دانشگاه بیرجند

دانشکده مهندسی

پایان نامه کارشناسی ارشد مهندسی برق- مخابرات(سیستم)

بهسازی گفتار به روش تک کاناله در محیطهای با نویز غیر ایستان

استاد راهنما

دکتر حسن فرسی

استاد مشاور

دکتر ناصر مهرشاد

پژوهش و نگارش

محمد حسین زارع بیدکی

آبان ۱۳۸۸

بسم الله الرحمن الرحيم

چکیده

بهسازی گفتار یکی از مهمترین موضوعات در زمینه مخابرات و پردازش سیگنال می باشد. در مخابرات صوتی ، ممکن است سیگنال صحبت با نویز محیط آلوده شده، و در نتیجه کیفیت ارتباط را به دلیل کاهش قابلیت فهم گفتار تحت تاثیر قرار دهد. همچنین فشرده سازی صحبت نویزی با کد کننده های نرخ پایین صحبت، باعث کاهش قابل توجه کیفیت گفتار خواهد شد،که این امر به دلیل خطای تخمین مکرر در پارامترهای مدل تولید کننده صحبت می باشد. این مشکل را می توان تا حد زیادی با استفاده از سیستم بهسازی گفتار (حذف نویز) ، که با حذف مولفه های نویز در سیگنال ورودی باعث افزایش کیفیت گفتار می شود، برطرف کرد.

در میان تکنیکهای بهسازی گفتار، روشهای مبتنی برتبدیل فوریه گسسته (یا STSA) بیشتر مورد بررسی قرار گرفته اند. رایج بودن بهسازی گفتار بر اساس STSA تنها به دلیل پیچیدگی محاسباتی کم آنها نمی باشد، بلکه اخیرا تکنیکهای ارائه شده در این روش، به طور قبل توجه قادر به افزایش بهبود کیفیت گفتار می باشند.در این پایان نامه جزئیات تکنیکهای معروف بهسازی گفتار بر اساسSTSA بررسی می‌شوند. و در انتها روشی برای بهبود عملکرد سیستم بهسازی گفتار در شرایط نویز غیر ایستان پیشنهاد می گردد.

تقدیم به خانواده‌ام

تقدیر و تشکر

اکنون که به یاری خداوند متعال موفق به ارائه این پایان نامه گردیده‌ام بر خود لازم می دانم تا از زحمات وتلاشهای بی شائبه استاد راهنمای گرامی‌ام، جناب آقای دکتر فرسی که از آغاز دورهٔ کارشناسی ارشد همواره از راهنمایی‌های علمی و مشورتهای ایشان استفاده نموده‌ام،کمال تشکر و قدردانی را داشته باشم.
همچنین رعایت ادب لازم می دارد تا از آقای دکتر مهرشاد که زحمت مشاوره و راهنمایی مرا در تکمیل این پایان نامه بر عهده داشته‌اند تشکر کنم.
در پایان از اساتید محترمم جناب آقایان دکتر ندا و دکتر فرخی که افتخار شاگردی آن عزیزان نصیب من گردید، خاضعانه و خالصانه تشکر و قدردانی می‌کنم.

فهرست مطالب

عنوان صفحه

فصل اول: مقدمه ۱

۱-۱) ساختار این پایان نامه ۴

فصل دوم: بهسازی گفتار با استفاده از روشهای مبتنی بر STSA ۵

۲-۱) مقدمه ۶

۲-۲) الگوریتم کلی بهسازی گفتار مبتنی بر STSA ۷

۲-۳) روش تفریق طیفی ۹

۲-۴) تخمین دامنه طیف با استفاده از معیار ۱۲

۲-۵) روش فیلتر وینر ۱۲

۲-۶) تخمین MMSE دامنه طیف ۱۴

۲-۷) بهسازی گفتار با توجه به عدم قطعیت حضور صحبت ۱۶

۲-۷-۱)تخمین MM-LSA دامنه طیف ۱۸

۲-۷-۱-۱) تخمین احتمال پیشین ۱۹

۲-۷-۲)تخمین OM-LSA ۲۱

۲-۷-۲-۱) تخمین Priori SNR ۲۲

۲-۷-۲-۲) تخمین احتمال غیاب صحبت پیشین(PSA) ۲۶

۲-۸)مقایسه عملکرد روشهای متداول بهسازی گفتار ۲۹

۲-۹)روشهایی برای کاهش نویز موزیکال و سطح نویز باقیمانده ۴۰

۲-۹-۱)تخمین posteriori SNR با استفاده از priori SNR ۴۰

۲-۹-۲)تخمین غیر علی priori SNR ۴۲

۲-۱۰)تخمین نویز ۴۹

۲-۱۰-۱)تخمین نویز به روش Hard-Decision ۴۹

۲-۱۰-۲)تخمین نویز به روش Soft-Decision ۵۰

۲-۱۰-۳)تخمین نویز به روش Mixed-Decision ۵۱

فصل سوم: تخمین نویز در شرایط نویز غیر ایستان ۵۳
۳-۱) مقدمه ۵۴
۳-۲)تخمین نویز به روش MS ۵۶
۳-۳)تخمین نویز به روش IMCRA ۵۸
۳-۳-۱)تخمین مینیمم کنترل شده ۶۱
۳-۳-۲)پیاده سازی الگوریتم ۶۶
۳-۵)تکنیک پیشنهادی برا بهبود روش IMCRA ۶۷
۳-۶)نتایج شبیه سازی روشهای مطرح شده ۶۸

نتیجه گیری ۷۴
پیشنهادهایی برای ادامه کار ۷۵

ضمیمه(۱) ۷۵
ضمیمه(۲) ۷۷

مراجع ۸۰

فهرست شکلها

عنوان **صفحه**

شکل۲-۱): بلوک دیاگرام کلی بهسازی گفتار بر پایه STSA ۹

شکل۲-۲): بلوک دیاگرام محاسبه بهره طیفی بهسازی گفتار به روشOM-LSA ۲۳

شکل۲-۳): احتمال حضور صحبت شرطی بر حسب priori SNR برای مقادیر مختلف احتمال غیاب صحبت پیشین و مقدار posteriori SNR ثابت($\gamma > 1$) ۲۵

شکل۲-۴): بلوک دیاگرام محاسبه P_{frame} ۲۹

شکل۲-۵): شکل موج زمانی و اسپکتوگرام سیگنال صحبت تمیز ۳۱

شکل۲-۶): شکل موج زمانی و اسپکتوگرام سیگنال صحبت نویزی ۳۲

شکل۲-۷): شکل موج زمانی و اسپکتوگرام سیگنال صحبت بهسازی شده به روش PSS ۳۳

شکل۲-۸): شکل موج زمانی و اسپکتوگرام سیگنال صحبت بهسازی شده به روش GBSS ۳۴

شکل۲-۹): شکل موج زمانی و اسپکتوگرام سیگنال صحبت بهسازی شده به روش ML ۳۵

شکل۲-۱۰): شکل موج زمانی و اسپکتوگرام سیگنال صحبت بهسازی شده به روش WF ۳۶

شکل۲-۱۱): شکل موج زمانی واسپکتوگرام سیگنال صحبت بهسازی شده به روش MMSE-LSA ۳۷

شکل۲-۱۲): شکل موج زمانی واسپکتوگرام سیگنال صحبت بهسازی شده به روش MM-LSA ۳۸

شکل۲-۱۳): شکل موج زمانی و اسپکتوگرام سیگنال صحبت بهسازی شده به روش OM-LSA ۳۹

شکل۲-۱۴): هیستوگرام تخمینهای priori SNR وpriori SNR (الف): SNR=۲۰ db (ب):SNR=۴ db ۴۱

شکل۲-۱۵): اسپکتگرام سیگنال بهسازی شده (الف): با استفاده از رابطه(۲-۱۸) (ب) با استفاده ازرابطه(۲-۸۴) ۴۳

شکل۲-۱۶):SNR ها درفریمهای متوالی- posteriori SNR (منحنی نقطه گذاری شده) - تخمین priori SNR به روش DD (منحنی خط چین) – تخمین غیر علی priori SNR (منحنی توپر)..... ۴۶

شکل۲-۱۷): اسپکتگرام سیگنال بهسازی شده (الف) : با استفاده از روش تخمین DD (ب) بااستفاده از روش غیر علی ۴۸

شکل۲-۱۸): بهبود SEGSNR برحسب نرخ خطای VAD در روشهای مختلف تخمین نویز ۵۲

شکل ۳-۱) پریودگرام هموار شده و سطح نویز تخمین زده سیگنال نویزی شده ۵۷

۳-۲): تخمین ایدآل نویز(خط پررنگ) و روشهای مورد برسی- (الف)تخمین ایده ال و IMCRA – (ب) تخمین ایده ال وMS-(ج) تخمین ایده آل و روش پیشنهاد شده ۷۰

۳-۳): تخمین ایدآل نویز(خط نازک) ، IMCRA (خط بریده) و روش پیشنهادی (خط پررنگ)که روی فرکانس میانگین گیری شده اند ۷۱

فهرست جداول

عنوان **صفحه**

جدول۲-۱):بهبود SEGSNR برای روشهای بهسازی گفتاربرپایهSTSA درشرایط نویز وسیله نقلیه ۳۰
جدول ۲-۲):بهبود SEGSNR با استفاده از تخمینهایDD و غیر علی priori SNR ۴۷
جدول ۳-۱): خطای تخمین نسبی segmental برای تخمینگرهای IMCRA ، MS و روش پیشنهادی ۷۱
جدول۳-۲): بهبود segmental SNR برای تخمینگرهای IMCRA ، MS و روش پیشنهادی ۷۲

حروف اختصاری

DD : Decision-Direct
DFT : Discrete Fourier transform
GSP : Generalized spectral subtraction
IMCRA : Improved Minima controlled Recursive Averaging
LSA : Log spectral amplitude
ML : Maximum-liklihood estimation
MMSE : Minimum Mean Square error estimation
MM-LSA : Mulltiplicativily-modified log spectral amplitude estimation
MS : Minimum statistics
MCRA : Minima controlled Recursive Averaging
NC : Noncausal
OM-LSA : Optimally-modified log spectral amplitude
PSA : Priori probability of speech presence
SD : Soft-decision
SEGSNR : Segmental SNR
SNR : Signal to Noise ratio
STSA : Short -time spectral amplitude
STFT : Short -time Fourier transform
SP : Spectral subtraction
VAD : Voice activity detection

فصل اول

مقدمه

در طول چند دهه اخیر با پیشرفت روشهای پردازش گفتار[1] و پیشرفت سیستمهای مبتنی بر سیگنال های صوتی (همچون سیستمهای بازشناسی گفتار[2])، مقابله با اثرات نامطلوب و مخرب عوامل محیطی یک امر بدیهی به نظر می رسد. همانطور که می دانیم سیگنال گفتار از بدو تولد ، دستخوش عوامل مختلف محیطی از جمله نویز[3]، انعکاس[4]، تداخل[5]، و اعوجاج[6] می شود. این پدیده ها، ضمن پایین آوردن کیفیت گفتار مورد نظر عملکرد سیستم پردازنده صوتی همچون سیستمهای بازشناسی گفتار، کدکننده ها[7] و . . . را نیز دچار مشکل می کنند.

منظور از بهسازی سیگنال گفتار ، تلاش برای بالا بردن قابلیت فهم این سیگنال و بهبود عملکرد سیستم وابسته به آن ، در شرایطی که تحت تاثیر عوامل تخریبی قرار گرفته باشد، می باشد. بر اساس تعداد میکروفنهای در دسترس بهسازی گفتار به دو دسته روشهای تک کاناله[8] (میکروفنه) و روشهای چند کاناله[9] تقسیم بندی می شوند .

در روشهای تک کاناله تنها یک میکروفن جهت دریافت سیگنال صحبت موجود می باشد. پس در این حالت تنها ورودی برای سیستم بهسازی گفتار یک سیگنال صحبت نویزی شده می باشد. در این روش با فرض ایستان[10] بودن نویز در طول نواحی فعال صحبت، مشخصات آماری نویز از طریق مقادیر آن در نواحی غیر فعال صحبت (سکوت) به دست می آید. این نواحی با استفاده از یک آشکار کننده سیگنال صحبت[11] مشخص می شوند.

روشهای مطرح شده در این دسته از سیستمهای بهسازی گفتار اکثرا براساس تکنیکهای حوزه تبدیل[12] ، فیلترهای وفقی[13] و روشهای مبتنی بر مدل[14] می باشند. تکنیکهای بر پایه حوزه تبدیل بعد از تبدیل سیگنال نویزی از حوزه زمان به حوزه دیگر ، مولفه های نویز را حذف کرده و سپس با اعمال معکوس تبدیل، سیگنال بهسازی شده را بازسازی می کنند. روشهای تبدیل حوزه معروف ، تبدیل فوریه گسسته[15] ، تبدیل کسینوسی

1- Speech processing
2- Speech recognition
3- Noise
4- Reflection
5- Interference
6- Distortion
7- Coder
8- Single channel algorithms
9- Multi channel algorithms
10- Stationary
11- Voice activity detection
12- Transform domain techniques
13- Adaptive filtering
14- Model-based methods
15- Discrete Fourier transform

گسسته[1] ، تبدیل KLT[2] و تبدیل موجک[3] می باشند. روشهای فیلتر وفقی حذف نویز را با استفاده از یک فیلتر وفقی مانند فیلتر کالمن[4] انجام می دهند. در تکنیکهای بر پایه مدل ، سیگنال نویزی را با استفاده از مدل صحبت ابتدایی[5] ، مانند مدل Voiced/Unvoiced دسته بندی کرده و سپس بر اساس مدلهای صحبت دسته بندی شده بهسازی گفتار انجام می شود.[۱]

در روشهای چند کاناله ، دو یا چند گیرنده در ورودی سیستم بهسازی گفتار مورد استفاده قرار می گیرد . مسلما افزایش تعداد میکروفن ورودی بر توانایی سیستم در حذف نویز خواهد افزود ، ولی در مقابل بر هزینه و پیچیدگی سیستم هم اضافه خواهد شد. در این زمینه نیز تاکنون روش های مختلفی با بهره گیری از فیلترهای وفقی ، فیلتر وینر چند کاناله و . . . ارائه شده که هر یک ، ضمن دارا بودن ویژگی های مطلوب ، ناکارآمدگیهای خاص خود را نیز به دنبال دارند.

بهسازی سیگنال گفتاربه منظور تحقق اهداف متنوعی انجام می گیرد. مهمترین کاربردهای این فرآیند ، عبارتند از :

۱) افزایش کیفیت سیگنال گفتار ، به منظور ارتقای عملکرد سیستم های مبتنی بر این سیگنال ، همچون سیستم بازشناسی گفتار و . . .
۲) بهبود ضبط صدا در محیط های پر نویز و پر انعکاس .
۳) بهبود گفتار دریافتی در گیرنده های مخابراتی که در اثر عواملی همچون محدودیت های پهنای باند ، اعوجاج های خطی و غیر خطی سیستم ها و . . . تخریب شده اند.
۴) استفاده در کاربردهای نظامی و جاسوسی.
۵) بهبود کیفیت صدا در سمعک .

از یکسو، تنوع عوامل مخرب محیطی واز سوی دیگر محدودیت های عملی موجود در بهسازی سیگنال گفتار ، طراحی سیستم واحدی را که قادر به حذف کامل اغتشاشات محیطی باشد ، با مشکل مواجه می نمایند. از این رو ، به منظور بهسازی موثر سیگنال گفتار ، لازم است که در شرایط مختلف ، بر اساس ضوابط موجود ، الگوریتم بهسازی مقتضی انتخاب و اجرا شود.

1- Discrete Cosine transform
2- Karhunen–Lo`eve transform
3- Wavelet transform
4- Kalman filter
5- Priori speech model

ساختار این پایان نامه

در این پایان نامه تنها روشهای مبتنی بر تبدیل فوریه گسسته، به دلیل بازدهی بالا در تخمین نویز بررسی شده است. این روشها بر اساس دامنه های طیف زمان کوتاه (STSA[1]) بررسی می شوند به طوری که ابتدا تبدیل STFT[2] سیگنال صحبت نویزی شده محاسبه شده سپس هریک از بینهای فرکانسی در هر فریم در یک بهره ضرب می شود، سیگنال صحبت بهسازی شده با انجام عکس تبدیل STFT از این طیف به دست می آید. در فصل دوم الگوریتم کلی بهسازی گفتار مبتنی بر STSA بیان گردیده و روشهای معروف آن که شامل تفریق طیفی ، فیلتر وینر، تخمین ML[3] ، تخمین MMSE[4] ، MM-LSA[5] و OM-LSA[6] می باشند، باشند، بیان شده است. همچنین نتایج شبیه سازی انجام شده برای هر یک از این روشها، که با استفاده از برنامه MATLAB انجام شده ، نمایش داده شده است. یکی از قسمتهای اصلی در بهسازی گفتار تخمین واریانس نویز می باشد. در انتهای این فصل روشهای ابتدایی تخمین واریانس نویز که در شرایط نویز ایستان عملکرد قابل قبولی دارند معرفی شده اند.

در فصل سوم ابتدا روشهای معروف تخمین واریانس نویز در شرایط نویز غیر ایستان، که شامل روشهای MS[7] و IMCRA[8] می باشند معرفی شده، سپس ایده ای برای بهبود روش IMCRA پیشنهاد شده است. در در نهایت عملکرد این روشها با استفاده از شبیه سازی در محیطهای با نویز مختلف مقایسه شده است.

1- Short -time spectral amplitude
2- Short -time Fourier transform
3- Maximum-liklihood estimation
4- Minimum Mean Square error estimation
5- Mulltiplicativily-modified log spectral amplitude estimation
6- Optimally-modified log spectral amplitude
7- Minimum statistics
8- Improved Minima controlled Recursive Averaging

فصل دوم

بهسازی گفتار با استفاده از روشهای مبتنی برSTSA

۱-۱) مقدمه

بهسازی گفتار مبتنی بر STSA شامل روشهای تفریق طیفی ، فیلتر وینر، تخمین ML، تخمینMMSE ، MM-LSA و OM-LSA می باشد. در تمام این روشها ابتدا تبدیل STFT سیگنال صحبت نویزی شده محاسبه شده سپس دامنه هریک از بینهای فرکانسی در هر فریم در یک بهره ضرب می شود، سیگنال صحبت بهسازی شده با انجام عکس تبدیل STFT از این طیف به دست می آید. روشهای MM-STSA و OM-LSA بر خلاف سایر روشها ، از احتمال عدم قطعیت حضور صحبت[1] نیز استفاده می کنند، به طوری که بهره اصلاح کننده دامنه طیف به احتمال حضور صحبت وابسته بوده و در زمانهایی که احتمال حضور صحبت افزایش می یابد برای جلوگیری از از بین رفتن سیگنال صحبت بهره افزایش و در شرایط سکوت تا حد امکان بهره کاهش می یابد.

در این فصل ابتدا اصول بهسازی گفتار مبتنی بر STSA بیان شده و پارامترهای موثر در آن معرفی می-گردد، سپس بهره اصلاح کننده دامنه طیف برای هریک از روشهای آن به دست می آید. درادامه تکنیکهایی برای کاهش نویز موزیکال[2] که یکی از نویزهای باقیمانده در سیگنال صحبت بهسازی شده می باشد، بیان می گردد. یکی از پارامترهای مهم در محاسبه بهره اصلاح کننده دامنه طیف تخمین واریانس نویز می باشد. در پایان این فصل روشهای ابتدایی برای تخمین واریانس نویز در شرایط نویز ایستان بیان خواهد شد.

۲-۲) الگوریتم کلی بهسازی گفتار مبتنی بر STSA

اگر سیگنال صحبت x(n) با سیگنال نویز جمع شونده ناهمبسته[3] d(n) جمع شود، سیگنال نویزی y(n) حاصل می شود، که n اندیس زمان گسسته را نشان می دهد.

$$y(n) = x(n) + d(n) \qquad (۱-۲)$$

سیگنال y(n) با استفاده از تابع پنجره به فریمهایی که همپوشانی دارند تقسیم شده سپس با استفاده از تبدیل STFT[4] تجزیه می شود.

1- Uncertainty of Speech Presence
2- Musical noise
3- Uncorrelated
4- Short -time Fourier transform

$$Y(k,l) = \sum_{n=0}^{N-1} y(n+lM)h(n)e^{-j\left(\frac{2\pi}{N}\right)nk} \qquad (۲-۲)$$

k اندیس بین فرکانسی ، l اندیس فریم ، h پنجره آنالیزکننده با طول N (مثلا پنجره Hamming) و M گام فریم بندی می باشدکه به میزان همپوشانی فریمها بستگی دارد. با فرض اینکه $X(k,l)$ و $D(k,l)$ نشان دهنده تبدیل STFT سیگنال x(n) وd(n) باشند، تبدیل STFT سیگنال نویزی به صورت زیر بیان می شود.

$$Y(k,l) = X(k,l) + D(k,l) \qquad (۳-۲)$$

تبدیلSTFT سیگنال بهبود یافته با ضرب تابع بهره در هر یک از مولفه های طیفی به دست می آید.

$$\hat{X}(k,l) = G(k,l)Y(k,l) \qquad ; \quad \leq 0\ G(k,l) \leq 1 \qquad (۴-۲)$$

با استفاده از عکس تبدیلSTFT و پنجره سنتزکننده $\tilde{h}$ که با پنجره آنالیز کننده h متعامد می باشد سیگنال صحبت تمیز تخمین زده شده به صورت زیر به دست می آید.

$$\hat{x}(n) = \sum_{l}\sum_{k=0}^{N-1} \hat{X}(k,l)\tilde{h}(n-lM)e^{j\left(\frac{2\pi}{N}\right)k(n-lM)} \qquad (۵-۲)$$

در ادامه این فصل جهت سادگی نوشتاری اندیس فریم l نمایش داده نشده است.به طوری که مثلا $Y(k,l)$ به صورت نمایش Y_k داده شده است. بنابراین رابطه (۲-۴) به صورت زیر نمایش داده می شود.

$$\hat{X}_k = G_k Y_k \quad \text{for } 0 \leq G_k \leq 1 \qquad (۶-۲)$$

ضریب اصلاح G_k تابع posteriori SNR و priori SNR می باشد که به صورت زیر بیان می شوند.

$$\text{Posteriori SNR} \quad : \gamma_k = \frac{|Y_k|^2}{E(|D_k^2|)} \tag{۷-۲}$$

$$\text{Priori SNR} \quad : \xi_k = \frac{E(|X_k|^2)}{E(|D_k^2|)} \tag{۸-۲}$$

$E(|D_k^2|)$ واریانس آماری k امین مولفه فرکانسی نویز و $E(|X_k|^2)$ واریانس آماری k امین مولفه فرکانسی سیگنال گفتار می باشند. posteriori SNR به راحتی از طیف ورودی نویزی محاسبه می شود. $E(|D_k^2|)$ طبق روشهای تخمین نویز که در پایان این فصل و فصل بعد مورد بررسی قرارمی گیرد محاسبه می گردد.

هر چند $E(|X_k|^2)$ برای تخمین Priori SNR دردسترس نمی باشد، Ephraim و Malah روش DD [۱] را برای محاسبه آن پیشنهاد دادند [۲] که به صورت زیر می باشد.

$$\xi_k = \alpha \frac{\left|\hat{X}_k^{(l-1)}\right|^2}{E(\left|D^{l2}_k\right|)} + (1-\alpha)Max(\gamma_k^l - 1,0) \tag{۹-۲}$$

$0 \leq \alpha < 1$ و l اندیس فریم می باشد.

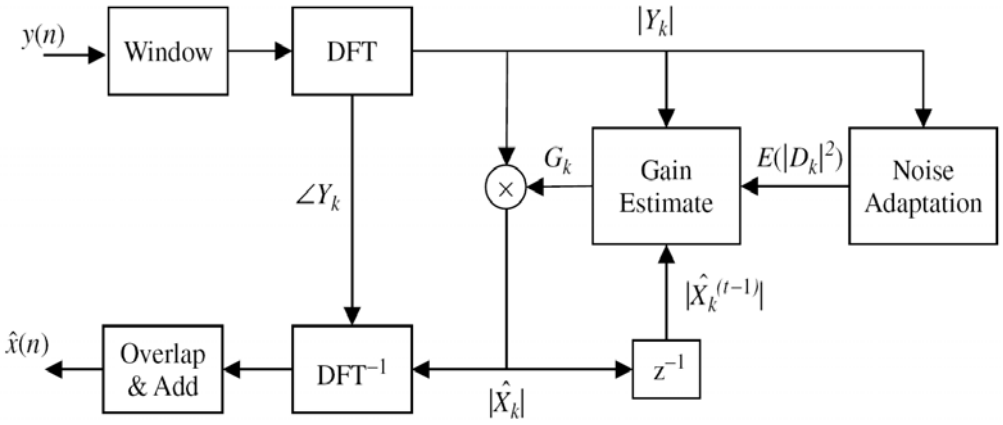
انگیزه اصلی در بهبود گفتارکم کردن نویز باقیمانده[۲] همزمان با ثابت بودن کیفیت گفتار می باشد. بنابراین روش بهینه نیاز به یک trade-off بین کاهش نویز[۳] وکیفیت گفتار[۴] دارد. برای مثال تخمین بیش از حد نویز ممکن است باعث کاهش کیفیت گفتار یا کاهش قابل فهم بودن آن می گردد. از سوی دیگر تخمین نویز ممکن است کم باشد که باعث ایجاد نویز باقیمانده قابل توجه خواهد شد. مهمترین نویز باقیمانده در بهبود گفتار نویز موزیکال یا tonal noise می باشد،که ترکیب سیگنالهای باند باریک بوده و با تغییر زمان وفرکانس قطع و وصل می شود.

بلوک دیاگرام کلی روشهای بهبود گفتار بر اساس STSA در شکل (۲-۱) نمایش داده شده است. سیگنالهای صحبت نویزی شده ابتدا با استفاده از پنجره کردن[۵] و تبدیل فوریه گسسته به STSA تبدیل می‌شود. دامنه طیف بهبود یافته با ضرب مولفه های طیفی سیگنال نویزی با بهره تخمین زده شده مطابق با آن به دست می آید. Overlap-add برای جبران اثر پنجره و حذف تغییرات ناگهانی بین دو فریم استفاده می

1- Decision-Direct
2- Residual noise
3- Noise reduction
4- Speech quality
5- windowing

شود. بخش مهم این روش تخمین دقیق بهره می باشد.

در ادامه روشهای بهسازی گفتار مبتنی برSTFT که شامل روشهای تفریق طیفی، تخمین با استفاده از معیار ML ، روش فیلتر وینر، تخمین MMSE دامنه طیف و روشهای بهسازی گفتار باتوجه به عدم قطعیت حضور صحبت که شامل روشهای MM-STSA و OM-LSA می باشند بیان گردیده و تابع بهره هر یک محاسبه می شود.



شکل۲-۱) : بلوک دیاگرام کلی بهسازی گفتار بر پایه STSA [۱]

۲-۳) روش تفریق طیفی[1]

قدرت طیف نویزی Y_k در معادله (۲-۳) را می توان به صورت زیر بیان کرد.

$$|Y_k|^2 = |X_k|^2 + |D_k|^2 + X_k^* D_k + X_k D_k^* \qquad (۱۰-۲)$$

X_k^* و D_k^* نشان دهنده مزدوج مختلط X_k و D_k می باشند. برای تخمین $|X_k|^2$ در معادله (۲-۱۰) با اعمال امید ریاضی به معادله از آنجایی که $|D_k|^2$ و $X_k^* D_k$ و $X_k D_k^*$ در دسترس نیستند خواهیم داشت.

$$|Y_k|^2 = \left|\hat{X}_k\right|^2 + E(|D_k|^2) + E(X_k^* D_k) + E(X_k D_k^*) \qquad (۱۱-۲)$$

1- Spectral subtraction

که $\left|\hat{X}_k\right|^2$ طیف قدرت بهبود یافته است.$E(|D_k|^2)$ را می توان از روش تخمین نویز که در بخش (۲-۱۰) بررسی شده به دست آورد. به دلیل اینکه x(n) با d(n) ناهمبسته است، $E(X_k^* D_k) = 0$ و $E(X_k D_k^*) = 0$ می باشند. بنابراین معادله (۲-۱۱) را می توان به صورت زیر بازنویسی کرد.

$$|Y_k|^2 = \left|\hat{X}_k\right|^2 + E(|D_k|^2) \qquad (۲-۱۲)$$

طیف توان بهبود یافته $\left|\hat{X}_k\right|^2$ را می توان با تفریق $E(|D_k|^2)$ از $|Y_k|^2$ به دست آورد. که تفریق طیف قدرت نامیده می شود. تفریق توان طیف را می توان به مرتبه طیفی دلخواه تعمیم داد.که GSS [1] نامیده می‌شود و به صورت زیر خواهد بود.

$$|Y_k|^\nu = \left|\hat{X}_k\right|^\nu + E(|D_k|^\nu) \qquad (۲-۱۳)$$

ν مرتبه طیفی[2] می باشد. اگر $\nu = 1$ در نظر گرفته شود GSS در معادله (۲-۱۳) تفریق دامنه و در صورتی که $\nu = 2$ باشد تفریق توان را نتیجه می دهد.

در عمل بهسازی گفتار بر اساس GSS به دلیل تغییرات سریع STSA سیگنال نویزی شده دارای نویز موزیکال می باشد. در بعضی از موارد اندازه نویز تخمین زده شده می تواند از دامنه طیف سیگنال ورودی بیشتر باشد در این صورت دامنه طیف بهبود یافته صفر در نظر گرفته می شود تا از منفی شدن آن جلوگیری شود، این امر باعث تولید نویز موزیکال می شود.

Berouti برای کاهش نویز موزیکال روش زیر را پیشنهاد نمود [۳]. اگر

$$\left|\tilde{X}_k\right|^\nu = |Y_k|^\nu - \alpha E(|D_k|^\nu) \qquad (۲-۱۴)$$

تخمین Berouti's GSS (BGSS) به صورت زیر به دست می آید.

$$\left|\hat{X}_k\right|^\nu = \begin{cases} \left|\tilde{X}_k\right|^\nu & \text{if } \left|\tilde{X}_k\right|^\nu > \beta E(|D_k|^\nu) \\ \beta E(|D_k|^\nu) & \text{otherwise} \end{cases} \qquad (۲-۱۵)$$

1- Generalized spectral subtraction
2- Spectral order