të selektuar është arritur. Një kriter tjetër ndalimi është duke përdorur energjinë e mbetur të sinjalit. (sharawankar & Thakare, 2013).

Tiparet e MP të selektuara nga procesi i mëposhtëm:

1. Windowing
2. Dekompozimi
3. Ndalimi i procesit pas marrjes së atomeve
4. Regjistrimi i frekuencave për çdo atom
5. Gjetja e mesatares dhe e devijimit standart që i korrespondon çdo parametri në mënyrë të ndarë dhe që korespondon në 4 vlera karakteristikash.
6. Zgjedhja e atomeve për të nxjerrë tiparet për trajnimin dhe testimin e të dhënave.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Efektiviteti i metodave për nxjerrjen e tiparave është konfirmuar tashmë në njohjen e zërit ose në eksperimentet për përforcimin e zërit por mbetet e vështirë për të njohur zërin në ambjentet rikumbuese. Siç përmëndet nga Sharawankar & Thakare, nëse përgjigja e impulsit në një dhomë është më e gjatë se gjatësia e transformimit me kohë të shkurtër Furie (DFT); efektet e jehonave janë multiplikative dhe mbledhëse në boshtin e fuqisë spektrale. Për më tepër bëhet e vështirë të vlersojmë efektet e jehonva në kohë ose në boshtin e frekuencave.

Një metodë e re për ‘nxjerrjen e tipareve’ është ‘integrated phoneme subspace (IPS)’ e cila bazohet në frekuencat logaritmike të bankes së filtrave Mel. Bazuar në eksperimentet e kryera ‘feature’ i propozuar krijohet nga matricat e transformimit që janë lineare dhe time-invariant. Metoda e propozuar mund të përdoret në kombinim me metodat e tjera siç janë proçesimi i sinjaleve të zërit ose përshtatja e modelit, në mënyrë që të përmirsojmë saktësinë e njohjes në ambjentet e jetës reale. Disa teknika të reja janë zhvilluar duke përdorur një kombinim të shumë teknkave. Ka një nevojë për të zhvilluar metoda të reja hibride që do të japin performancë më të mirë në fushën e njohjes së zërit (2013).

5.3 Konkluzionet

Gjuha e folur nuk është thjesht një mënyrë për të pasur akses në informacione por është një informacion në vetvete. E folura është mënyra primare e komunikimit midis qënieve njerzore dhe forma më natyrale dhe më efiçente për këmbimin e informacionit midis njerzve në të folur. Njohja e zërit mund të përcaktohet si proçesi i konvertimit të sinjalit të zërit në një varg fjalësh nëpërmjet algoritmeve të implementuara si programe kompjuterike. Për më tepër është e natyrshme për njerzit të kenë pritshmëri në komunikimin me anë të zërit me kompjuterat të cilat mund të njohin zërat në shume gjuhë të ndryshme. India ka një gjuhë shumë të pasur e cila ka 18 gjuhe zyrtare, të cilat shkruhen ne 10 mënyra të ndryshme shkrimi. Makina e njohjes së zërit paraqit një

METODAT E ANALIZËS SPEKTRALE PËR ASR

gjenerim vargje fjalësh që i përshtaten më mirë sinjalit të zërit. Disa nga programet e njohura përfshijnë realitetin virtual, kërkimin e shumfishtë, auto-attendants, informacioni i udhëtimit dhe rezervimit, përkthyesit, kuptimi i gjuhës natyrale dhe shumë aplikime të tjera.

Gjetjet e bëra më lartë tregojnë se ‘përkulja’ e spektrit të DFT ose FFT në mënyrë të drejtpërdrejtë na siguron një përafrim më të saktë në shkallën perceptuale në krahasim me mesatarizimin e bankës së filtrave. Gjithashtu është e rëndësishme të kemi parasysh që aplikimi i drejtpërdrejt i spektrit të DFT të treguar si në figuren 20 shmang përdorimin e bankës së filtrave në mënyrë tradicionale. Për më tepër si një prove e kërkimit primar të kryer nga Parinam, është treguar së përdorimi i drejtpërdrejtë i vlerave spektrale të FFT dhe përdorimi i tipareve që integrojnë një frekuencë Mel ose të ngjashme me Mel; mund të shmang hapin e bankës së filtrave në procesimin e zërit. Megjithatë gjetjet e Parinam nuk na kanë treguar ndonjë disavantazh në përdorimin e bankës së filtrave, do të kete një kursim kohe duke shkurtuar dhe përmirësuar në mënyrë domethënëse sistemin. Përmirësimet në sistemet e procesimit të zërit janë tepër të dëshirueshme meqenëse ekziston një oqean i pafund mundësish në sistemin e ASR.

Bazuar në kërkimet e kryera unë kam arritur në përfundimin se ‘përkulja’ e spektrit të FFT në mënyrë të drejtpërdrejtë në vënd që të përdorim mesatarizimin e bankës së filtrave të cilat na sigurojnë një përafrim më preciz të shkallës perceptuale. Përdorimi i drejtpërdrejtë i vlerave spektrale të FFT janë po aq efektive sa përdorimi i bankës së filtrave Mel ose Gammatone, të cilat na garantuan se tiparet e nxjerra nga vlerat spektrale të FFT marrin parasysh një frekuencë Mel ose të ngjashme me të siç janë koeficientët cepstral të frekuencës Mel (MFCCs). Llogaritja e sinjaleve të zërit duke përdorur FFT ose tiparet spektrale të bankës së filtrave dhe përdorimi i një teknike të bazuar në ‘sliding block for spectral features’ shikohet që është më efektive në termat e

METODAT E ANALIZËS SPEKTRALE PËR ASR

saktësisë së ASR. Unë i konsideroj efektet e një procesi të përmirësuar ASR në fushat e diskutuara që të jenë një hap më përpara në fushën e ASR.

5.4 Referencat

Arisoy, E., Sainath, T.N., Kingsbury, B., Ramabhadran, B. (2012) Deep Neural Network Language Models. In Proc. NAACL-HLT workshop, Montreal, Canada. pp. 20–28.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Rossell, M. (2006) An Introduction to Front-End Processing and Acoustic Features for Automatic Speech Recognition.

Besacier, L., B, Barnard, E., Karpov, A., Schultz, T. (2014) Automatic speech recognition for under-resourced languages: A survey. Elsevier.

Bhanuprasad, K., Svenson. M. (2008) Errgrams - A way to Improving ASR for Highly Inflective Dravidian Languages. In Proc. 3rd International Joint Conf. on Natural Language Processing IJCNLP'08, India, pp. 805-810.

Cetin, O. (2008) Unsupervised adaptive speech technology for limited resource languages: a case study for Tamil *SLTU'08*, Hanoi, Vietnam

Chan, H. Y., Rosenfeld. R. (2012) Discriminative pronunciation learning for speech recognition for resource scarce languages. In *Proceedings of the 2nd ACM Symposium on Computing for Development*, p. 12.

Cucu, H., Besacier, L., Burileanu, C., Buzo, A. (2011) Investigating the Role of Machine Translated Text in ASR Domain Adaptation: Unsupervised and Semi-supervised Methods. In Proc. ASRU 2011, Haëaii, USA

Cucu, H., Besacier, L., Burileanu, C., Buzo, A. (2012) ASR Domain Adaptation Methods for Low-Resourced Languages: Application to Romanian Language. EUSIPCO'2012, Bucarest, Romania

METODAT E ANALIZËS SPEKTRALE PËR ASR

Cucu, H., Besacier, L., Burileanu, C., Buzo, A. (2013) SMT-Based ASR Domain Adaptation Methods for Under- Resourced Languages: Application to Romanian. *Speech Communication Journal, Special Issue on Processing Under-Resourced Languages*

Denoual, E., Lepage, Y. (2006) The character as an appropriate unit of processing for non-segmenting languages. NLP Annual Meeting, Tokyo, Japan, pp. 731-734

Davis, S. Mermelstein, P. (1980) Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp. 357-366

Evangelho, J (2013) NVIDIA's Volta GPU Launches In 2016, Delivers 1TB/s Of Memory Bandwidth. *Forbes*. Retrieved from: <http://www.forbes.com/sites/jasonevangelho/2013/03/19/nvidias-volta-gpu-launches-in-2016-delivers-1tbs-of-memory-bandwidth/>

Gemmeke, J. F., Van hamme, H. (2011) A hierarchical exemplar-based sparse model of speech with an application to ASR. *IEEE ASRU 2011*. Hawaii, USA.

Ghoshal, A., Jansche, M., Khudanpur, S., Riley, M., Ulinski, M. (2009) Webderived Pronunciations. *IEEE ICASSP*

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., Senior, A. W., Vanhoucke, V., Nguyen, P., Sainath, T. & B. Kingsbury (2012) Deep neural networks

METODAT E ANALIZËS SPEKTRALE PËR ASR

for acoustic modeling in speech recognition. *Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97

Hibare, R., Vibhute, A. (2014) Feature extraction techniques in speech processing: a survey. *International journal of computer applications* Vol: 107 (5)

Huang, X., Acero, A., Hon, H. (2001) *Spoken Language Processing: A guide to theory, algorithm, and system development*. Prentice Hall

Institute of Electrical and Electronics Engineers (IEEE)

Dictionary.com, "cepstrum," in *The Free On-line Dictionary of Computing*. Source location: Denis Howe. W<http://dictionary.reference.com/browse/cepstrum>. Available: <http://dictionary.reference.com>. Accessed: March 11, 2015.

Irino, T. & Patterson, R. D. (1997) A time-domain, Leveldependent auditory filter: The gammachirp. *J. Acoust.Soc. Am.* 101(1): 412-419

Jing, Z., Min, Z. (2010) Speech recognition system based improved DTW algorithm, in *Proc. Int. Conf. on Computer, Mechatronics, Control and Electronic Engineering. CMCE-2010*, Vol. 5, pp. 320-323.

Karanasou, P., Karanasou, Lamel, L. (2010) Comparing SMT Methods for Automatic Generation of Pronunciation Variants,” *IceTAL 2010*, pp. 167, Reykjavik, Iceland

METODAT E ANALIZËS SPEKTRALE PËR ASR

Karpov, A., Kipyatkova, I., Ronzhin, A. (2011) Very Large Vocabulary ASR for Spoken Russian with Syntactic and Morphemic Analysis, in Proc. Interspeech'2011, Florence, Italy, 2011, pp. 3161–3164.

Laurent, A., Deleglise, P., Maignier, S. (2009) Grapheme to phoneme conversion using an SMT system. Interspeech 2009, pp. 708-711

Le, V., Besacier, L. (2009) Automatic Speech Recognition for Under-Resourced Languages: Application to Vietnamese Language. IEEE Transactions on Audio, Speech and Language Processing, Vol: 17, Iss: 8, pp: 1471 – 1482

Marple, S.L. Jr. (1987) Digital Spectral Analysis With Applications, *Prentice-Hall*, Englewood Cliffs, Neë Jersey, USA.

Markel, J. and Gray, A.H. (1980) Jr., Linear Prediction of Speech. *Springer-Verlag*, Neë York, New York, USA

Motlíček, P. (2003) Feature Extraction in Speech Coding and Recognition. Oregon Graduate Institute of Science and Technology, pp. 1-50, 2002

NIST. (2009) The US NIST 2009 (RT-09) Rich Transcription Meeting Recognition Evaluation Plan

Parinam, V. N. D. (2013) Spectral analysis methods for automatic speech recognition applications. State University of New York at Binghamton

METODAT E ANALIZËS SPEKTRALE PËR ASR

Patterson R.D. & Moore, B.C.J. (1986) Frequency Selective in Hearing, chapter Auditory filters and excitation patterns as representations of frequency resolution, pp. 123–177, Academic Press Ltd., London

Piccone, J. (1992) Signal Modelling Techniques in Speech Recognition. *Texas Instruments*

Shrawankar, U., Thakare, V. (2013) Techniques for feature extraction in speech recognition system: a comparative study.

Siniscalchi, S. M., Reed, J., Svendsen, T., Lee, C. (2013) Universal attribute characterization of spoken languages for automatic spoken language recognition. *Computer Speech & Language* 27(1): 209-227

Silveira, M. A., Schroeder, C. P., Paulo, J., Lustosa da Costa, C., De Oliveira, C. D., Apolinario Jr, J. A., Serrano, A. M. R., Quintilliano, P., De Sousa Jr, R. T. (2013) Convolutional ICA based speaker identification using Mel Frequency Cepstral Coefficients and Gaussian Mixture Models. *International journal of forensic computer science* Vol: 1 pp: 23-34

Sainath, T. N, Kingsbury, B., Mohamed, A., Ramabhadran, B. (2013) Learning filter banks within a deep neural network framework. *IEEE*.

Seide, F., Li, G., Chen, X., Yu, D. (2011) Feature engineering in Context-Dependent Deep Neural Networks for conversational speech transcription. In Proc. ASRU-2011 International Workshop, pp. 24-29.

METODAT E ANALIZËS SPEKTRALE PËR ASR

Schlippe, T., Ochs, S., Schultz, T. (2013) web-based tools and methods for rapid pronunciation dictionary creation. *Speech Communication Journal, Special Issue on Processing Under- Resourced Languages*

Shikano, K. (1986) Evaluation of LPC Spectral Matching Measures for Phonetic Unit Recognition. TM No. CMU-CS-86-108, *Computer Science Department*, Carnegie-Mellon University, Pittsburgh, PA, US, 15213

Trentin, E., Gori, M. (2001) A Survey of Hybrid ANN/HMM Models for Automatic Speech

Recognition, *Neurocomputing* 37(1), pp. 91-126.

Tachbelie, M., Abate, T., Besacier, L. (2013) Using different acoustic, lexical and language modeling units for ASR of an under-resourced language – Amharic”. ; *Speech Communication Journal, Special Issue on Processing Under-Resourced Languages*

Tarjan, B. and Mihajlik, P. (2010) On Morph-Based LVCSR

Vu, N.T., Kraus, F., Schultz. T. (2010) Multilingual A-stabil: A new confidence score for multilingual unsupervised training, In Proc. SLT, USA

Vu, N.T., Kraus, F., Schultz. T. (2011) Rapid building of an ASR system for Under-Resourced Languages based on Multilingual Unsupervised Training. In Proc. Interspeech, Italy

METODAT E ANALIZËS SPEKTRALE PËR ASR

Vu, N.T., Metze, F., Schultz, T. (2012) Multilingual bottle-neck feature for under resourced languages. In Proc. SLTU, South Africa

Vu, N.T., Breiter, W., Metze, F., Schultz, T. (2012) An Investigation on Initialization Schemes for Multilayer Perceptron Training Using Multilingual Data and their Effect on ASR Performance. In Proc. Interspeech, USA,

van Heerden, C., Kleynhans, N., Barnard E., Davel, M. Pooling ASR data for closely related languages. (2010) In Proceedings of the Workshop on Spoken Languages Technologies for Under-Resourced Languages SLTU 2010. Penang, Malaysia, pp. 17-23.

Young, S. (2008) HMMs and Related Speech Recognition Technologies. In Springer Handbook of Speech Processing, *Springer-Verlag Berlin Heidelberg* pp. 539-557.

Yapanel, U. H. & Hansen, J.H.L. (2008) A new perceptually motivated MVDR-based acoustic front-end (PMVDR) for robust automatic speech recognition. *Speech Comm.*, Vol. 50, pp. 142-152

Whittaker, E.W.D. (2000) Statistical language modelling for automatic speech recognition of Russian and English. *Cambridge University PhD thesis* 140 p.

Yu, D., Siniscalchi, S. M., Deng, L., Lee, C (2012) Boosting attribute and phone estimation

accuracies with deep neural networks for detection-based speech recognition. In Proc. ICASSP-2012, pp. 4169-4172.