



دانشکده مهندسی برق و رباتیک
گروه الکترونیک

استفاده از روش PLP در کد کننده CELP

دانشجو:

جعفر نجفی کسلانی

استاد راهنما:

دکتر حسین مروی

استاد مشاور:

دکتر علی سلیمانی

پایان نامه جهت اخذ درجه کارشناسی ارشد

بهمن ماه ۱۳۸۷

چکیده

در طی سال های اخیر تلاشهای زیادی برای بهبود ارتباطات تلفنی و گسترش آن انجام گرفته و می گیرد. در این میان استفاده از کد کننده های صحبت به عنوان کد کردن یک سیگنال خاص که از صدای انسان و خصوصیات آن بهره مند می شود مورد توجه بسیاری از دانشمندان و موسسات مربوطه بوده است و به دلیل ویژگی های مذکور، روش های مختلفی برای بهره گیری از آن ویژگی ها بیان شده است. در این پایان نامه سعی شده است تا با مطالعه ای اجمالی بر روش های معرفی شده، اصول کارکرد این روش ها و مقایسه آنها، دید کلی نسبت به این روش ها ارائه گردد. همچنین کد کننده CELP بعنوان یکی از مشهورترین کد کننده ها به طور خاص بررسی و استاندارد FS1016 که مربوط به این کد کننده می باشد پیاده سازی شده است. به علاوه با اعمال الگوریتم PLP که در واقع بیشتر برای سیستم های تشخیص گفتار اتوماتیک مورد استفاده می گیرد بر روی کد کننده سلپ، کد کننده تغییر یافته ای طراحی گردید که در برخی از تست ها نتایج قابل قبولی بدست آمد که نتایج آن در این پایان نامه ذکر شده است.

کلمات کلیدی :

کد کننده های سیگنال صحبت، ضرایب تخمین خطی، فریم بندی، تخمین خطی ادراکی، کد کننده سلپ، کد کننده FS1016، تخمین خطی بلند مدت، کد کننده ملپ

عنوان	صفحه
فصل ۱ مقدمه	۱
۱-۱. معرفی.....	۲
۱-۲. تولید صوت توسط انسان.....	۳
۱-۳. مقدمه ای بر اصول کد کردن.....	۵
۱-۴. خصوصیات لازم یک کد کننده سیگنال گفتار.....	۷
۱-۵. طبقه بندی کد کننده های مختلف.....	۱۰
۱-۵-۱. طبقه بندی بر مبنای نرخ بیت.....	۱۰
۱-۵-۲. طبقه بندی بر مبنای نحوه کد کردن.....	۱۱
۱-۵-۳. کد کننده های شکل موج.....	۱۱
۱-۵-۴. کد کننده های پارامتریک.....	۱۱
۱-۵-۵. کد کننده های مرکب.....	۱۲
فصل ۲ معرفی کد کننده های متداول	۱۳
۲-۱. ساختار کلی یک کد کننده و دیکد کننده سیگنال صحبت.....	۱۴
۲-۲. کد کننده PCM.....	۱۶
۲-۳. مدل تخمین خطی تولید سیگنال.....	۱۷
۲-۴. روش تخمین خطی.....	۱۸
۲-۴-۱. انواع روش های تخمین خطی.....	۱۸
۲-۴-۲. روش محاسبه ضرایب LP.....	۱۹
۲-۵. کد کننده با روش تخمین خطی.....	۲۲
۲-۶. کد کننده با تخمین خطی و ورودی مرکب.....	۲۵
۲-۶-۱. مدل MELP تولید سیگنال.....	۲۶
۲-۶-۲. محاسبه و کوانتیزه کردن دامنه فوریه.....	۲۷
۲-۶-۳. فیلتر تولید کننده پالس.....	۲۸
۲-۶-۴. فیلتر های شکل دهنده.....	۲۸
۲-۶-۵. انکدر MELP.....	۳۱
۲-۶-۶. دیکدر MELP.....	۳۲
۲-۶-۷. نحوه تخصیص بیت در کد کننده MELP استاندارد.....	۳۴
۲-۷. کد کننده باتحریک پالس های منظم.....	۳۵
فصل ۳ کد کننده سلپ	۳۷

۳۸	۱-۳. مقدمه
۳۹	۲-۳. خود بستگی نمونه ها
۴۰	۳-۳. روش آنالیز به وسیله ساخت
۴۱	۴-۳. ساختار کلی سلپ
۴۱	۱-۴-۳. پیشگویی کوتاه مدت، فیلتر $\frac{1}{A(z)}$
۴۲	۲-۴-۳. پیشگویی بلند مدت، فیلتر $\frac{1}{B(z)}$
۴۲	۳-۴-۳. فیلتر وزن دهنده خطا $W(Z)$
۴۵	۴-۴-۳. انکدر کد کننده CELP
۴۶	۵-۴-۳. دیکدر کد کننده CELP
۴۷	۶-۴-۳. روش یافتن بهترین سیگنال تحریک
۵۰	۷-۴-۳. محاسبه میزان خطا و ضریب بهره
۵۲	۸-۴-۳. فیلتر نهایی

فصل ۴ کد کننده های سلپ معروف

۵۶	۱-۴. کد کننده سلپ با تاخیر زمانی کم
۵۶	۱-۱-۴. مقدمه
۵۷	۲-۱-۴. روش های مختلف کاهش تاخیر زمانی در LD-CELP coder
۵۷	۳-۱-۴. انکدر کد کننده سلپ با تاخیر زمانی کم
۵۸	۴-۱-۴. دیکدر کد کننده سلپ با تاخیر زمانی کم
۶۰	۲-۴. کد کننده سلپ با نرخ بیت متغیر و منبع کنترل شده
۶۰	۱-۲-۴. مقدمه
۶۰	۲-۲-۴. بیت دهی قابل تغییر
۶۱	۳-۲-۴. قانون تخصیص بیت
۶۲	۳-۴. سلپ جبری
۶۲	۱-۳-۴. مقدمه
۶۳	۲-۳-۴. ساختار منبع تحریک کننده سلپ جبری
۶۴	۴-۴. کد کننده سلپ بهبود یافته با استفاده از آنالیز کپستروم

فصل ۵ پردازش شهودی و تخمین خطی ادراکی (PLP)

۶۶	۱-۵. مقدمه
۶۶	۲-۵. نحوه کارکرد غشای پایه به عنوان آنالیز گر طیف
۶۷	۳-۵. باند های اساسی
۶۸	۵-۴. سطح تشخیص صدا
۷۰	۵-۵. کد کننده شهودی سیگنال صحبت
۷۰	۱-۵-۵. کد کننده عمومی ادراکی سیگنال صحبت
۷۱	۶-۵. تخمین خطی ادراکی (PLP)

۷۱	 ۱-۶-۵ مقدمه
۷۲	 ۲-۶-۵ پردازش کپسترال
۷۳	 ۳-۶-۵ شکل دهی سیگنال ورودی
۷۴	 ۴-۶-۵ روش استخراج ضرایب PLP

۷۸ فصل ۶ پیاده سازی کد کننده سلب استاندارد

۷۹	 ۱-۶ مقدمه
۸۰	 ۲-۶ ارزیابی کیفیت سیگنال ساخته شده
۸۰	 ۱-۲-۶ معیار های بررسی کیفیت سیگنال
۸۰	 ۱-۱-۲-۶ اندازه گیری سیگنال به نویز
۸۱	 ۲-۱-۲-۶ روش های اندازه گیری ذهنی
۸۳	 ۲-۲-۶ جنبه های مختلف کیفیت دریافت ادراکی سیگنال
۸۳	 ۳-۲-۶ شرایط مختلف اندازه گیری کیفیت سیگنال
۸۵	 ۳-۶ پیاده سازی کردن CELP استاندارد
۸۵	 ۱-۳-۶ تخمین بهبود یافته پارامترهای فیلتر ساخت دور تناوب گام
۸۹	 ۲-۳-۶ نحوه کار کتاب کد تطبیق پذیر
۹۱	 ۳-۳-۶ به روز رسانی کتاب کد
۹۱	 ۴-۶ ساختار منبع تحریک تصادفی
۹۵	 ۵-۶ جستجوی کتاب کد تطبیقی
۹۵	 ۱-۵-۶ دوره تناوب گام صحیح
۹۷	 ۲-۵-۶ دوره تناوب گام غیر صحیح
۹۸	 ۳-۵-۶ انکدر دوره تناوب
۹۸	 ۱-۳-۵-۶ راهکار انکدر کردن زیرفریم
۹۸	 ۴-۶ انکد کردن بهره کتاب کد تطبیقی
۹۹	 ۷-۶ روش جستجوی کد تحریک تصادفی
۱۰۰	 ۱-۷-۶ بهره کتاب کد تصادفی بهبود یافته
۱۰۱	 ۸-۶ آنالیز LP
۱۰۱	 ۹-۶ فیلتر وزن دار ادراکی
۱۰۲	 ۱۰-۶ انکدر و دیکدر
۱۰۴	 ۱۱-۶ پاسخ کلی
۱۰۴	 ۱۲-۶ کارکرد دیکدر
۱۰۵	 ۱۳-۶ تخصیص بیت

۱۰۶ فصل ۷ کد کننده پیشنهادی و نتایج عملی

۱۰۷	 ۱-۷ مقدمه
۱۰۸	 ۲-۷ برخی موارد استفاده شده از PLP در کد کننده ها
۱۱۲	 ۳-۷ کد کننده پیشنهادی

۱۱۵.....	۴-۷. دیتا بیس مورد استفاده.....
۱۱۶.....	۵-۷. نتایج عملی.....
۱۱۷.....	۱-۵-۷. نتایج برای سیگنال صحبت بدون نویز و موسیقی.....
۱۲۰.....	۲-۵-۷. نتایج برای سیگنال صحبت با افزودن نویز و بدون موسیقی.....
۱۲۳.....	۳-۵-۷. نتایج برای سیگنال صحبت همراه با موسیقی بدون نویز.....
۱۲۶.....	۴-۵-۷. نتایج برای سیگنال صحبت همراه با موسیقی و نویز.....
۱۳۰.....	۶-۷. نرخ بیت مورد استفاده.....
۱۳۰.....	۷-۷. بررسی اثر تغییر برخی پارامترها بر روی کیفیت و نرخ بیت.....
۱۳۱.....	۱-۷-۷. بررسی اثر تغییر پارامترهای γ, β, α
۱۳۲.....	۲-۷-۷. بررسی اثر کاهش ضرایب LPC.....
۱۳۳.....	۳-۷-۷. اثر تغییر طول پنجره.....
۱۳۶.....	۸-۷. نتیجه گیری و پیشنهادات.....

عنوان	صفحه
-------	------

فصل ۱ مقدمه ۱

- شکل ۱-۱. ساختار ساده تولید صدا توسط انسان..... ۴
- شکل ۱-۲. بلوک دیاگرام کلی یک کد کننده سیگنال صحبت..... ۵
- شکل ۱-۳. بلوک دیاگرام کلی یک کد کننده سیگنال صحبت..... ۶
- شکل ۱-۴. مقایسه کیفیت شنیداری کد کننده های مختلف..... ۱۲

فصل ۲ معرفی کد کننده های متداول ۱۳

- شکل ۲-۱. شکل کلی انکدر یک کد کننده سیگنال صحبت..... ۱۵
- شکل ۲-۲. شکل کلی دیکدر یک کد کننده سیگنال صحبت..... ۱۵
- شکل ۲-۳. کد کننده PCM..... ۱۶
- شکل ۲-۴. مدل ساده تولید سیگنال..... ۱۷
- شکل ۲-۵. مدل LP تولید سیگنال..... ۱۹
- شکل ۲-۶. انکدر یک کد کننده LPC..... ۲۲
- شکل ۲-۷. دیکدر یک کد کننده LPC..... ۲۴
- شکل ۲-۸. مدل MELP تولید سیگنال..... ۲۷
- شکل ۲-۹. فیلتر تولید کننده پالس..... ۲۹
- شکل ۲-۱۰. بلوک دیاگرام کلی فیلتر شکل دهنده پالس..... ۳۰
- شکل ۲-۱۱. پاسخ زمانی و فرکانسی فیلتر شکل دهنده پالس و نویز..... ۳۰
- شکل ۲-۱۲. بلوک دیاگرام کلی انکدر کد کننده MELP..... ۳۲
- شکل ۲-۱۳. بلوک دیاگرام کلی دیکدر کد کننده MELP..... ۳۳
- شکل ۲-۱۴. انکدر حلقه باز کد کننده با تحریک پالس نمونه های غیر صفر..... ۳۶
- شکل ۲-۱۵. دیکدر کد کننده با تحریک پالس نمونه های غیر صفر..... ۳۶

فصل ۳ کد کننده سلپ ۳۷

- شکل ۳-۱. مدل CELP تولید سیگنال صحبت..... ۳۹
- شکل ۳-۲. بلوک کلی نحوه آنالیز (شکل بالا) و ساخت (شکل پایین) سیگنال در کد کننده سلپ..... ۴۱
- شکل ۳-۳. پاسخ فرکانسی فیلتر وزن دار..... ۴۴
- شکل ۳-۴. بلوک کلی نحوه آنالیز در انکدر CELP برای کاهش محاسبات..... ۴۴
- شکل ۳-۵. مدل کلی انکدر یک کد کننده CELP..... ۴۶
- شکل ۳-۶. مدل کلی دیکدر یک کد کننده CELP..... ۴۷
- شکل ۳-۷. حالت پشت سر هم فیلتر های ساخت..... ۴۸
- شکل ۳-۸. اندازه دامنه یک مدل تمام قطب برای فیلتر نهائی..... ۵۳
- شکل ۳-۹. پاسخ فرکانسی مدل صفر - قطب فیلتر نهائی..... ۵۴

فصل ۴ کد کننده‌های سلف معروف

۵۵

- شکل ۴-۱. شکل کلی انکدر سلف کدر با تاخیر زمانی کم..... ۵۹
 شکل ۴-۲. شکل کلی دیکدر سلف کدر با تاخیر زمانی کم..... ۵۹
 شکل ۴-۳. دیاگرام حالت برای تخصیص نرخ بیت ۶۱
 شکل ۴-۴. مختصات پالس های کد کننده جبری در استاندارد G.729 (مستطیل های مشکی)..... ۶۳
 شکل ۴-۵. مدل تولید سیگنال در کد کننده سلف بهبود یافته با استفاده از آنالیز کپستروم..... ۶۴

فصل ۵ پردازش شهودی و تخمین خطی ادراکی (PLP)

۶۵

- شکل ۵-۱. محدوده فرکانسی باند های اساسی بر اساس تابعی از فرکانس مرکزی باند..... ۶۸
 شکل ۵-۲. سطح تشخیص صدا در استاندارد آمریکائی..... ۶۹
 شکل ۵-۳. بلوک دیاگرام کلی کد کننده شهودی سیگنال صحبت..... ۷۱
 شکل ۵-۴. پردازش PLP..... ۷۵
 شکل ۵-۵. تبدیل فرکانسی باند های اساسی..... ۷۶

فصل ۶ پیاده سازی کد کننده سلف استاندارد

۷۸

- شکل ۶-۱. شکل کلی نحوه پیدا کردن پارامترهای فیلتر تخمین بلندمدت کد کننده سلف..... ۸۶
 شکل ۶-۲. استخراج بردار کتاب کد تطبیقی بر حسب زمان های گذشته..... ۸۸
 شکل ۶-۳. ساختار کتاب کد تطبیقی..... ۹۰
 شکل ۶-۴. ساختار کتاب کد بدون هیچ همپوشانی..... ۹۲
 شکل ۶-۵. منبع تحریک تصادفی با همپوشانی..... ۹۳
 شکل ۶-۶. نمونه های تصادفی موجود در منبع تحریک FS1016..... ۹۵
 شکل ۶-۷. بلوک کلی انکدر کاربردی شده..... ۱۰۳
 شکل ۶-۸. بلوک کلی دیکدر کاربردی شده..... ۱۰۴

فصل ۷ کد کننده پیشنهادی و نتایج عملی

۱۰۶

- شکل ۷-۱. بلوک کلی کد کننده اصلاح شده..... ۱۱۳
 شکل ۷-۲. بانک فیلتر بارک استفاده شده..... ۱۱۴
 شکل ۷-۳. سیگنال اصلی (شکل اول) سیگنال ساخته شده با کد کننده عمومی سلف (شکل دوم) و کد کننده بهبود یافته (شکل سوم)..... ۱۱۸
 شکل ۷-۴. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلف (نقطه چین)..... ۱۱۹
 شکل ۷-۵. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلف (نقطه چین)..... ۱۱۹
 شکل ۷-۶. سیگنال اصلی (شکل اول) سیگنال ساخته شده با کد کننده عمومی سلف (شکل دوم) و کد کننده بهبود یافته (شکل سوم) با سیگنال به نویز ۱۷ دسی بل..... ۱۲۱
 شکل ۷-۷. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلف (نقطه چین)..... ۱۲۲
 شکل ۷-۸. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلف (نقطه چین)..... ۱۲۲
 شکل ۷-۹. سیگنال اصلی صدای همراه با موسیقی..... ۱۲۴
 شکل ۷-۱۰. سیگنال ساخته شده سیگنال همراه با موسیقی کد کننده عمومی سلف (شکل بالا) کد کننده بهبود یافته (شکل پایین)..... ۱۲۵
 شکل ۷-۱۱. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلف (نقطه چین) سیگنال همراه با موسیقی..... ۱۲۵
 شکل ۷-۱۲. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلف (نقطه چین) سیگنال همراه با موسیقی..... ۱۲۶

شکل ۷-۱۳. سیگنال اصلی سیگنال همراه با موسیقی و نویز ۱۷ دسی بل (شکل بالا) و سیگنال ساخته شده با کد کننده عمومی سلپ (شکل دوم) و کد کننده بهبود یافته (شکل سوم).....	۱۲۸
شکل ۷-۱۴. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلپ (نقطه چین) سیگنال همراه با موسیقی و نویز ۱۷ دسی بل.....	۱۲۹
شکل ۷-۱۵. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلپ (نقطه چین) سیگنال همراه با موسیقی و نویز ۱۷ دسی بل.....	۱۲۹
شکل ۷-۱۶. سیگنال اصلی (شکل بالا) سیگنال ساخته شده با کد کننده عمومی (شکل دوم) و کد کننده بهبود یافته (شکل سوم) با طول فریم ۲۹۰.....	۱۳۴
شکل ۷-۱۷. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلپ (نقطه چین) با طول فریم ۲۹۰.....	۱۳۵
شکل ۷-۱۸. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلپ (نقطه چین) با طول فریم ۲۹۰.....	۱۳۵

فهرست جداول

عنوان	صفحه
جدول ۱-۱. استاندارد کدکنندها.....	۹
جدول ۱-۲. تقسیم بندی کد کننده های سیگنال گفتار بر مبنای نرخ بیت.....	۱۰
جدول ۱-۲. بیت های تخصیص داده شده برای پارامتر های کد کننده LPC در کد کننده استاندارد FS1015.....	۲۴
جدول ۲-۲. بیت های تخصیص داده شده برای پارامتر های کد کننده MELP در کد کننده استاندارد FS MELP CODER.....	۳۴
جدول ۵-۱. رابطه بین بارک و مل و باند های اساسی فرکانسی.....	۶۹
جدول ۶-۱. نمره دهی گروهی مطلق.....	۸۱
جدول ۶-۲. نمره دهی گروهی بر مبنای انحراف.....	۸۲
جدول ۶-۳. تخصیص بیت کد کننده کاربردی شده.....	۱۰۵
جدول ۷-۱. نتایج سیگنال بدون نویز و موسیقی.....	۱۱۷
جدول ۷-۲. نتایج سیگنال با سیگنال به نویز ۱۷ دسی بل و بدون موسیقی.....	۱۲۱
جدول ۷-۳. نتایج سیگنال صحبت به همراه موسیقی.....	۱۲۴
جدول ۷-۴. نتایج سیگنال صحبت به همراه موسیقی و نویز.....	۱۲۷
جدول ۷-۵. تغییرات کیفیت و نرخ بیت بر اساس تغییر ضرایب LPC.....	۱۳۲
جدول ۷-۶. تغییرات کیفیت و نرخ بیت بر اساس تغییر طول فریم.....	۱۳۳

فصل ۱ مقدمه

۱-۱. معرفی

امروزه در ارتباط بین انسان ها گفتار نقش بسیار مهمی دارد و هر انسان با توجه به ساختار سیستم صوتی خود صدای خاص خود را دارد. در این راستا با پیشرفت تکنولوژی، ارتباطات تلفنی گسترش زیادی پیدا کرده اند و تحقیقات، استانداردها و تقاضاهای زیادی برای استفاده از این تکنولوژی بوجود آمده است. هدف تحقیقات جدید دستیابی به روشهایی است که بتوانند بر محدودیتهای موجود غلبه کنند و روشهای بهینه را با کیفیت مناسب ارائه دهند. برای افزایش بهره وری از تکنولوژی های امروزی استفاده از تمامی اطلاعات و دانش های مربوطه، نقش بسیار عمده ای را ایفا می کنند. مثلا شناخت فیزیولوژی صحبت کردن و شنوایی انسان در تکنولوژی سیستم های تلفنی بسیار سودمند می باشد. با توجه به این موضوع بررسی های دانشمندان نشان می دهد، ساختار سیستم صحبت کردن انسان ها و صداهائی که توسط انسان تولید می گردد دارای خصوصیات و پارامترهای خاصی می باشد که در اصوات دیگر وجود ندارد. همچنین سیستم شنوایی انسان نیز دارای ویژگی های خاصی می باشد که از ساختار ذاتی سیستم شنوایی انسان ناشی می گردد و خصوصیات و ویژگی های خاص خود را دارد. به همین دلیل دانشمندان مختلف این ویژگی ها را در مواردی که تکنولوژی در زمینه صدای انسان و زمینه های وابسته به آن مورد استفاده قرار می گیرد، است مورد استفاده قرار می دهند.

به طور کلی کد کردن سیگنال گفتار روشی است که در آن سیگنال ورودی را بعد از نمونه برداری دیجیتال کرده و آن را با کمترین بیت ممکن به نحوی که کیفیت اولیه سیگنال حفظ گردد توصیف کنیم. همچنین در برخی منابع از واژه فشرده سازی سیگنال صحبت به جای کد کردن آن استفاده می شود چرا که در کد کردن سیگنال گفتار هدف اصلی کاهش بیت های موجود نشان دادن سیگنال صحبت ضمن حفظ کیفیت آن سیگنال می باشد.

در حالت کلی روشهای مختلف ارائه شده برای کد کننده های سیگنال گفتار مبتنی بر الگوریتم هایی هستند که طی آن یک مجموعه از اطلاعات (سیگنال اصلی) به صورت ورودی وارد شده و مجموعه متناظر با آن اطلاعات (کد های بدست آمده از سیگنال اصلی) در خروجی تولید می گردد. در حقیقت این الگوریتم ها مجموعه ای از دستورالعمل ها هستند که بصورت یک مجموعه محاسبات کامپیوتری بر روی اطلاعات ورودی انجام می شوند و اطلاعاتی از سیگنال ورودی را برای

ساخت مجدد سیگنال در گیرنده، استخراج و برای آن ارسال می کنند. این دستورالعمل ها به صورت سخت افزار مداری پیاده سازی می شود و با کمک نرم افزارهای مربوط الگوریتم های لازم بر روی اطلاعات اجرا می گردد.

از جمله کاربرد کدکننده ها میتوان به موارد زیر اشاره نمود:

- ارتباطات تلفنی
- ارتباطات نظامی
- تلفنهای سلولی
- پست الکترونیکی صوتی
- کاربردهای چند رسانه ای

در این فصل ابتدا مدل تولید سیگنال گفتار توسط انسان بیان شده و سپس خلاصه ای از سیستم کد کردن بیان می گردد. خصوصیات مورد نظر در کد کننده های سیگنال گفتار و مقایسه کد کننده ها در قسمت بعد ارائه شده است.

۱-۲. تولید صوت توسط انسان

برای شناخت کامل سیستم کد کننده های صحبت، آشنائی با روند تولید صدا و پردازش سیستم شنوائی انسان لازم به نظر می رسد. در همین زمینه استفاده از اطلاعات واج شناسی، فیزیواوژیکی و سطوح شنوائی انسان نیز می توانند کمک های زیادی داشته باشند.

در این بخش به معرفی اجمالی سیستم صوتی در تولید صدا که از دو بخش عمده سیستم تحریک^۱ و لوله صوتی^۲ تشکیل شده است می پردازیم.

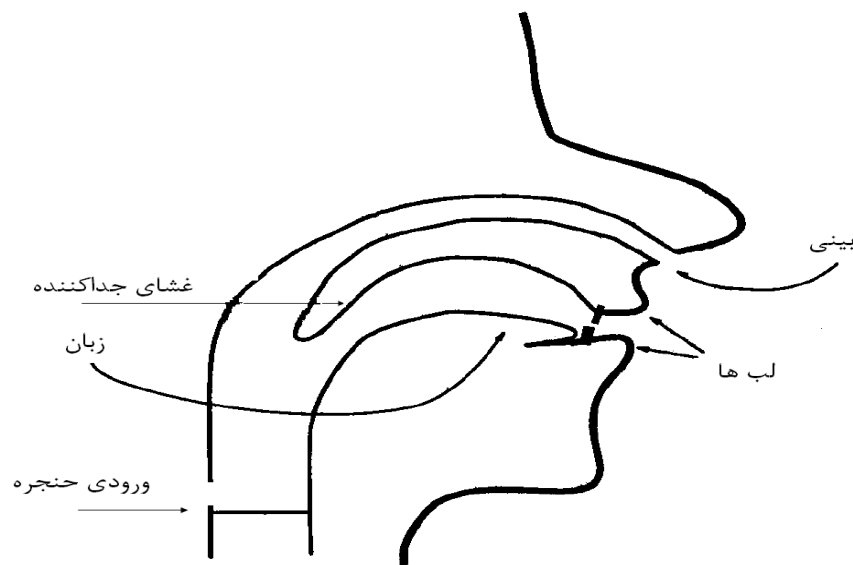
لوله صوتی از ابتدای شش شروع شده و به لب و بینی ختم می شود و مهمترین قسمت در تولید مشخصه های صحبت به شمار می آید چرا که صدای انسان در اثر عبور هوا از لوله صوتی بوجود می-

^۱ Excitation

^۲ Vocal-tract

آید و مشخصه های فرکانسی سیگنال به وجود آمده بستگی به مشخصه لوله صوتی بستگی دارد. شکل ۱-۱[1]، ساختار ساده ای از چگونگی تولید صدا را نشان می دهد. به طور خلاصه و با توجه به شکل ۱-۱ با حرکت پرده دیافراگم^۱ به سمت بالا و منبسط شدن آن، هوای داخل شش ها با فشار به سمت نای حرکت می کند و پس از برخورد با تارهای صوتی و عبور از مسیر گلو به دهان و بینی ختم می شود و با توجه به حالت زبان، لب ها و دندان ها صدا های مختلف توسط افراد مختلف ادا می شوند.

با توجه به آزمایشات انجام شده، به طور کلی تفاوت بین صدای افراد مختلف به دلیل متفاوت بودن مشخصه و ویژگی های مسیر حرکت هوای خارج شده از شش تا خروج از دهان و بینی می باشد[1].

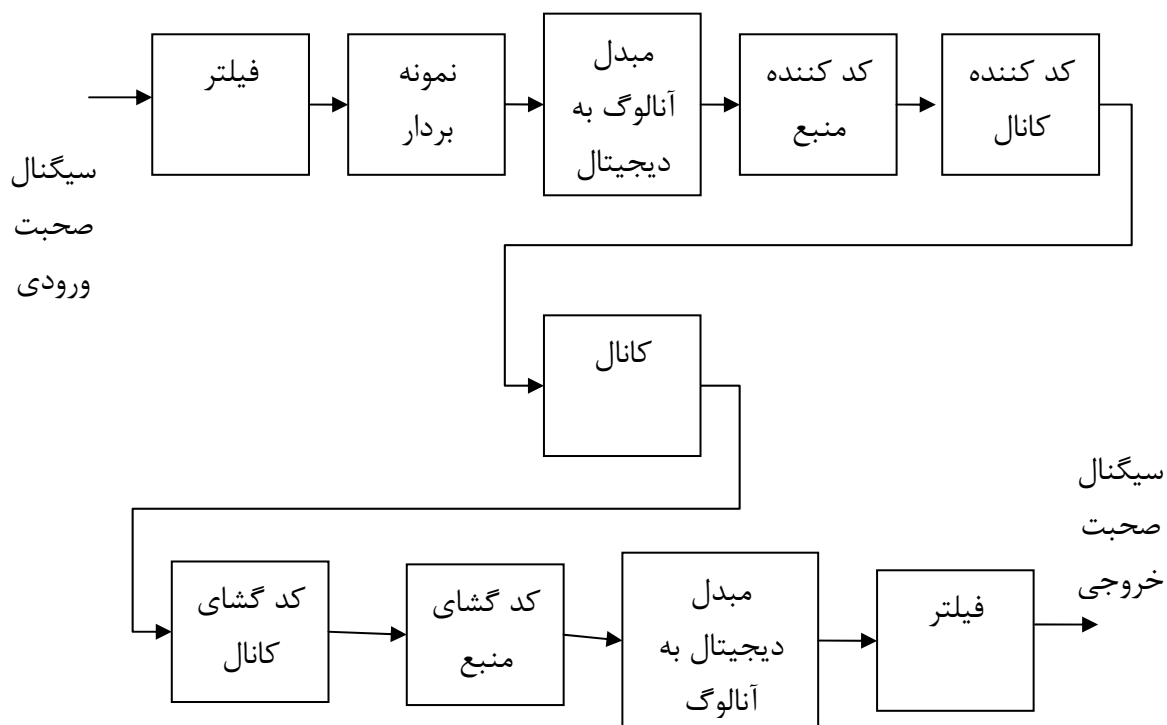


شکل ۱-۱. ساختار ساده تولید صدا توسط انسان

¹ Diaphragm

۳-۱. مقدمه ای بر اصول کد کردن

شکل ۱-۲ [2] بلوک دیاگرام کلی یک کدکننده سیگنال گفتار را نشان می دهد. سیگنال گفتار پیوسته در زمان از یک فیلتر پایین گذر معمولاً با فرکانس قطع 3.5KHz برای حذف مولفه های فرکانس بالا (که معمولاً برای سیگنال گفتار اطلاعاتی در آن موجود نیست) عبور داده شده و نمونه گیری از آن (با توجه به فرکانس نایکوئیست) با فرکانس 8KHz انجام می شود. سپس نمونه ها با کوانتیزه کننده یکنواخت به معادل دیجیتال آنها که به آن سیگنال گفتار دیجیتال شده می گویند، تبدیل می شوند. بیت های تخصیص داده شده برای هر نمونه بطور استاندارد برای حفظ کیفیت سیگنال باید بالاتر از ۸ بیت باشند که در اکثر موارد برای هر نمونه ۱۶ بیت تخصیص داده می شود که به نسبت کیفیت خوبی برای سیگنال دیجیتال شده دارند.

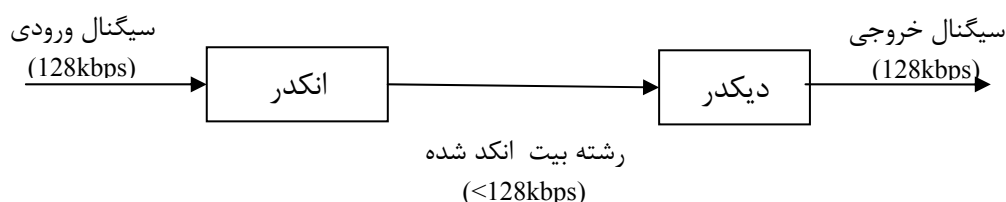


شکل ۱-۲. بلوک دیاگرام کلی یک کد کننده سیگنال صحبت

سیگنال دیجیتال شده توسط کد کننده منبع کد شده یعنی به ازای رشته نمونه های ورودی، بسته به کار کرد سیستم، رشته بیت های متناظر با نمونه های وارد شده در خروجی آن تولید می گردد. رشته بیت های تولید شده توسط کد کننده منبع، معرفی کننده سیگنال ورودی متناظر می باشد و هدف از انجام کدینگ بر روی اطلاعات ورودی بالا بردن آنتروپی^۱ اطلاعات سیگنالی می باشد. سیگنال کد بندی شده توسط کد کننده منبع به منظور محافظت در برابر خطاهای موجود در کانال که به محیط انتقال و ماتریس خطای انتقال وابسته است و موجب تخریب سیگنال می شوند و همچنین در صورت نیاز به افزودن اطلاعات امنیتی از کد کننده کانال عبور داده می شود. (مثلاً میتوان با افزودن بیت های پریته، محل اتفاق خطای بیت های ارسالی را تشخیص داد و تصحیح کرد). این سیگنال پس از عبور از کانال (ارسال برای گیرنده) در مقصد دریافت شده و پس از انجام پردازش های مربوط که عکس عمل های انجام گرفته در فرستنده می باشد، سیگنال گفتارابتدایی با مقداری خطا که بستگی به سیستم پردازش، نویز کوانتیزه کردن و خطای کانال دارد، در خروجی گیرنده تولید می گردد.

به عمل پردازش و استخراج اطلاعات برای ارسال در کانال، انکدینگ^۲ و به عمل دریافت اطلاعات از کانال و تبدیل آنها به سیگنال اولیه به طور کلی دیکدینگ^۳ می گویند و هدف اصلی کد کننده های سیگنال صحبت کاهش بیت های لازم برای نشان دادن سیگنال صحبت، در عین حفظ کیفیت سیگنال ساخته شده در گیرنده می باشد.

شکل ۱-۳ [3] بلوک بسیار ساده عمل کد کردن را نشان می دهد.



شکل ۱-۳. بلوک دیاگرام کلی یک کد کننده سیگنال صحبت

^۱ میزان اطلاعات مفید موجود در سیگنال [2]

^۲ Encoding

^۳ Decoding

در طول این پایان نامه از مفهوم نرخ بیت^۱ استفاده زیادی می شود که معنای آن میزان بیت استفاده شده برای نمایش سیگنال گفتار در یک ثانیه می باشد. برای مثال میزان نرخ بیت سیگنال گفتار دیجیتال شده ورودی برابر با $\text{Bit-rate} = 8\text{kHz} * 16\text{bits} = 128\text{kbps}$ می باشد که هدف کد کننده منبع (شکل ۱-۳) کاهش این نرخ بیت می باشد. معمولاً نرخ بیت خروجی این واحد خیلی کمتر از نرخ بیت ورودی آن می باشد. به طور مثال نرخ بیت کد کننده LPC که بعداً معرفی خواهد شد برابر 2.4kbps می باشد که حدوداً ۵۳ برابر کمتر از نرخ بیت ورودی است.

باید توجه داشت که سیگنال صحبت، سیگنالی با تغییرات دینامیکی زیاد، غیر ساکن^۲ و دارای پیچیدگی زیاد است و به همین دلیل برای پردازش آن ابتدا آن را به فریم های کوچک زمانی تقسیم کرده و سپس پردازش بر روی آن انجام می پذیرد.

۴-۱. خصوصیات لازم یک کد کننده سیگنال گفتار

هدف اصلی کد کننده های سیگنال گفتار افزایش کیفیت سیگنال خروجی در یک نرخ بیت ثابت و یا کاهش نرخ بیت کد کننده در یک کیفیت خاص می باشد. نرخ بیت مناسب برای ارسال یک سیگنال گفتار به میزان هزینه ارسال، حافظه، هزینه کد کردن، سیگنال دیجیتال شده ورودی و میزان کیفیت مورد نظر بستگی دارد و در بسیاری از کد کننده های معروف سیگنال خروجی تولید شده متفاوت از سیگنال ورودی می باشد چرا که بسیاری از اضافات سیگنال اصلی در کد کردن مورد استفاده قرار نمی گیرد و در بسیاری از کاربردها مانند کاربرد های نظامی و مخابرات امن کیفیت بالای سیگنال نسبت به حفظ امنیت سیگنال در درجه اهمیت کمتری قرار دارد و کیفیتی در حد شناخت صحبت کافی می باشد.

^۱ Bit-rate

^۲ Non stationary

به طور کلی خصوصیات یک کد کننده خوب را می توان در عوامل زیر خلاصه نمود:

- نرخ بیت پائین
- کیفیت بالا
- کارایی خوب در قبال وجود خطاهای کانال
- کارایی خوب به ازای زبان ها، گویش ها و گوینده های مختلف.
- کارایی خوب در حضور سیگنال های غیر سیگنال گفتار (مانند سیگنالهای موزیک در خطوط تلفن)
- استفاده از حافظه و پیچیدگی محاسباتی کم.
- تاخیر کم در کد کردن^۱.

تاخیر در کد کردن خود از چهار عامل زیر ناشی می شود:

(الف) میزان زمان مورد نیاز برای بافر کردن اطلاعات ورودی برای پردازش در کد کننده

(ب) زمان مورد نیاز برای پردازش اطلاعات در کد کننده.

(ج) زمان مورد نیاز برای ارسال اطلاعات توسط خط انتقال.

(د) زمان مورد نیاز برای پردازش اطلاعات و تولید سیگنال خروجی در کد گشا^۲

طراحی و استفاده از هر کد کننده معمولاً بر اساس ملاحظات کیفیت مورد نیاز، میزان هزینه قابل قبول، قابلیت استفاده در سیستم مورد نظر و مسائل دیگری که مد نظر می باشد انجام می گیرد و همواره تعادلی بین تمام ملاحظات برای انتخاب کد کننده مورد نظر انجام می گیرد. برای آشنائی بیشتر با کد کننده های استفاده شده در استاندارد های مختلف، برخی از این کد کننده ها در جدول ۱-۴ [4] و [32]، آورده شده است. در این جدول برخی از کد کننده های استاندارد و کاربرد های آنها به همراه سال و میزان نرخ بیت مورد استفاده در کد کننده آورده شده است.

^۱ مدت زمان لازم برای پردازش اطلاعات در قسمت کد کننده و کد گشا برای ساخت سیگنال را تاخیر کد کننده می گویند.
^۲ دیکدر

فصل اول - مقدمه

جدول ۱-۱. استاندارد کدکنندها

سال کاربرد	نوع کد کننده	کاربرد	نرخ بیت (kbps)
۱۹۷۲	PCM	شبکه سوئیچ تلفن عمومی (نسل اول)	۶۴
۱۹۸۴	ADPCM	شبکه سوئیچ تلفن عمومی (نسل اول)	۳۲
۱۹۹۲	LD-CELP (سلپ با تاخیر کم)	شبکه سوئیچ تلفن عمومی (نسل اول)	۱۶
۱۹۸۵	APC (کد کننده تخمینی تطبیقی)	استاندارد INMARSAT (ارتباط دریایی)	۱۶
۱۹۹۱	(RPE-LTP) تحریک پالس منظم با تخمین بلند مدت	سیستم های موبایل اروپایی و سیستمهای سلولی	۱۳
۱۹۹۰	MPLPC (تخمین خطی چند پالسی)	سیستم های فضایی	۹/۶
۱۹۹۲	VSELP (تخمین خطی تحریک جمع برداری)	سیستم های موبایل آمریکای شمالی	۸
۱۹۹۳	VSELP	سیستم موبایل ژاپن	۶/۷
۱۹۹۳	CELP	موبایل با نرخ نیمه	۵-۶
۱۹۹۳	MBE (تحریک چند بانندی)	استاندارد INMARSAT (موبایل زمینی)	۶/۴
۱۹۹۱	CELP	استاندارد ملی آمریکا	۴/۸
۱۹۹۱	VDPC (تخمین تطبیقی برداری)	ماهواره های موبایل	۴/۸
۱۹۷۷	LPC (تخمین خطی)	استاندارد ملی آمریکا	۲/۴

۱-۵. طبقه بندی کدکننده های مختلف

معمولا کد کننده های سیگنال گفتار بر اساس معیار های مختلفی دسته بندی می شوند که مهمترین این تقسیم بندی ها بر مبنای نرخ بیت و نحوه کد کردن می باشد که مختصرا توضیح داده می شود.

۱-۵-۱. طبقه بندی بر مبنای نرخ بیت

تمامی کد کننده های سیگنال گفتار برای کاهش نرخ بیت 128kbps طراحی شده اند و به طور معمول کد کننده های مختلف را مطابق جدول ۱[3]، بر مبنای نرخ بیت کد کننده ها، طبقه بندی می کنند.

جدول ۱-۲. تقسیم بندی کد کننده های سیگنال گفتار بر مبنای نرخ بیت

گروه	محدوده نرخ بیت
نرخ بیت بالا	>15 kbps
نرخ بیت متوسط	5 to 15 kbps
نرخ بیت پائین	2 to 5 kbps
نرخ بیت خیلی پائین	<2 kbps

۱-۵-۲. طبقه بندی بر مبنای نحوه کد کردن

کدکننده های موجود برحسب اینکه به چه نحو، سیگنال اصلی را توسط بیت ها نشان می دهند بر سه دسته تقسیم بندی می شوند.

۱-۵-۲-۱. کدکننده های شکل موج

این کدکننده ها برای کدبندی همه نوع سیگنال به کار می روند و در حقیقت نمونه های دیجیتال شده را به طور مستقیم برای نشان دادن سیگنال استفاده می کنند به این معنی که عدد دیجیتال شده معادل هر نمونه را تولید کرده و توسط کانال انتقال می دهد از جمله اینگونه کدکننده ها میتوان به موارد زیر اشاره نمود که در نرخ بیت بالای 32kpbs کار می کنند:

^۱PCM (مدولاسیون کد پالس)

^۲ADPCM (مدولاسیون کد پالس تطبیقی)

۱-۵-۲-۲. کد کننده های پارامتریک

در اینگونه کد کننده ها مشخصه هایی از سیگنال به جای خود سیگنال ارسال می گردد و به همین دلیل فقط برای سیگنال گفتار که مشخصه های خاص خود (دوره تناوب گام^۳، فرکانس فرمنت^۴ و...) را دارد مورد استفاده قرار می گیرند و در حقیقت با داشتن مشخصه هایی از سیگنال می توان سیگنال را باز سازی نمود به همین دلیل نرخ بیت پایینی دارند. در این گونه کدکننده ها به دلیل استفاده از مشخصه های سیگنال، مفهوم سیگنال به نوبه کار نمی رود و باید از طریق گوش کردن به صدای ساخته شده میزان کیفیت سیگنال را تعیین نمود.

کد کننده ^۵LPC که نرخ بیت آن 2.4Kbps می باشد از این دسته از کد کننده ها می باشد.

^۱ pulse code modulation

^۲ Adaptive Differential PCM

^۳ Pitch Period

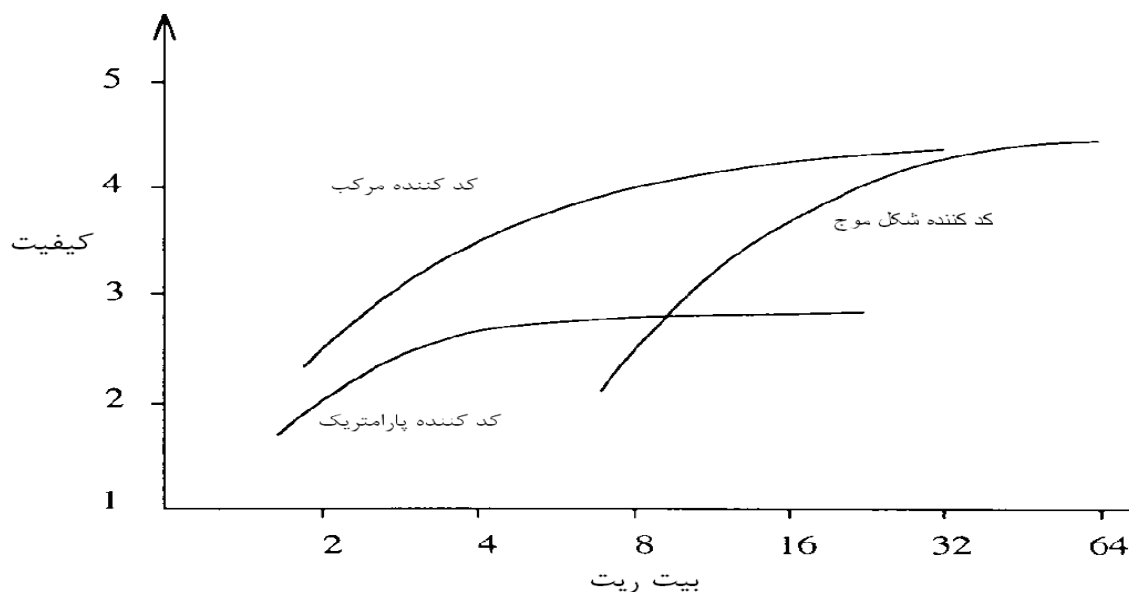
^۴ Formant Frequency

^۵ linear Prediction Coding

۱-۵-۲-۳. کد کننده های مرکب

این کد کننده حالت ترکیب یافته کد کننده های شکل موج و پارامتریک می باشند به این مفهوم که علاوه بر مشخصه های سیگنال اصلی، چندین نمونه یا پارامترهایی از شکل سیگنال اصلی نیز برای ساخت سیگنال به کد گشا ارسال می گردد. به همین دلیل نرخ بیت آنها از کد کننده های پارامتریک بالاتر و از کد کننده های شکل موج کمتر می باشد همچنین کیفیت آن از کد کننده های پارامتریک بالاتر و از کد کننده های شکل موج کمتر است. از جمله این کد کننده ها می توان به کد کننده CELP^۱ اشاره نمود.

شکل ۴-۱ [۴]، نمودار مقایسه کد کننده های فوق را نشان می دهد که در آن محور افقی نرخ بیت استفاده شده و محور عمودی نشان دهنده کیفیت شنیداری سیگنال بر مبنای معیار MOS [توضیح در فصل ۶ می باشد].



شکل ۴-۱. مقایسه کیفیت شنیداری کد کننده های مختلف

¹ Code Excited Linear Prediction Coding

فصل ۲ معرفی کد کننده های

متداول

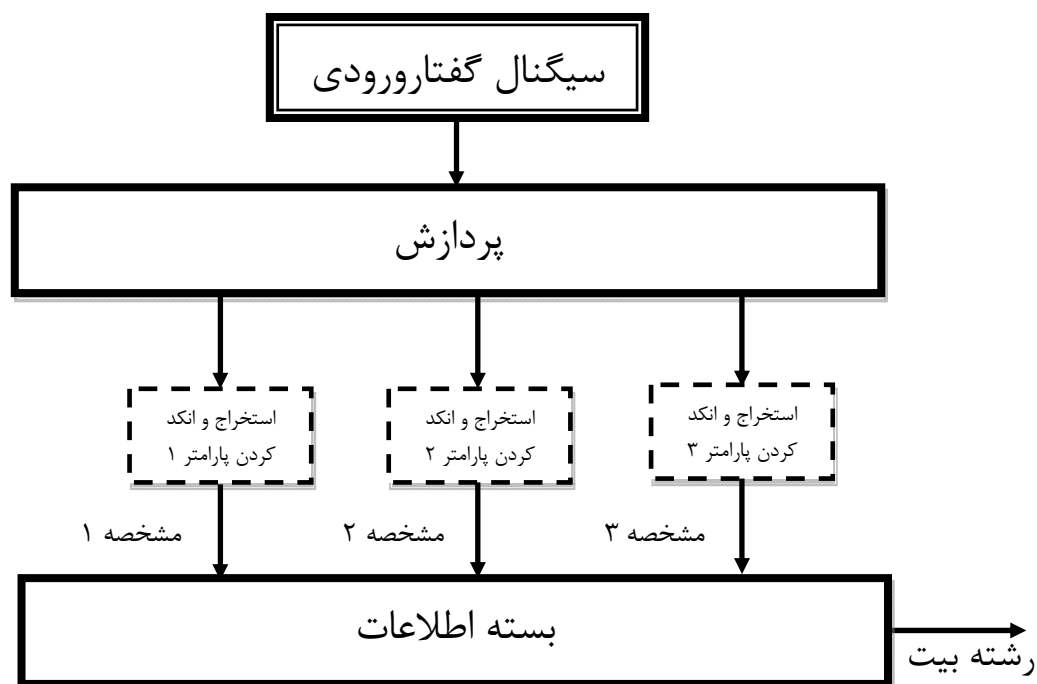
۲-۱. ساختار کلی یک کد کننده و دیکد کننده سیگنال صحبت

همانطور که پیش تر بیان گردید در بررسی کد کننده های سیگنال گفتار باید این نکته را در نظر گرفت که در این کد کننده های از روشهای پردازش سیگنال دیجیتال و روشهای آماری استفاده گسترده ای می شود. همچنین از آنجا که خصوصیات آماری سیگنال گفتار با زمان تغییرات زیادی دارد برای پردازش آماری مناسب، این سیگنال ها به تکه های کوچک (فریم) که معمولا تعداد نمونه هایی بین ۲۰-۲۴۰ در هر فریم وجود دارد تقسیم بندی می شود و پردازش فریم به فریم انجام می گیرد.

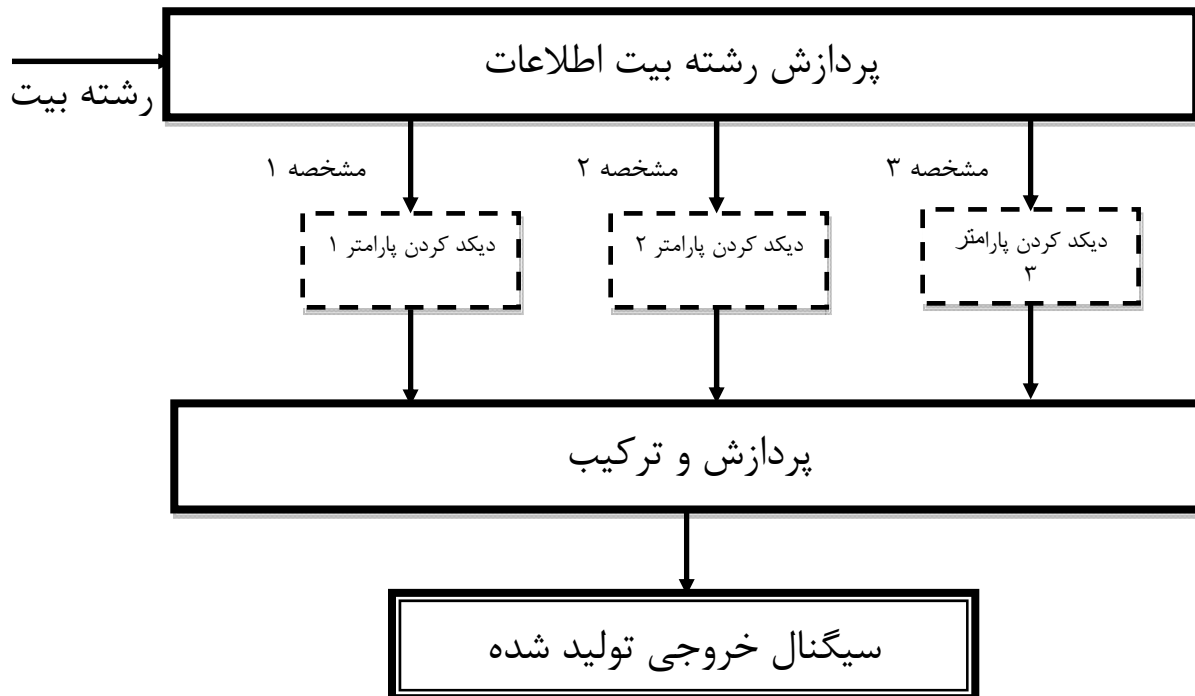
شکل ۲-۱ و ۲-۳ [3]، بلوک دیاگرام کلی کد کننده و کد گشا یک کد کننده سیگنال گفتار را نشان می دهد. برای کد کننده، سیگنال گفتار ورودی وارد شده و پس از پردازش، پارامترهای مورد نظر از فریم موجود استخراج شده و به صورت یک بسته رشته بیت با نظم مشخص پشت سر هم قرار گرفته و برای کد گشا فرستاده می شود. در کد گشا این رشته بیتها دریافت شده و با توجه به شیوه قرار گرفتن آنها، پارامترهای ارسال شده شناسایی و استخراج می شود و سیگنال مورد نظر به کمک آنها دوباره تولید می گردد.

کد کننده های مختلف بنا به نرخ بیت مورد نظر و کیفیت دلخواه پارامتر های گوناگون را با الگوریتم های مختلفی در کد کننده بدست آورده و برای کد گشا ارسال می نمایند تا با استفاده از پارامتر های موجود بتواند سیگنال مورد نظر را با توجه به مدل تعریف شده برای آن کد کننده مشخص تولید نمایند [32] و [13].

در این فصل کد کننده های متداول سیگنال گفتار معرفی شده و معایب و مزایای هر یک به طور خلاصه بیان شده و مدل تولید سیگنال گفتار هر مدل، کد گشا و کد کننده مربوط به کد کننده معرفی می گردد.



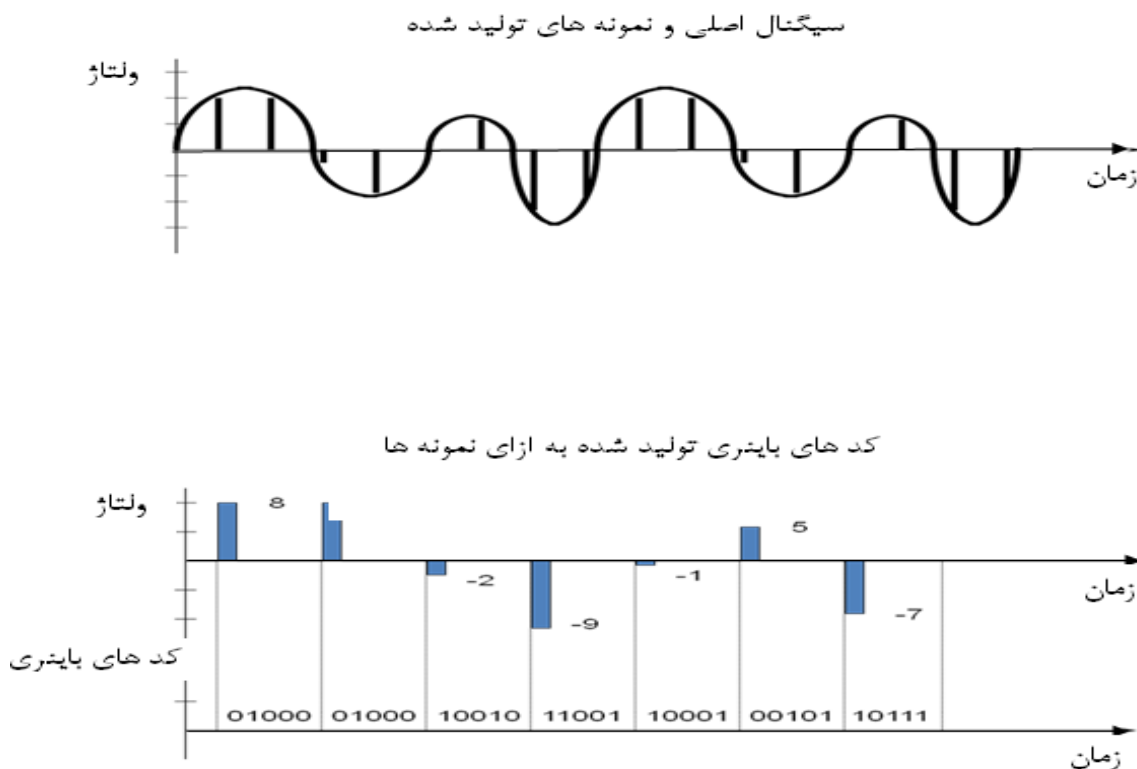
شکل ۲-۱. شکل کلی انکدر یک کد کننده سیگنال صحبت



شکل ۲-۲. شکل کلی دیکدر یک کد کننده سیگنال صحبت

۲-۲. کد کننده PCM^۱

این کد کننده ساده ترین و در عین حال معروف ترین کد کننده سیگنال می باشد. در این کد کننده از نمونه های سیگنال که از طریق نمونه برداری مطابق قاعده نایکوئیست به دست می آیند، مقدار باینری متناظر با آنها تولید و ارسال می گردد و از آنجا که تمامی نمونه ها ارسال می گردند، دارای نرخ بیت بالائی هستند و سیگنال تولید شده در گیرنده به نسبت سیگنال اصلی دارای کیفیت خوبی می باشد. از این کد کننده میتوان برای کد کردن انواع سیگنال های مختلف استفاده نمود. شکل ۲-۱ [2]، شکل کلی این کد کننده را نشان می دهد.

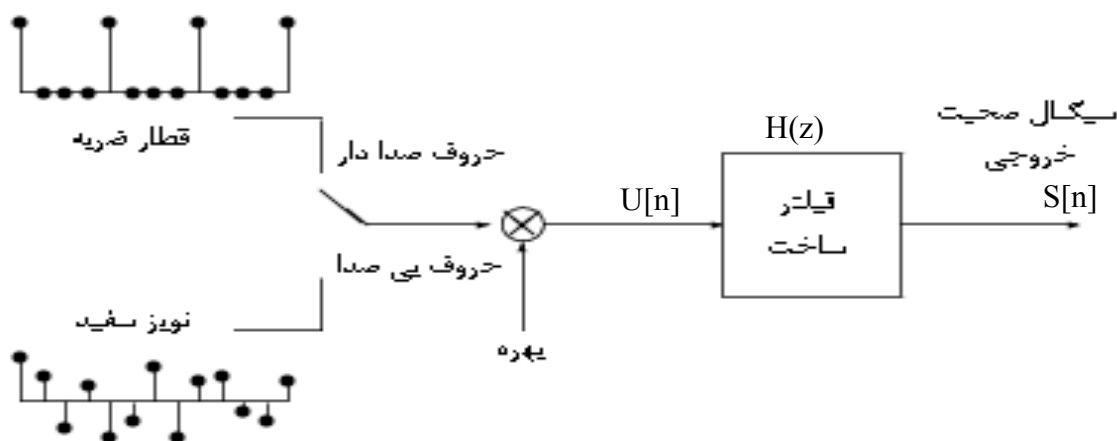


شکل ۲-۳. کد کننده PCM

^۱ Pulse Code Modulation

۲-۳. مدل تخمین خطی تولید سیگنال

این مدل که اساس کار کد کننده با تخمین خطی می باشد بر اساس مدل ساده تولید سیگنال گفتار که در شکل ۲-۴ [۱]، آورده شده است کار می کند. در این مدل که ساده ترین مدل تولید صوت توسط انسان می باشد، هوای تولید شده توسط شش ها با دومنبع تحریک قطار ضربه برای حروف صدا دار^۱ و منبع نویز سفید برای حروف بی صدا^۲ مدل سازی می شود و میزان قدرت این هوا با پارامتر بهره مدل سازی می شود و مدل سازی لوله صوتی توسط یک فیلتر تمام قطب انجام می شود که ضرایب آن از روش تخمین خطی که در ادامه توضیح داده می شود بدست می آیند و معادله مربوط به آن در معادله ۱-۲ آورده شده است. معادله مربوط به تشکیل سیگنال خروجی بر حسب ورودی و فیلتر ساخت در معادله (۲-۲) آورده شده است.



شکل ۲-۴. مدل ساده تولید سیگنال

^۱ Voice

^۲ Unvoice

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{10} z^{-10}} \quad (1-2)$$

$$s(n) + \sum_{i=1}^{10} a_i s(n-i) = u(n) \quad (2-2)$$

۲-۴. روش تخمین خطی

این روش در اکثر کدر های پیشرفته امروزی کاربرد دارد و ایده کار اینست که هر نمونه سیگنال صحبت بر حسب ترکیب خطی نمونه های قبلی سیگنال طبق معادله $\hat{s}[n] = -\sum_{i=1}^M a_i s[n-i]$ تخمین زده می شود. هدف اصلی در این تخمین پیدا کردن ضرایب a_i می باشد به طوری که سیگنال تخمین زده شده $\hat{s}[n]$ بیشترین شباهت را با سیگنال اصلی $s[n]$ داشته باشد و معیار این شباهت را با استفاده از تابع ارزش مینیمم مربع خطا^۱ اندازه می گیرند [12] و [37].

ضرایب بدست آمده برای این تقریب را ضرایب تخمین خطی^۲ میگویند. این ضرایب برای هر فریم محاسبه شده و به وسیله آنها سیگنال اصلی تخمین زده می شود. نمایش سیگنال بدست آمده با استفاده از تخمین خطی در حوزه فرکانس شباهت زیادی با سیگنال اصلی دارد.

فایده اصلی این روش در کد کننده های سیگنال صحبت در این است که برای ساختن سیگنال صحبت در دیکدر نیازی به داشتن کلیه نمونه های سیگنال اصلی نداریم و با استفاده از ضرایب LP و تعداد کمی از نمونه های اول سیگنال اصلی (به تعداد ضرایب LP) می توانیم سیگنال اصلی را باز سازی کنیم.

۲-۴-۱. انواع روش های تخمین خطی

به طور کلی دو روش تخمین خطی وجود دارد :

¹ Minimum Mean Square Error (MMSE)

² Linear Prediction Coefficient (LPCs)

• روش تخمین داخلی :

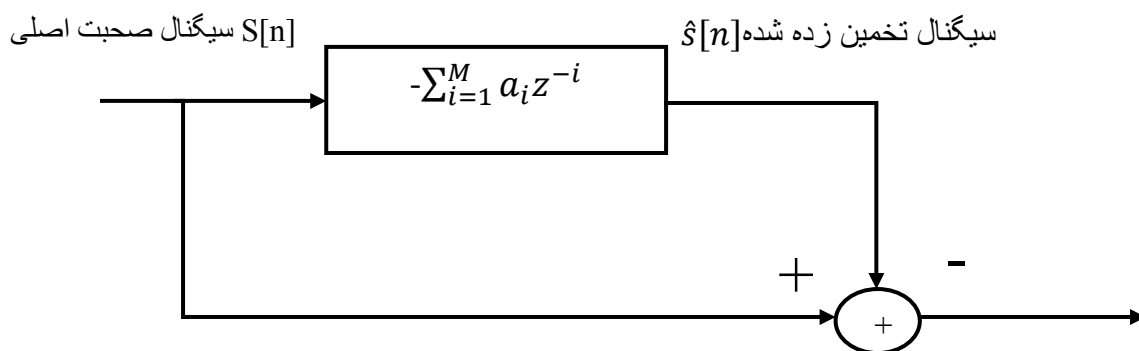
در این روش ضرایب LP مربوط به هر فریم از همان فریم مورد پردازش محاسبه و بدست می آید و به همین دلیل از هر فریم باید حداقل به تعداد ضرایب LP، نمونه از سیگنال صوت وجود داشته باشد.

• روش تخمین خارجی :

در این روش ضرایب LP مربوط به هر فریم از فریم قبلی محاسبه و بدست می آید. باید دقت داشت از آنجا که خواص سیگنال با زمان تغییر می کند، باید در این روش اندازه فریم ها کوچک در نظر گرفته شوند تا اثر تغییرات سیگنال مقداری خنثی شود. این روش باعث افزایش سرعت پردازش می شود چرا که نیازی نیست تا کل فریم ذخیره شده و مورد پردازش قرار گیرد.

۲-۴-۲. روش محاسبه ضرایب LP

روش اصلی محاسبه ضرایب LP استفاده از روش های آماری می باشد. همان طور که بیان شد سیگنال $\hat{s}[n]$ باید از لحاظ $MMSE^1$ کمترین خطا را با سیگنال اصلی $s[n]$ داشته باشد و با توجه به این نکته و با توجه به مدل تولید سیگنال به روش تخمین خطی که در شکل ۲-۵ [3]، آورده شده است می توانیم ضرایب تخمین خطی را با استفاده از روابط زیر بدست آوریم :



شکل ۲-۵. مدل LP تولید سیگنال

¹ Minimum Mean Square Error

$$\hat{s}[n] = - \sum_{i=1}^M a_i s[n-i] \quad (3-2)$$

$$\hat{S}(Z) = -S(Z) \sum_{i=1}^M a_i z^{-i} \quad (4-2)$$

$$e[n] = s[n] - \hat{s}[n] \quad (5-2)$$

که در این معادلات M تعداد ضرایب LP است و هرچه تعداد آن بالاتر باشد تخمین بهتری از سیگنال بدست می آید. با استفاده از روابط بالا و با توجه به اینکه هدف از این تخمین اینست که مقدار خطا به صفر برسد روابط زیر را خواهیم داشت:

$$J = E \{e^2[n]\} = E \left\{ \left(s[n] + \sum_{i=1}^M a_i s[n-i] \right)^2 \right\} \quad (6-2)$$

تابع ارزش فوق با انتخاب مناسب ضرایب LPC باید مقدار می نیمم را داشته باشد بنابر این نسبت به این ضرایب از معادله (۶-۲) مشتق می گیریم:

$$\frac{\partial J}{\partial a_k} = 2E \left\{ \left(s[n] + \sum_{i=1}^M a_i s[n-i] \right) s[n-k] \right\} = 0 \quad (7-2)$$

$$E \{s[n]s[n-k]\} + \sum_{i=1}^M a_i E \{s[n-i]s[n-k]\} = 0 \quad (8-2)$$

اگر داشته باشیم:

$$R_s[i-k] = E \{s[n-i]s[n-k]\} \quad (9-2)$$

$$R_s[k] = E \{s[n]s[n-k]\} \quad (10-2)$$

که در آن تابع R نشان دهنده همبستگی بین متغیرهای تصادفی ممان های متقابل متغیرهای تصادفی می باشد. بنابر این از معادله ۸-۲ خواهیم داشت:

$$\sum_{i=1}^M a_i R_s [i - k] = R_s [k] \quad (11-2)$$

بنابر این با توجه به معادلات بالا خواهیم داشت:

$$R_s a = -r_s \quad (12-2)$$

که در آن داریم R_s ماتریس همبستگی و a ماتریس ضرایب LPC به فرم زیر می باشند:

$$R_s = \begin{bmatrix} R_s[0] & R_s[1] & \dots & R_s[M-1] \\ R_s[1] & R_s[0] & \dots & R_s[M-2] \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ R_s[M-1] & R_s[M-2] & \cdot & R_s[0] \end{bmatrix} \quad (13-2)$$

$$a = [a_1 \quad a_2 \quad \dots \quad a_M]^T \quad (14-2)$$

$$r_s = [R_s[1] \quad R_s[2] \quad \dots \quad R_s[M]]^T \quad (15-2)$$

بنابر این برای محاسبه ضرایب LPC از معادله ۱۲-۲ داریم:

$$a = -R_s^{-1} r_s \quad (16-2)$$

بلوک تشخیص حروف صدا دار با استفاده از الگوریتم های موجود [۱] (تعداد عبور از صفر، مقدار انرژی سیگنال در فریم و ...) صدا دار یا بی صدا بودن حروف را تشخیص می دهد. سپس ضرایب LPC برای هر فریم محاسبه شده و برای دیکدر ارسال می شود. همچنین از این ضرایب با استفاده از روش آنالیز- ساخت میزان بهره مناسب تقویت توان در یک حلقه بسته محاسبه می گردد. مقدار دوره تناوب گام^۱ نیز در صورت صدا دار بودن حرف محاسبه می گردد.

در نهایت پس از پردازش سیگنال ورودی، به ازای هر فریم، پارامترهای ضرایب LPC، بهره تقویت، صدا دار یا بی صدا بودن حروف و اندازه دوره تناوب در صورت صدا دار بودن حروف به صورت یک رشته بیت در انکدر در آمده و برای دیکدر ارسال می گردد.

در دیکدر LPC که شکل کلی آن در شکل ۲-۷ [3] آمده است بیت های ارسال شده از انکدر دریافت شده و پس از استخراج پارامترهای ارسال شده، با استفاده از اطلاعات موجود و مدل LPC ساخت سیگنال، سیگنال مورد نظر ساخته می شود. در این مدل از یک فیلتر نهایی^۲ برای از بین بردن اثر فیلتر پیش تاکید که در انکدر از آن استفاده شده بود، در انتهای پردازش دیکدر استفاده می شود.

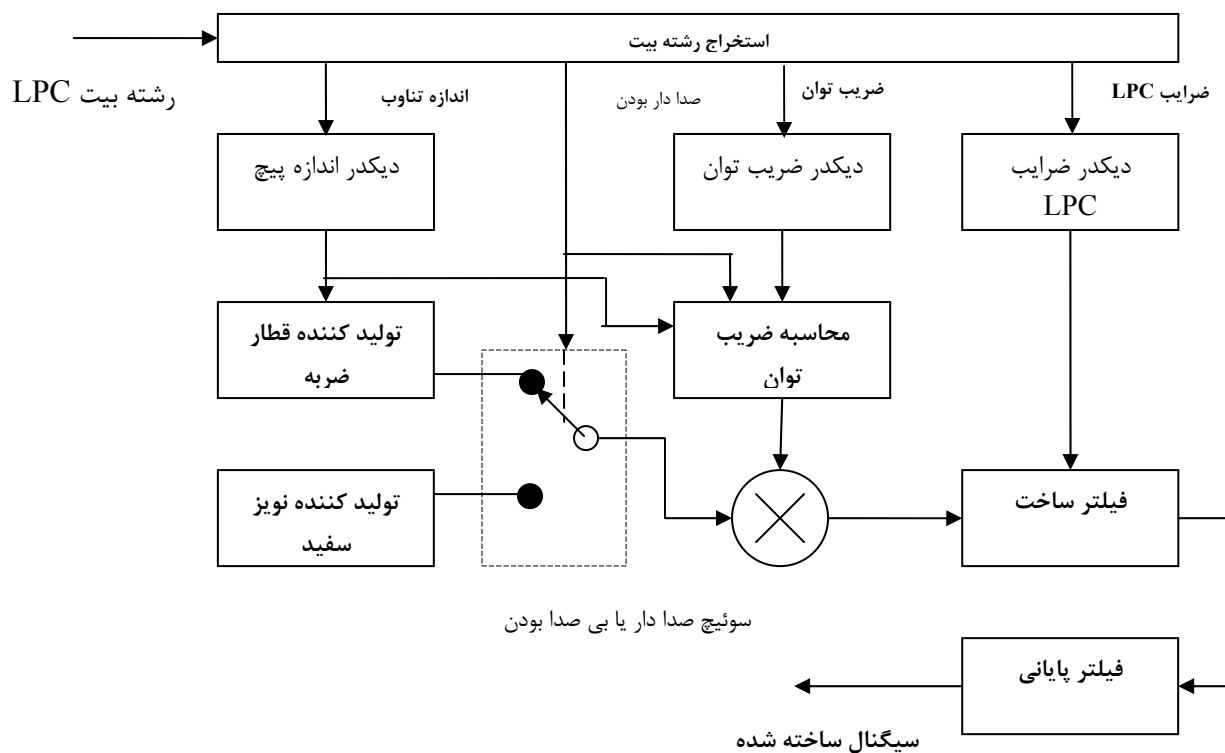
در استاندارد FS1015 [3]، که مربوط به کدکننده LPC می شود سیگنال صحبت ورودی با فرکانس ۸ کیلو هرتز نمونه برداری شده و به ازای هر نمونه ۱۲ بیت تخصیص داده می شود و فیلتر پیش تاکید با تابع تبدیل $(1-0.9375z^{-1})$ بوجود می آید. در این استاندارد تخمین دوره تناوب (گام) از بین ۶۰ مقدار انجام می گیرد که این ۶۰ مقدار به ترتیب عبارتند از ۲۰ و ۲۱ و ۲۹ و ۴۰ و ۴۲ و ۷۸ و ۸۰ و ۸۴ و ۱۵۶ این مقادیر در حقیقت معادل فرکانس بین ۵۰ تا ۴۰۰ هرتز می باشد. همچنین نحوه تخصیص بیت به پارامترهای استخراج شده مطابق جدول ۱-۲ [3] می باشد.

مطابق جدول ۵۴ بیت به ازای یک فریم به طول ۲۲/۵ میلی ثانیه ارسال می گردد که برابر ۲۴۰۰ بیت به ازای یک ثانیه (نرخ بیت) می باشد.

¹ Pitch period

² De-emphasis

فصل دوم - معرفی کد کننده های متداول



شکل ۲-۷. دیکدر یک کد کننده LPC

جدول ۲-۱. بیت های تخصیص داده شده برای پارامتر های کد کننده LPC در کد کننده استاندارد FS1015		
	بیت تخصیص داده شده	
پارامتر	حروف صدا دار	حروف بی صدا
دوره تناوب و مشخص کننده صدا دار بودن	۷	۷
ضرب توان	۵	۵
ضرایب LPC	۴۱	۲۰
بیت همزمانی	۱	۱
کاهش خطا	-	۲۱
جمع	۵۴	۵۴

➤ محدودیت های LPC

محدودیت ها و معایب استفاده از کد کننده های LPC عبارتند از:

در بسیاری موارد تشخیص حروف صدا دار و بی صدا برای استفاده در مدل LPC ساخت سیگنال دشوار می باشد.

- استفاده از مدل نویز سفید و قطار پالس به عنوان منبع تحریک فیلتر با محیط واقعی تفاوت دارد و در حقیقت سیگنال ورودی حالت های مختلفی دارد.
- اطلاعات فاز سیگنال اصلی در این کد کننده حفظ نمی گردد و فقط از اطلاعات اندازه فرکانس برای تخمین ضرایب LPC استفاده می گردد.
- در استفاده از این مدل برای voice ها به دلیل عدم یکنواخت بودن طیف قطار پالس، سیگنال خوبی تولید نمی گردد.
- سیگنال ساخته شده خروجی به نسبت سایر کد کننده ها کیفیت شنیداری کمتری دارد و دارای زبری و ناهمواری های شنیداری نامناسب می باشد.

۲-۶. کد کننده با تخمین خطی و ورودی مرکب^۱

این کد کننده برای بهبود کد کننده LPC به وجود آمد و تفاوت مدل آن با LPC در ورودی تحریک کننده فیلتر می باشد چرا که این ورودی ترکیبی است از پالس های پریودیک و نویز سفید که هر کدام از یک فیلتر شکل دهنده عبور کرده اند می باشد.

این مدل بسیار پیچیده تر از مدل LPC می باشد و برای عملی کردن این کد احتیاج به DSP های قدرتمند و محاسبات زیاد داریم که هزینه زیادی را تحمیل می کند در عین حال در سال ۱۹۹۶ به عنوان یکی از گزینه ها برای استاندارد دولتی آمریکا معرفی گردید [3].

¹ MELP (MIXED EXCITATION LINEAR PREDICTION CODER)

۲-۶-۱. مدل MELP تولید سیگنال

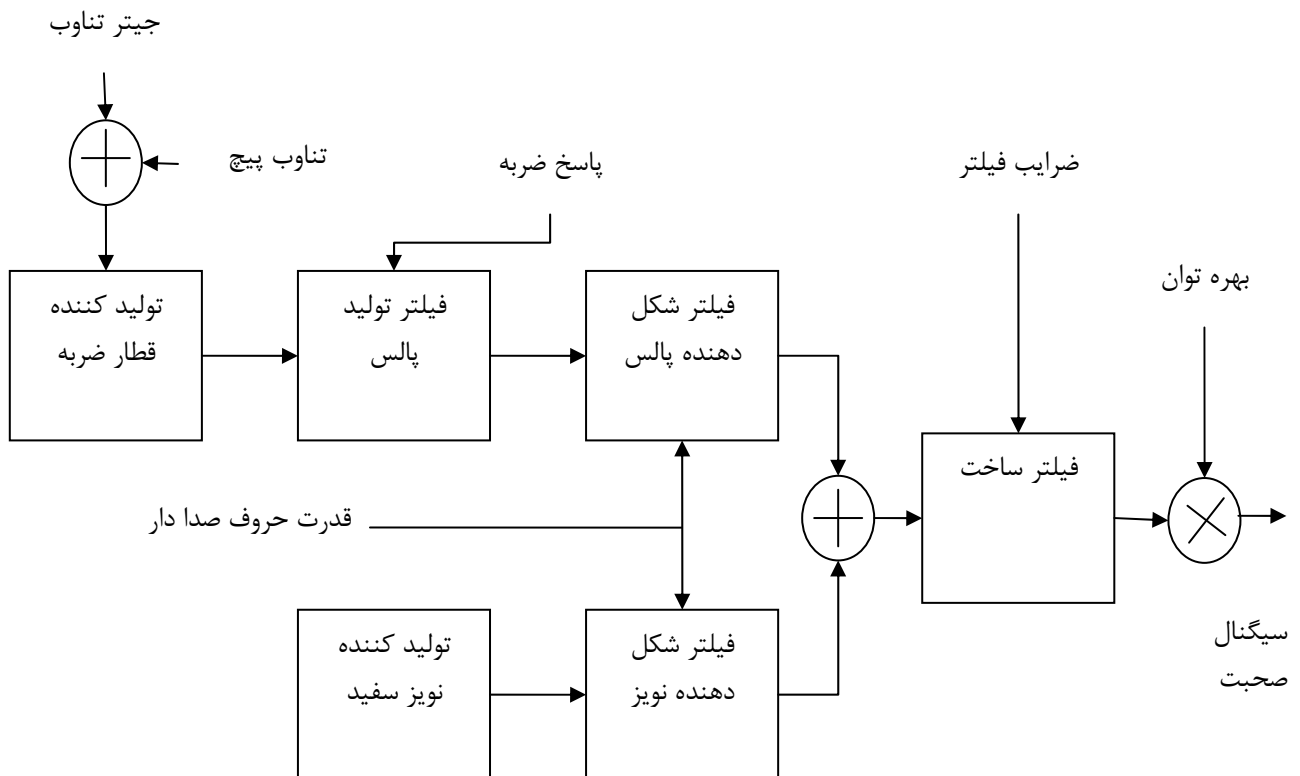
شکل کلی تولید سیگنال در مدل MELP در شکل ۲-۸ [3] آورده شده است. در این مدل با استفاده از بلوک های افزوده نسبت به LPC سعی بر آن شده است که منبع تحریک فیلتر ساخت طوری باشد که سیگنال تولید شده مشابهت بیشتری به سیگنال اصلی داشته باشد و کار فیلتر های شکل دهنده پالس و نویز ساختن چنین ورودی تحریکی می باشد.

از مهمترین مزیت های این کد کننده تبدیل اصوات به سه دسته voice، unvoice و jittery voice می باشد که با این مزیت بر مشکل تشخیص voice از unvoice که پیچیدگی های خاص خود را در مدل LPC دارد غلبه می کند. حالت جیتیتری زمانی به وجود می آید که سیگنال تحریک غیر پریودیک بوده ولی کاملاً تصادفی نمی باشد. در این حالت سیگنال در محدوده Voice ها قرار می گیرد و معمولاً این مدل تحریک با جیتر گام سیگنال اصلی کنترل می شود و مقدار تصادفی با توجه به مقدار گام بوجود می آید. (با استفاده از آزمایش نشان داده شده ۲۵٪ دوره تناوب نتایج خوبی دارد [3]).

در این کد کننده شکل پالس تحریک، برای تحریک پریودیک، از سیگنال ورودی استخراج می شود و به عنوان اطلاعات ساخت سیگنال در فریم اطلاعات ارسال می گردد. در مدل LPC سیگنال تحریک voice ها تشکیل شده است از قطار ضربه که با دنیای واقعی متفاوت می باشد ولی در مدل MELP شکل پالس های تحریک که اطلاعات مهمی را دارد از تبدیل فوری خطای تخمین بدست می آید و از کمیت های بدست آمده برای تولید فیلتر تولید کننده پالس های تحریک استفاده می شود.

با توجه به شکل ۲-۸ تحریک پریودیک و نویزی بعد از عبور از فیلتر شکل دهنده پالس و نویز با یکدیگر جمع شده و سیگنال تحریک مرکب را برای فیلتر ساخت بوجود می آورد که این مسئله هسته اصلی MELP را تشکیل می دهد. پاسخ فرکانسی فیلترهای شکل دهنده پالس و نویز با زمان تغییر می کند و پارامترهایی که قدرت voice بودن نامیده می شوند شکل این فیلتر ها را کنترل می کنند این پارامتر ها با توجه به سیگنال ورودی تخمین زده می شود و به عنوان اطلاعات در هر فریم ارسال می گردد.

در بخش بعد اجزای مختلف تولید سیگنال توسط MELP بطور مختصر معرفی می گردد.



شکل ۸-۲. مدل MELP تولید سیگنال

۲-۶-۲. محاسبه و کوانتیزه کردن دامنه فوریه

کد کننده MELP بر مبنای محاسبه دامنه ی فوریه، از سیگنال خطای تخمین برای بدست آوردن شکل پالس تحریک استفاده می کند. کمیت های این پردازش کوانتیزه می شود و در فریم های اطلاعات ارسال می شوند . هدف اصلی این است که در دیکدر سیگنالی پر یودیکی مشابه سیگنال تحریک اصلی (در حالت voice) تولید گردد.

در این بخش روش محاسبه دامنه فوریه بیان می گردد و باید توجه داشت که این محاسبه زمانی انجام می گیرد که فریم سیگنال voice یا jittery voice باشد.

۲-۶-۳. فیلتر تولید کننده پالس

همان طور که در شکل ۲-۸ نشان داده شد فیلتر شکل دهنده ی پالس برای سیگنال های پریودیک یکی از بلوک های اصلی MELP می باشد. ایده اصلی کار این است که پاسخ ضربه این فیلتر در انکدر پیدا شده و برای تولید قطار پالس (تحریک) به دیکدر ارسال شود. شکل ۲-۹ روند تولید پالس برای حالتی که قطار ضربه با دوره تناوب T ، فیلتر را برای تولید پالس در خروجی تحریک میکند را نشان می دهد. با توجه به دامنه ی تبدیل فوریه مشاهده می شود که طیف فرکانسی قطار پالس با ضرب طیف فرکانسی قطار ضربه و طیف پاسخ ضربه فیلتر تولید پالس بدست می آید. بنابر این با داشتن ارتفاع نقاط ماکزیمم طیف دامنه قطار پالس خروجی می توان پاسخ فرکانسی فیلتر تولید پالس را به طور تقریبی بدست آورد بنابر این با داشتن نقاط ماکزیمم طیف فرکانسی خروجی فیلتر تولید کننده پالس و ارسال آن برای دیکدر می توان این فیلتر را در دیکدر بوجود آورد.

با توجه به توضیحات بالا، در انکدر دامنه DFT برای خطای تخمین محاسبه می گردد و نقاط ماکزیمم طیف دامنه متناظر با هارمونیک های فرکانس $pitch$ اندازه گیری می شود و اندازه این نقاط ماکزیمم برای دیکدر به منظور تشکیل پالس تحریک با تشکیل فیلتر ساخت پالس فرستاده می شود.

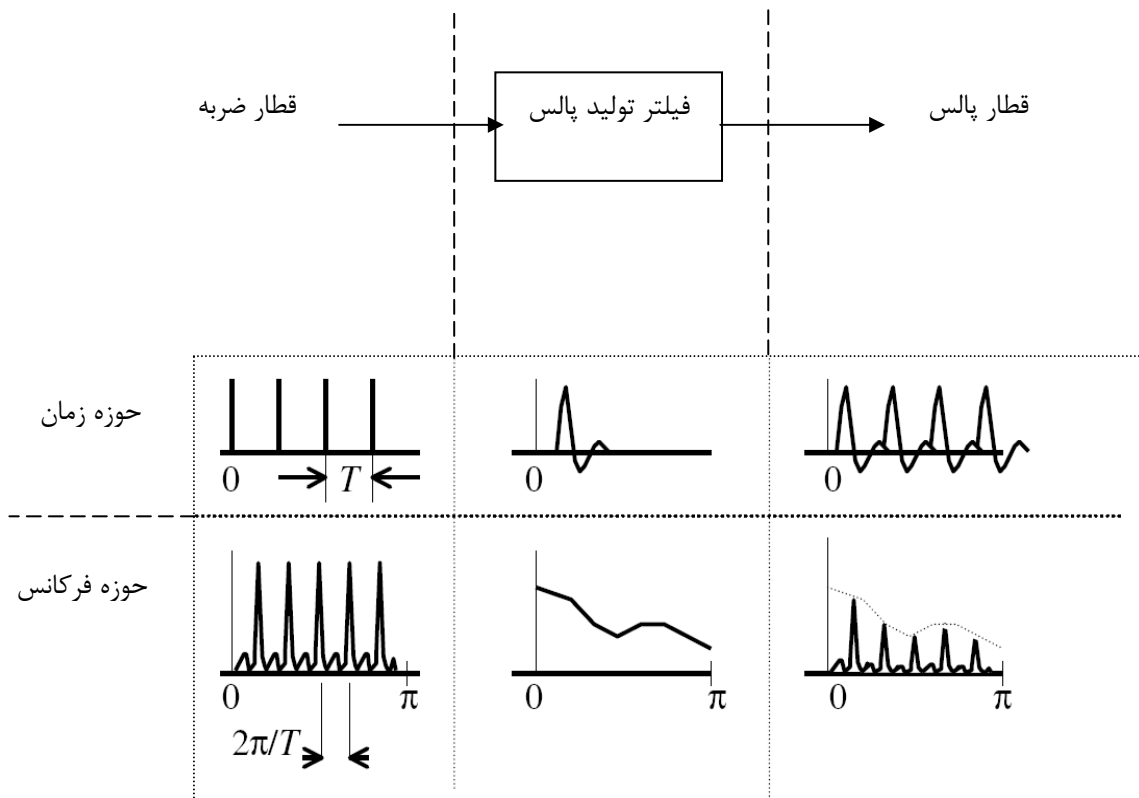
۲-۶-۴. فیلتر های شکل دهنده

همان طور که در شکل ۲-۸ نشان داده شده است در مدل MELP از دو فیلتر شکل دهنده پالس و نویز که پارامتر های آن با سیگنال ورودی (میزان $voice$ یا $unvoice$ بودن) تغییر می کند، استفاده شده است. این فیلترها از ترکیب موازی چند فیلتر در باندهای فرکانسی خاص (در MELP استاندارد ۵ باند) تشکیل شده اند و برای ساخت سیگنال تحریک مرکب در دیکدر استفاده می شوند.

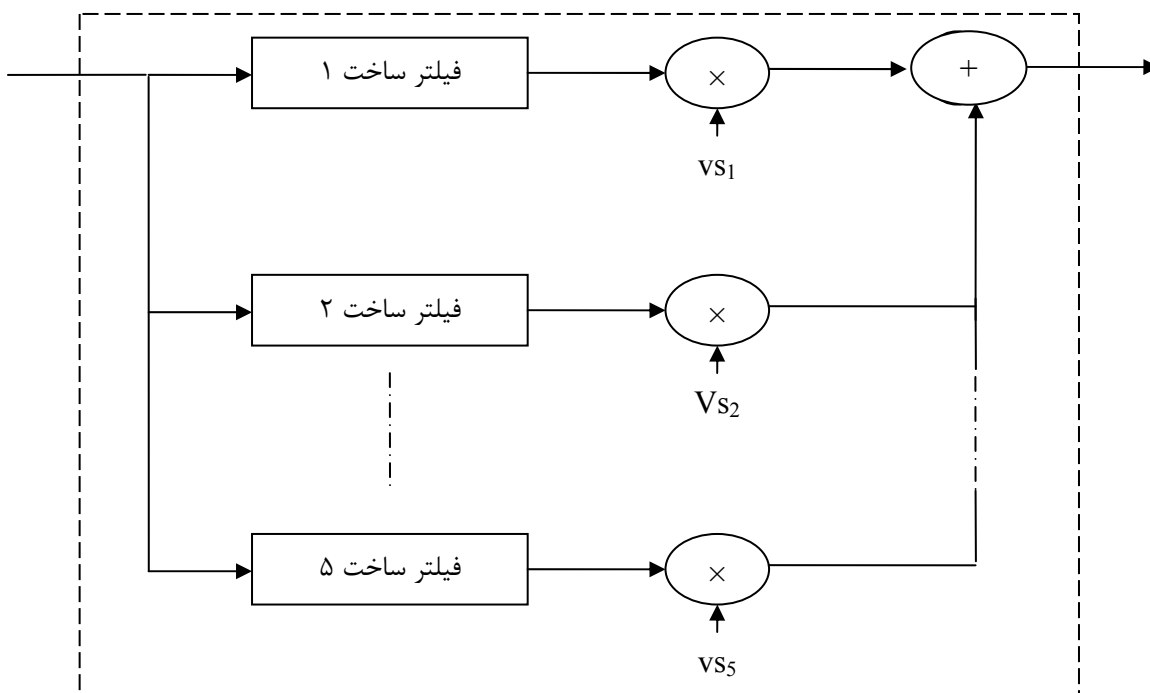
شکل ۲-۱۰ [3] بلوک دیاگرام کلی فیلتر شکل دهنده پالس را نشان می دهد. این فیلترها، فیلتر ساخت نامیده می شوند و هر فیلتر ساخت یک باند مشخص فرکانسی را کنترل می کنند. این باندها به طور استاندارد از بازه های $0-500$ ، $500-1000$ ، $1000-2000$ ، $2000-3000$ ، $3000-4000$ ، $4000-20000$ هرتز تشکیل شده است و با در نظر گرفتن پاسخ ضربه هر فیلتر ساخت به صورت $h_i[n]$ و $i=1,2,3,4,5$ پاسخ نهائی فیلتر شکل دهنده برابر است با

$$h_p[n] = \sum_{i=1}^5 v s_i h_i[n], \dots \dots \dots 0 \leq v s_i \leq 1 \quad (۱۷-۲)$$

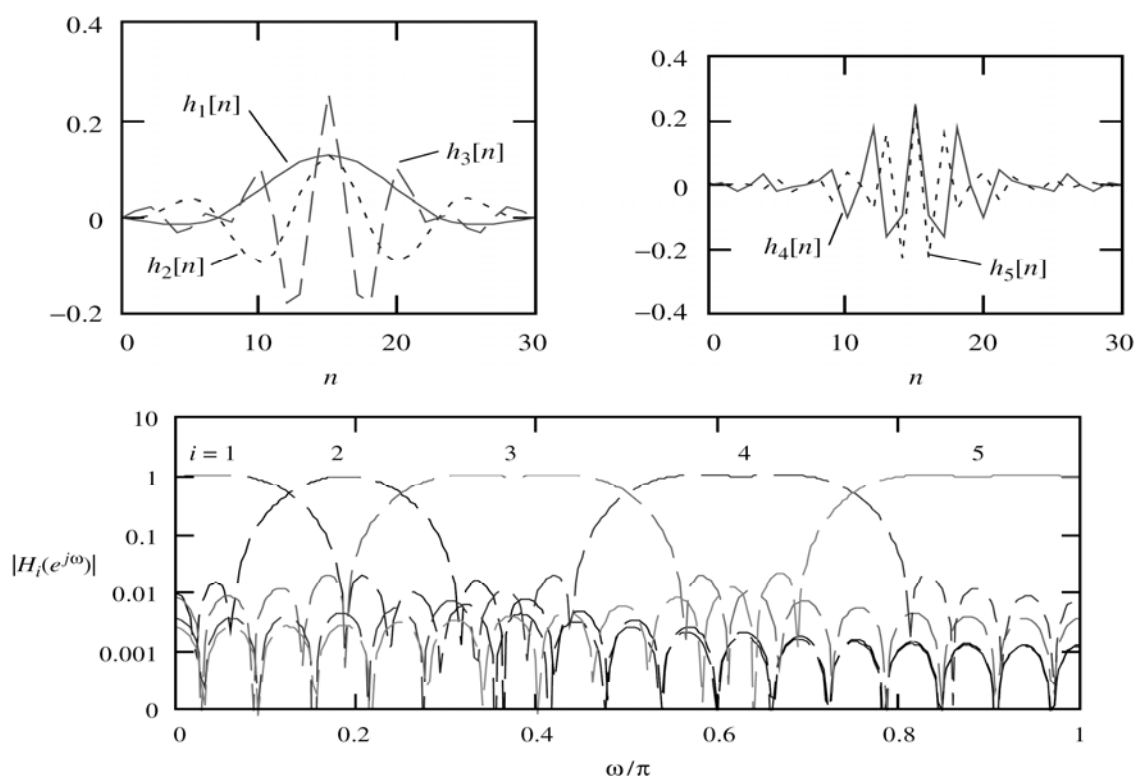
که در آن ضریب (voice strengths) $v s_i$ به میزان voice بودن سیگنال در هر باند بستگی دارد. در کد کننده MELP استاندارد به لحاظ مزیت های فیلتر های FIR، در طراحی فیلتر شکل دهنده از این فیلترها استفاده می شود. شکل ۱۱-۲ شکل زمانی و فرکانسی نمونه ای از این فیلترها را نشان می دهد [3].



شکل ۹-۲. فیلتر تولید کننده پالس



شکل ۲-۱۰. بلوک دیاگرام کلی فیلتر شکل دهنده پالس



شکل ۲-۱۱. پاسخ زمانی و فرکانسی فیلتر شکل دهنده پالس و نویز

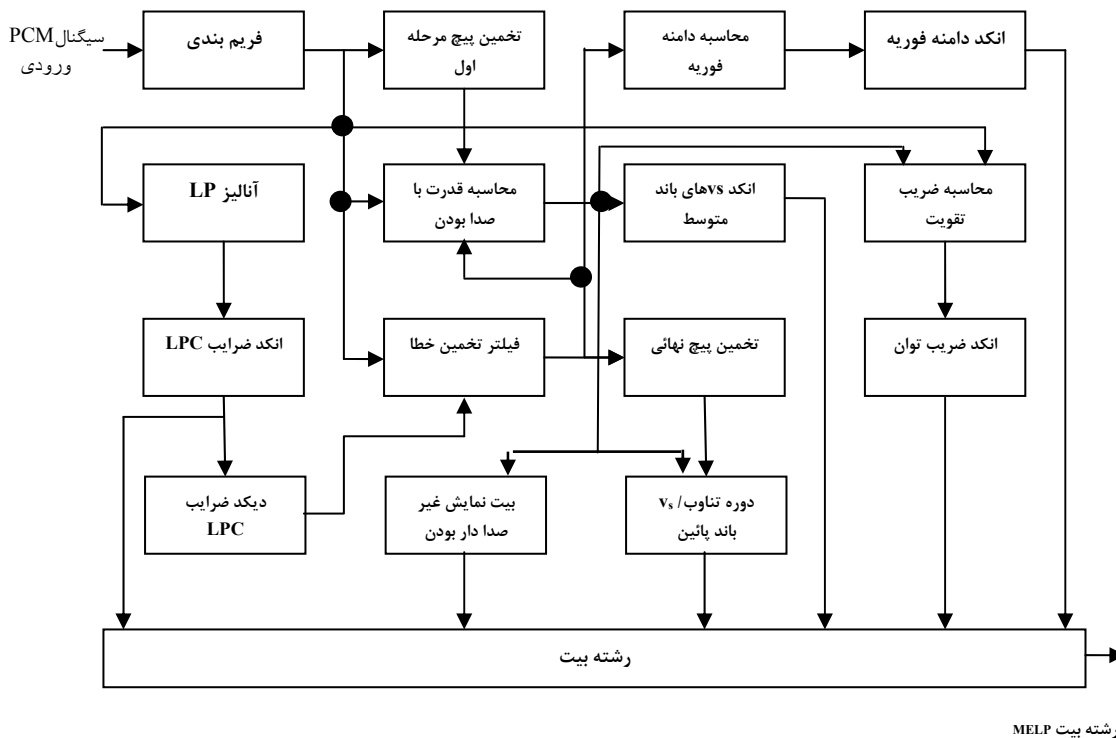
۲-۶-۵. انکدر MELP

بلوک دیاگرام کلی انکدر یک کد کننده MELP در شکل ۲-۱۲ [3]، آورده شده است. همانگونه که در شکل دیده می شود سیگنال صحبت ورودی ابتدا فریم بندی می شود و دوره ی تناوب (pitch) اولیه آن توسط الگوریتم های مربوط تخمین زده می شود تا برای پردازش های بعدی مورد استفاده قرار گیرد. پارامتر های فیلتر شکل دهنده پالس که دامنه فوریه می باشند از روی محاسبه تخمین میزان خطا محاسبه می شوند و مقدار آنها در نقاط ماکزیمم برای تشکیل این فیلتر در دیکدر به صورت رشته بیت در می آیند. همچنین ضرایب LPC سیگنال اصلی نیز برای هر فریم محاسبه گشته و به منظور تشکیل فیلتر ساخت فرمنت (vocal tract) برای دیکدر ارسال می گردد. در نهایت با استفاده از روش آنالیز- ساخت، پارامترهای فیلترهای شکل دهنده (FIR filters) برای پائین ترین باند که اهمیت بیشتری دارد و ۴ باند بالاتر و همچنین ضریب قدرت صدا دار بودن هر کدام محاسبه گشته و با محاسبه گین تقویت این پارامتر ها به صورت بیت های نمایش دهنده در یک رشته بیت اطلاعات با الگوریتم خاص در آمده و برای دیکدر ارسال می گردد.

در پردازش هر فریم یک بیت برای نشان دادن میزان قدرت voice یا jittery voice بودن در باند پائین با عنوان پرچم غیر پریودیک بودن^۱ استفاده می شود که مقدار آن از رابطه زیر محاسبه می گردد.

$$a_f = \begin{cases} 1, & \text{if } v s_1 \leq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (2-18)$$

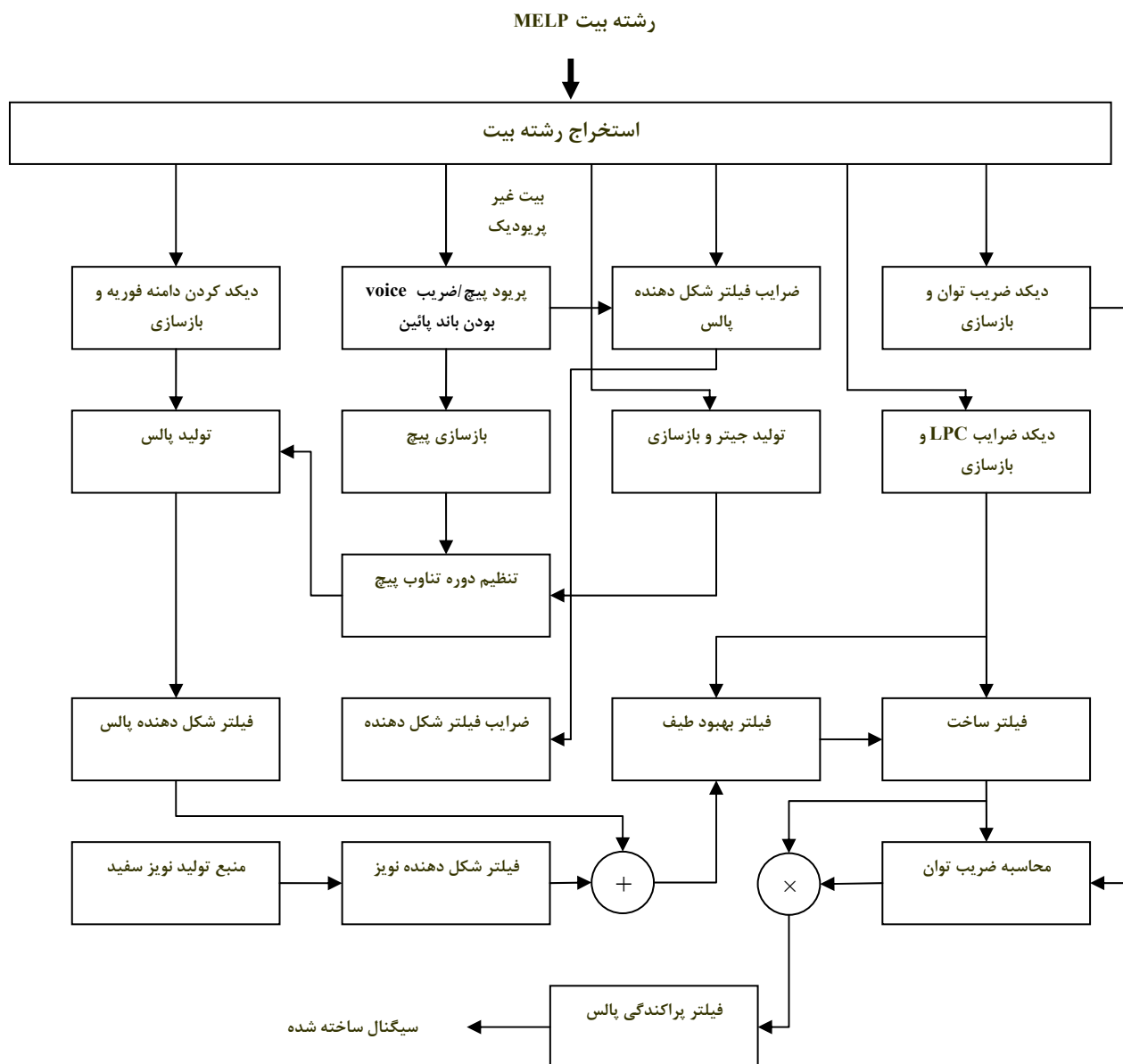
¹ Aperiodic flag



شکل ۲-۱۲. بلوک دیاگرام کلی انکدر کد کننده MELP

۲-۶-۶. دیکدر MELP

شکل ۲-۱۳ [3]، بلوک دیاگرام دیکدر MELP را نشان می دهد. مطابق شکل، رشته بیت ارسال شده از انکدر در یافت شده مطابق الگوریتم ارسال بیت های مربوط به هر پارامتر شناسائی شده و پارامتر های مختلف آن که شامل ضرایب LPC، بیت نمایش گر حرف غیر صدا دار، دوره تناوب پیچ، قدرت صدا دار بودن باند پائین، قدرت صدا دار بودن باندهای میانی و بهره توان می باشد از آن استخراج می شود و با استفاده از مدل MELP تولید سیگنال که در شکل ۲-۸ آورده شده است، و توضیحات مربوطه که در قسمت قبل بیان گردید سیگنال مورد نظر ساخته می شود. در این شکل از دو فیلتر افزوده شده فیلتر بهبود طیف و فیلتر پراکندگی پالس برای بهبود کیفیت شنیداری سیگنال استفاده شده است [3].



شکل ۲-۱۳. بلوک دیاگرام کلی دیکدر کد کننده MELP

۲-۶-۷. نحوه تخصیص بیت در کد کننده MELP استاندارد

جدول ۲-۲ [3]، نحوه تخصیص بیت در کد کننده MELP را بطور خلاصه بیان می کند. مطابق جدول برای هر فریم ۵۴ بیت تخصیص داده شده است که با طول فریم استاندارد ۲۲/۵ میلی ثانیه نرخ بیتی برابر ۲۴۰۰ بیت در ثانیه خواهیم داشت.

جدول ۲-۲. بیت های تخصیص داده شده برای پارامتر های کد کننده MELP در کد کننده استاندارد FS MELP CODER		
	بیت تخصیص داده شده	
پارامتر	حروف صدا دار	حروف بی صدا
دوره تناوب و قدرت voice بودن باند پائین	۷	۷
ضریب توان	۸	۸
ضرایب LPC	۲۵	۲۵
بیت همزمانی	۱	۱
کاهش خطا	-	۱۳
قدرت voice بودن باند میانی	۴	-
بیت نشان دهنده غیر پررودیک بودن	۱	-
دامنه فوریه	۸	-
جمع	۵۴	۵۴

۲-۷. کد کننده با تحریک پالس های منظم

این نوع کد کننده بر مبنای استفاده از روش مدولاسیون کد پالس تفاضلی^۱ (DPCM) استوار است که در آن به جای ارسال نمونه های اصلی برای باز سازی سیگنال، اختلاف نمونه ها ارسال می گردد تا رنج دینامیکی و پراکندگی نمونه های ارسال کمتر گردد.

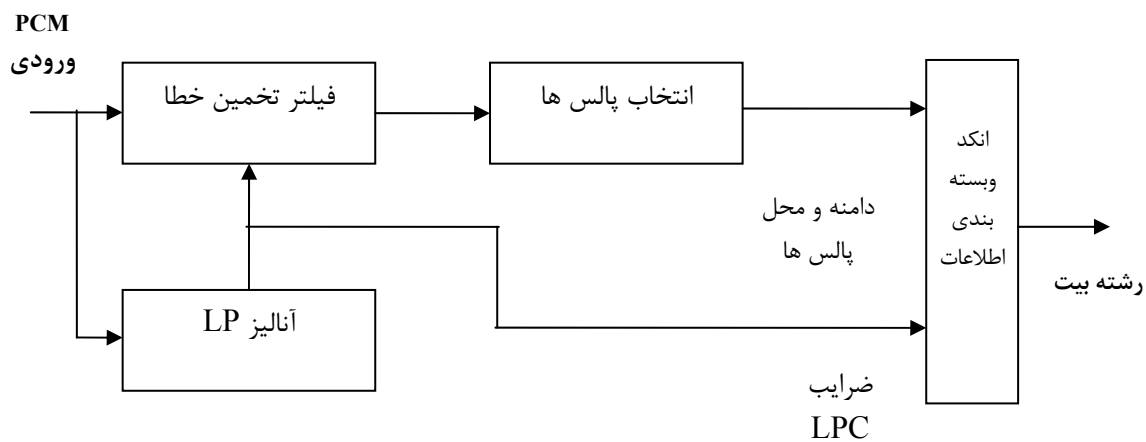
در سال ۱۹۸۲ ثابت شد که برای بدست آوردن کیفیت خوب در سیگنال صحبت نیازی نیست تمامی نمونه های تفاضلی شده را برای بازسازی سیگنال ارسال نمود به این معنی که بسیاری از نمونه های تفاضلی شده دارای دامنه کوچکی هستند و برای بازسازی سیگنال می توان از آنها صرف نظر نمود.

طی آزمایش های انجام شده ثابت شد که با بکار بردن فقط ۱۰٪ نمونه ها می توان سیگنال را با کیفیت مناسب بازسازی نمود و مسئله ای که در این زمینه اهمیت ویژه ای داشت اتخاذ روشی مناسب بود که طی آن این مقدار نمونه ها بدرستی انتخاب شوند تا سیگنال بازسازی شده از نظر درکی دچار اعوجاج نشود [3].

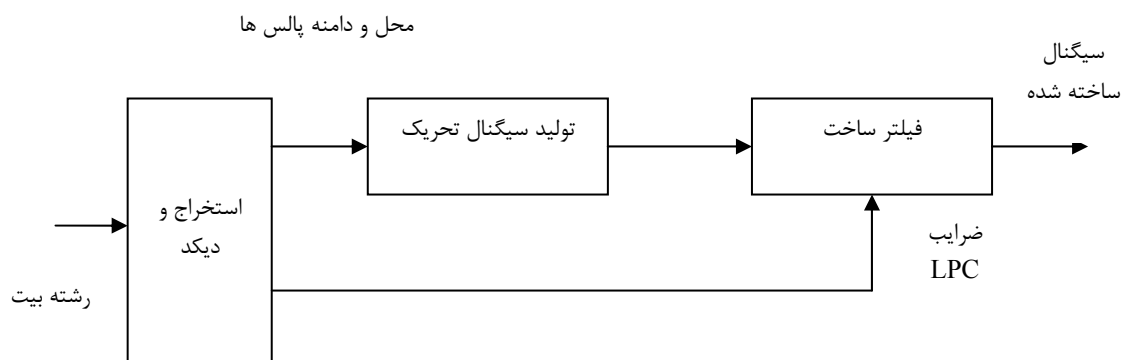
در کد کننده با تحریک پالس های منظم، از پالس تحریک مرکب از درصدی از نمونه های غیر صفر به عنوان ورودی فیلتر ساخت استفاده می گردد. اندازه دامنه و مکان نمونه های غیر صفر در انکدر انکد می شود و برای دیکدر ارسال می گردد. تعداد نمونه های انتخاب شده در هر فریم نیز به نرخ بیت مورد قبول بستگی دارد و هرچه تعداد این نمونه ها بیشتر باشد کیفیت سیگنال ساخته شده بهتر می گردد.

شکل کلی انکدر و دیکدر یک کد کننده بر مبنای سیستم تحریک با پالس های غیر صفر و روش حلقه باز برای بدست آوردن پارامترهای مورد نیاز برای ساخت سیگنال در شکل ۲-۱۴ و ۲-۱۵ [3] آورده شده است. نرخ بیت مورد نیاز برای ارسال این کد کننده با توجه به تعداد نمونه های ارسالی تعیین می گردد و معمولاً به نسبت کد کننده های معرفی شده نرخ بیت بالاتری دارد.

¹ Differential Pulse Code Modulation



شکل ۲-۱۴. انکدر حلقه باز کد کننده با تحریک پالس نمونه های غیر صفر



شکل ۲-۱۵. دیکدر کد کننده با تحریک پالس نمونه های غیر صفر

فصل ۳ کد کننده سلپ

۳-۱. مقدمه

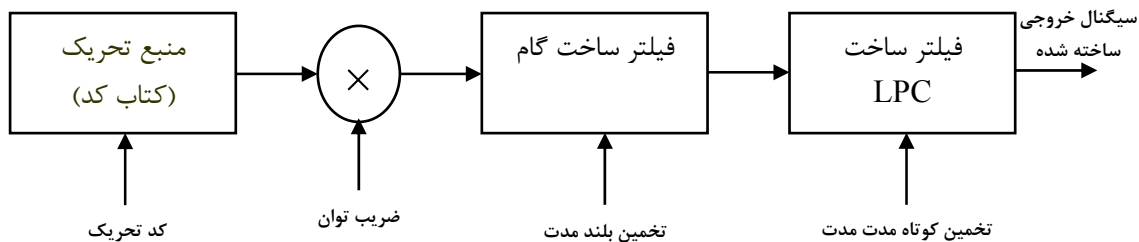
ایده اصلی کد کننده با تخمین خطی و کد تحریک^۱ (سلپ) از بهبود LPC گرفته شده است و این ایده اولین بار در سال ۱۹۸۵ توسط آقای Schroder و همکاران مطرح گردید [3]. در این کد کننده برای تولید سیگنال هم از آنالیز LPC و هم از آنالیز پیش گوئی بلند مدت برای تولید سیگنال استفاده می گردد و به همین دلیل برای سیگنال های صدادار مشکل روش LPC را برای تولید سیگنال ندارد. نام گذاری این کد کننده از آنجا ناشی می شود که در انکدر و دیکدر، منبع تحریک ورودی فیلتر ساخت معادل لوله صوتی، تشکیل شده است از تعدادی مشخص سیگنال که در یک حافظه ذخیره شده اند (معمولا ۱۰۲۴ سیگنال) و وظیفه انکدر انتخاب سیگنال مناسب از بین این سیگنال ها برای تحریک فیلتر ساخت و ارسال شماره (کد) سیگنال مذکور برای دیکدر می باشد. این سیگنال به گونه ای انتخاب می گردد که سیگنال ساخته شده توسط خروجی فیلتر ساخت بیشترین شباهت را به سیگنال اصلی داشته باشد. به این ترتیب با دانستن موقعیت این سیگنال با عدد مشخص کننده آن در حافظه ذخیره سازی سیگنال تحریک و ارسال آن و سایر پارامتر های ساخت سیگنال در مدل سلپ تولید صدا برای دیکدر، که در دیکدر نیز همان حافظه وجود دارد، سیگنال خروجی ساخته می شود.

در طی سال های اخیر بنا به ماهیت این کد کننده که جزو کد کننده های مرکب می باشد و ساختار خاصی دارد، ایده های زیادی برای بهبود آن پیشنهاد و اجرائی شده است که این تحقیقات و تلاش ها همچنان ادامه دارند.

مدل تولید صحبت برای این کد کننده در شکل ۳-۱ [6] آورده شده است که در این مدل منبع تحریک^۲ می تواند منبع ثابت یا تطبیق پذیر، پالسهای زمانی مشخص یا نویز تصادفی باشد. برای حالت کلی این منبع یک سیگنال نویز سفید با نمونه های مختلف که ۱۰۲۴ سیگنال به وجود

^۱ Code Excited Linear Prediction Coder (CELP)

^۲ Excitation Codebook



شکل ۳-۱. مدل CELP تولید سیگنال صحبت

می آورد در نظر گرفته می شود. تعداد نمونه ها در هر پالس به تعداد نمونه های هر زیر فریم که یک چهارم فریم سیگنال اصلی می باشد در نظر گرفته می شود (۶۰ نمونه در هر سیگنال) که این پالس به عنوان سیگنال تحریک ورودی فیلتر ساخت سیگنال که خود متشکل از دو فیلتر پشت سر هم فیلتر LPC و فیلتر بلند مدت است، استفاده می شود.

۳-۲. خود بستگی نمونه ها

در سیگنال صحبت دو نوع خود بستگی را می توان بیان نمود. نوع اول خود بستگی کوتاه مدت است که بین نمونه های مجاور وجود دارد و در حقیقت از همین بستگی برای محاسبه ضرایب تخمین خطی استفاده می کنند. نوع دوم خود بستگی بلند مدت است که در فریم های صدا دار سیگنال صحبت بین نمونه هایی با فاصله پریود گام وجود دارد. این خود بستگی از طبیعت پریودیک آواهای صدادار نشأت می گیرد.

پیشگویی بلند مدت، فیلتریست که این نوع خود همبستگی را در سیگنال تحریک القا می کند تا طبیعت پریودیک مورد نیاز در فریم های صدادار برای سیگنال سنتز شده بوجود آید. این فیلتر، همانند پیشگویی کوتاه مدت توسط آنالیز پیشگویی خطی بدست می آید. اما این پیشگویی نه بر اساس نمونه هایی مجاور، بلکه بر اساس نمونه هایی با فاصله یک پریود گام، انجام می گیرد. بهمین دلیل آنرا پیشگویی بلند مدت گوئید. تخمین بلند مدت سیگنال صحبت در ساده ترین حالت به صورت $\hat{y}[n] = by(n - T)$ است که در آن T تأخیر بلند مدتی است که به ازای آن خود بستگی سیگنال صحبت به مقدار بیشینه خود می رسد. برای فریم های صدادار T متناظر با پریود گام و یا

مضاربی از آن است و در فریم های بی صدا مقادیر تصادفی اختیار می کند. b ضریب پیشگویی بلند مدت است و به گونه ای محاسبه می شود که انرژی خطای پیشگویی کمینه شود. خطای پیشگویی از اعمال سیگنال صحبت به فیلتر آنالیز $B(z) = 1 - bz^{-T}$ بدست می آید. روش بدست آوردن ضریب b همانند بدست آوردن ضرایب LPC می باشد بطوریکه داریم:

$$J = \sum_n (y[n] - \hat{y}[n])^2 = \sum_n (y[n] - b y[n-T])^2 \quad (1-3)$$

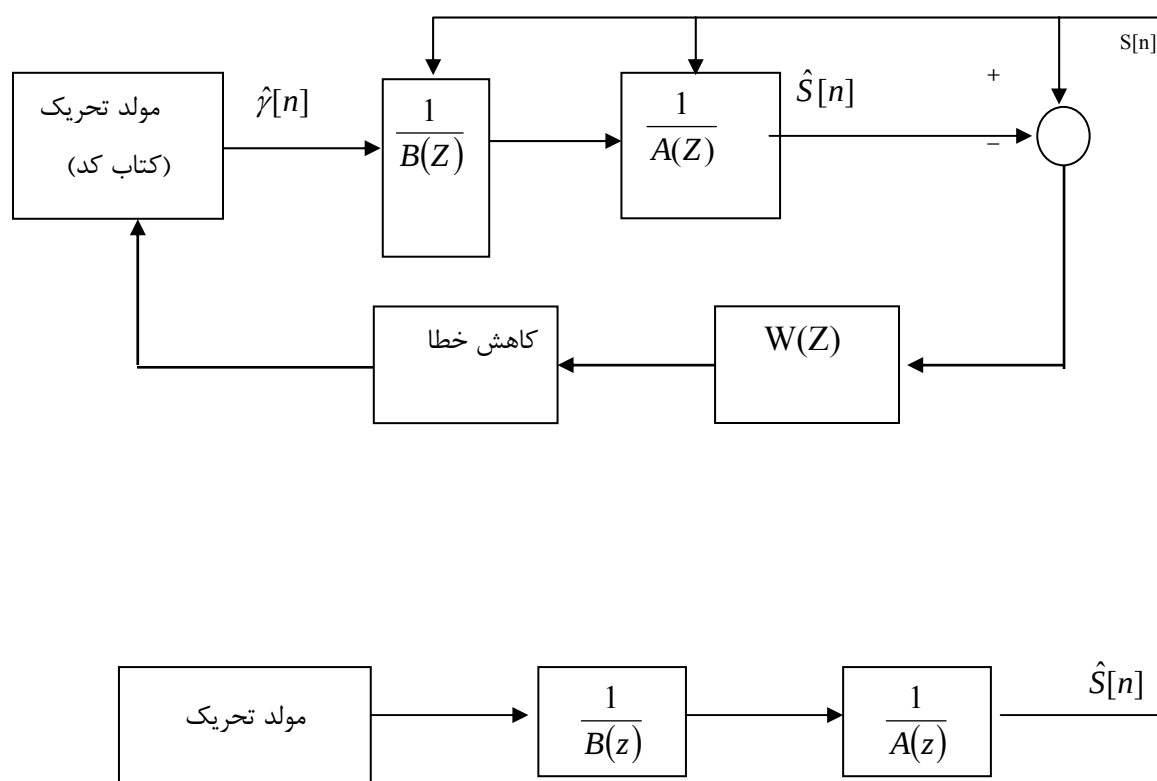
برای مینیمم شدن معادله (۱-۳) که نشان دهنده خطای پیش گویی است، نسبت به پارامتر b معادله فوق مشتق گرفته و برابر صفر قرار می دهیم و بهترین بهره از رابطه (۲-۳) بدست می آید:

$$b = - \frac{\sum_n y[n] - y[n-T]}{\sum_n y^2[n-T]} \quad (2-3)$$

۳-۳. روش آنالیز به وسیله ساخت^۱

روش آنالیز - ساخت بر این مبنا استوار است که برای تعیین برخی پارامتر های یک سیگنال برای ارسال به دیکدر ابتدا مدل دیکدر در خود انکدر ایجاد می گردد و سیگنالی که قرار است در دیکدر بازسازی شود در خود انکدر ساخته می شود و با تغییر پارامتر ها موجود بهینه ترین پارامتر ها برای ساخت شبیه ترین سیگنال انتخاب می گردد و برای دیکدر ارسال می گردد. البته به دلیل وقت گیر بودن این روش به جهت مورد آزمایش قرار دادن همه پارامتر های موجود، برای تعیین پارامترهای فیلتر ساخت از روش حلقه باز (روش مستقیم) استفاده می شود و تنها برای مشخص کردن سیگنال تحریک مناسب و بهره سیگنال از روش حلقه بسته (روش آنالیز - ساخت) استفاده می گردد.

¹ Analysis-By-Synthesis (ABS)



شکل ۳-۲. بلوک کلی نحوه آنالیز (شکل بالا) و ساخت (شکل پایین) سیگنال در کد کننده سلپ

۳-۴. ساختار کلی سلپ

بلوک دیاگرام کلی و اصلی سلپ که بر مبنای مدل تولید آن است در شکل ۳-۲ [6]، آورده شده است که اجزای آن در بخش بعدی توضیح داده می شوند.

۳-۴-۱. پیشگویی کوتاه مدت، فیلتر $\frac{1}{A(z)}$

فیلتر $\frac{1}{A(z)}$ همان فیلتر ساخت فرمنت است که با استفاده از ضرایب تخمین خطی بدست می آید.

این فیلتر که یک فیلتر تمام قطب درجه M است همان فیلتر استفاده شده در کد کننده LPC می باشد و از طریق آنالیز پیشگویی خطی بدست می آید. در این آنالیز، پیشگویی بر اساس M نمونه قبلی سیگنال انجام می شود و از این جهت این فیلتر را پیشگویی کوتاه مدت گویند.

این فیلتر مدلی از لوله صوتی نتیجه می دهد که بطور مستقیم در ساخت طیف سیگنال ساخته

شده تاثیر دارد. پارامترهای این فیلتر برای هر فریم سیگنال که بطور معمول بین ۲۰ تا ۳۰ میلی ثانیه است محاسبه می گردد. این پارامتر ها برای ساخت سیگنال، کد بندی شده و به دیکدر سلپ فرستاده می شوند. تعداد این ضرایب (M) هر چه بیشتر باشد سیگنال تولید شده خروجی کیفیت بهتری خواهد داشت و معمولاً تعداد این ضرایب ۱۰ عدد در نظر گرفته می شود [3].

۳-۴-۲. پیشگویی بلند مدت، فیلتر $\frac{1}{B(Z)}$

فیلتر $\frac{1}{B(Z)}$ فیلتری است که برای کم کردن محدودیت های فیلتر ساخت برای سیگنال های صدا دار استفاده می گردد و پارامتر های آن که شامل ضریب b و دوره تناوب T می گردد این پارامتر ها در هر فریم ۴ بار محاسبه می گردند یعنی به ازای هر زیر فریم که ۲۵٪ یک فریم کامل است این پارامتر محاسبه شده و برای دیکدر ارسال می گردد.

۳-۴-۳. فیلتر وزن دهنده خطا $W(Z)$

فیلتر $W(Z)$ تطبیق دهنده سیگنال خطا با مشخصات ادراکی گوش است این عمل با شکل دادن به طیف سیگنال خطا انجام می گیرد. معمولاً از خاصیت حذف کردن برخی مولفه های سیگنال به عنوان مشخصه اصلی شنوایی انسان در بسیاری از کاربرد ها استفاده می شود. مثلاً میزان ادراک انسان از سیگنالی که میزان سیگنال به نویز بالایی دارد بسیار بالاست و می توان در فرکانس هایی که میزان انرژی سیگنال در آن فرکانس ها زیاد است مقداری نویز به سیگنال افزود بدون اینکه اعوجاجی در آن ایجاد گردد. و برعکس اگر نسبت سیگنال به نویز در نواحی فرکانسی پائین باشد کوچکترین تغییری به سرعت اعوجاج ایجاد خواهد کرد. وظیفه فیلتر $W(Z)$ تمرکز انرژی خطا، در محدوده هایی از فرکانس است که سیگنال اصلی دارای انرژی بیشتری می باشد. این عمل باعث می شود که نسبت سیگنال به نویز در محدوده هایی از فرکانس که مقدار کمی دارد، افزایش یابد و برعکس. باید توجه داشت که شکل این فیلتر باید مطابق شکل تابع تبدیل اصلی فیلتر ساخت باشد تا شکل اصلی سیگنال تغییر نکند.

در عمل و با استفاده از آزمایش های مختلف، تابع تبدیل فیلتر $W(Z)$ بصورت زیر معرفی

می گردد :

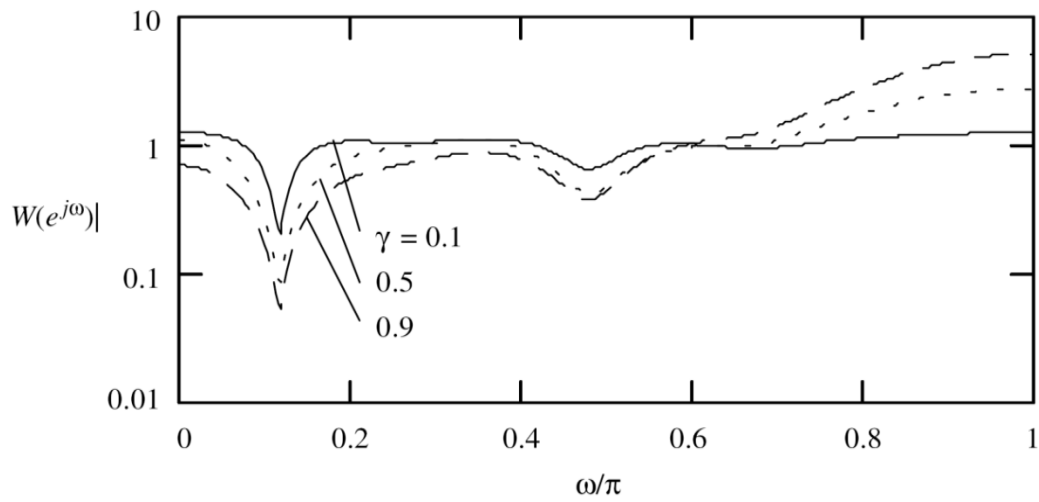
$$W(Z) = \frac{A(Z)}{A(\frac{Z}{\gamma})} = \frac{1 + \sum_{i=1}^M a_i z^{-i}}{1 + \sum_{i=1}^M a_i \gamma^i z^{-i}} \quad (3-3)$$

که در معادله فوق $A(Z)$ همان فیلتر آنالیز کوتاه مدت و γ عددی بین صفر و یک می باشد که به صورت تجربی بدست می آید.

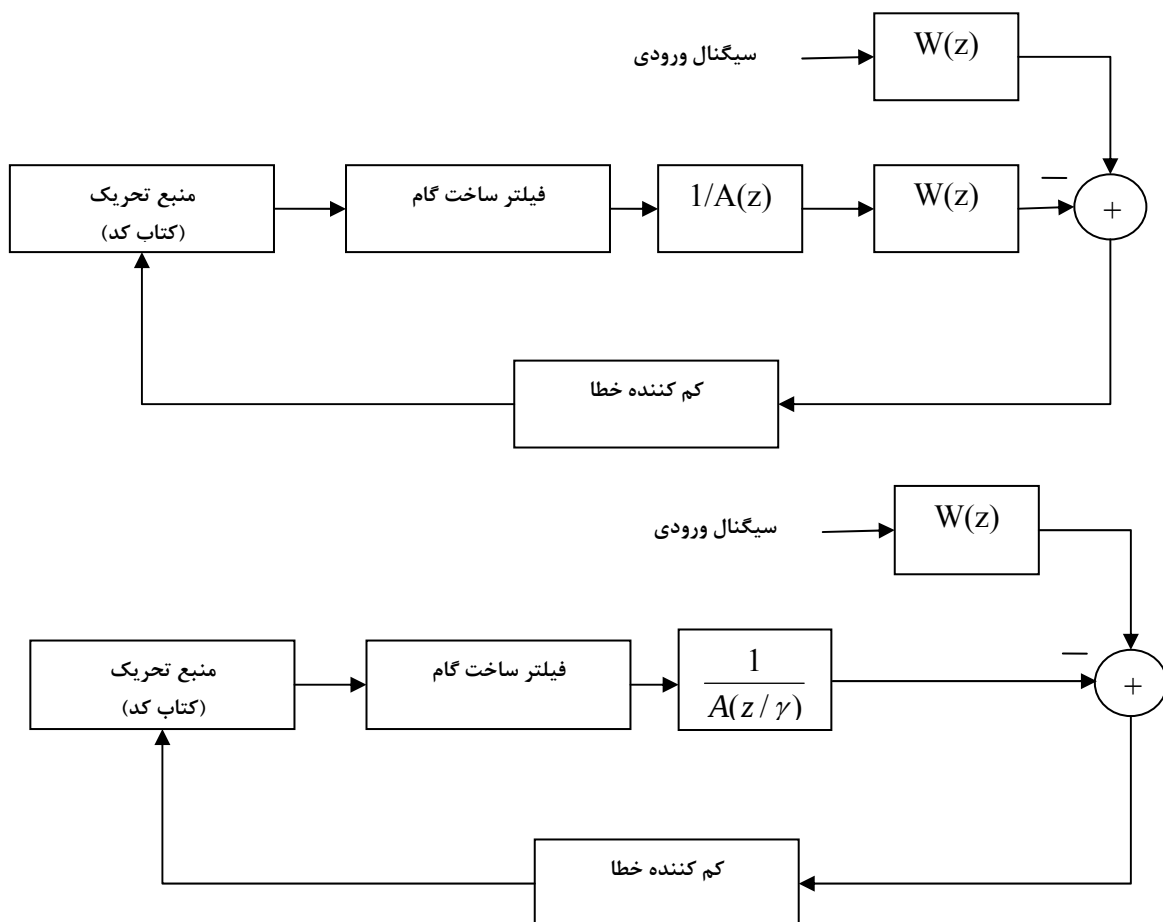
آزمایشات نشان می دهد که برای نمونه گیری با فرکانس ۸ کیلو هرتز مقدار ۰/۸-۰/۹ برای γ مقدار مناسبی است همچنین از آنجا که سیستم فوق یک سیستم خطی است برای کاهش محاسبات کامپیوتری شکل ۳-۴ را می توان به جای آنالیز گر انکدر CELP استفاده نمود. البته باید توجه داشت از آنجا که کلیه بردارهای تحریک موجود در منبع تحریک (کتاب کد) برای آزمون مناسب ترین سیگنال تحریک مورد استفاده قرار می گیرند، زمان صرف شده برای یافتن مناسب ترین سیگنال تحریک در این کد کننده بیشترین زمان تاخیر را به خود صرف می کند.

نمونه ای از اندازه تابع تبدیل فیلتر وزن دهنده خطا به ازای مقادیر مختلف γ در شکل ۳-۳ آورده شده است که در آن مقدار ضرایب LPC برای فریم مورد استفاده شده در این مثال عبارتند از :

$$a_1 = -1.286, a_2 = 1.138, a_3 = -1.047, a_4 = 0.691, a_5 = -0.304, a_6 = 0.373, a_7 = -0.071, \\ a_8 = 0.012, a_9 = 0.048, a_{10} = 0.064$$



شکل ۳-۳. پاسخ فرکانسی فیلتر وزن دار



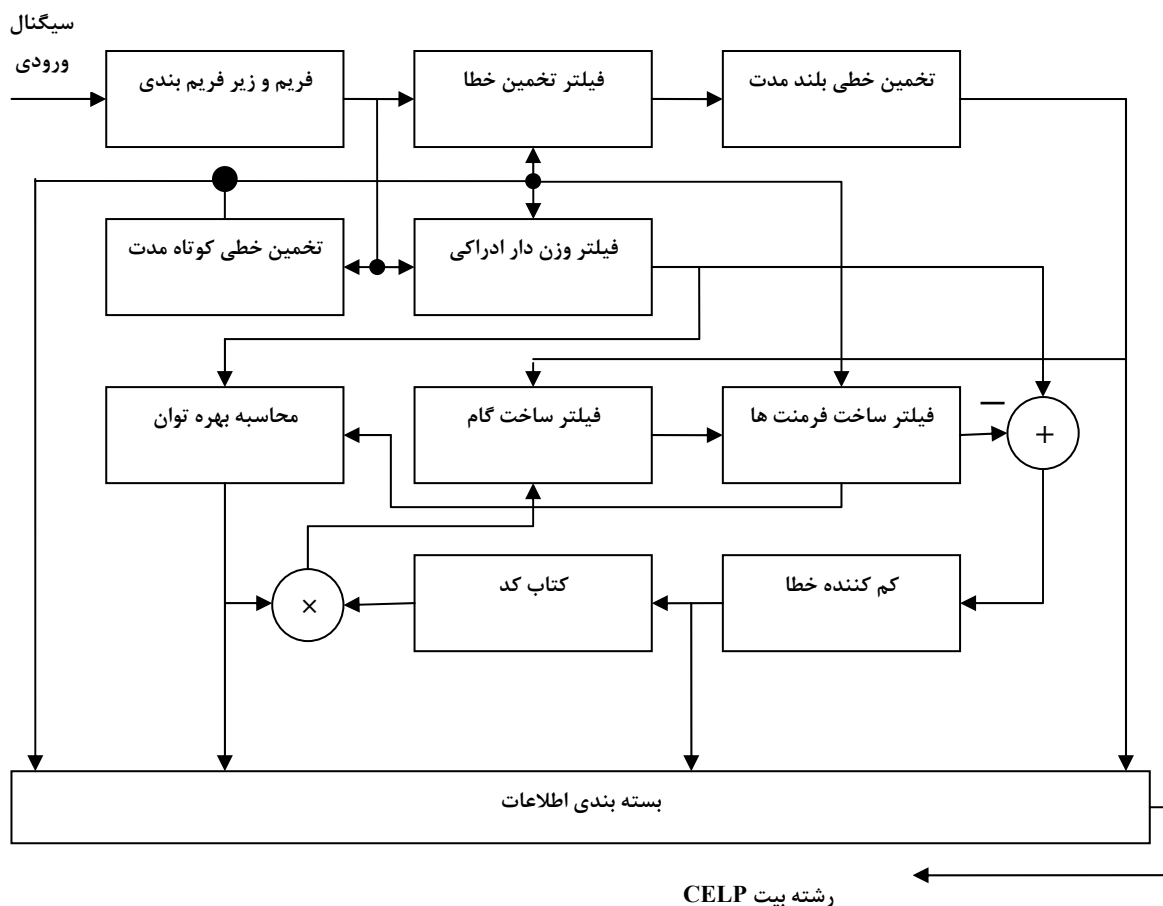
شکل ۳-۴. بلوک کلی نحوه آنالیز در اندکدر CELP برای کاهش محاسبات

۳-۴-۴. انکدر کد کننده CELP

شکل کلی انکدر یک کد کننده CELP در شکل ۳-۵[3]، آورده شده است. مطابق شکل سیگنال PCM ورودی ابتدا به فریم های ۲۰ تا ۳۰ میلی ثانیه و زیر فریم های ۵ تا ۷.۵ میلی ثانیه تقسیم بندی می شود و پردازش کوتاه مدت برای استخراج ضرایب تخمین خطی بر روی هر فریم و سپس تخمین بلند مدت برای استخراج پارامترهای فیلتر ساخت گام بر روی هر زیر فریم انجام می گیرد. بنابر این ضرایب فیلتر ساخت گام، فیلتر ساخت فرمنت ها و فیلتر وزن دار ادراکی در این مرحله بدست می آیند.

سیگنال تحریک که طول آن (تعداد نمونه های آن) برابر یک زیر فریم سیگنال است در مرحله بعد تخمین زده می شود و به ازای هر زیر فریم یک سیگنال تحریک انتخاب می گردد. نحوه انتخاب به این صورت است که هر سیگنال تحریک با بهره تخمین زده شده برای آن به فیلترهای ساخت اعمال شده و هر کدام از این سیگنال ها که مقدار مینیمم مربع خطای آن با سیگنال اصلی کمتر باشد انتخاب گشته و کد مربوط به آن به همراه بهره توان برای دیکدر ارسال می گردد. نحوه انتخاب کلی این سیگنال تحریک در ادامه توضیح داده می گردد.

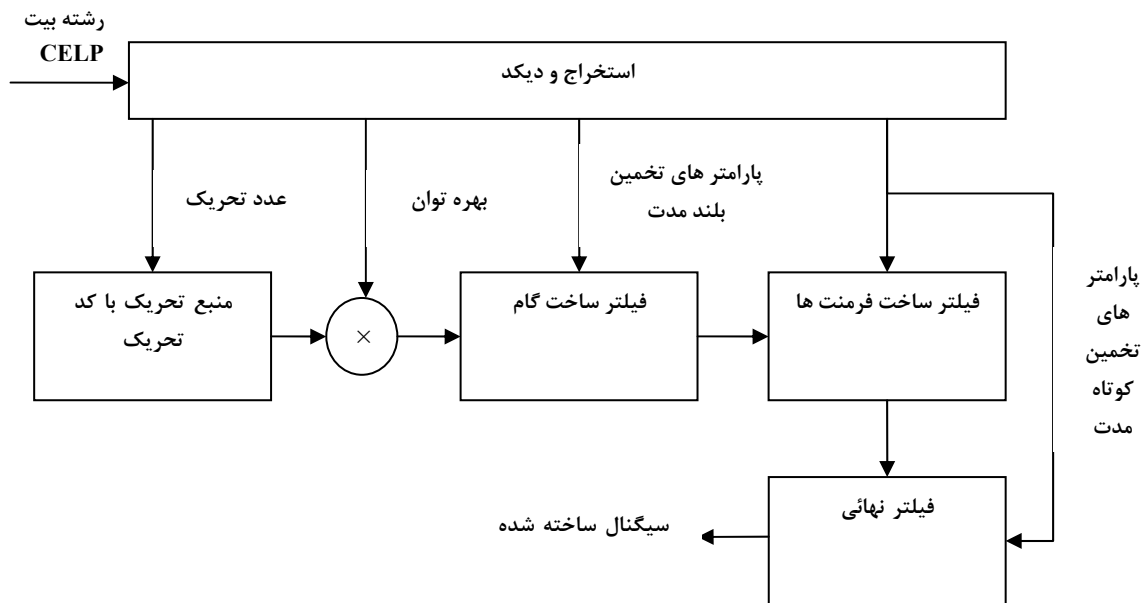
رشته بیت سلپ به طور کلی شامل بیت هائی می شود که با الگوریتم خاصی که پشت سر هم قرار گرفته اند و شامل پارامترهای ضرایب LPC، کد سیگنال تحریک، بهره تقویت سیگنال تحریک، ضریب و دوره تناوب فیلتر تخمین بلند مدت می باشد.



شکل ۳-۵. مدل کلی انکدر یک کد کننده CELP

۳-۴-۵. دیکدر کد کننده CELP

بلوک دیاگرام کلی دیکدر یک کد کننده CELP در شکل ۳-۵ آورده شده است. بطور کلی این بلوک اطلاعات و رشته بیت دریافت شده از انکدر را دریافت نموده و پارامترهای موجود در آن را با الگوریتم موجود استخراج می نماید پس از استخراج و دیکد کردن مطابق مدل CELP تولید سیگنال، سیگنال خروجی را می سازد. در این بلوک از فیلتر نهائی استفاده شده است که در ادامه توضیح داده می شود.



شکل ۳-۶. مدل کلی دیکدر یک کد کننده CELP

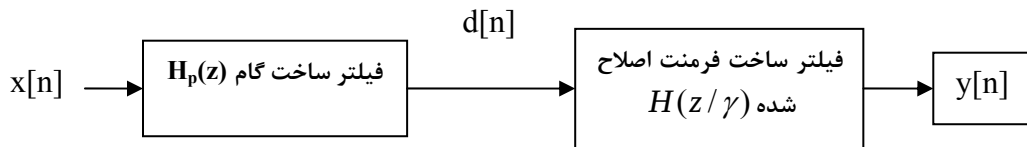
۳-۴-۶. روش یافتن بهترین سیگنال تحریک

در این بخش به شکل جزئی تر نحوه پیدا کردن سیگنال تحریک و محاسبه ضریب بهره آن بیان می گردد. به طور کلی همانطور که قبلا بیان شد روش گام به گام پیدا کردن سیگنال تحریک مناسب به طریق زیر می باشد:

۱- سیگنال ورودی به زیر فریم ها تقسیم شده و توسط فیلتر وزن دار ادراکی شکل دهی می شود.

۲- برای هر بردار سیگنال تحریک منبع مراحل زیر را انجام می دهیم:

- مقدار بهره مربوط به هر بردار را پیدا می کنیم (روش پیدا کردن بهره بیان خواهد شد) و بردار تحریک را در آن ضرب می کنیم.
- سیگنال حاصل از مرحله قبل را از دو فیلتر ساخت گام و فرمنت عبور می دهیم.
- با محاسبه جمع مربع اختلاف نمونه سیگنال خروجی فیلتر ها و نمونه های سیگنال اصلی عبور داده شده از فیلتر ادراکی وزن دار میزان خطا را محاسبه می کنیم.
- بردار مربوط به کمترین میزان خطا به عنوان بهترین بردار انتخاب شده و کد آن را انتخاب می کنیم.



شکل ۷-۳. حالت پشت سر هم فیلتر های ساخت

در عمل برای پاسخ فیلتر ها دو پاسخ حالت صفر و پاسخ به ازای ورودی صفر نیز در بررسی فیلترها مورد توجه قرار می گیرد. با توجه به شکل ۷-۳ معادلات زیر را می توانیم بنویسیم:

$$y[n] = d[n] - \sum_{i=0}^M a_i r^i y[n-i] \quad (۴-۳)$$

$$d[n] = x[n] - b d[n-T] \quad (۵-۳)$$

که در آن M درجه تخمین، a_i ضرایب LPC، b بهره تخمین بلند مدت، T پریود گام می باشد. با توجه به معادلات (۴-۳) و (۵-۳) روابط زیر را می توانیم بنویسیم که در آن r شماره زیر فریم است

$$d_r[n] = d_{r-1}[n+N], \dots, -T \leq n \leq -1 \quad (۶-۳)$$

$$d_r[n] = x[n] - b d_r[n-T], \dots, 0 \leq n \leq N-1 \quad (۷-۳)$$

$$y_r[n] = y_{r-1}[n+N], \dots, -M \leq n \leq -1 \quad (۸-۳)$$

$$y_r[n] = d_r[n] - \sum_{i=0}^M a_i r^i y_r[n-i], \dots, 0 \leq n \leq N-1 \quad (۹-۳)$$

در این مرحله پاسخ حالت صفر و پاسخ ورودی صفر برای هر فیلتر را محاسبه می کنیم:

- پاسخ حالت صفر فیلتر ساخت پریود گام (d1):

$$d_l[n] = 0, \dots, -T \leq n \leq -1 \quad (۱۰-۳)$$

$$d_l[n] = x_r[n] - b d_l[n-T], \dots, -T \leq n \leq N-1 \quad (۱۱-۳)$$

با توجه به معادلات ۱۰-۳ و ۱۱-۳ و با توجه به اینکه در معادله ۱۱-۳ تا زمان T حالت اولیه فیلتر برابر صفر است بطور خلاصه می توانیم بنویسیم:

$$dl_r[n] = x_r[n], \dots, 0 \leq n \leq T-1 \quad (12-3)$$

$$dl_r[n] = x_r[n] - bdl_r[n-T], \dots, T \leq n \leq N-1 \quad (13-3)$$

در این معادلات فرض بر آن شده است که تعداد N از T بزرگتر است و اگر $N < T$ باشد در این صورت خروجی فیلتر با ورودی آن برابر است.

- پاسخ به ورودی صفر فیلتر ساخت پریود گام (d2):

$$d2_r[n] = d_{r-1}[n+N], \dots, -T \leq n \leq -1 \quad (14-3)$$

$$d2_r[n] = -bd_2[n-T], \dots, 0 \leq n \leq N-1 \quad (15-3)$$

- پاسخ نهائی فیلتر ساخت پریود گام برابر است با:

$$d_r[n] = dl_r[n] + d2_r[n], \dots, 0 \leq n \leq N-1 \quad (16-3)$$

- پاسخ حالت صفر فیلتر پریود گام که در حالت پاسخ حالت صفر فیلتر ساخت فرمنت فیلتر می شود (y1):

$$y1_r[n] = 0, \dots, -M \leq n \leq -1 \quad (17-3)$$

$$y1_r[n] = dl_r[n] - \sum_{i=0}^M a_i \gamma^i y1_r[n-i], \dots, 0 \leq n \leq N-1 \quad (18-3)$$

- پاسخ ورودی صفر فیلتر پریود گام که در حالت پاسخ حالت صفر فیلتر ساخت فرمنت فیلتر می شود (y2):

$$y_{2r}[n]=0, \dots, -M \leq n \leq -1 \quad (19-3)$$

$$y_{2r}[n]=d_{2r}[n]-\sum_{i=0}^M a_i \gamma^i y_{2r}[n-i], \dots, 0 \leq n \leq N-1 \quad (20-3)$$

- پاسخ ورودی صفر فیلتر ساخت فرمنت :

$$y_{3r}[n]=y_{r-1}[n+N], \dots, -M \leq n \leq -1 \quad (21-3)$$

$$y_{3r}[n]=\sum_{i=0}^M a_i \gamma^i y_{3r}[n-i], \dots, 0 \leq n \leq N-1 \quad (22-3)$$

- پاسخ نهائی (فیلتر ساخت فرمنت):

$$y_r[n]=y_{1r}[n]+y_{2r}[n]+y_{3r}[n], \dots, 0 \leq n \leq N-1 \quad (23-3)$$

۷-۴-۳. محاسبه میزان خطا و ضریب بهره

همانطور که قبلا بیان گردید برای پیدا کردن بهترین بردار تحریک باید مقدار مینیمم خطا را پیدا کنیم. اگر فرض کنیم تعداد بردارهای موجود در منبع تحریک L و تعداد نمونه های هر بردار برابر N باشد برای منبع تحریک داریم:

$$v^{(l)}[n]; l=0, \dots, L-1; n=0, \dots, N-1 \quad (24-3)$$

که ضریب توان هر کدام از این بردارها با $g^{(l)}$ مشخص می گردد و هر بردار بعد از اعمال ضریب مناسب خود به فیلتر های ساخت اعمال می گردد. با استفاده از معادلات پاسخ حالت صفر و پاسخ به ورودی صفر از قسمت قبل داریم:

$$\varepsilon^{(l)} = \sum_{n=0}^{N-1} (u[n] - y_1^{(l)}[n] + y_2[n] + y_3[n])^2 \quad (25-3)$$

که در آن $\varepsilon^{(l)}$ مشخص کننده میزان خطا به ازای بردار تحریک l و $u[n]$ سیگنال ورودی وزن دار شده می باشد. کمترین مقدار این خطا به ازای بردار و کد تحریک مربوط انتخاب می گردد. اما نکته مهم پیدا کردن ضریب $g^{(l)}$ مناسب می باشد و داریم :

$$y_{10}^{(l)}[n] = \frac{y_1^{(l)}[n]}{g^{(l)}}, \dots, n = 0, \dots, N-1 \quad (26-3)$$

و اگر داشته باشیم:

$$u_0[n] = u[n] - y_2[n] - y_3[n] \quad (27-3)$$

با توجه به دو معادله ۲۶-۳ و ۲۷-۳ خواهیم داشت:

$$\varepsilon^{(l)} = \sum_{n=0}^{N-1} (u_0[n] - g^{(l)} y_{10}^{(l)}[n])^2 \quad (28-3)$$

مقدار خطای فوق به ازای $g^{(l)}$ مناسب باید کمترین باشد بنابراین از معادله ۲۸-۳ نسبت به $g^{(l)}$ مشتق گرفته و برابر صفر قرار می دهیم و مقدار مناسب $g^{(l)}$ برابر است با:

$$g^{(l)} = \frac{\sum_{n=0}^{N-1} (u_0[n] - y_{10}^{(l)}[n])^2}{\sum_{n=0}^{N-1} (y_{10}^{(l)}[n])^2} \quad (29-3)$$

و با استفاده از مقدار فوق مقدار می نیمم خطا برابر است با:

$$\varepsilon^{(l)} = \left(\sum_{n=0}^{N-1} (u_0[n])^2 \right) - p^{(l)} \quad (30-3)$$

که در آن داریم:

$$p^{(l)} = \frac{\sum_{n=0}^{N-1} (u_0[n] - y_0^{(l)}[n])^2}{\sum_{n=0}^{N-1} (y_0^{(l)}[n])^2} \quad (3-31)$$

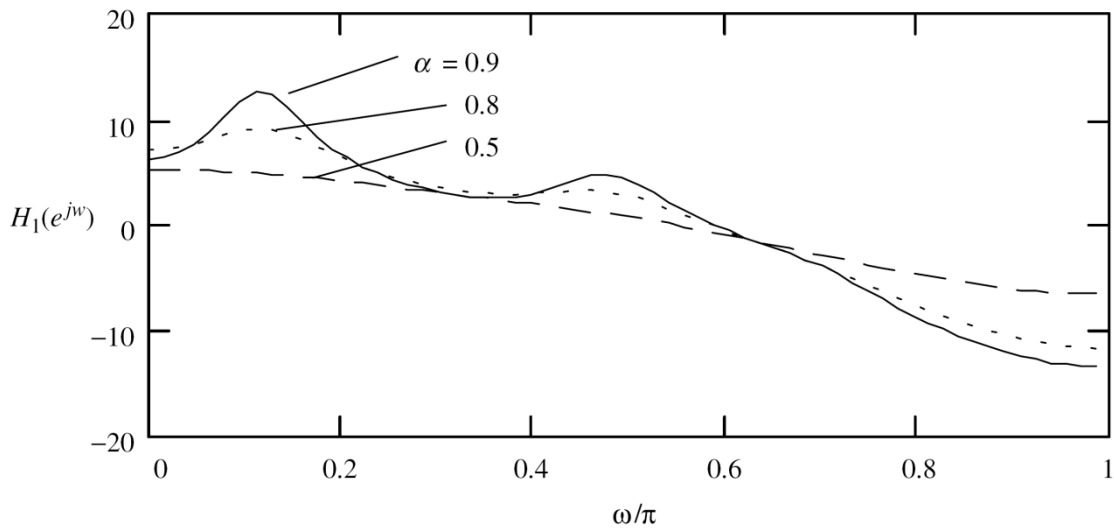
۳-۴-۸. فیلتر نهائی

مطابق شکل ۳-۶، فیلتر نهائی^۱ در انتهای دیکدر برای بهبود کیفیت سیگنال بدست آمده و همچنین کاهش اثر فیلتر وزن دار استفاده شده در انکدر مورد استفاده قرار می گیرد [3]. این فیلتر باید طوری انتخاب گردد که نقاط ماکزیمم و مینیمم طیف اصلی سیگنال را حفظ نماید و همچنین طیف سیگنال در نقاط دیگر فرکانسی را تحت تاثیر قرار ندهد و از آنجا که هدف، ساختن سیگنال مشابه ورودی است، از تابع تبدیل به دست آمده از ضرایب LPC استفاده می شود تا طیف فرکانسی آن شباهت زیادی به طیف سیگنال اصلی داشته باشد.

برای تعیین تابع تبدیل این فیلتر اگر دقیقاً از یک فیلتر ساخته شده از مدل LPC با تابع تبدیل $(H_1(z))$ مطابق معادله ۳-۳۲ استفاده کنیم، با توجه به پاسخ فرکانسی آن به ازای $0 < \alpha < 1$ های که در شکل ۳-۸ آورده شده است مشاهده می شود که این تابع ماهیتی پائین گذر دارد و طیف فرکانس بالای سیگنال صحبت موجود که مربوط به اثر لب ها می باشد را تضعیف می کنند ولی برای نقاط ماکزیمم و می نیمم طیف اثر مطلوبی دارد.

$$H_1(z) = \frac{1}{1 + \sum_{i=1}^M a_i \alpha^i z^{-i}} \quad (3-32)$$

¹ Post filter



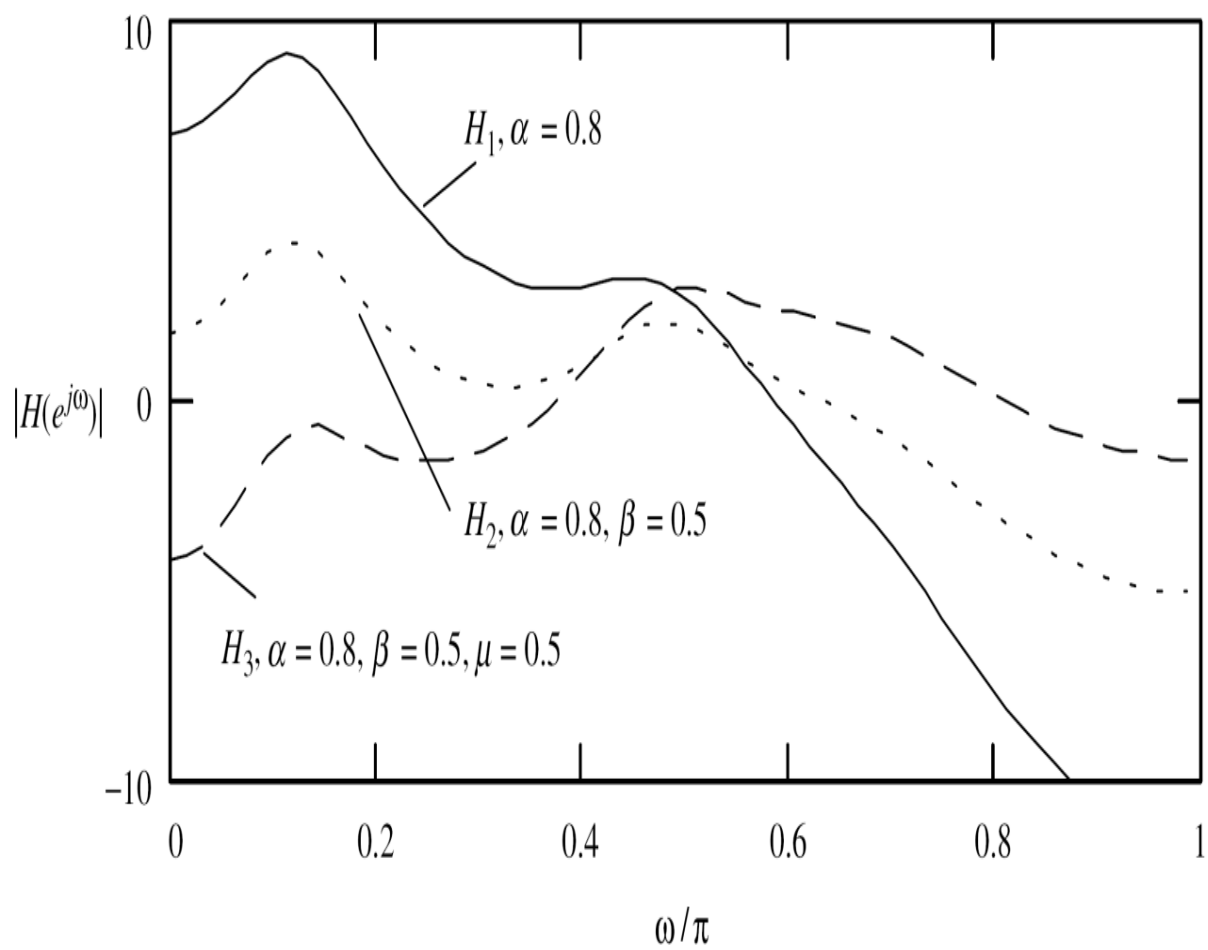
شکل ۳-۸. اندازه دامنه یک مدل تمام قطب برای فیلتر نهائی

برای بهبود این اثر فیلتر $H_1(z)$ ابتدا از یک مدل صفر و قطب با تابع تبدیل ($H_2(z)$) مطابق معادله ۳-۳۳ استفاده گردید که در آن $0 < \beta < \alpha < 1$ و مقادیر α و β توسط نتایج حاصل از گوش دادن به دست می آمد.

اگر چه با افزودن صفر ها به تابع اولیه اثر تضعیف فرکانس های بالا کاهش یافت اما مقداری تضعیف دیگر وجود داشت که برای جبران آن در نهایت از تابع تبدیل ($H_3(z)$) مطابق معادله ۳-۳۴ به عنوان فیلتر نهائی استفاده گردید که طیف فرکانسی مربوط در شکل ۳-۹ آورده شده است.

$$H_2(z) = \frac{1 + \sum_{i=1}^M a_i \beta^i z^{-i}}{1 + \sum_{i=1}^M a_i \alpha^i z^{-i}} \quad (3-33)$$

$$H_3(z) = (1 + \mu z^{-z}) \frac{1 + \sum_{i=1}^M a_i \beta^i z^{-i}}{1 + \sum_{i=1}^M a_i \alpha^i z^{-i}} \quad (3-34)$$



شکل ۳-۹. پاسخ فرکانسی مدل صفر-قطب فیلتر نهایی

فصل ۴ کد کننده‌های سلپ

معروف

طی سال‌های اخیر تلاش‌های زیادی برای بهبود کد کننده سلپ انجام گرفته است که منجر به معرفی انواع زیادی از کد کننده‌های خانواده سلپ شده است که تعداد آنها متنوع می‌باشد. با توجه به گستردگی تنوع سلپ‌ها ارائه تمام این کد کننده‌ها در این پایان‌نامه مقدور نبود و برای آشنائی مختصر با برخی از این سلپ‌ها ۳ مورد از این کد کننده‌ها در این فصل به طور مختصر معرفی می‌گردد.

۴-۱. کد کننده سلپ با تاخیر زمانی کم^۱

۴-۱-۱. مقدمه

مقدار فاصله زمانی بین ورود سیگنال اصلی به انکدر و ساخته شدن سیگنال در خروجی دیکدر هنگامی که خروجی انکدر مستقیماً به ورودی دیکدر متصل شده باشد را تاخیر کد کردن می‌گویند. در سیستم‌های PCM به علت ارسال مستقیم نمونه‌ها این مشکل زیاد به چشم نمی‌خورد ولی در کد کننده‌هایی مانند CELP به دلیل پردازش فریم به فریم، مراحل پردازشی مختلف، مخصوصاً زمان مربوط به انتخاب سیگنال تحریک مناسب، این مسئله اهمیت ویژه‌ای پیدا می‌کند. لازم به ذکر است در مکالمه‌های دو طرفه این مسئله بسیار مهم است چرا که اگر زمان این تاخیر بالاتر از 150ms [3]، باشد سیستم‌های صحبت دو طرفه را دچار اختلال می‌کند. کد کننده معروفی که برای کم کردن تاخیر کد کردن مورد استفاده قرار می‌گیرد کد کننده سلپ با تاخیر زمانی کم (LD-CELP) می‌باشد.

باید یاد آور شد که برای بدست آوردن این ویژگی (تاخیر کم) باید مقداری از ویژگی‌های مطلوب کد کننده معمولی مانند نرخ بیت پائین، کیفیت خوب، هزینه محاسبات کامپیوتری و مقاومت در برابر خطای کانال را هزینه نمائیم.

زمان تاخیر در این کد کننده ۱.۲۵ تا ۱.۸۷۵ میلی ثانیه می‌باشد که نسبت به سلپ معمولی که زمان تاخیر ۲۰ تا ۳۰ میلی ثانیه دارد بسیار کمتر می‌باشد [3].

^۱Low-Delay CELP Coder (L-D CELP)

۴-۱-۲. روش های مختلف کاهش تاخیر زمانی در LD-CELP coder

- کاهش طول فریم ها به ۲۰ نمونه در هر فریم : با اتخاذ این تدبیر زمان مورد نیاز برای پردازش هر فریم کاهش می یابد چرا که مدت زمان کمتری برای بافر شدن نمونه ها نیاز داریم.
- استفاده از تخمین خود همبستگی بازگشتی : در این روش برای محاسبه ضرایب LPC از روش پنجره chen برای محاسبه مقادیر خود همبستگی (استفاده از تداخل پنجره ها) استفاده می گردد [3].
- استفاده از روش تخمین خارجی : از آنجا که طول فریم ها کوچک است می توان از این روش برای تخمین پارامتر های LPC استفاده نمود.
- استفاده از سیگنال ساخته شده در خروجی برای تخمین ضرایب خطی : در کد کننده های معمولی از سیگنال ورودی نمونه گیری شده و ضرایب LPC از آنها استخراج شده و به وسیله رشته بیت برای دیکدر فرستاده می شوند ولی در LD-CELP ضرایب LPC از روی سیگنال ساخته شده قبلی به روش تخمین خارجی استخراج شده و مورد استفاده قرار می گیرد.
- استفاده از ضرایب LPC با درجه بالا (درجه ۵۰) : از آنجا که ضرایب LPC در دیکدر تخمین زده می شود و دیگر احتیاجی به تخصیص بیت برای ارسال آنها نداریم بنابر این با بالا بردن درجه تخمین کیفیت سیگنال ساخته شده نیز بهتر می گردد.
- استفاده از سیگنال ساخته شده در خروجی برای محاسبه میزان بهره تحریک لازم : در این روش بهره سیگنال تحریک نیز با استفاده از سیگنال ساخته شده قبل تخمین زده می شود.

۴-۱-۳. انکدر کد کننده سلپ با تاخیر زمانی کم

شکل کلی انکدر سلپ کدر با تاخیر زمانی کم در شکل ۴-۱ [3] آورده شده است. در این کد کننده برای شروع کار، انکدر نیاز به ۵ نمونه اول سیگنال اصلی (یک زیر فریم) دارد تا با ارسال پارامتر های این نمونه ها برای دیکدر و با روش تخمین خارجی بتواند کار ساخت سیگنال را انجام دهد. در مرحله اول ارسال، ۵ نمونه اول سیگنال گرفته شده و پارامتر های LP آن محاسبه می گردد تا از طریق روش

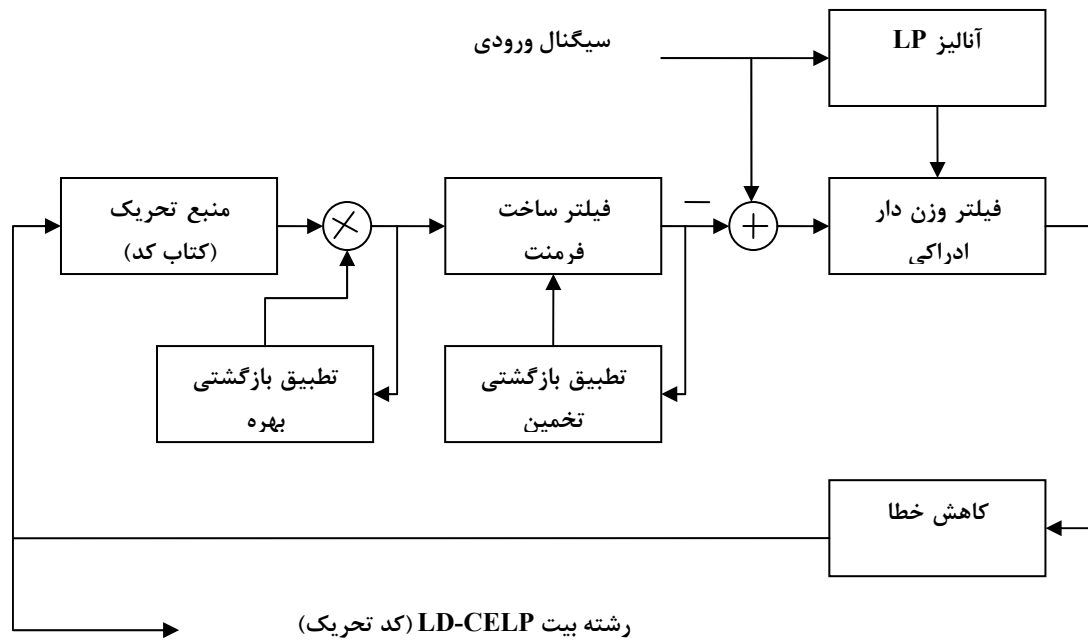
آنالیز - ساخت کد سیگنال تحریک مناسب انتخاب می شود. این کد کننده کار کردی مانند کد کننده CELP دارد با این تفاوت که از روش تخمین خارجی از روی سیگنال ساخته شده قبلی برای ساخت سیگنال استفاده می کند.

پارامترهای فیلتر وزن دار (۱۰ ضریب LPC) در انکدر به ازای هر فریم و از روش تخمین خارجی سیگنال اصلی محاسبه می گردد. پارامترهای LPC فیلتر ساخت (۵۰ ضریب) نیز با روش تخمین خارجی از روی سیگنال ساخته شده قبلی در دیکدر محاسبه می گردند و پارامتر بهره سیگنال تحریک نیز به ازای هر زیر فریم و با استفاده از سیگنال قبلی ساخته شده محاسبه می گردد.

۴-۱-۴. دیکدر کد کننده سلپ با تاخیر زمانی کم

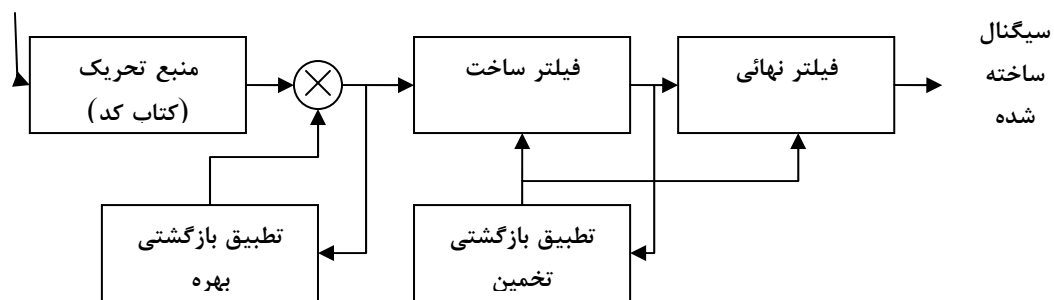
شکل کلی دیکدر سلپ کدر با تاخیر زمانی کم در شکل ۴-۲ [3]، آورده شده است. همانطور که در قسمت قبل بیان گردید این دیکدر برای ساخت سیگنال کد سیگنال تحریک را از انکدر دریافت کرده و با استفاده از مقادیر اولیه سیگنال که به آن داده می شود از طریق تخمین خارجی و با توجه به شکل ۴-۲ سیگنال را تولید می نماید.

فصل چهارم - کد کننده‌های سلپ معروف



شکل ۴-۱. شکل کلی انکدر سلپ کدر با تاخیر زمانی کم

رشته بیت دریافتی



شکل ۴-۲. شکل کلی دیکدر سلپ کدر با تاخیر زمانی کم

۴-۲. کد کننده سلپ با نرخ بیت متغیر و منبع کنترل شده^۱

۴-۲-۱. مقدمه

در بسیاری از کد کننده‌های سلپ معمولی به تمامی سیگنال فریم‌های یکسانی تعلق می‌گیرد. در حالی که در گفتگوی دو طرفه یکی از گوینده‌ها در ۶۳٪ مواقع ساکت می‌باشد و هیچ سخنی نمی‌گوید و این بدین معنی است که می‌توان تعداد بیت تخصیص داده شده در این زمان را کاهش داد. همچنین با توجه به رنج دینامیکی سیگنال صحبت می‌توان تعداد بیت‌های تخصیص داده شده را کنترل نمود.

این بخش بر مبنای استاندارد TIA IS96 [3]، که برای کد کننده سلپ با نرخ بیت قابل تغییر استفاده می‌شود استوار است. در این استاندارد از انرژی سیگنال و نویز زمینه برای تعیین نرخ بیت تخصیص داده شده است، استفاده می‌گردد.

۴-۲-۲. بیت دهی قابل تغییر

چون اساس این کد کننده بر مبنای نحوه تخصیص بیت است، الگوریتم این تخصیص بسیار مهم است و باید تا حد امکان درست طراحی شود. در استاندارد TIA IS96 به طور معمول از چهار نرخ بیت زیر به ازای ۲۰ میلی ثانیه (۱۶۰ نمونه) استفاده می‌گردد [3]:

- نرخ بیت کامل: 8.55 kbps
- نرخ بیت نیمه: 4 kbps
- نرخ بیت یک چهارم: 2 kbps
- نرخ بیت یک هشتم: 0.8 kbps

معمولاً سیگنال صحبت در نواحی موجود با نرخ بیت کامل و در حالت بی‌صدایی و یا نویز زمینه با نرخ بیت کمتری کد می‌گردد.

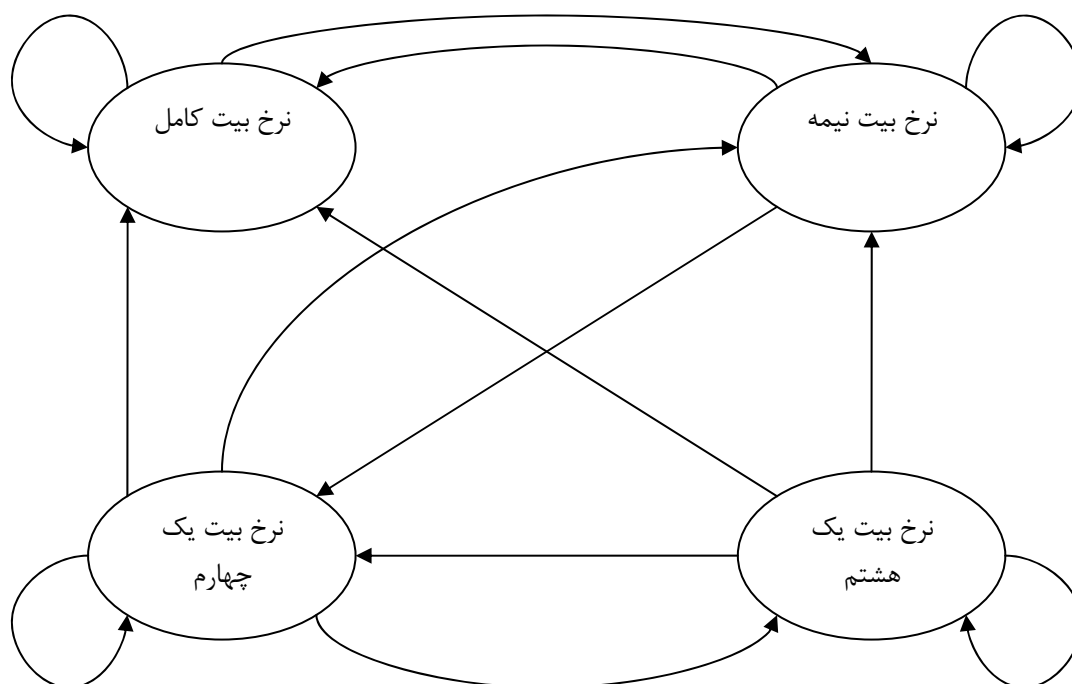
این کد کننده ساختاری مشابه با کد کننده سلپ معمولی دارد با این تفاوت که پارامترهای تولید سیگنال در بازه‌های زمانی مشخص تعیین می‌گردند و با توجه به نرخ بیت تخصیص داده شده، زیر فریم‌ها در طول بازه فریم ۲۰ میلی ثانیه به اندازه‌های مختلف تقسیم بندی می‌شوند و دو نوع زیر فریم اصلی در نظر گرفته می‌شود:

¹ Source-Controlled Variable Bit-Rate CELP

۱. زیر فریم پرپود گام که بازه زمانی را دارد که به ازای آن پارامترهای فیلتر ساخت پرپود گام را استخراج می کنند.
۲. زیر فریم کد تحریک که به ازای آن سیگنال تحریک مناسب استخراج می گردد.

۳-۲-۴. قانون تخصیص بیت

یک الگوریتم تغییر پذیر برای تعیین نرخ بیت هر فریم مورد استفاده قرار می گیرد. این روش یک الگوریتم مبتنی بر تخمین نویز زمینه است و انتخاب نرخ بیت برای هر فریم بر مبنای اختلاف بین انرژی فریم سیگنال اصلی و انرژی نویز زمینه می باشد. شکل ۳-۴ نحوه انتخاب نرخ بیت را نشان می دهد. مطابق این شکل برای هر فریم دریافتی نرخ بیت تخصیص داده شده یا باید به سطوح بالاتر برود یا در همان سطح باقی بماند و یا اینکه به یک سطح پایین تر مجاز است برود و هیچگاه نمی تواند به چند سطح پایین تر از یک فریم به فریم بعدی کاهش پیدا کند.



شکل ۳-۴. دیاگرام حالت برای تخصیص نرخ بیت

۴-۳. سلپ جبری^۱

۴-۳-۱. مقدمه

این نوع سلپ برای کاهش هزینه های محاسباتی سلپ استاندارد ارائه گردید [28]. همانطور که بیان شد بیشترین زمان تاخیر در انکدر سلپ معمولی مربوطه به پیدا کردن کد تحریک مناسب می باشد به همین منظور تحقیقات زیادی برای پیدا کردن ساختاری ساده برای جستجو و یافتن کد تحریک مناسب انجام گرفته است. برای مثال کد کننده IS54 بر مبنای بردارهای کوچک پایه که کل فضای برداری منبع تحریک را اشغال می کند استوار است.

کد کننده سلپ جبری بر این مبنا استوار است که دیگر نیازی نیست تمامی سیگنال های تحریک موجود را در حافظه منبع ذخیره شود. بنابراین حافظه زیادی برای ذخیره این کدها نیاز نمی باشد و در هزینه ها صرفه جویی می گردد. ایده این کد کننده در سال ۱۹۸۷ توسط آقای Lambine و همکاران مطرح گردید و از آن پس تحقیقات زیادی برای بهبود آن انجام گرفت که در این میان پنج استاندارد از آن استخراج گردید [3] که عبارتند از:

- ITU-T G.723.1 Multipulse Maximum Likelihood Quantization (MP-MLQ)/ACELP (1995).
- ITU-T G.729 Conjugate Structure (CS)-ACELP (1995).
- TIA IS641 ACELP (1996).
- ETSI GSM Enhanced Full Rate (EFR) ACELP (1996).
- ETSI Adaptive Multirate (AMR) ACELP (1999).

واژه جبری در این کد کننده به معنی استفاده از روش جبری ساده یا روش های ریاضی برای تولید بردارهای تحریک می باشد. کد کننده ای که بر این مبنا استوار است G.729 CS-ACELP می باشد و در اصل برای سیستم های ارتباط شخصی، سیستم های ماهواره دیجیتال و بطور کلی سیستم ها گفتاری که صحبت در آن بصورت بسته های دیجیتال فرستاده می شود استفاده می شود [3].

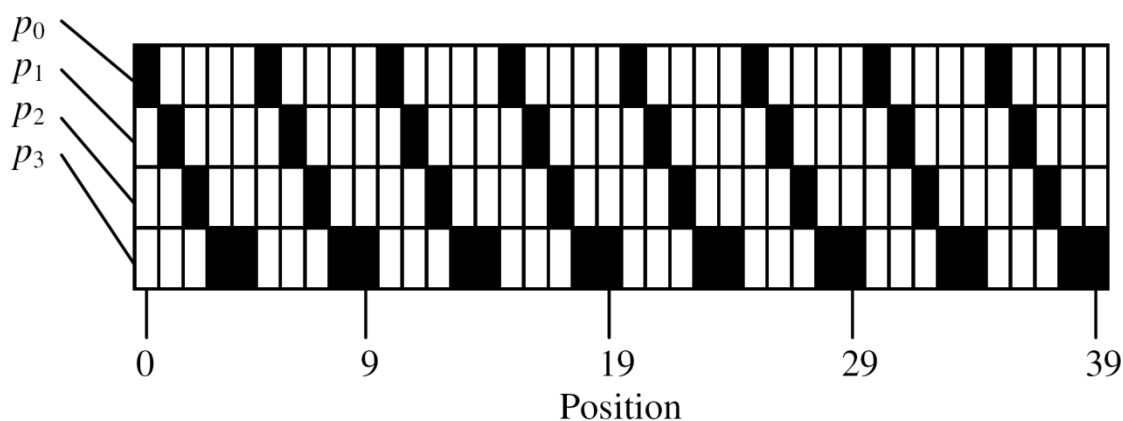
¹ Algebraic CELP

۴-۳-۲. ساختار منبع تحریک کننده سلپ جبری

همانطور که در سلپ معمولی بیان گردید منبع تحریک ثابت شامل بردارهایی است که به عنوان ورودی فیلترهای ساخت استفاده می گردد و مناسب ترین بردار تحریک در انکدر برای هر زیر فریم سیگنال انتخاب می گردد. کد کننده G.729 ساختار جبری و ریاضی برای منبع تحریک ثابت دارد و بر مبنای طراحی جایگشت تک پالس بین سیگنال^۱ استوار است. در این کد کننده هر بردار دارای ۴ مقدار غیر صفر می باشد. هرپالس می تواند مقدار ۱ یا -۱ داشته باشد و می تواند جایگاهی مانند شکل ۴-۴ [3]، داشته باشد. هر بردار تحریک شامل ۴۰ نمونه می باشد که معادل زیر فریم می باشد و بردار تحریک $v[n]$ از جمع ۴ پالس مطابق معادله زیر تولید می شود:

$$v[n] = \sum_{i=0}^3 p_i[n] = \sum_{i=0}^3 s_i \delta[n - m_i]; n = 0, \dots, 39; \quad (1-4)$$

که در آن $s_i = \pm 1$ علامت پالس و m_i مختصات پالس می باشد بنابر این هر پالس ۱ بیت برای علامت، برای p_0, p_1, p_2 هر کدام ۳ بیت برای نشان دادن موقعیت و برای p_3 ، ۴ بیت مورد نیاز است بنابر این برای نشان دادن کل کتاب کد ۱۷ بیت مورد نیاز می باشد [3].



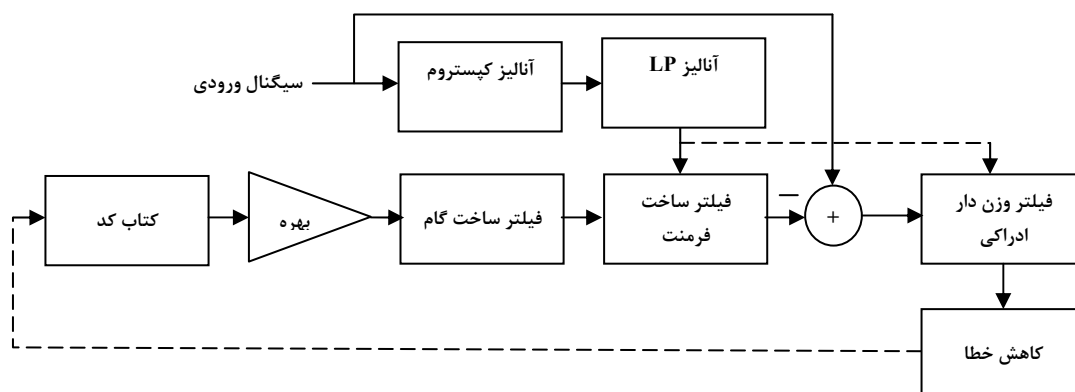
شکل ۴-۴. مختصات پالس های کد کننده جبری در استاندارد G.729 (مستطیل های مشکی)

¹ Interleaved Single Pulse Permutation Design

۴-۴. کد کننده سلپ بهبود یافته با استفاده از آنالیز کپستروم^۱

این کد کننده کار کردی همانند کد کننده سلپ معمولی دارد با این تفاوت که از آنالیز کپستروم برای کاهش نرخ بیت کد کننده سلپ استفاده می کند در این روش ابتدا آنالیز کپستروم بر روی سیگنال ورودی اعمال شده و سپس ضرایب کم ارزش کپستروم حذف و ضرایبی که ارزش بیشتری دارند برای کد کردن در کد کننده سلپ مورد استفاده قرار می گیرند و به این ترتیب نرخ بیت ارسالی توسط کد کننده کاهش می یابد. شکل ۴-۵ بلوک دیاگرام تولید سیگنال در این کد کننده را نشان می دهد [8]. در دیکدر این کد کننده نیز کپستروم معکوس بر روی ضرایب کپستروم ارسال شده از دیکدر اعمال شده و سیگنال گفتار بازسازی می شود.

بطور کلی اساس کار این کد کننده بر این مبنا استوار است که در حوزه کپستروم اطلاعات لوله صوتی در فرکانس های پایین و اطلاعات مربوط به سیگنال تحریک و دوره تناوب در فرکانس های بالا قرار دارند و معمولاً فرکانس های میانی اطلاعات زیادی ندارند و با توجه به این نکته می توان از حذف مولفه های فرکانسی میانی برای کاهش نرخ بیت استفاده نمود.



شکل ۴-۵. مدل تولید سیگنال در کد کننده سلپ بهبود یافته با استفاده از آنالیز کپستروم

¹ MODIFIED CELP CODER USING CEPSTRAL ANALYSIS

فصل ۵ پردازش شهودی و تخمین خطی ادراکی (PLP)

۵-۱. مقدمه

برای فهمیدن اصول کد کردن سیگنال صحبت علاوه بر اصول ساخت سیگنال دانستن تئوری سیستم شنوایی انسان نیز ضروری است چرا که برای دیکد کردن، می توان اطلاعات اضافی که توسط گوش انسان تشخیص داده نمی شود را حذف نمود.

میدانیم که ویژگی های متنوعی در سیگنال بدست آمده دخیل هستند و این حقیقت تاثیر زیادی در فشرده سازی صوت و صدا دارد چرا که این گونه پردازش می تواند با کد کننده های سیگنال صحبت ادغام شود و نرخ بیت مورد نیاز برای ساخت سیگنال با کیفیت را کاهش دهد. در این فصل نحوه کارکرد گوش در فرکانس های مختلف، باند های مهم فرکانسی در شنوایی انسان، مینیمم مقدار صوت برای شناسایی از سکوت و محیط نویزی و در نهایت کد کننده شهودی سیگنال صحبت و الگوریتم PLP معرفی میگردد.

۵-۲. نحوه کارکرد غشای پایه^۱ به عنوان آنالیز گر طیف

غشای پایه عضو اصلی گوش داخلی می باشد چرا که لرزش صوت باعث تحریک و حرکت غشای پایه در اثر عبور از گوش میانی می شود و این تحریک باعث تحریک سلول های مو می شود که تحریک این سلول ها باعث تولید پالس هایی توسط فیبر های عصبی و شنوایی شده و صدا حس می گردد. در سال ۱۹۵۴ آقایان Ohm and Von Helmholtz برای اولین بار نشان دادند که غشای پایه مانند یک آنالیز گر طیف عمل می کند. در سال ۱۹۶۰ آقای Von Békésy این تئوری را گسترش داد و نشان داد که غشای پایه به صورت ناحیه ای لرزش می کند و محل لرزش متناسب با فرکانس سیگنال صوت تحریک کننده می باشد. Von Békésy ثابت کرد که غشای پایه یک آنالیز گر طیف است و ثابت کرد که این غشا آرایه ای از فرکانس های تشدید منظم نیست بلکه یک خط انتقال غیر خطی (بصورت لگاریتمی) محدود با رزولوشن فرکانسی مشخص می باشد. آزمایش های بیشتر نشان داد که این رزولوشن فرکانسی محدود با باند های اساسی مشخص می شوند [6].

¹ Basilar Membrane

۵-۳. باندهای اساسی^۱

به عبارت ساده باندهای اساسی به محدوده فرکانسی یا دسته فرکانسی گفته می شود که در آنها صدا با کیفیت یکنواخت درک می شود. این فرکانس ها با آزمایش بدست می آیند و در حقیقت باندهای فرکانسی اساسی بازه ای از فرکانسهای خاص می باشد که در آن تغییرات صدا به ازای سیگنال خاص احساس نمی گردد و در صورت وجود نویز مشخص در این بازه فرکانسی سطح شنیداری سیگنال تغییر نمی کند مگر اینکه به باند اساسی فرکانسی بعدی برویم.

پدیده اشباع برای این باندهای اساسی به معنی پر شدن از صدا می باشد به طوری که اگر اصوات با سطح سیگنال پائین تر از سطح مشخص شده به این باند افزوده شود شنیده نخواهند شد و با توجه به این موضوع می توان سیگنال هایی که در هر باند قابل شنیده شدن نیستند حذف نمود و تنها محدوده سیگنال های قابل شنیدن توسط گوش انسان را برای کد نمودن استفاده نمود.

آزمایشهای زیادی برای تعیین باند های فرکانسی اساسی صورت گرفته است برای مثال می توان این باند ها را بر اساس تابعی از فرکانس مرکزی هر باند نشان داد که نتیجه یکی از این آزمایشات برای شنوایی گوش در حالت تک صدایی در شکل ۵-۱ [6] آورده شده است. همانطور که دیده می شود به ازای فرکانس مرکزی 500hz به بالا پهنای باند اساسی آن فرکانس نیز به طور خطی با افزایش لگاریتمی فرکانس مرکزی افزایش می یابد. این شکل پایه و اساس حوزه مل^۲ و بارک^۳ که هر دو برای تخصیص عدد ثابت برای هر باند اساسی به وجود آمده اند، می باشد. حوزه بارک به گونه ای نرمالیزه شده است که به ازای هر باند حساس یک بارک مشخص داشته باشیم. این دو حوزه به گونه ای طراحی شده اند که به طور لگاریتمی با افزایش فرکانس افزایش می یابند. به طور معمول برای هر باند فرکانسی اساسی یک بارک مشخص و 150-200 گروه مل وجود دارد. رابطه بین مل و بارک تقریباً خطی است و این تقریبی بودن به علت این موضوع است که نتایج بدست آمده از آزمایشات مختلف تقریبی است و دقیق نمی باشد. جدول ۵-۱ رابطه بین بارک، مل و فرکانس را نشان می دهد.

¹ Critical Bands

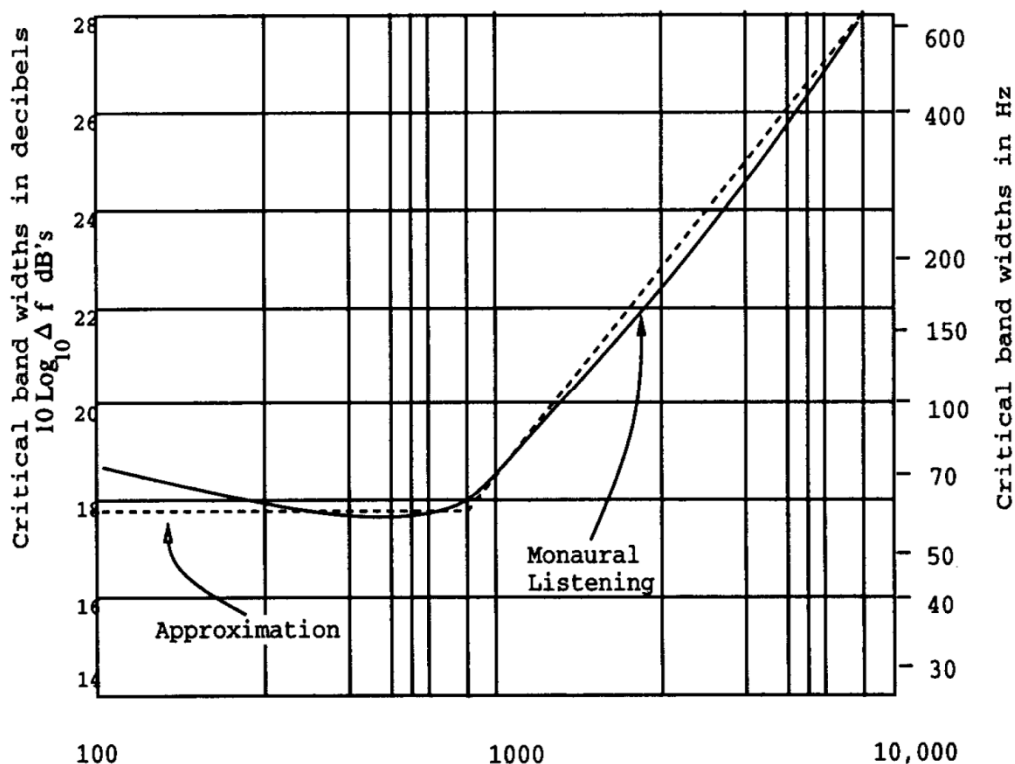
² Mel

³ Bark

۴-۵. سطح تشخیص صدا

سطح تشخیص یا شنیده شدن صدای سیگنال صوت مشخص به میزان فشار مینیمم موثر صدا برای تحریک سیستم شنوایی بدون حضور نویز گفته می شود که اغلب اوقات با واحد دسی بل بر ۰.۰۰۰۲ میکرو بار بیان می گردد.

سطح شنوایی در استاندارد آمریکائی برای صداهای خالص با یک شنونده معمولی در محیط بدون انعکاس با گذاشتن گوشی در شکل ۵-۲ [6]، آورده شده است. این شکل بیان گر این موضوع است که سیستم شنوایی انسان های عادی صوتی که قدرت آنها زیر این منحنی قرار می گیرد (اصوات ضعیف تر) را نمی تواند بشنود اما صوتی که قدرت آنها بالای این منحنی باشد (اصوات قوی تر) را می تواند بشنود.

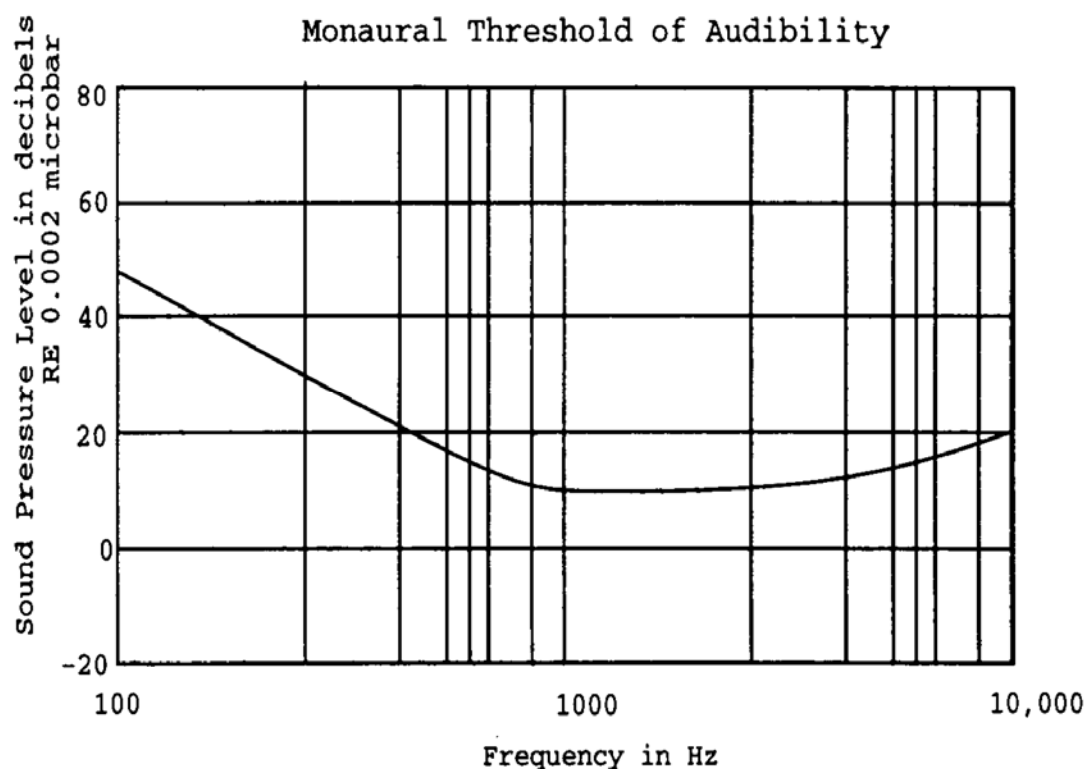


شکل ۵-۱. محدوده فرکانسی باند های اساسی بر اساس تابعی از فرکانس مرکزی باند

فصل پنجم - پردازش شهودی و تخمین خطی ادراکی (PLP)

جدول ۵-۱. رابطه بین بارک و مل و باند های اساسی فرکانسی

Critical Band No. (Barks)	Frequency (Hz)	Mels
1	20-100	0-150
2	100-200	150-300
3	200-300	300-400
4	300-400	400-500
5	400-510	500-600
6	510-630	600-700
7	630-770	700-800
8	770-920	800-950
9	920-1080	950-1050
10	1080-1270	1050-1150
11	1270-1480	1150-1300
12	1480-1720	1300-1400
13	1720-2000	1400-1550
14	2000-2320	1550-1700
15	2320-2700	1700-1850
16	2700-3150	1850-2000
17	3150-3700	2000-2150
18	3700-4400	2150-2300
19	4400-5300	2300-2500
20	5300-6400	2500-2700
21	6400-7200	2700-2850
22	7200-9500	2850-3050



شکل ۵-۲. سطح تشخیص صدا در استاندارد آمریکائی

۵-۵. کد کننده شهودی سیگنال صحبت^۱

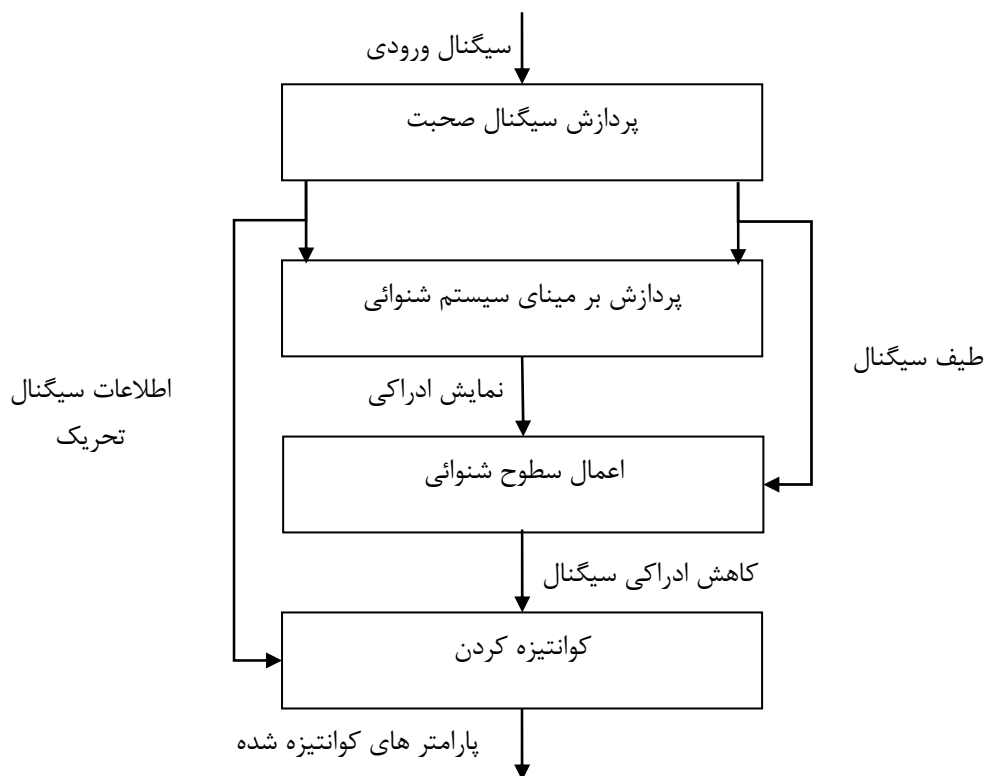
هدف کد کننده شهودی کم کردن بیت های نمایش دهنده سیگنال با حذف قسمت هائی از سیگنال که قابل شنیده شدن با گوش انسان نیستند می باشد. به این ترتیب کیفیت سیگنالی که با گوش انسان شنیده می شود در دو حالت اولیه و ثانویه یکسان می باشد.

۵-۵-۱. کد کننده عمومی ادراکی سیگنال صحبت

بسیاری از مراحل پردازش کد کننده شهودی سیگنال صحبت همانند کد کننده های متداول سیگنال صحبت می باشد با این تفاوت که هدف اولیه این کد کننده تشخیص و تعیین مولفه هائی از سیگنال که قابل شنیده شدن با گوش انسان نیستند می باشد. شکل ۵-۵[6]، بلوک دیاگرام کلی یک کد کننده شهودی سیگنال صحبت را نشان می دهد.

با توجه به شکل سیگنال ورودی برای بدست آوردن مشخصه های فرکانسی لوله صوتی و مشخصه های تحریک (مدل LPC) پردازش می شود و پارامتر های بدست آمده در این مرحله توسط پردازش گری بر مبنای سیستم شنوائی انسان به حوزه ادراکی تبدیل می شوند (تبدل حوزه فوریه به باند های اساسی). در این حوزه پارامتر هائی که قابل حذف شدن می باشند حذف می شوند و تخصیص بیت بر مبنای اهمیت پارامترها انجام می شود. به این ترتیب سیگنال خروجی دارای بیت ها و پارامترهای کمتری نسبت به سیگنال ورودی است اما کیفیت شنیداری آن تغییری نکرده است. مسئله اساسی در این پردازش پیدا کردن سیگنال هائی است که می توان آنها را حذف نمود.

¹ Perceptual Speech Coding



شکل ۵-۳. بلوک دیاگرام کلی کد کننده شهودی سیگنال صحبت

۵-۶. تخمین خطی ادراکی (PLP)

۵-۶-۱. مقدمه

تخمین خطی ادراکی^۱ (PLP) یکی از روش های مشهور و پر کاربرد برای استخراج ویژگی های صوت می باشد و برای آنالیز صوت به کار می رود. این روش اولین بار در سال ۱۹۸۹ توسط HERMAN SKY مطرح گردید. در این روش پارامتر های سیگنال صوت را به نحوی که بیشترین شباهت را به سیستم شنوایی انسان داشته باشد استخراج می کند. ویژگی های استخراج شده باید بر مبنای شنوایی انسان باشند به همین منظور پردازش طیفی سیگنال صحبت بر روی تکه های فرکانسی با دقت و تاکید خاص انجام می گیرد به طوری که در نواحی فرکانسی که حساسیت گوش

^۱ Perceptual Linear Prediction (PLP)

انسان بیشتر است پردازش با دقت و تاکید بیشتری انجام می پذیرد.

هدف از استخراج ضرایب PLP در اکثر کاربرد های امروزی برای تشخیص گفتار اتوماتیک^۱ (ASR) می باشد و در تمامی این سیستم ها میزان دقت استخراج و نحوه استخراج این پارامتر ها بسیار مهم است چراکه اصلی ترین و اولین قسمت در تشخیص صوت استخراج پارامتر های آن می باشد و اگر پارامتر ها کیفیت پائینی داشته باشند قسمت های بعدی نمی توانند به نحو مناسب کار کنند. همچنین ویژگی های استخراج شده باید نسبت به تغییرات کانال مخابراتی، گوینده های مختلف، نویز زمینه و مبدل های مختلف مقاومت کافی را داشته باشند [10].

۵-۶-۲. پردازش کیسترال^۲

سیستم های تشخیص اتوماتیک گفتار امروزی بر پایه مدل سازی لوله صوتی در حالی که اطلاعات سیگنال تحریک جدا شده باشد استوار می باشد. همانطور که قبلا بیان گردید این عمل به خاطر این موضوع می باشد که هر لوله صوتی خصوصیات متمایز کننده ای برای سیگنال های صحبت و ویژگیهای خاص خود را دارد [39].

برای جدا کردن اثر لوله صوتی و سیگنال تحریک آن، از پردازش کیسترال استفاده می شود. به این ترتیب که اگر تابع $g(n)$ به عنوان سیگنال تحریک و $h(n)$ به عنوان پاسخ ضربه لوله صوتی باشد $s(n)$ سیگنال صحبت خروجی می باشد و روابط زیر را برای آن داریم:

$$s(n) = g(n) \otimes h(n) \quad (۱-۵)$$

وبا گرفتن تبدیل فوریه و سپس لگاریتم از طرفین داریم:

$$S(f) = G(f).H(f) \quad (۲-۵)$$

$$\text{Log}(S(f)) = \text{Log}(G(f).H(f)) = \text{Log}(G(f)) + \text{Log}(H(f)) \quad (۳-۵)$$

¹ Automatic Speech Recognition (ASR)

² Cepstral processing

که در این صورت رشته خروجی تفکیکی از سیگنال تحریک و پاسخ ضربه لوله صوتی می باشد. با گرفتن عکس تبدیل فوریه از طرفین معادله بالا این دو قسمت از هم جدا می شوند چرا که خواص لوله صوتی با زمان تغییرات زیادی ندارد و بنابر این فرکانس های پائین استخراج شده مربوط به لوله صوتی است و فرکانسهای بالا مربوط به سیگنال تحریک می باشد [8].

باید توجه نمود، تبدیل کپستروم از اپراتور غیر خطی استفاده می کند، این روش نسبت به نویز جمع شونده حساسیت بیشتری دارد ولی در محیط بدون نویز کارایی مناسبی دارد. همچنین در حضور نویز سفید گوسی، در حالتی که سیگنال به نویز کاهش می یابد توزیع کپستروم باعث می گردد این شکل از حالت گوسی دور گردد [10].

همچنین تابع لگاریتمی می تواند پدیده روی هم افتادگی را بوجود بیاورد. از آنجا که لگاریتم هارمونیک هایی ایجاد می کند توصیه می شود که با افزودن صفرهایی در مکان های مناسب برای بالا بردن سرعت نمونه گیری DFT استفاده شود. به هر حال در سیستم های تشخیص اتوماتیک گفتار از روش کپستروم استفاده می گردد.

۵-۶-۳. شکل دهی سیگنال ورودی

در مرحله اول از الگوریتم استخراج ویژگی ها، ابتدا سیگنال ورودی گرفته شده، نمونه برداری، دیجیتال، فیلتر و فریم بندی می شود.

سیگنال دیجیتال شده با استفاده از یک فیلتر دیجیتال^۱ درجه یک بالا گذر فیلتر می گردد چرا که سیگنال صدا دار بطور طبیعی یک حلقه فرکانس کاهش در آن قسمت دارد و برای یکنواخت کردن سیگنال از یک سیگنال بالا گذر استفاده می شود. همچنین گوش انسان برای فرکانس های بالای 1 kHz حساسیت بیشتری دارد. این فیلتر باعث می شود که بتوانیم اطلاعات ادراکی مهمی از سیگنال اصلی بگیریم. تابع تبدیل فرکانس این فیلتر برابر است با:

$$h_{pre}(z) = 1 + a_{pre} z^{-1} \quad (۴-۵)$$

که معمولا a_{pre} بین ۱- تا ۰/۴- تغییر می کند [10].

بعد از عمل فیلترینگ اولیه، سیگنال صحبت دیجیتالی شده به فریم های کوچک تقسیم بندی می شود. معمولا طول فریم بر مبنای میزان و نوع تغییرات لوله صوتی انتخاب می گردد بطوریکه

¹ Pre-emphasis filter

سیگنال صوت از نظر آماری به ازای هر فریم ساکن یا ساکن گوسی شود. بطور معمول در عمل طول هر فریم 20-40ms انتخاب می شود.

پس از فریم بندی پنجره همینگ برای از بین بردن حالت غیر پیوستگی در مرز فریم ها به هر فریم اعمال می گردد. این ناپیوستگی ها در حالت طیفی ناشی و حالت اعوجاج ایجاد می کند که باعث کاهش محدوده دینامیکی طیفی می گردد که می تواند در تشخیص فرمت های با انرژی کمتر تاثیرگذار و عمل استخراج ویژگی ها را دچار اختلال نماید.

۵-۶-۴. روش استخراج ضرایب PLP

همانطور که بیان گردید روش PLP بر مبنای فیزیولوژی ساختار شنوایی انسان بنا شده است. سیستم شنوایی انسان برای فرکانس های زیر 1 kHz حساسیت کمی دارد و در بازه میانی طیف شنوایی حساسیت زیاد دارد.

مراحل اجرای الگوریتم PLP به صورت مرحله به مرحله در شکل ۵-۶ [7]، آورده شده است که به طور خلاصه هر مرحله توضیح داده می شود:

• آنالیز طیفی^۱

بعد از عبور از فیلتر پیش تاکید و فریم بندی سیگنال برای هر فریم طیف توان زمان کوتاه محاسبه می گردد.

• آنالیز باند های اساسی^۲

طیف توان بدست آمده به حوزه بارک برده می شود تا بر اساس سیستم شنوایی انسان یک سری از فرکانس ها اهمیت بیشتری پیدا کند. تابع انتقال از حوزه فرکانس به حوزه بارک به فرم های زیر امکان پذیر است :

$$F_{bark} = 6 \log(f / 600 + \sqrt{(f / 600)^2 + 1}) \quad (5-5)$$

$$F_{bark} = 13 \tan^{-1}(0.00076 f) + 3.5 \tan^{-1}((f / 7500)^2) \quad (6-5)$$

¹ Spectral Analysis

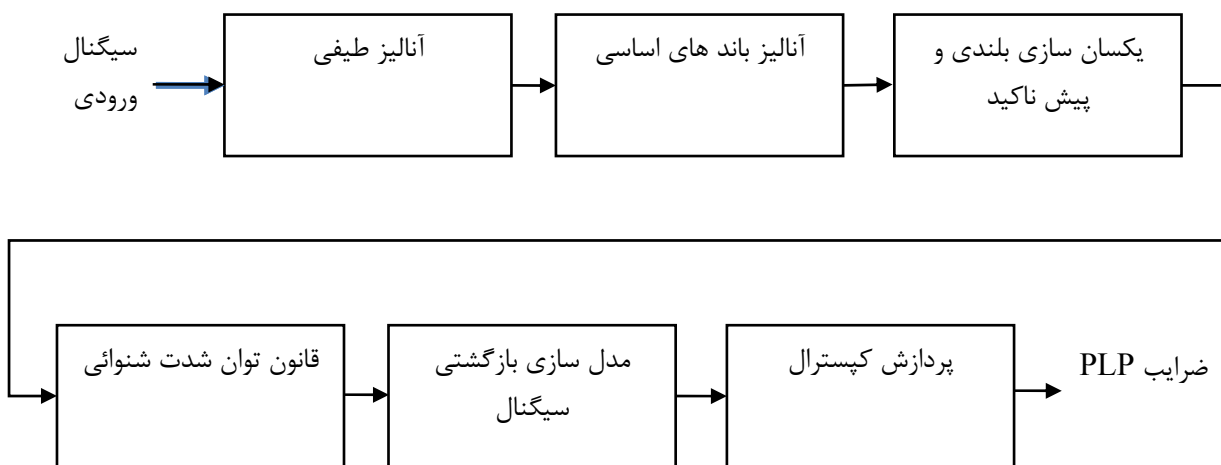
² Critical Band Analysis

که این تابع برای شبیه سازی اثر ماسک کردن فرکانسی می باشد و شکل تابع آن در شکل ۵-۷ آورده شده است. بطور کلی در بیشتر سیستم ها از معادله اول به علت معادله راحت تر استفاده می گردد.

• یکسان سازی بلندی و پیش تاکید^۱

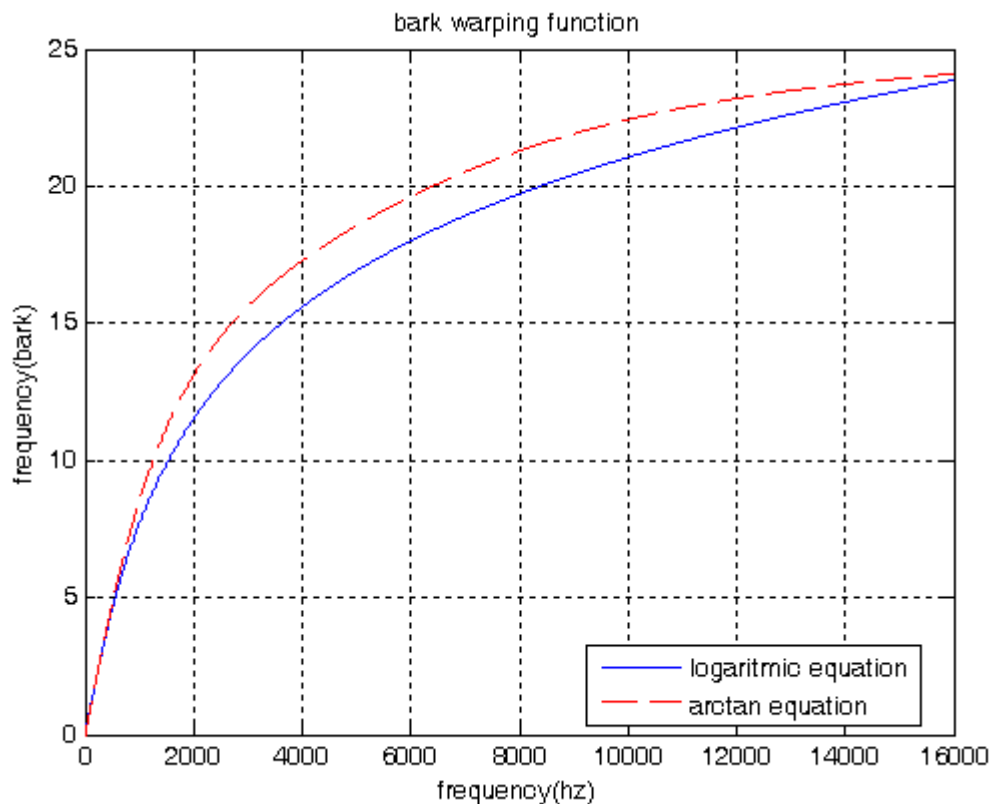
این تابع برای تخمین شنوایی انسان در 40db به کار می رود و در حقیقت کار اصلی آن همسان کردن شدت شنوایی سیگنال ها می باشد و معادله تبدیل آن به فرم زیر است :

$$E(\omega) = \frac{(\omega^2 + 56.8 * 10^6) \omega^4}{(\omega^2 + 6.3 * 10^6)^2 (\omega^2 + 0.38 * 10^9)} \quad (۷-۵)$$



شکل ۵-۴. پردازش PLP

¹ Equal loudness / Pre-emphasis



شکل ۵-۵. تبدیل فرکانسی باند های اساسی

• قانون توان شدت شنوائی^۱

در این مرحله یک تابع برای گرفتن ریشه سوم از طیف موجود به منظور تخمین قانون توانائی شنوائی انسان به سیگنال موجود اعمال می گردد که معادله آن به شکل زیر است:

$$\ell(\omega) = \sqrt[3]{E(\omega)} \quad (۸-۵)$$

• مدل سازی بازگشتی سیگنال^۱

^۱ Intensity loudness Power Law

در این مرحله یک تبدیل معکوس فوریه بر روی سیگنال انجام می گیرد و یک رشته مشابه خود همبستگی تولید می گردد و با استفاده از مقادیر همبستگی ساخته شده و مدل های آماری مدل تمام قطب سیگنال پیدا شده و با استفاده از ضرایب مدل تمام قطب، ضرایب طیفی تخمین خطی که روابط آن در زیر آورده شده است استخراج می گردد:

اگر فرض کنیم مدل تمام قطب سیگنال تابع تبدیلی به فرم زیر داشته باشد:

$$\hat{S}(z) = \frac{A}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (9-5)$$

که در آن تمامی پارامترها (ضرایب تخمین خطی به همراه A) مشخص شده است، ضرایب تخمین طیفی از روابط زیر بدست می آیند:

$$c(n) = \begin{cases} 0 & n < 0 \\ \ln(A) & n = 0 \\ a_n + \sum_{k=1}^{n-1} (k/n) c(k) a_{n-k} & 0 < n < p \\ \sum_{k=n-p}^{n-1} (k/n) c(k) & n > p \end{cases} \quad (10-5)$$

با توجه به معادله های بالا دستگاه فوق می تواند بینهایت جواب داشته باشد ولی در عمل ۱۲-۲۰ ضریب کارائی مناسبی دارد [10].

هارمانسکی ثابت کرد که مزیت الگوریتم PLP در اینست که حساسیت بیشتری به فرکانس اصلی نسبت به پردازش Lp دارد و وابستگی کمتری به گوینده های مختلف دارد [9].

¹ Autoregressive Modeling

فصل ۶ پیاده سازی کد کننده

سلپ استاندارد

۶-۱. مقدمه

در سال ۱۹۸۴ وزارت دفاع ایالات متحده^۱ برنامه ای را شروع نمود تا به وسیله آن ساختار سیستم مخابراتی گفتار را در راستای ساختار کد کننده FS 1015 LPC بهبود بخشد. در بین سال ۱۹۸۸-۱۹۸۹ کد کننده CELP با نرخ بیت 4.8 kbps بوسیله این وزارت خانه و آزمایشگاه بل به طور مشترک توسعه پیدا کرد و معرفی گردید. این کد کننده در سال ۱۹۹۱ بعنوان استاندارد دولتی رسمی در ایالات متحده با نام استاندارد FS 1016 معرفی گردید [3]. باید متذکر شد که علاوه بر ساختار کلی کد کننده CELP که در فصل سوم معرفی گردید، تغییرات و بهبودهایی در این کد کننده اعمال گردید تا پاسخ های بهتری از سیگنال خروجی بدست آید که این تغییرات و نحوه پیاده سازی آن در این فصل بیان می گردد.

همچنین با معرفی این کد کننده، تحول عمیقی در کد کننده های سیگنال صحبت ایجاد گشت و با اینکه نرخ بیتی حدود ۲ برابر کد کننده MELP داشت به سرعت جای را آن را گرفت چرا که کیفیتی تقریباً مشابه با آن داشت و لی پیچیدگی های آن را نداشت.

در این فصل ابتدا معیارهای کارکرد مناسب کد کننده های مختلف به طور خلاصه بیان شده و سپس مختصری راجع به نحوه کارکرد کد کننده CELP استفاده شده در نتیجه گیریهای این پایان نامه و پارامترهای آن بحث می گردد. لازم به یادآوری است که در اجرائی کردن کد کننده CELP از مراجع مختلف بهره گیری شده است تا بتوان کد کننده ای استاندارد را برای کاربردی شدن استفاده نمود به همین دلیل بسیاری از پارامترهای استفاده شده مطابق استاندارد FS 1016 می باشد. باید متذکر شد که علاوه بر ساختار کلی کد کننده CELP در فصل سوم معرفی گردید تغییرات و بهبودهایی در این کد کننده اعمال گردید تا پاسخ های بهتری از سیگنال خروجی بدست آید که این بهبود ها در کد کننده پیاده سازی شده اعمال شده است.

¹ U.S Department of Defense

۲-۶. ارزیابی کیفیت سیگنال ساخته شده

کیفیت سیگنال ساخته شده یکی از مسایل بسیار مهم در کد کننده های سیگنال صحبت می باشد. در سال های اول استفاده از سیستم های مخابراتی محاسبه نسبت سیگنال به نویز (SNR) سیگنال دریافتی در گیرنده یکی از روش های غالب برای محاسبه کیفیت کارکرد سیستم مخابراتی بود که در سیستم های کدینگ شکل موج مانند PCM, ADPCM مورد استفاده قرار گرفت. اما با معرفی کد کننده LPC استاندارد در سال ۱۹۸۴ ملاحظه گردید که مقدار سیگنال به نویز سیگنال ساخته شده عددی بسیار ناچیز می باشد و به همین دلیل احساس نیاز به روش های جدید برای ارزش دادن به قابلیت های و توانمندی کد کننده های صحبت که به صورت پارامتری کار می کردند به وجود آمد. یکی از ساده ترین روش هایی که مورد استفاده قرار گرفت و هم اکنون نیز از آن استفاده می گردد اینست که سیگنال خروجی ساخته شده توسط شنونده های مختلف از نظر کیفیت شنیداری مورد ارزشیابی قرار گیرد و معدل نمره های گرفته شده به عنوان کیفیت سیگنال ساخته شده بعنوان میزان توانمندی کد کننده معرفی گردد.

در این بخش روش های استاندارد بررسی کیفیت ساخت سیگنال های صحبت مورد بررسی قرار می گیرد.

۱-۲-۶. معیار های بررسی کیفیت سیگنال

۱-۱-۲-۶. اندازه گیری سیگنال به نویز

در این روش با داشتن سیگنال اصلی $x[n]$ و سیگنال خروجی ساخته شده $y[n]$ ، مقدار سیگنال به نویز از رابطه ۱-۶ محاسبه می گردد که در آن n بازه اندازه گیری را نشان می دهد و می تواند کل بازه سیگنال یا فریم هایی از آن را در بر بگیرد. این روش در کد کننده های شکل موج کاربرد دارد.

$$SNR = 10 \log \left(\frac{\sum_n x[n]^2}{\sum_n (x[n] - y[n])^2} \right) \quad (1-6)$$

۶-۲-۱-۲. روش های اندازه گیری ذهنی

در این روش، سیگنال ساخته شده صحبت به شنونده ها ارائه می گردد و از آن ها خواسته می شود تا سیگنال شنیده شده را توسط نمره هایی که از نظر خود به آن می دهند ارزش دهی نمایند. این تست به طور معمول در شرایط مختلف انجام می گیرد تا به وسیله آن کیفیت و جنبه های مختلف یک کد کننده مورد ارزیابی قرار گیرد. در این بخش برخی از روش های مشهور برای اجرای این روش ارائه می گردد. در این روش ها هرچه تعداد شنونده ها بیشتر باشد تست معتبرتر است و معمولاً شنونده ها باید بیش از ۱۶ نفر باشند.

الف) درجه بندی گروهی مطلق^۱

در این تست از شنونده خواسته می شود که برای کیفیت سیگنال ساخته شده یکی از ۵ گزینه موجود در جدول ۶-۱ [3] انتخاب نماید و میزان معدل این نمره ها را با میانگین نمره های داده شده^۲ (MOS) بیان می کنند. که هرچه به عدد ۵ نزدیک تر باشد کدکننده توانمندی بیشتری را از نظر کیفیت سیگنال ساخته شده دارا می باشد.

جدول ۶-۱. نمره دهی گروهی مطلق	
نمره	کیفیت سیگنال
۵	عالی
۴	خوب
۳	متوسط
۲	ضعیف
۱	بد

¹ Absolute Category Rating

² Mean Opinion Score

ب) درجه بندی گروهی بر مبنای انحراف

در این درجه بندی ابتدا سیگنال اصلی موجود را برای شنونده پخش نموده و سپس سیگنال ساخته شده را نیز برای آنها پخش می نمایند تا با سیگنال اصلی مقایسه گردد. شنونده با مقایسه میزان اختلاف شنیداری سیگنال خروجی ساخته شده با سیگنال اصلی مطابق جدول ۶-۲ [3] اقدام به ارزش دهی به کدکننده می نماید. میانگین نمره های داده شده با انحراف از سیگنال اصلی^۱ DMOS معرفی می گردد و هرچه عدد مرتبط به ۵ نزدیک تر باشد نشان دهنده کیفیت شنیداری مناسب تر سیگنال ساخته شده و در نتیجه توانمندی کیفی کدکننده می باشد.

جدول ۶-۲. نمره دهی گروهی بر مبنای انحراف	
نمره	میزان انحراف
۵	انحرافی حس نمی شود.
۴	انحراف آزار دهنده نیست
۳	مقدار کمی انحراف
۲	انحراف آزار دهنده
۱	انحراف خیلی آزار دهنده

¹ Degradation MOS

۲-۲-۶. جنبه های مختلف کیفیت دریافت ادراکی سیگنال

کیفیت سیگنال دریافتی از جنبه های بسیار زیادی قابل بررسی است که چند نمونه از آن بطور خلاصه به شرح زیر است:

- **قابلیت فهم**
پیغام سیگنال دریافت شده بطور واضح قابل فهم باشد.
- **حفظ حالت عادی و طبیعی**
این مساله زمانی مطرح می گردد که اختشاشات مانند نویز، اکو، غیر همزمانی وجود داشته باشد.
- **قابلیت تشخیص گوینده**
سیگنال ساخته شده اطلاعات گوینده را در خود دارد و می تواند از روی آن گوینده را تشخیص داد.

۳-۲-۶. شرایط مختلف اندازه گیری کیفیت سیگنال

کیفیت سیگنال ساخته شده توسط کد کننده های مختلف در شرایط مختلف اندازه گیری می شود، شرایطی مانند:

- **وابستگی به گوینده**
کد کننده نباید نسبت به گوینده های مختلف وابسته باشد اما بسیاری از کد کننده ها با نرخ بیت پایین نسبت به گوینده های مختلف کیفیت مختلفی دارند.
- **وابستگی به زبان**
کد کننده باید نسبت به زبان های مختلف وابسته نباشد. همانطور که می دانیم هر زبان آواهای مختلف خود را دارد و می توان برای زبان های مختلف میزان کیفیت آن را بررسی نمود.
- **وابستگی به قدرت سیگنال ورودی**
کیفیت سیگنال بدست آمده برای سیگنال های ورودی با توان و قدرت مختلف قابل تغییر می باشد و این پارامتر نیز باید برای ارزش دهی به کارکرد کد کننده مدنظر باشد. به این منظور می توان از دو سیگنال با توان خیلی پایین و خیلی بالا برای تست استفاده نمود.

- **نویز زمینه**

مساله نویز زمینه اهمیت بسیار زیادی برای کد کننده های امروزی دارد چرا که برای دستگاه های قابل حمل، محیط های مختلف بوجود می آید و شرایط نویزی زیادی ممکن است بوجود بیاید. بنابراین کد کننده در شرایط نویزی مانند، داخل ماشین، خیابان، نویز گذرا، موسیقی زمینه و گوینده های همزمان باید مورد بررسی قرار گیرد.

• کارکرد پشت سر هم

در یک ارتباط مخابراتی، انکدر و دیکدر مدام در حال کار هستند و باید کد کننده از نظر کارکرد پشت سر هم و اینکه در کارکرد با سایر کد کننده ها قابل استفاده می باشد یا نه باید مورد بررسی قرار گیرد.

• نویز کانال

بسیاری از کانال های مخابراتی دارای نویز کانال می باشند و کیفیت کد کننده را می توان با میزان خطای مختلف در کانال بررسی نمود.

• سیگنال غیر صحبت

میزان بازسازی سیگنال صدا مانند موزیک یا تن های برقراری ارتباط در برخی موارد مهم می باشند و باید کیفیت کد کننده در این موارد بررسی شوند.

۶-۳. پیاده سازی کردن CELP استاندارد

برای پیاده سازی کد کننده CELP باید مفاهیم مطرح شده در ساختار کلی CELP که در فصل سوم بررسی گردید را مورد استفاده قرار داده و با توجه به این موضوع که اطلاعات کد کننده CELP استاندارد و ساختار اصلی استفاده شده در سازمان های دولتی ایالت متحده در اختیار نبود، سعی بر آن شد تا با استفاده از منابع موجود در سایت ها و کتاب های مختلف قسمت های مختلف این کد کننده توسط نرم افزار MATLAB شبیه سازی شده و تا حد امکان از قسمتهای استاندارد موجود در سایت ها و منابع مختلف به منظور بهره گیری و شبیه سازی مناسب استفاده گردد.

در این پایان نامه سعی شده است با استفاده از ترکیب مناسب فایل های موجود و نوشتن برنامه های مذکور به صورت توابع MATLAB حداکثر بهره وری از اطلاعات انجام گیرد و پارامتری به گونه ای طراحی شوند که اخلاقی در کارکرد کلی مجموعه و استفاده از منابع استاندارد صورت نگیرد.

همچنین در این فصل با معرفی پارامترهای استفاده شده در کد کننده پیاده سازی شده میزان بیت تخصیص داده شده به آنها نیز بیان می گردد.

۶-۳-۱. تخمین بهبود یافته پارامترهای فیلتر ساخت دور تناوب گام

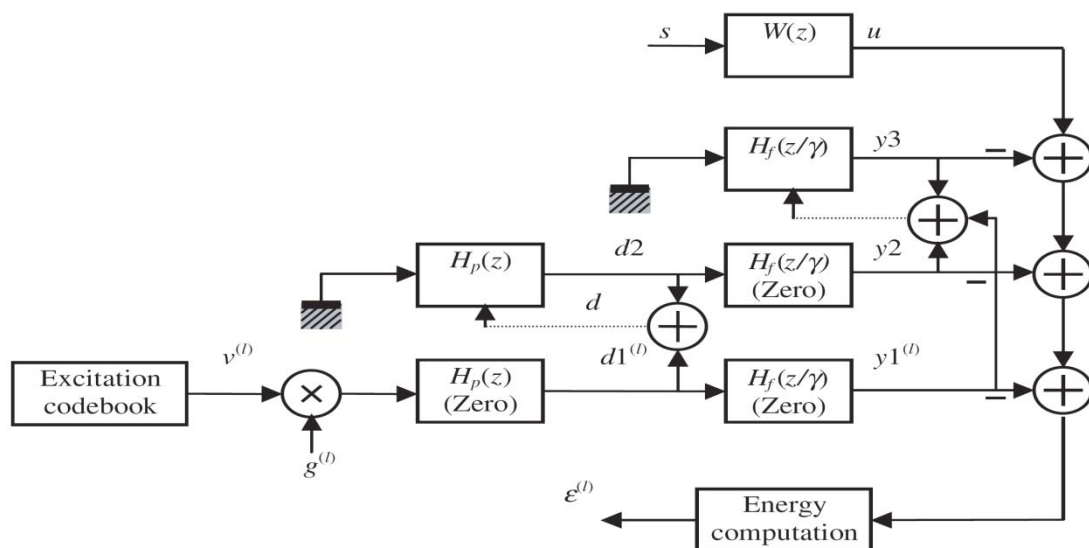
همانطور که قبلاً بیان گردید یکی از بلوک های اصلی کد کننده CELP فیلتر ساخت دور تناوب $(1/B(z))$ می باشد که دارای دو پارامتر مهم ضریب تخمین بلند مدت خطی (b) و دور تناوب (T) می باشد که از خاصیت مهم تناوبی بودن حروف صدادار برای تخمین آنها استفاده می گردد.

یکی از راههای پیدا کردن پارامترهای تخمین بلند مدت فیلتر ساخت پرپودگام استفاده از سیگنال تخمین خطای کوتاه مدت سیگنال می باشد. در این روش پارامترها به گونه ای انتخاب می گردند تا سیگنال خطای بوجود آمده کمترین مقدار را داشته باشد و در نتیجه پارامترهای بدست آمده مقداری از خواص آماری سیگنال اصلی را نشان می دهد. به هر حال هدف از تخمین پارامترها به طور کلی اینست که سیگنال خروجی تولید شده در دیکدر از نظر ادراکی با سیگنال ورودی انکدر، همخوانی داشته باشد و این هدف بطور مستقیم می تواند در تخمین این پارامترها در انکدر CELP بکار گرفته شود. بنابراین برای بدست آوردن این پارامتر در خود انکدر از اختلاف خروجی فیلتر وزن دار ادراکی و خروجی فیلترهای ساخت استفاده می گردد. با توجه به مطالب ذکر شده معادلات زیر را مطابق شکل ۶-۱ [3] میتوانیم برای محاسبه پارامترهای ذکر شده داشته باشیم:

$$\begin{aligned}\varepsilon_r^{(\ell)} &= \sum_{n=0}^{N-1} (u_r[n] - y_r^{(\ell)}[n])^2 \\ &= \sum_{n=0}^{N-1} (u_r[n] - y_{1_r}^{(\ell)}[n] - y_{2_r}[n] - y_{3_r}[n])^2\end{aligned}\quad (۲-۶)$$

برای مینیمم شدن مقدار $\varepsilon_r^{(\ell)}$ که نشان دهنده مقدار انحراف از سیگنال اصلی می باشد ، باید پارامترهای بهینه انتخاب کردند که شامل بردار تحریک شماره 1 $(v^{(\ell)})$ ، بهره بردار تحریک 1 $(g^{(\ell)})$ ، دور تناوب T و ضریب فیلتر ساخت گام b می باشد و همچنین ضرایب فیلتر ساخت فرمنت که از طریق تخمین LP بدست می آیند.

روش معرفی شده گرچه قابل اجراست ولی محاسبات کامپیوتری را به شدت بالا می برد و به همین علت از روش های جایگزین استفاده می شود. یکی از روش های پیشنهادی اینست که ابتدا به ازای بهره تحریک صفر، مقدار پارامترهای فیلتر ساخت پریودگام (b,T) محاسبه گردند و برای محاسبه ضریب تقویت پالس تحریک این مقادیر ثابت فرض شوند و بهره پالس تحریکی مورد نظر بدست آید. معادلات زیر این روش را روشن تر می کند.



شکل ۶-۱. شکل کلی نحوه پیدا کردن پارامترهای فیلتر تخمین بلندمدت کد کننده سلپ

$$j_r = \sum_{n=0}^{N-1} (u_r[n] - y_{2r}[n] - y_{3r}[n])^2 \quad (3-6)$$

که پارامترهای T و b باید بگونه ای انتخاب گردند تا معادله بالا (3-6) کمترین مقدار را به خود بگیرد و به طور خلاصه و با مشتق گیری از رابطه 3-6، رابطه 4-6 منجر می گردد:

$$\frac{\partial y_{2r}[n]}{\partial b} = \sum_{k=0}^n h[k] \frac{\partial}{\partial b} d_{2r}[n-k], \dots, 0 \leq n \leq N-1 \quad (4-6)$$

که در آن $h(n)$ فرم بهبود یافته فیلتر ساخت فرمنت می باشد. اما باید دانست که برای پیدا کردن d_2 ما به نمونه های $(r-1)$ ام زیر فریم نیازمندیم. بنابراین d_2 برای فریم حاضر مشخص نمی باشد چرا که برای دانستن مقادیر d_2 ما به مقادیر T و b نیازمندیم که مقادیرشان مشخص نمی باشد. از فصل سوم و معادله 3-15 داریم:

$$d_{2r}[n] = -bd_{2r}[n-T], \dots, 0 \leq n \leq N-1 \quad (5-6)$$

که مقدار پاسخ ورودی صفر فیلتر ساخت پرپودگام می باشد و برای دانستن مقادیر d_2 برحسب نمونه های گذشته ما باید محدوده T را مشخص کنیم و با توجه به آن مقادیر مختلف d_2 بدست می آیند. به طور مثال اگر $T \geq N$ فرض شود و مقادیر داخل [] سمت راست معادله، مقادیر زمانی منفی خواهند بود بنابراین تمامی مقادیر d_2 را می توانیم از مقادیر گذشته بدست آوریم اما اگر داشته باشیم $N/2 \leq T \leq N$ آنگاه داریم:

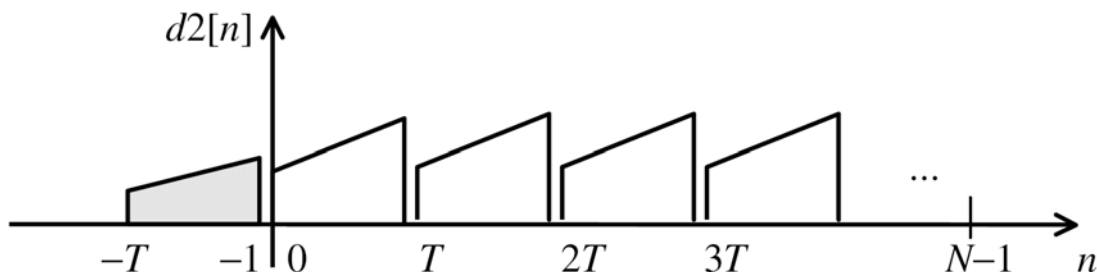
$$d_{2r}[n] = \begin{cases} -bd_{2r}[n-T], \dots, 0 \leq n \leq N-1 \\ b^2 d_{2r}[n-2T], \dots, T \leq n \leq N-1 \end{cases} \quad (6-6)$$

و اگر داشته باشیم $N/3 \leq T \leq N/2$ آنگاه برای d_2 داریم:

$$d_{2r}[n] = \begin{cases} -bd_{2r}[n-T], & 0 \leq n \leq T-1 \\ b^2 d_{2r}[n-2T], & T \leq n \leq 2T-1 \\ -b^3 d_{2r}[n-3T], & 2T \leq n \leq N-1 \end{cases} \quad (7-6)$$

مشاهده می گردد که با کاهش مقدار T تعداد معادلات بیشتر شده و درجه آنها بیشتر می گردد بنابراین در عمل برای پیاده سازی آن الگوریتم های مختلفی ارائه شده است که در کد کننده CELP مورد استفاده در این پایان نامه از روش کتاب کد تطبیق پذیر برای محاسبه پارامترهای فیلتر ساخت پرپود گام که در کننده استاندارد CELP مورد استفاده قرار می گیرد بهره گرفته شده است. در این روش اصول کلی پیدا کردن پارامترها مانند روش قبل می باشد با این تفاوت که برای غلبه بر محدودیت محاسبات کامپیوتری، کارکرد فیلتر ساخت گام به شکل زیر تعریف می گردد.

$$d_{2r}[n] = -d \begin{cases} d_{2r}[n-T], & 0 \leq n \leq T \\ d_{2r}[n-2T], & T \leq n \leq 2T \\ d_{2r}[n-3T], & 2T \leq n \leq 3T \\ \vdots \end{cases} \quad (8-6)$$



شکل ۶-۲. استخراج بردار کتاب کد تطبیقی بر حسب زمان های گذشته

که اثر آن در شکل ۶-۳ [3]، نشان داده شده است. واضح است که این تعریف بطور کامل با فیلتر ساخت پرپودگام CELP مطابقت ندارد و فقط برای حالتی که $T \geq N$ باشد تعاریف هر دو فیلتر یکی می باشد اما برای زمانهای دیگر یعنی زمانی که $N \geq T$ باشد، پاسخ مقداری تغییر خواهد کرد و بجای استفاده از توان های بالا برای بهره فیلتر از توان یک آن استفاده می گردد و پاسخ ورودی صفر فیلتر در هر زمان بر حسب پاسخ زمان های گذشته آن مشخص می گردد. با این تعریف مقدار b به سادگی با پیدا کردن مقادیر T بدست می آید. باید توجه داشت که دامنه پالس ها به ازای $T \leq N$ از هر پالس به پالس دیگر ثابت می ماند اما در عمل ثابت شده است که با استفاده از روش فوق، تاثیر زیادی بر روی کارکرد سیستم اعمال نمی شود [3].

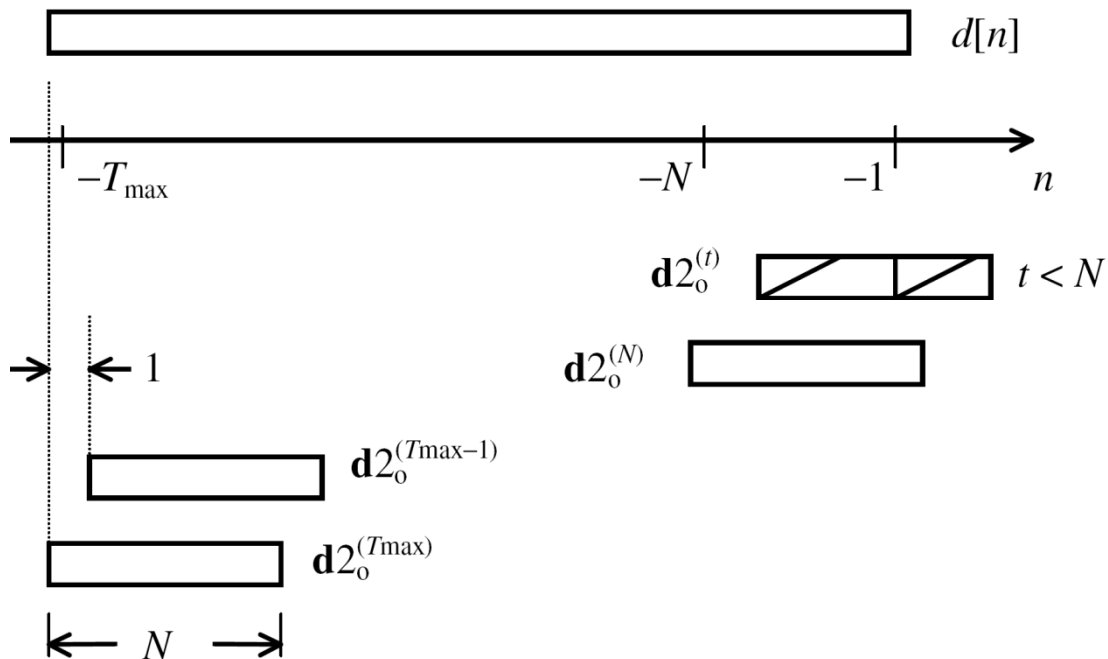
۶-۳-۲. نحوه کار کتاب کد تطبیق پذیر

علت نام گذاری این روش به نام کتاب کد تطبیق پذیر به این علت است که در حقیقت روش پیدا کردن پرپودگام بهینه می تواند مانند کتاب کدی در نظر گرفته شود که بردارهای آن با یکدیگر هم پوشانی دارند به این معنی که بردارهای نزدیک به هم دارای نمونه های مشترکی هستند که هر کدام با پرپود T نام گذاری می شوند و کتاب کد به این علت تطبیق پذیر نامیده شده است که از هر زیر فریم به زیر فریم بعدی تغییر می کند. برای روشن تر شدن مساله می توانیم ساختار کتاب کد را مانند شکل ۶-۳ [3] مورد بررسی قرار دهیم.

در این ساختار کتاب کد یک آرایه یک بعدی به تعداد T_{max} عضو می باشد که در آن T_{max} ماکزیمم دور را تناوب موجود برای سیگنال ها می باشد. برای راحتی کتاب کد را با $d[-T_{max}]$ تا $d[-1]$ نمایش می دهیم و شماره زیر فریم را حذف می کنیم بنابر این آرایه های این آرایه به فرم زیر خواهند بود:

$$d 2_0^{(t)} = [d 2_0^{(t)}[0], \dots, d 2_0^{(t)}[N-1]]^T, t = T_{min} \text{ to } T_{max} \quad (9-6)$$

حال معادله ۶-۸ می تواند برای تعریف عناصر بردار به شکل زیر تعریف گردد و در صورتی که $t < N$ باشد. خواهیم داشت:



شکل ۳-۶. ساختار کتاب کد تطبیقی

$$d2_0^{(t)}[n] = \begin{cases} d[n-t], & 0 \leq n \leq t \\ d[n-2t], & t \leq n \leq 2t \\ d[n-3t], & 2t \leq n \leq 3t \\ \vdots \end{cases} \quad (10-6)$$

و اگر $t \geq N$ باشد آنگاه خواهیم داشت:

$$d2_0^{(t)}[n] = d[n-t] \quad (11-6)$$

بنابراین می توان برای تمامی مقادیر t بردار $d2_0^{(t)}$ همراه با ضریب b بهینه را انتخاب نمود و با استفاده از آن بردار تقویت شده مناسب $d2^{(t)} = b \cdot d2_0^{(t)}$ را که مناسب ترین پاسخ ورودی صفر را نشان می دهد، بدست آورد.

۳-۳-۶. به روز رسانی کتاب کد

بعد از پیدا کردن $d_2^{(t)}$ پاسخ کلی فیلتر ساخت پرلودگام به شکل زیر است:

$$d[n] = d_1^{(l)}[n] + d_2^{(t)}[n], \dots, n = 0 \text{ to } N - 1 \quad (۱۲-۶)$$

که در آن $d_1^{(l)}[n]$ پاسخ حالت صفر با جستجو در کتاب کد پیدا می شود و کتاب کد با این سیگنال به فرم زیر به روز رسانی می شود:

$$d[n] \leftarrow d[n + N], n = -T_{\max}, -T_{\max} + 1, \dots, -1 \quad (۱۳-۶)$$

که به معنی شیفیت به سمت چپ N نمونه می باشد و جستجو برای زیر فریم بعدی انجام می پذیرد. برای استاندارد FS1016، $T_{\max}=147$ ، $N=60$ و t بین ۱۴۷ تا ۲۰ تغییر می کند که برای انکد آن از ۷ بیت استفاده می گردد که برای پیاده سازی آن در این پایان نامه از همین مقادیر استفاده شده است.

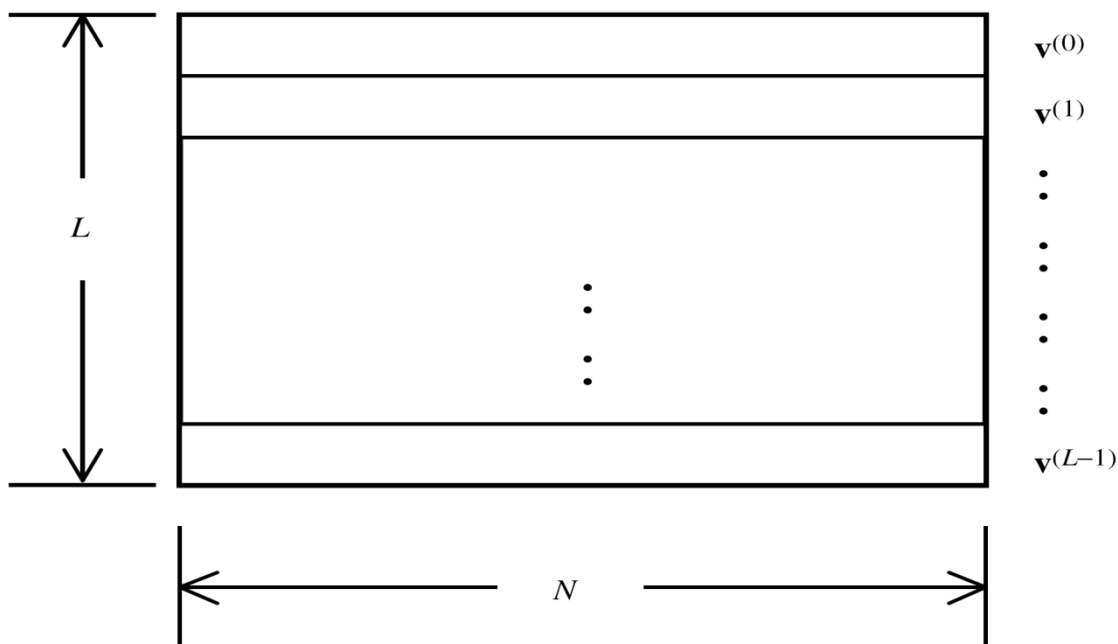
۴-۶. ساختار منبع تحریک تصادفی

ساختار یک کتاب کد استاندارد CELP تشکیل شده است از یک ماتریس $L \times N$ که در آن L تعداد بردارهای تحریک در منبع و N تعداد نمونه های موجود در هر بردار (N به تعداد زیر نمونه های هر زیر فریم) می باشد. به این نمونه ساختار، ساختار غیر همپوشان گفته می شود و ساختاری همانند یک ماتریس ساده که در آن بردارها در زیر هم قرار گرفته اند دارند. شکل ۴-۶ [3]، نمایانگر این ساختار می باشد. براین اساس اگر نمونه های ساختار کد را به صورت باینری در نظر بگیریم حافظه ای به اندازه $L \times N$ نیاز خواهیم داشت. در این ساختار برای پیدا کردن سیگنال تحریک مناسب برای کل سیستم باید تمامی بردارهای موجود درون ماتریس یکی یکی برای یافتن بهترین سیگنال مورد آزمایش قرار گیرند که این آزمایش زمان و محاسبات کامپیوتری زیادی را می طلبد.

برای غلبه بر این مشکل نمونه ها را در یک آرایه یک بعدی به طوریکه برای L بردار تحریک،

بسیاری از نمونه های آنها با یکدیگر مشترک باشند قرار می دهیم. در حقیقت نمونه های مشترک همپوشانی زیادی با یکدیگر دارند و می توانیم از روش محاسبه همپوشانی برای کاهش حجم محاسبات استفاده نمائیم.

برای مثال اگر تعداد نمونه هایی که در هر بردار منحصر بفرد و هیچ همپوشانی با بردار قبل یا بعدی خود را ندارد S در نظر بگیریم، می توانیم با یک آرایه یک بعدی به طول $S(L-1) + N$ نمونه یک ساختار کتاب کد تشکیل دهیم که این ساختار حجم حافظه و محاسبات مربوط به ساختار معمولی کتاب کد را کاهش می دهد. این ساختار در شکل ۶-۵ [3] نشان داده شده است با توجه به شکل می توان ساختار کل کتاب کد را مطابق معادلات زیر توصیف نمود:



شکل ۶-۴. ساختار کتاب کد بدون هیچ همپوشانی

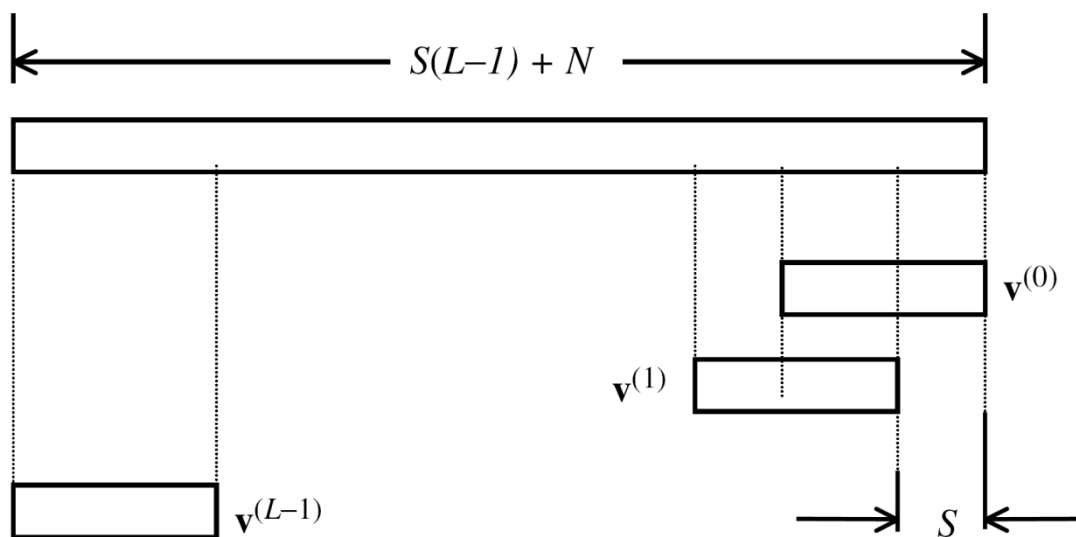
$$V[n]; n = 0, 1, \dots, S(L-1) + N - 1 \quad (۱۴-۶)$$

به این ترتیب هر کدام از L بردار به شکل زیر خواهند بود:

$$\begin{aligned} v^{(0)}[n] &= v[n + (L-1)s] \\ v^{(1)}[n] &= v[n + (L-2)s] \\ v^{(L-1)}[n] &= v[n] \end{aligned} \quad (۱۵-۶)$$

و میتوانیم برای هر بردار که شامل N نمونه است معادله زیر را بنویسیم:

$$v^{(l-1)}[n] = v^{(l)}[n-s], \dots, s \leq n \leq N-1 \quad (۱۶-۶)$$



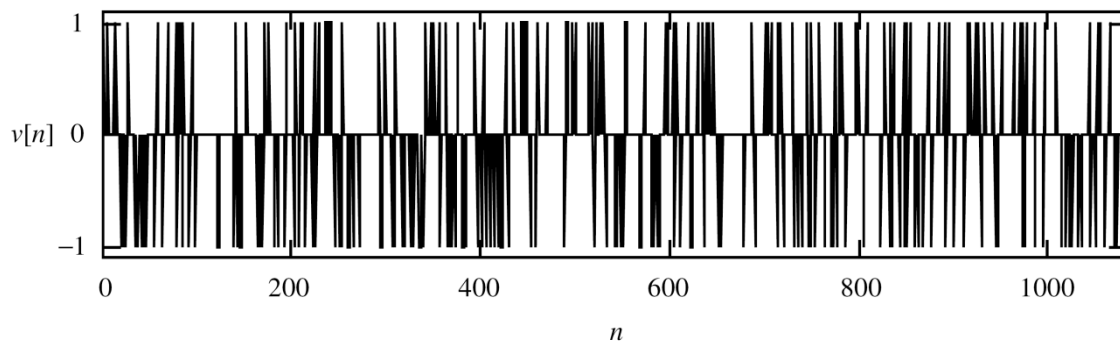
شکل ۶-۵. منبع تحریک تصادفی با همپوشانی

این ساختار که در کد کننده پیاده سازی شده این پایان نامه از آن استفاده شده است، اختلاف زیادی با ساختار CELP اصلی که ساختارش کاملاً تصادفی است، ندارد چرا که در ساختار CELP اصلی نمونه ها به صورت نمونه های نویز سفید غیر وابسته هستند و اگر همان نویز به صورت شیفت یافته نیز اعمال گردد ساختار غیر وابسته آن از بین نخواهد رفت.

در کد کننده استاندارد FS1016 استفاده شده، ۹ بیت برای نشان دادن آدرس محل قرار گیری بردار تحریک ($2^9=512$) و تعداد نمونه های غیر همپوشان $S=2$ در نظر گرفته شده است. بردارها به صورت از راست به چپ شماره گذاری می گردند. بنابراین با داشتن ۶۰ نمونه استاندارد در هر زیر فریم و ۵۱۲ بردار تحریک به طور کلی تعداد $(512-1) \times 2 + 60 = 1082$ نمونه خواهیم داشت که در مقایسه با $512 \times 60 = 30720$ نمونه، مقدار کمتری می باشد و حافظه کمتری نیاز دارد. همچنین مساله مهم دیگری که اهمیت این روش را نسبت به استفاده از روش عادی مشخص می کند، میزان کاهش محاسبات کامپیوتری بخاطر استفاده از روش کانولوشن بازگشتی برای استفاده از نمونه ها می باشد و براحتی می توان محاسبات مربوط به پاسخ حالت صفر را با استفاده از نمونه های قبلی بدست آورد. همانطور که بیان گردید نمونه های استفاده شده در کد کننده CELP عمومی برای ساختار منبع تحریک و بردارهای تحریک از یک منبع تولید نویز سفید بوجود می آیند.

اما در کد کننده FS1016 که در این پایان نامه از آن استفاده شده است از نمونه های تصادفی ویژه ای استفاده می شود. این نمونه ها از یک منبع تولید نویز سفید گوسی با واریانس یک و متوسط صفر تولید می گردند. نمونه ها به صورت سه حالت ۱، ۰ و ۱- تولید می گردند. این نمونه ها که تعداد آنها ۱۰۸۲ می باشد در واقع ساختار این منبع تحریک را نشان می دهد. شکل ۵-۶ [3]، نمونه های موجود در این ساختار که در سیم مخابرات ملی استفاده شده است را نشان می دهد.

۷۷٪ از ۱۰۸۲ نمونه موجود در این ساختار مقدار صفر به خود می گیرند و از آنجا که این نمونه ها فشرده هستند مقدار کمی انحراف نسبت به ساختارهای دیگر کتاب کد بوجود می آورند، اما نسبت محاسبات به خاطر محاسبه کانولوشن برای انتخاب و وجود بردار تحریک مناسب بسیار کاهش پیدا می کند. ذکر این نکته لازم است که آزمایشات نشان داده اند که ساختار کتاب کد با تعداد زیادی از مقادیر نمونه صفر تقریباً مانند ساختار کتاب کد با درصد کمتری از صفرها کار می کند [3].



شکل ۶-۶. نمونه های تصادفی موجود در منبع تحریک FS1016

۵-۶. جستجوی کتاب کد تطبیقی

در قسمت ابتدای فصل ساختار کلی کتاب تطبیقی معرفی گردید در این بخش روش جستجو برای بهترین بردار کد تطبیقی و نحوه کد کردن زمان تاخیر غیر صحیح (کسری) و پارامترهای مرتبط معرفی می گردد.

۱-۵-۶. دوره تناوب گام صحیح

همانطور که بیان شد برای پیدا کردن دور تناوب پریودگام، در یک پردازش آنالیز- ساخت در کتاب کد تطبیق پذیر مقدار این دور تناوب در باز $t = T_{\min}, T_{\min}+1, \dots, T_{\max}$ جستجو می گردد که مقادیر $T_{\min}=20$ و $T_{\max}=147$ می باشد. در این روش همانند روش پیدا کردن سگنال تحریک مناسب عمل می شود و مناسب ترین دوره تناوب به نحوی انتخاب می گردد که میزان خطا مینیمم گردد. بنا بر این برای پیدا کردن مناسب ترین پارامترها می توانیم از معادلات زیر استفاده کنیم.

پاسخ حالت صفر فیلتر ساخت فرمنت بهینه شده برابر است با :

$$y_2^{(t)} = H d_2^{(t)} = b^{(t)} H d_0^{(t)} \quad (۱۷-۶)$$

که در آن H ماتریس پاسخ ضربه فیلتر ساخت می باشد بنابراین با توجه به معادله ۶-۱۸ که برای اندازه گیری میزان خطا استفاده می شود میتوانیم معادله ۶-۱۷ را بازنویسی کنیم و معادله ۶-۱۹ را خواهیم داشت:

$$J^{(t)} = \sum_{n=0}^{N-1} (u[n] - y^{(t)}[n] - y_3[n])^2 \quad (۱۸-۶)$$

$$J^{(t)} = \|u - y^{(t)} - y_3\|^2 \quad (۱۹-۶)$$

هدف از یافتن پارامترها مینیمم کردن معادله ۶-۱۹ می باشد بنابر این با تعریف $u_3 = u - y_3$ خواهیم داشت:

$$J^{(t)} = \|u_3 - b^{(t)} H \cdot d_0^{(t)}\|^2 \quad (۲۰-۶)$$

با حل معادله فوق بهترین بهره کتاب کد تطبیقی از رابطه ۶-۲۱ بدست می آید و مناسب ترین سیگنال استفاده شده سیگنالی است که معادله ۶-۲۲ را ماکزیمم کند [3].

$$b^{(t)} = \frac{u_3^T H \cdot d_0^{(t)}}{\|H \cdot d_0^{(t)}\|^2} \quad (۲۱-۶)$$

$$Q^{(t)} = \frac{(u_3^T H \cdot d_0^{(t)})^2}{\|H \cdot d_0^{(t)}\|^2} \quad (۲۲-۶)$$

و از آنجا که کتاب کد تطبیقی یک کتاب کد با حالت همپوشانی با یک شیفت است کانولوشن بازگشتی برای محاسبات فوق مورد استفاده قرار می گیرد.

۲-۵-۶. دوره تناوب گام غیر صحیح

دوره تناوب غیر صحیح که مقداری دقیق تر به نسبت دوره تناوب صحیح سیگنال ارائه می نماید برای کاهش اثر انعکاس سیگنال صحبت و همچنین اثر تندی و ناهمواری سیگنال ساخته شده استفاده می گردد. همچنین استفاده از این دوره تناوب در کاهش اثر نویزی سیگنال تحریک در وجود و تولید نویز در سیگنال ساخته شده اثر دارد. در این پایان نامه از این روش برای استفاده به عنوان تاخیر غیر صحیح استفاده شده است که در استاندارد FS1016 نیز از آن استفاده می شود. این استاندارد از یک روش بازسازی^۱ بر مبنای پنجره همینگ^۲ تابع سینک که به شکل زیر تعریف می شود استفاده می کند.

$$w_0(n) = 0.54 + 0.46 \cos\left(\frac{2\pi n}{40}\right)$$

$$w(n, f) = w_0(n + f) \frac{\sin((n + f)\pi)}{(n + f)\pi} \quad (23-6)$$

که در آن ۱۹ و ۱۸ و و ۱۹- و ۲۰- n و f قسمت غیر صحیح دوره تناوب می باشد و می تواند یکی از ۵ مقدار $\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}$ را بخود بگیرد. برای بدست آوردن دوره تناوب t+f که در آن قسمت صحیح f قسمت اعشاری و t قسمت صحیح است ابتدا سیگنال را به شکل زیر می نویسیم:

$$dd^{(t)}[n] = \begin{cases} d[n], & \dots\dots\dots -147 \leq n < 0 \\ d_{20}^{(t)}[n], & \dots\dots\dots 0 \leq n < 60 \end{cases} \quad (24-6)$$

و برای هر سیگنال داده شده خواهیم داشت:

¹ Interpolation

² Hamming window

$$d20^{(t+f)}[n] = \sum_{k=-20}^{19} w(k, f) dd^{(t)}[n-t+k] \quad (25-6)$$

۳-۵-۶. انکدر دوره تناوب

استاندارد FS1016 برای مقدار دوره تناوب ۲۵۶ مقدار مختلف در بازه (۲۰-۱۴۷) تخصیص داده است که دارای دقت های مختلفی در بازه های مختلف است. برای بازه $[20-25 \frac{2}{3}]$ و $[34-79 \frac{2}{3}]$ دارای دقت $\frac{1}{3}$ برای بازه $[26-33 \frac{3}{4}]$ دارای دقت $\frac{1}{4}$ و برای بازه $[80-147]$ دارای دقت ۱ می باشد به این تعریف دقت مناسبی برای دوره های پایین تر بدون اینکه بیت اضافی مصرف بکنیم خواهیم داشت چرا که همیشه ۸ بیت برای دوره تناوب مطابق جدول جستجو تخصیص داده می شود همچنین دقت $\frac{1}{4}$ در بازه فرکانسی قرار دارد که معمولا صدای مربوط به خانم ها در آن بازه قرار می گیرد به این معنی که دوره تناوب با دقت ۲۵٪ تغییر می کند.

۱-۳-۵-۶. راهکار انکدر کردن زیرفریم

استاندارد FS1016 از تعداد مختلفی بیت برای انکد کردن دوره ی تناوب استفاده می کند که در آن تعداد بیت ها به جایگاه زیر فریم بستگی دارد. برای زیر فریم اول و سوم، ۸ بیت و برای زیر فریم دوم و چهارم ۶ بیت تخصیص داده می شود که این ۶ بیت در حقیقت نشانگر میزان کیفیت نسبت به زیر فریم قبلی می باشد که می توانند به اندازه ۳۱ مقدار کمتر یا ۳۲ مقدار بیشتر باشد که البته مقدار صحیح و غیر صحیح را شامل می شود. این روش باعث کارایی بیشتر سیستم شده و مقدار محاسبات کامپیوتری و حافظه های مورد نیاز را کاهش می دهد.

۶-۶. انکد کردن بهره کتاب کد تطبیقی

بهره کتاب کد تطبیقی بین ۲ و ۱- تغییر می کند و برای آن ۵ بیت تخصیص داده می شود.

۶-۷. روش جستجوی کد تحریک تصادفی

در جستجوی کد تحریک مناسب تصادف روشی مانند روش پیدا کردن کد تحریک تطبیقی استفاده شده است. معادلات مربوط به پیدا کردن سیگنال تحریک مناسب و در نتیجه کد آن به شرح زیر می باشد.

پاسخ حالت صفر فیلتر بهبود یافته ساخت فرمنت به فرم زیر می باشد:

$$y_1^{(l)} = g^{(l)} H v^{(l)} \quad (26-6)$$

که در آن $g^{(l)}$ بهره سیگنال تحریک l ام منبع تحریک $v^{(l)}$ و H ماتریس پاسخ ضربه فیلتر بهبود یافته ساخت فرمنت می باشد و با بازنویسی معادله ۶-۲ خواهیم داشت:

$$\varepsilon^{(l)} = \|u - y_1^{(l)} - y_2 - y_3\|^2 = \|u_0 - g^{(l)} H v^{(l)}\|^2 \quad (27-6)$$

که در آن $u_0 = u - y_2 - y_3$ با حل معادله فوق بهره تحریک مناسب از رابطه ۶-۲۸ بدست آمده و بهترین سیگنال تحریک سیگنالی است که معادله ۶-۲۹ را ماکزیمم می کند [3].

$$g^{(l)} = \frac{u^T H v^{(l)}}{\|H v^{(l)}\|^2} \quad (28-6)$$

$$p^{(l)} = \frac{(u^T H v^{(l)})^2}{\|H v^{(l)}\|^2} \quad (29-6)$$

برای $L = 0-512$ که تعداد سیگنال های تحریک هستند که دارای همپوشانی با یکدیگر ندارند این محاسبات با عمل کانولوشن بازگشتی محاسبه شده و کارآیی سیستم بالاتر می رود.

۶-۷-۱. بهره کتاب کد تصادفی بهبود یافته

یکی از معایب سیگنال ساخته شده در FS1016 وجود مقداری نویز در سیگنال ساخته شده خروجی می باشد. اگرچه سیگنال خروجی ساخته شده دارای کیفیت نسبتاً خوبی است اما تلاشهایی برای بهبود این پدیده انجام گرفته است که منجر به استفاده از روش های مختلف برای استفاده شده است.

یکی از راههای پیشنهاد شده که در داخل FS1016 نیز استفاده شده است استفاده از بهره بهینه سیگنال تحریک می باشد. این روش کاهش تطبیقی بهره سیگنال تحریک در زمانی که تخمین بلند مدت فیلتر بهینه باشد را در بر می گیرد.

روش پیشنهادی اینست که بهره فیلتر ساخت پرلودگام افزایش و بهره سیگنال تحریک کاهش می یابد. در این حالت به علت کاهش زبری و ناهمواری سیگنال و نویز کوانیزه شده کیفیت سیگنال ساخته شده در خروجی بهتر می گردد و هنگامی که تخمین بلند مدت بقدر کافی موثر نباشد که بیشتر در مواقع بی صدا بودن سیگنال اتفاق می افتد بهره سیگنال تحریک افزایش می یابد و بهره تخمین بلند مدت کاهش می یابد.

میزان موثر و خوب بودن تخمین بلند مدت از رابطه ۶-۳۰ [3] بدست می آید.

$$R = \frac{(u - y_3)^T (u - y_3 - y_2^{(t)})}{\|u - y_3\|} \quad (۶-۳۰)$$

هرچه میزان این بهره (R) موثرتر (کوچکتر باشد) باشد، یعنی مقدار کوچکتر بردار $(u - y_3 - y_2^{(t)})$ به ضریب بهره کوچکتری منتهی می شود که با رابطه زیر تعریف می شود:

$$sf(R) = \begin{cases} 0.2, \dots, |R| < 0.04 \\ 1.4\sqrt{|R|}, \dots, |R| > 0.81 \\ \sqrt{|R|}, \dots, otherwise \end{cases} \quad (۶-۳۱)$$

با استفاده از رابطه ۳۱-۶ که به طور تجربی به دست آمده است مقدار بهینه شده بهره تقویت سیگنال تحریک از رابطه ۳۲-۶ محاسبه میشود و خواهیم داشت:

$$g' = sf(R).g^{(l)} \quad (۳۲-۶)$$

برای مقدار کوچکتر R که تخمین بلند مدت بهینه است، بهره اصلی سیگنال تحریک کاهش می یابد و اگر R افزایش یابد بهره سیگنال تحریک نیز افزایش می یابد. این معادله به صورت تجربی و مشاهدات بدست آمده و کیفیت سیگنال را بهبود داده است. باید متذکر شد از آنجا که این بهبود بعد از یافتن سیگنال تحریک مناسب اعمال می شود، مقدار محاسبات به مقدار خیلی کم تحت تاثیر قرار می دهد. بهره سیگنال تحریک با ۵ بیت در این کد کننده انکد شده است.

۸-۶. آنالیز LP

آنالیز LP برای هر فریم به طول ۲۴۰ نمونه با دوره زمانی 30ms انجام می شود که در آن از پنجره همینگ استفاده شده است و یازده مقدار خود همبستگی محاسبه می گردد. این مقادیر برای استفاده بر ازای هر زیر فریم ۶۰ نمونه ای مورد استفاده قرار می گیرد.

۹-۶. فیلتر وزن دار ادراکی

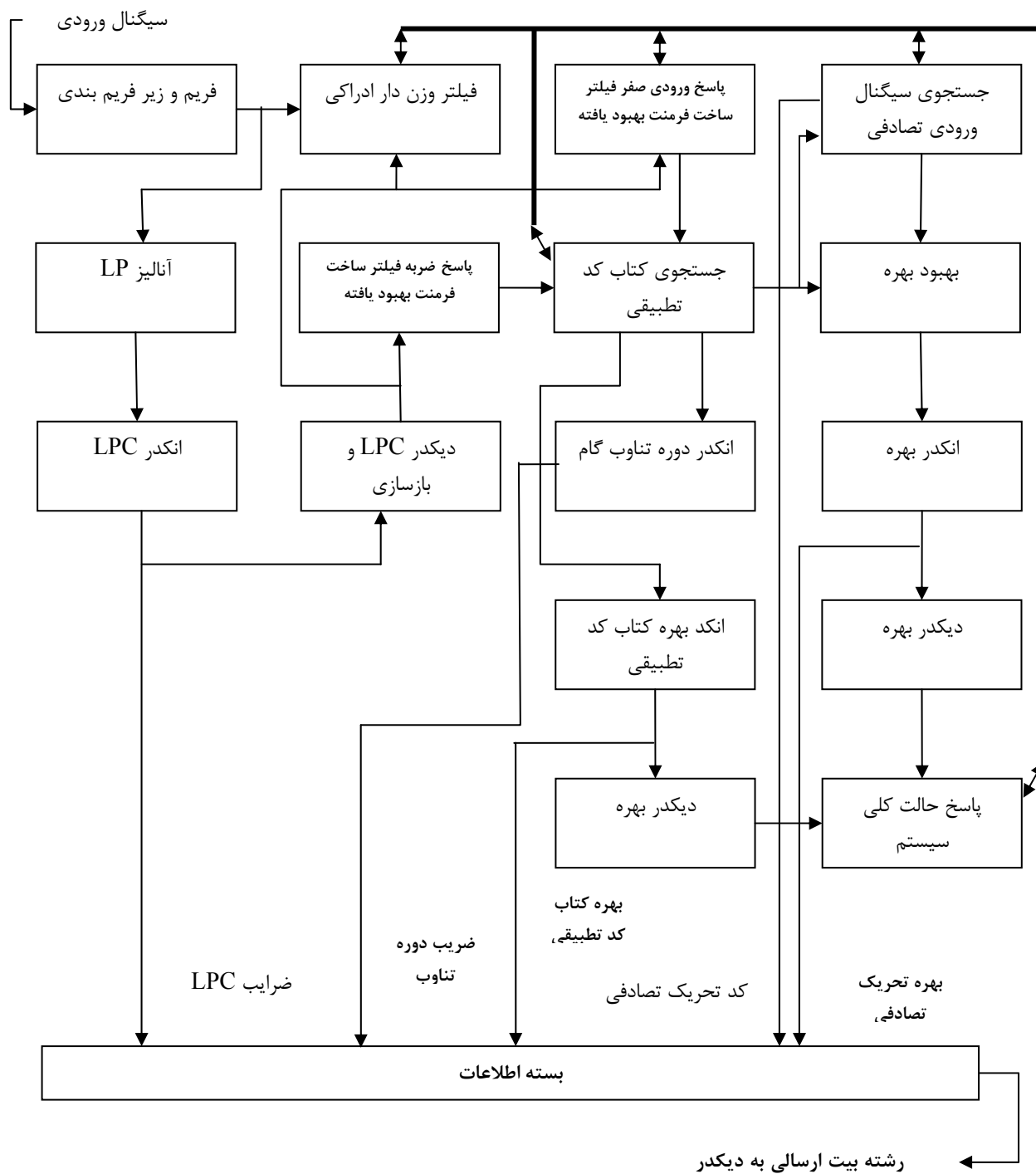
مقدار وزن $\gamma = 0.8$ با مقادیر کوانیزه شده LPC برای فیلتر وزن دار استفاده می شود. که این مقادیر برای هر زیر فریم به روز می شوند.

۶-۱۰. انکدر و دیکدر

بلوک دیاگرام کلی انکدر و دیکدر FS1016 که در این پایان نامه پیاده سازی شده است در شکل های ۶-۷ و ۶-۸ [3] نشان داده شده است.

مطابق شکل ۶-۷ در این کد کننده برای سیگنال ورودی بعد از فریم و زیر فریم بندی ضرایب LPC هر فریم محاسبه می گردد و این ضرایب برای ساخت فیلتر بهبود یافته ساخت فرمنت و فیلتر وزن دار ادراکی برای استفاده در الگوریتم آنالیز – ساخت در خود انکدر و ارسال برای دیکدر مورد استفاده قرار می گیرند.

در این بلوک با استفاده از الگوریتم آنالیز – ساخت و با توجه به روش های گفته شده کد تحریک منبع و بهره تقویت هر پالس انتخاب شده همچنین کد و بهره منبع کد تطبیقی برای ساخت فیلتر ساخت دوره تناوب گام انتخاب شده و به صورت رشته بیت با توجه به جدول تخصیص بیت که در جدول ۶-۳ آورده شده است برای دیکدر فرستاده می شود.

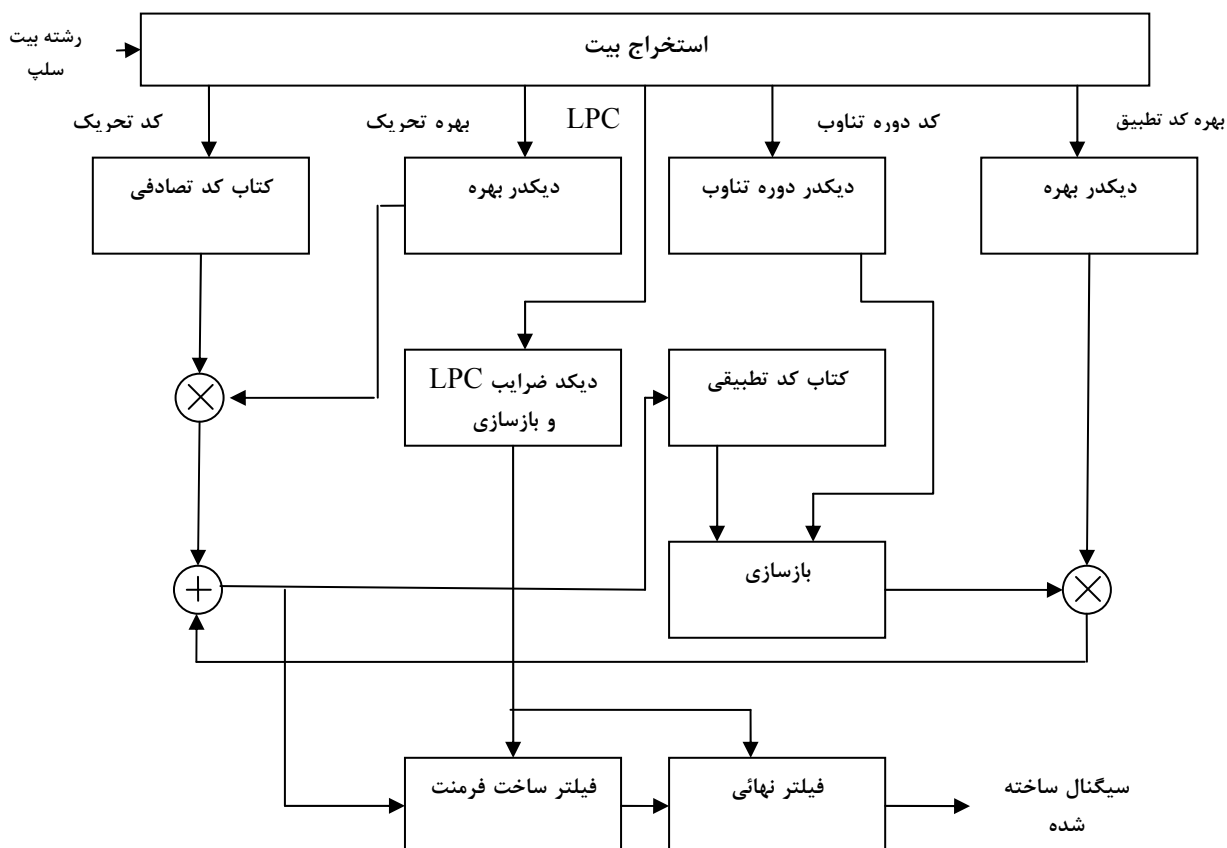


۶-۱۱. پاسخ کلی

این بلوک پاسخ کلی سیگنال تحریک و پاسخ فیلتر ساخت فرمنت ($y_1+y_2+y_3$) که برای استفاده در زیر فریم بعدی در سیستم ذخیره می شود را محاسبه می کند.

۶-۱۲. کارکرد دیکدر

سیگنال تحریک مشخص شده در بهره کوانیزه شده مربوطه خود ضرب می شود و شماره دوره تناوب مربوط برای پیدا کردن قسمت صحیح و اعشاری دوره تناوب برای ساخت فیلتر ساخت دوره تناوب گام مورد استفاده قرار می گیرد و با سیگنال تحریک مناسب به ورودی فیلتر ساخت فرمنت اعمال می گردد. فیلتر نهایی استفاده شده همان فیلتر معرفی شده در قسمت CELP کددر عمومی می باشد.



شکل ۶-۸. بلوک کلی دیکدر کاربردی شده

۶-۱۳. تخصیص بیت

جدول ۶-۳ تخصیص بیت کد کننده مورد استفاده را نشان می دهد. بیت همزمانی به ازای هر فریم بین ۱ و صفر تغییر می کند که به ازای فریم اول ارسالی صفر می باشد. ۴ بیت در هر فریم برای تصحیح خطا با تاکید بر روی عدد کتاب کد تطبیقی به علت اهمیت آن بر روی کیفیت سیگنال و یک بیت برای رزرو برای بهبودهای آینده در نظر گرفته شده است. به طور کلی ۱۴۴ بیت برای هر فریم در نظر گرفته می شود که برای طول فریم 30ms نرخ بیتی برابر 4800bps خواهد بود.

جدول ۶-۳. تخصیص بیت کد کننده کاربردی شده

تعداد بیت در فریم	دقت بیت اختصاصی	تعداد در فریم	پارامتر ارسالی
۳۴	۳ و ۳ و ۳ و ۳ و ۳ و ۳ و ۴ و ۴ و ۴ و ۴ و ۳	۱۰	ضرایب LPC
۲۸	۸ و ۶ و ۸ و ۶	۴	دوره تناوب کتاب تطبیقی
۲۰	۵	۴	بهره کتاب کد تطبیقی
۳۶	۹	۴	کد تحریک تصادفی
۲۰	۵	۴	بهره تحریک تصادفی
۱	۱	۱	بیت همزمانی
۴	۱	۴	تصحیح خطا
۱	۱	۱	بیت رزرو
جمع بیت به ازای هر فریم: ۱۴۴			

فصل ۷ کد کننده پیشنهادی و

نتایج عملی

۷-۱. مقدمه

ایده استفاده از الگوریتم PLP در کد کننده ها و به طور خاص در کد کننده سلپ ابتدا از آنجا نشأت گرفت که در بسیاری از سیستم های تشخیص گفتار، خواص سیگنال صحبت مورد پردازش قرار گرفته و استخراج ویژگی ها یکی از بلوک ها اصلی و مهم این سیستم ها محسوب می گردد و از طرفی اساس کار کد کننده ها سیگنال نیز استخراج ویژگی های سیگنال صحبت می باشد از این رو استفاده از الگوریتم های استفاده شده در سیستم های تشخیص گفتار در کد کننده ها مورد توجه قرار گرفت. همچنین در تحقیق انجام گرفته برای بررسی نوآوری روش پیشنهادی، مرجع شماره [11] که در آن از MFCC^۱ در کد کننده سلپ استفاده شده است به عنوان یکی از مراجع که مشابهت هایی با روش پیشنهادی این پایان نامه داشت مورد توجه قرار گرفت. در مرجع فوق از الگوریتم استخراج ضرایب MFCC که مشابهت های زیادی با استخراج ضرایب PLP دارد استفاده شده است. الگوریتم اجرا شده در آن پایان نامه بر مبنای استخراج ضرایب MFCC و ارسال مستقیم این ضرایب برای دیکدر میباشد. کد کننده بهبود یافته معرفی شده در آن مرجع برای افزایش کیفیت شناسایی سیگنال گفتار در سیگنال ساخته شده در دیکدر، بجای ارسال ضرایب LPC برای فیلتر ساخت فرمنت، فیلتر وزن دار و فیلتر نهائی، ضرایب MFCC محاسبه شده و برای دیکدر ارسال می گردد. در دیکدر با استفاده از الگوریتم های موجود این ضرایب به ضرایب LPC تبدیل شده و برای ساخت سیگنال مورد استفاده قرار می گیرد از طرفی چون این ضرایب در تشخیص اتوماتیک گفتار بکار می روند، برای تشخیص سیگنال گفته شده نیز مورد استفاده قرار می گیرند. از معایب الگوریتم ارائه شده در آن مرجع می توان به افزایش نرخ بیت از ۴.۸ کیلو بیت بر ثانیه به ۸.۷ کیلو بیت بر ثانیه اشاره نمود که حدود ۲ برابر نرخ بیت کد کننده استاندارد سلپ می باشد.

¹ Mel Frequency Cepstral Coefficient

۷-۲. برخی موارد استفاده شده از PLP در کد کننده ها

در مطالعه و تحقیق انجام شده برای بررسی اثر استفاده از الگوریتم PLP در کد کننده ها، در بیشتر موارد، مراجع مربوطه از روش استخراج و استفاده از ضرایب PLP به عنوان اطلاعات ارسالی در رشته بیت انکدر برای افزایش دقت تشخیص اتوماتیک گفتار استفاده نموده اند که البته استفاده از این ضرایب به طور کلی باعث افزایش نرخ بیت کد کننده ها شده است. لازم به ذکر است که معمولاً با کاهش بیت ریت مورد استفاده توسط کد کننده ها که از اهداف کد کننده ها می باشد، از میزان اطلاعات مربوط به گوینده و تشخیص گفتار کاسته می شود و استفاده از ضرایب PLP معمولاً برای کاهش این اثر مورد استفاده قرار گرفته است.

برای آشنایی با برخی از موارد استفاده شده از PLP در کد کننده ها، در این بخش به طور مختصر به برخی مراجع استفاده کننده از الگوریتم PLP و نحوه استفاده در آنها معرفی می گردند.

• استفاده از PLP در کد کننده سلپ [40]

در مرجع شماره ۴۰ که در سال ۲۰۰۶ به عنوان ایده ثبت^۱ ارائه گردیده است، کد کننده سلپی بر مبنای استفاده از PLP معرفی شده است که الگوریتم معرفی شده در این مرجع با الگوریتم استفاده شده در پایان نامه بسیار متفاوت است و به نظر می رسد در مرجع فوق تنها ایده استفاده از PLP معرفی شده است و هیچ گونه نتایج عملی، روش پیاده سازی، نحوه استخراج پارامترهای معرفی شده، تعداد ضرایب PLP استخراج شده، میزان بیت تخصیص داده شده و ... مشخص نمی باشد. با توجه به شکل ۳ معرفی شده در مرجع مربوط، در این روش از یک بلوک پردازش PLP برای استخراج ضرایب PLP و پیدا کردن دوره تناوب برای استفاده در فیلتر ساخت دوره تناوب گام $(B(z))$ از روی سیگنال اصلی استفاده شده است و همچنین ضرایب PLP استخراج شده بطور مستقیم در تولید طیف

¹ Patent

فرکانسی سیگنال ساخته شده (فیلتر ساخت فرمنت) مورد استفاده قرار می گیرد و این ضرایب به طور مستقیم به عنوان اطلاعات رشته بیت ارسالی سلپ برای دیکدر سلپ فرستاده می شود. به طور کلی رشته بیت ارسالی در کد کننده معرفی شده در این مرجع شامل پارامتر های ضرایب PLP، اطلاعات تحریک و دوره تناوب گام می باشد. اما در روش پیشنهادی در این پایان نامه از ضرایب PLP استفاده نشده است و فقط از الگوریتم یافتن این ضرایب تا مرحله پیدا کردن ضرایب LPC برای استفاده در بلوک های فیلتر ساخت فرمنت، فیلتر وزن دار ادراکی و فیلتر نهایی استفاده شده است و رشته بیت ارسالی برای دیکدر شامل ضرایب LPC بدست آمده از روش PLP و سایر پارامتر های معرفی شده می باشد. همچنین برای پیدا کردن دوره تناوب مناسب از روش کتاب کد تطبیقی استفاده شده است که این دوره تناوب با دقت غیر صحیح معرفی شده از روی کتاب کد تطبیقی و نه از روی خود سیگنال استخراج می شود.

• استفاده از PLP در کد کننده صوت سینوسی [41]

در این مرجع سیگنالی که قرار است کد شود برای کاهش بیت ریت ارسالی به صورت جمعی از سیگنال های سینوسی با دامنه، فاز و فرکانس مشخص نمایش داده می شود و کار اصلی انکدر ارسال این دامنه، فاز و فرکانس ها برای دیکدر می باشد. اما ارسال کل اطلاعات به دلیل محدودیت بیت ریت امکان پذیر نمی باشد چرا که قسمت های شبه نویز تعداد این سیگنال های سینوسی را بالا می برد و معمولاً با توجه به بیت ریت قابل قبول، تعداد مشخصی از این اطلاعات برای ساخت سیگنال در دیکدر ارسال می گردد. میزان اختلاف سیگنال اصلی و سیگنال ساخته شده در دیکدر با استفاده از اطلاعات دامنه، فاز و فرکانس از نظر شنیداری دارای اعوجاج می باشد و به همین دلیل از یک کد کننده برای کد کردن سیگنال خطا که اختلاف سیگنال اصلی و سیگنال ساخته شده با اطلاعات سیگنال های سینوسی در دیکدر می باشد استفاده می شود.

در حالت معمول برای کد کردن این سیگنال باقی مانده از باند های برابر فرکانسی و محاسبه انرژی این باند ها و کوانتیزه کردن و ارسال آن برای دیکدر استفاده می شود اما در این مرجع از الگوریتم PLP برای کد کردن این سیگنال باقی مانده که از نظر شنیداری و ادراکی اهمیت ویژه ای دارد استفاده شده است. استفاده از الگوریتم PLP در این روش این مزیت را داراست که از اثر ماسک گذاری [3] یعنی حذف سیگنال هایی که شنیده نمی شوند استفاده شده است و به این ترتیب در کد کردن سیگنال باقی مانده فقط سیگنال هایی کد می شوند که از نظر ادراکی در سیگنال ساخته شده در دیکدر شنیده می شوند. بنابراین با حذف سیگنال هایی که شنیده نمی شوند نرخ بیت مورد نیاز برای کد کردن سیگنال کاهش می یابد. در این روش سیگنال باقی مانده توسط ضرایب LPC که از طریق الگوریتم PLP استخراج شده اند کد بندی می شوند. البته استخراج و پیدا کردن سطح ماسک کردن در این روش از پیچیدگی های آن می باشد. باید یاد آور شد که بیت ریت مورد استفاده در این کد کننده به طور کلی بالای ۲۳ کیلو بیت بر ثانیه می باشد و الگوریتم معرفی شده گرچه مشابهت هایی با برخی قسمت های استفاده شده در این پایان نامه دارد اما روش استخراج استفاده شده برای کد بندی سیگنال صوت و نه سیگنال گفتار می باشد و همچنین از پیچیدگی های استخراج و استفاده از اثر ماسک گذاری در این پایان نامه استفاده نشده است.

• استفاده از ضرایب PLP برای افزایش کیفیت تشخیص اتوماتیک گفتار

در مرجع ۴۲ از پارامتر های استخراج شده از طریق الگوریتم PLP و سایر الگوریتم های استخراج ویژگی برای تشخیص اتوماتیک گفتار به منظور استفاده در فشرده سازی سیگنال و بالا بردن کیفیت سیگنال دریافتی در دیکدر استفاده شده است. در این مقاله از ۶ ضریب PLP در استخراج ضرایب LPC و ضرایب^۱ LSF استفاده شده و اثر استفاده از آن بر روی میزان افزایش دقت سیستم تشخیص

^۱ Line Spectral Frequency

اتوماتیک گفتار بررسی شده است. همچنین در مرجع ۴۳ برای افزایش کارایی تشخیص اتوماتیک گفتار، روش استفاده از ضرایب MFCC وزن دار شده استفاده شده است که در این روش علاوه بر پارامترهای مربوط به کد کننده ضرایب فوق نیز مورد استفاده قرار گرفته اند و میزان بهبود کیفیت تشخیص گفتار در سیگنال خروجی تولید شده بررسی گردیده است.

در مرجع ۴۴ برای ارسال ضرایب PLP و LPC به منظور افزایش کیفیت سیگنال ساخته شده و کاهش بیت ریت مورد نیاز برای ارسال، از یک مالتی پلکسر برای ارسال همزمان این ضرایب استفاده شده است و افزودن ضرایب PLP به رشته بیت ارسالی به میزان 400bps میزان نرخ بیت را افزایش داده است. در این مرجع از ۵ ضریب PLP برای هر فریم استفاده شده است. در مرجع ۴۵ از روش ^۱PLP که یک روش وقت گیر به نسبت روش های معمولی است برای افزایش کیفیت سیگنال ساخته شده با استفاده از تخصیص بازه فرکانسی با دقت های خاص معرفی شده است. از دیگر مراجعی که در آنها از الگوریتم و ضرایب PLP و ویژگی های مربوط به تشخیص اتوماتیک گفتار در کد کننده ها مورد استفاده قرار گرفته است، می توان به مراجع ۴۶، ۴۷، ۴۸ و ۴۹ اشاره نمود که در آنها برای بهبود کیفیت شنیداری سیگنال و بهبود ضریب تشخیص اتوماتیک سیگنال ساخته شده در دیکدر از الگوریتم PLP و الگوریتم های مشابه، به شیوه های مختلف استفاده شده است.

¹ Pure Linear Prediction

۷-۳. کد کننده پیشنهادی

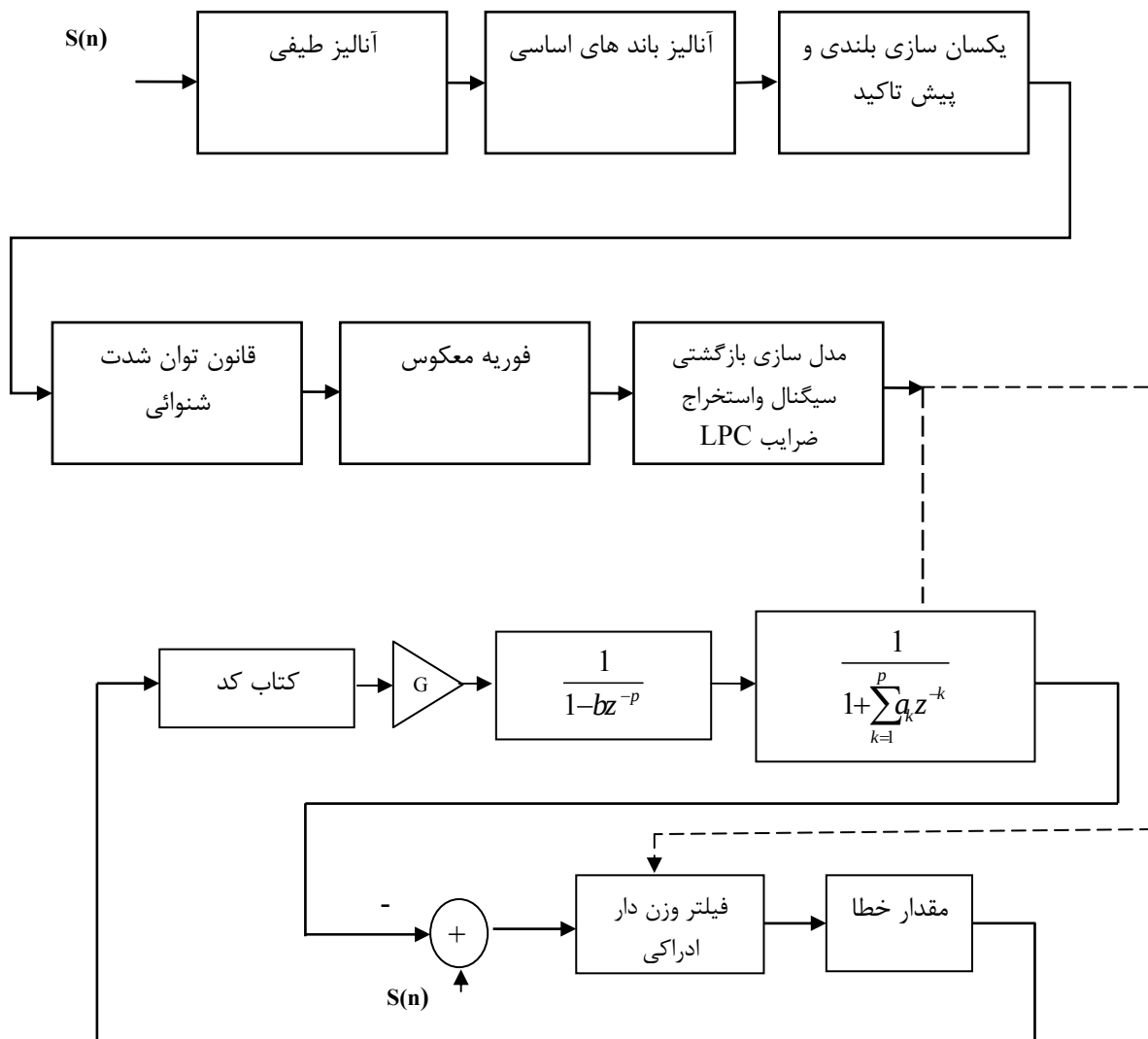
همان طور که قبلا بیان شد به طور کلی سیگنال صحبت را می توان با مدل یک منبع تحریک و یک لوله صوتی که به طور سری با یکدیگر قرار دارند مدل سازی نمود. در این مدل سازی خواص لوله صوتی به عنوان مهمترین پارامتر نقش بسزایی در کیفیت سیگنال ساخته شده را دارد و به همین دلیل استخراج پارامتر های آن از اهمیت ویژه ای برخوردار می باشد.

در کد کننده بهبود یافته کاربردی شده از الگوریتم PLP برای جداسازی سیگنال تحریک و خواص لوله صوتی و تاکید بر روی فرکانس های با اهمیت بیشتر برای تخمین پارامترهای LPC لوله صوتی استفاده شده است که معادلات مربوط به پردازش کپسترال که برای جدا سازی سیگنال تحریک از طیف لوله صوتی در فصل پنجم ارائه شده است. علت استفاده از این روش و ایده بکارگیری آن در این است که در استفاده از الگوریتم PLP اثرات فیزیولوژی گوشی انسان و خواص ادراکی آن لحاظ شده بود و در اکثر مقالات از اثر حذف نویز این الگوریتم استفاده زیادی شده بود.

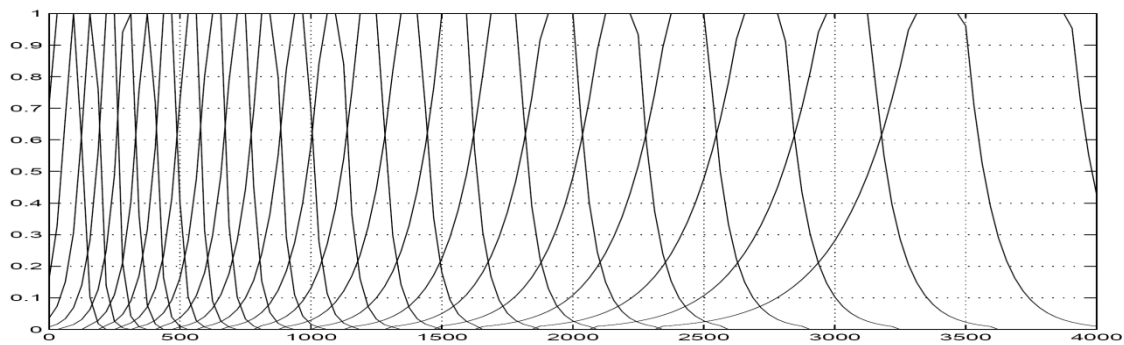
در این کد کننده ما پردازش های اولیه روی سیگنال را بر مبنای پردازش PLP انجام داده و برای مدل سازی فیلتر ساخت فرمنت، فیلتر وزن دار ادراکی، فیلتر نهائی که جزو قسمت های اصلی کد کننده سلپ می باشند، و همچنین سایر پردازش هایی که از ضرایب تخمین خطی استفاده می شود از روش PLP برای تخمین پارامتر های LPC آن بهره می گیریم. باید توجه داشت بسیاری از پارامترهای استفاده شده در کد کننده استاندارد اعم از پارامترهای استفاده شده در فیلتر ساخت فرمنت بهینه شده، فیلتر وزن دار ادراکی و فیلتر نهائی با انجام آزمایش های تجربی بر روی کد کننده عمومی سلپ بدست آمده اند و ما در این پایان نامه آزمایش های مختلف را برای پیدا کردن پارامترهای بهینه انجام داده ایم که نتیجه آنها در قسمت نتایج ارائه می گردد.

به طور کلی بلوک دیاگرام استفاده شده برای پردازش سیگنال برای استخراج ویژگی های سیگنال از طرق این الگوریتم مطابق شکل ۷-۱ می باشد. مطابق شکل ارائه شده در کد کننده بهبود یافته سیگنال صحبت ورودی ابتدا با پنجره هائی به طول ۲۴۰ نمونه (۳۰ میلی ثانیه) فریم بندی شده و پنجره همینگ به آن اعمال میگردد و پس از عبور از فیلتر بالا گذر درجه اول مورد پردازش طیفی قرار گرفته و طیف انرژی آن برای هر فریم محاسبه می گردد. پردازش باند های اساسی برای تاکید بر روی فرکانس های دارای اهمیت بیشتر در سیستم شنوایی انسان بر روی آن انجام می گیرد. فیلتر

بانک هایی که برای تبدیل باند های اساسی در PLP استفاده شده است مطابق شکل ۷-۲ می باشد که این نمونه از فیلتر بانک برای تبدیل به حوزه بارک در این پایان نامه استفاده شده است.



شکل ۷-۱. بلوک کلی کد کننده اصلاح شده



شکل ۷-۲. بانک فیلتر بارک استفاده شده

در مراحل بعدی قوانین یکسان سازی بلندی و شنوایی انسان بر روی سیگنال انجام شده و فوریه معکوس بر روی سیگنال موجود اعمال می گردد و ضرایب LPC از آن استخراج می گردد و در بلوک های ساخت فرمنت و فیلتر وزن دار اعمال می گردد. تعداد ضرایب LPC استخراج شده برای استفاده در حالت کارکرد عادی ۱۰ ضریب می باشد که البته با تعداد ضرایب کمتر مورد آزمایش قرار گرفته شده است و نتایج تست در قسمت نتایج آورده شده است. همچنین این ضرایب به طور مستقیم در دیکدر کد کننده بهبود یافته در قسمت فیلتر ساخت فرمنت و فیلتر نهائی مورد استفاده قرار گرفته است. همانطور که بیان گردید از آنجا که پارامترهای استخراج شده با پارامترهای استخراج شده در کد کننده استاندارد در بسیاری موارد دارای اختلاف هایی می باشد که برای بهبود کیفیت سیگنال با پارامترهای موجود برخی از پارامترهای کد کننده استاندارد مورد باز نگری قرار گرفت و نتایج تست ها آورده شده است.

روش پیاده سازی کد کننده بهبود یافته با استفاده از شبیه سازی ها و پیاده سازی های استفاده شده در کد کننده فصل قبل بوده است و در تخصیص بیت ها روش استاندارد تخصیص رعایت شده است و در آزمایش های خاص انجام شده در برخی موارد پارامترهای کد کننده و تعداد بیت تخصیص داده شده تغییر داده شده است که در قسمت نتایج آورده شده است.

۴-۷. دیتا بیس مورد استفاده

برای بررسی کارایی کد کننده بهبود یافته، استفاده از صدا های مختلف که بتواند کارایی کد کننده مذکور را مورد ارزیابی قرار داده و میزان کیفیت آن را در مقایسه با کد کننده سلپ عمومی مشخص نماید مورد توجه قرار گرفت.

دیتا بیس مورد نظر باید جنبه های مختلف بررسی کیفیت سیگنال تولید شده را در شرایط مختلف در بر می گرفت و به همین منظور ابتدا تصمیم گرفته شد تا از دیتا بیس های استاندارد که به زبان های خارجی بودند استفاده گردد اما در مراحل اولیه پیاده سازی کد کننده بهبود یافته، مشخص گردید با استفاده از زبان های غیر از زبان فارسی برای مقایسه و نمره دهی، مقایسه کیفیت شنیداری سیگنال ساخته شده براحتی امکان پذیر نمی باشد چرا که متوجه شدن جملات گفته شده به زبان دیگر و تمییز جملات خود جزو مشکلات شنیداری می باشد و نمی توان مقایسه درستی در مواردی که مقداری مشابهت وجود دارد انجام داد. برای مثال در دیتا بیس orator جملات به زبان آلمانی هستند و نمی توان آنها را برای شنیدن و نمره دادن به افراد فارسی زبان که شناخت کافی از زبان آلمانی ندارند ارائه نمود از طرف دیگر تهیه خود دیتا بیس ها هم براحتی امکان پذیر نبود به همین دلیل تصمیم گرفته شده تا دیتا بیسی برای بررسی کیفیت کد کننده بهبود یافته در شرایط مختلف و مقایسه آن با کد کننده عمومی سلپ تهیه گردد. در تهیه این دیتا بیس از ۵۰ نفر درخواست گردید تا صدای ضبط شده خود را که حاوی کلمات و جملات مختلف طولانی و کوتاه، با صدای بلند و صدای آرام، شرایط هیجان زده و شاد، حالت ناراحت در اختیار قرار دهند تا برای بررسی کیفیت کد کننده مورد استفاده قرار گیرد.

این افراد که جنسیت مذکر و مونث دارند دارای سنین مختلف هستند و از همه آنها خواسته شده است تا کلمه سلام را به طور مشترک و کلمات و جمله های مختلف دیگر را بیان نمایند.

همچنین در این دیتا بیس از قسمت های مختلف کلیپ های موسیقی فارسی و انگلیسی که در آن گوینده به همراه پخش موسیقی وجود دارد برای تست در شرایط موزیکال استفاده شده است که این قسمت ها نیز با طول های مختلف زمانی ضبط گردیده اند و شامل ۳۰ قسمت ضبط شده مختلف فارسی و ۳۰ قسمت ضبط شده انگلیسی می باشد که در این قسمت ها شرایط موسیقی زمینه به صورت کم رنگ و پر رنگ وجود دارد.

در تست های انجام گرفته بر روی این دیتا بیس، شرایط مختلف و اجزای مختلف دیتا بیس مورد تست قرار گرفته اند و همچنین در تست هایی که اثر نویز بررسی شده است توسط نرم افزار نویز با درصد های مختلف بر روی این اجزا اعمال شده است و مورد استفاده قرار گرفته است.

۷-۵. نتایج عملی

برای بررسی کیفیت ساخت سیگنال و دیگر پارامتر های کد کننده بهبود یافته، سیگنال های ساخته شده و نرخ بیت استفاده شده در این کد کننده با کد کننده پیاده سازی شده سلپ عمومی مقایسه می شود. شبیه سازی های انجام شده بر روی کلیه فایل های صوتی دیتا بیس انجام گرفته است و نتیجه گیری های انجام شده برای میانگین نتایج می باشد. در این کد کننده بنا به ماهیت آن از معیار DMOS برای بررسی کیفیت سیگنال ساخته شده استفاده می شود و جدول نتایج DMOS جدولی است که میانگین نمره های داده شده برای سیگنال ساخته شده توسط گرفتن میانگین نمره های است که توسط ۴۰ نفر به سیگنال ساخته شده داده شده است. در تست DMOS انجام شده برای نتایج تست از ۴۰ نفر خواسته شد تا برای سیگنال هایی که برای آنها پخش می شود نمره هایی بنا به کیفیت سیگنال شنیده شده در مقایسه با سیگنال اصلی پخش شده، بدهند. البته آنها هیچ اطلاعی از کد کننده استفاده شده برای ساخت سیگنال نداشتند و میانگین نمره های آنها که بین ۵-۰ بود، به عنوان نتیجه تست در جدول های ارائه شده مورد استفاده قرار گرفته است. همچنین از آنجا که تعداد فایل های صوتی موجود و شکل های مربوط به این فایل ها نسبتا زیاد است، برای مشخص شدن شکل سیگنال های کد شده، شکل سیگنال اصلی، شکل سیگنال خروجی ساخته شده توسط کد کننده سلپ عمومی و سلپ بهبود یافته، FFT سیگنال های اصلی و ساخته شده فقط برای برخی از سیگنال های موجود در دیتا بیس در این پایان نامه آورده شده است.

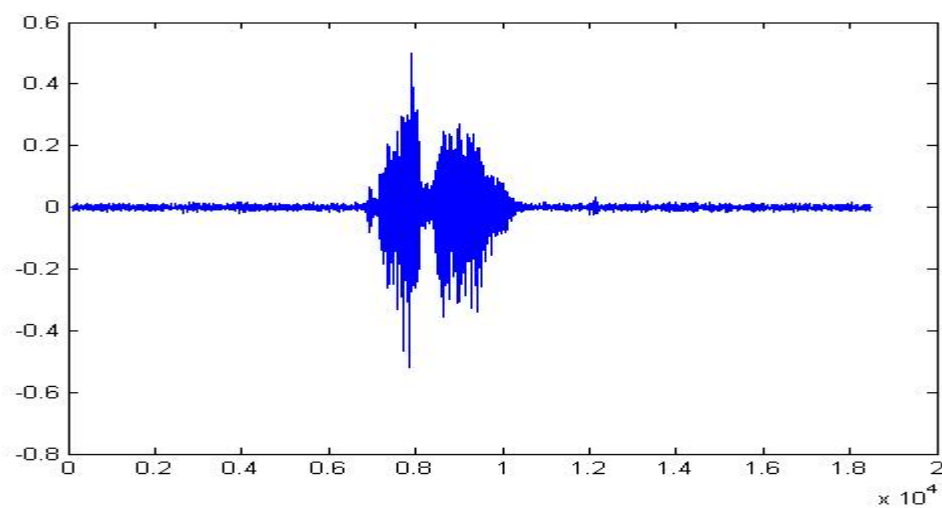
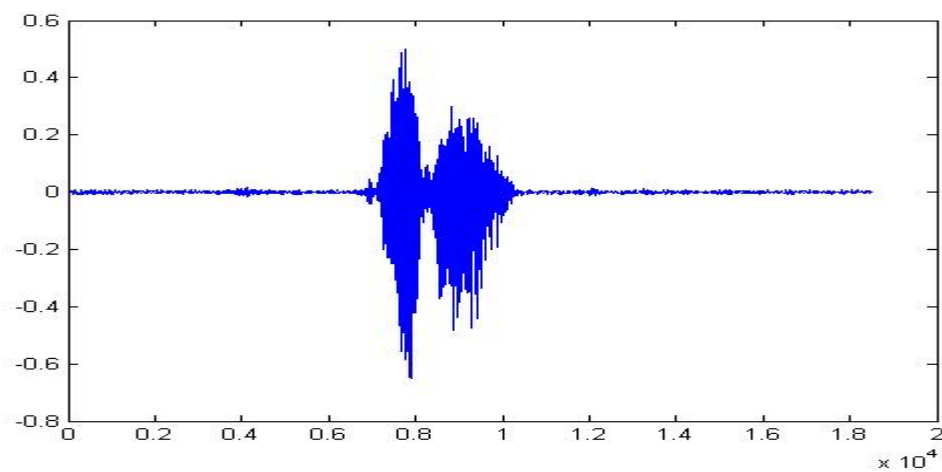
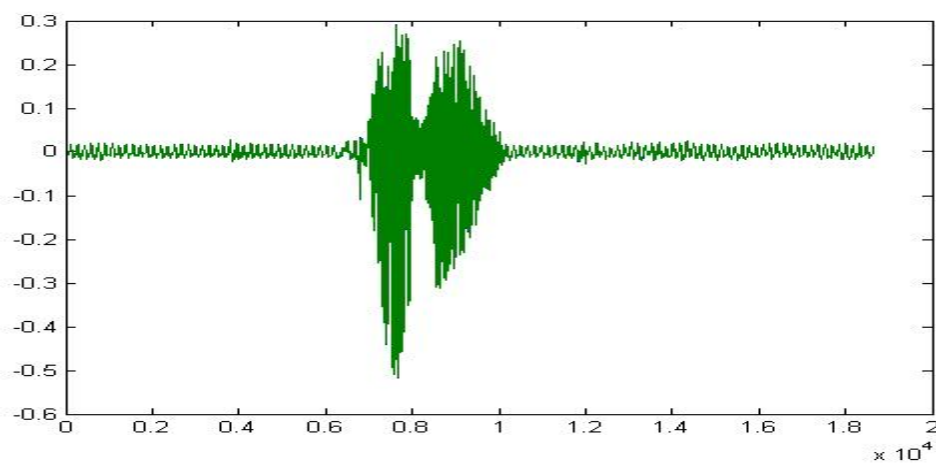
۷-۵-۱. نتایج برای سیگنال صحبت بدون نویز و موسیقی

در تست انجام شده برای ۵۰ فایل صوتی ضبط شده در دیتا بیس با مشخصات پارامتر های معرفی شده در کد کننده استاندارد با مشخصات $\gamma=0.8$ برای فیلتر وزن دار ادراکی و $\alpha=0.8$ و $\beta=0.5$ برای فیلتر نهایی در دیکدر سیگنال های خروجی ساخته شده مورد تست ۴۰ نفر قرار گرفت و میانگین نمره های داده شده برای این ۵۰ فایل صوتی مطابق جدول ۷-۱ می باشد. لازم به ذکر است طبق اظهار نظر افرادی که در تست شرکت کرده بودند سیگنال ساخته شده توسط کد کننده بهبود یافته، گرچه دارای مقدار بسیار کمی نویز بود اما به نسبت سیگنال ساخته شده توسط کد کننده سلپ استاندارد دارای وضوح بیشتری بود به طوری که کلمات گفته شده واضح تر بودند و همچنین از نظر شناسائی گوینده کیفیت مناسب تری داشت.

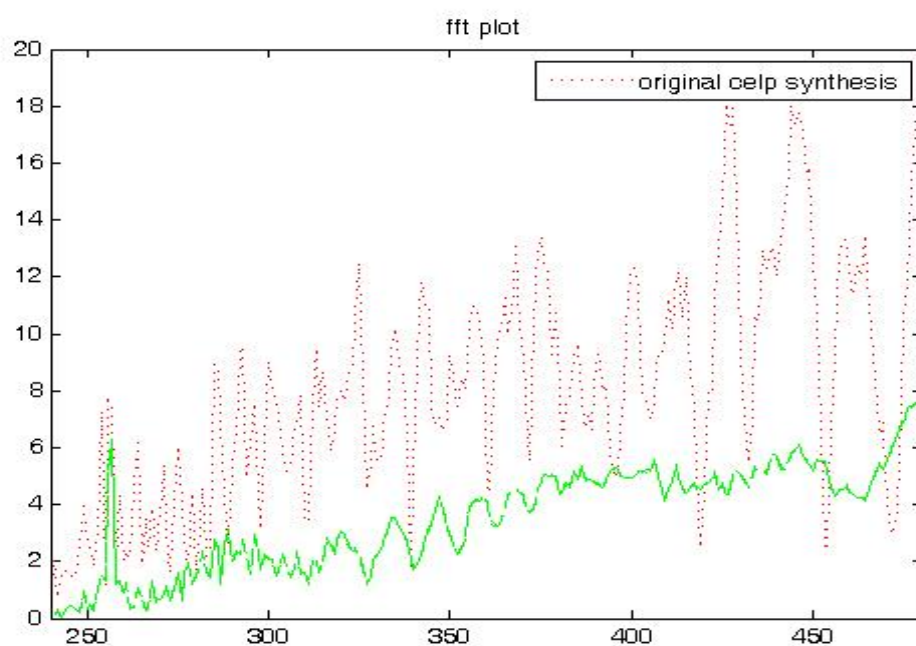
شکل زمانی سیگنال اصلی و سیگنال ساخته شده در کد کننده سلپ عمومی و سلپ بهبود یافته و شکل فرکانسی سیگنال اصلی و سیگنال ساخته شده با سلپ عمومی و سلپ بهبود یافته برای بازه [240-480] به ترتیب در شکل های ۷-۳، ۷-۴، ۷-۵ آورده شده است. همانطور که در شکل زمانی سیگنال ها در شکل ۷-۳ دیده می شود شکل کلی سیگنال اولیه در هر دو کد کننده به طور کلی حفظ می گردد. در این شکل برای ساخت سیگنال خروجی دامنه نمونه ها برای شنیده شدن راحت تر دو برابر شده است. همچنین محور افقی تعداد نمونه ها و محور عمودی دامنه نمونه ها را نمایش می دهد. در حوزه فرکانس در نقاط ماکزیمم فرکانسی (فرمنت ها) مشاهده می شود که در بسیاری از نقاط برای سیگنال اصلی و سیگنال ساخته شده این نقاط با یکدیگر مطابقت دارند.

جدول ۷-۱. نتایج سیگنال بدون نویز و موسیقی

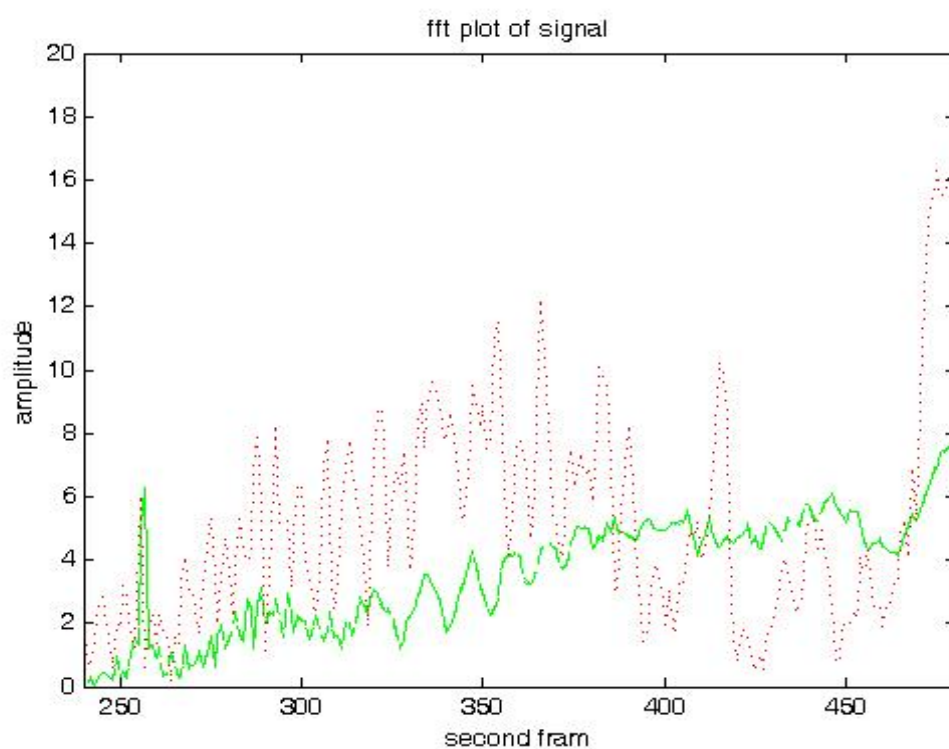
میانگین نمره های DMOS	کد کننده سلپ عمومی	کد کننده سلپ بهبود یافته
	۴/۶۴	۴/۵۵



شکل ۳-۷. سیگنال اصلی (شکل اول) سیگنال ساخته شده با کد کننده عمومی سلپ (شکل دوم) و کد کننده بهبود یافته (شکل سوم)



شکل ۷-۴. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلپ (نقطه چین)



شکل ۷-۵. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلپ (نقطه چین)

۷-۵-۲. نتایج برای سیگنال صحبت با افزودن نویز و بدون موسیقی

تست انجام شده در این مرحله به گونه ای بود که برای فایل های صوتی تست قبل که شامل ۵۰ فایل صوتی ضبط شده بود از طریق دستور افزودن نویز سفید گوسی جمع شونده با نرم افزار MATLAB (دستور AWGN) با درصد سیگنال به نویز مختلف کار کرد کد کننده عمومی سلپ و کد کننده بهبود یافته مورد تست قرار گرفت. در این تست ها نیز از پارامتر های تست قبل برای تست استفاده گردید. ($\gamma=0.8$ برای فیلتر وزن دار ادراکی و $\alpha=0.8$ و $\beta=0.5$ برای فیلتر نهایی در دیکدر و استفاده از ۱۰ ضریب LPC برای هر فریم) در تست انجام شده از مقادیر سیگنال به نویز از ۱۴ دسی بل تا ۲۰ دسی بل مورد تست قرار گرفت چرا که در سیگنال به نویز های پائین تر سیگنال صحبت خیلی به سختی در سیگنال حاوی نویز قابل تشخیص بود و یا هیچ صدای برای بررسی کیفیت کار کد کننده وجود نداشت. در این تست سیگنال های خروجی ساخته شده با سیگنال به نویز ۱۷ دسی بل مورد تست ۴۰ نفر قرار گرفت و میانگین نمره های داده شده برای این ۵۰ فایل صوتی مطابق جدول ۷-۲ می باشد. در این تست نیز طبق اظهار نظر افرادی که در تست شرکت کرده بودند سیگنال ساخته شده توسط کد کننده بهبود یافته، نسبت سیگنال ساخته شده توسط کد کننده سلپ استاندارد دارای وضوح بیشتری بود و طبق جدول نتایج، کارکرد کد کننده بهبود یافته در شرایط افزودن نویز سفید جمع شونده به نسبت کد کننده عمومی سلپ نتایج بهتری داشت و اختلاف قابل تشخیصی در نتایج شنیده می شد و کیفیت شنیداری بهتر و همچنین شناسایی بهتر گوینده در نتایج دیده می شد.

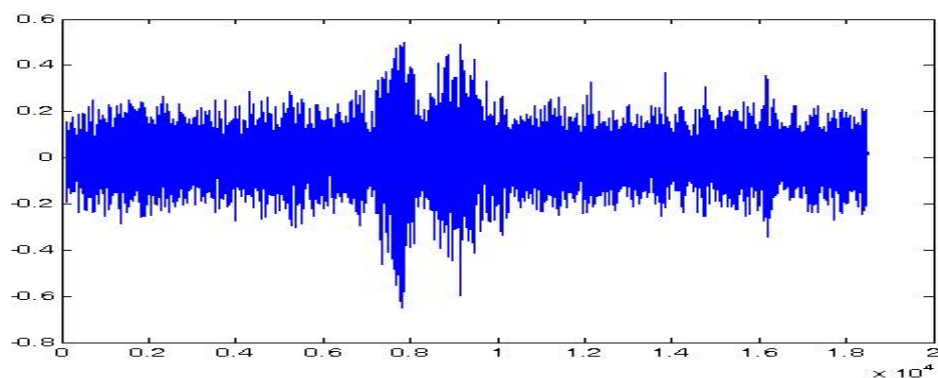
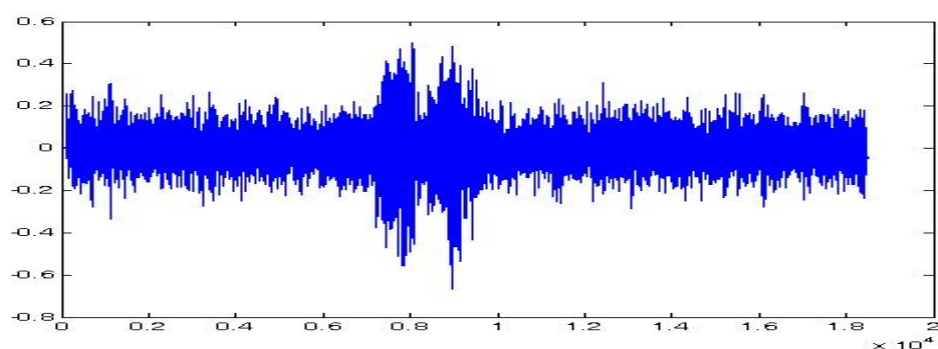
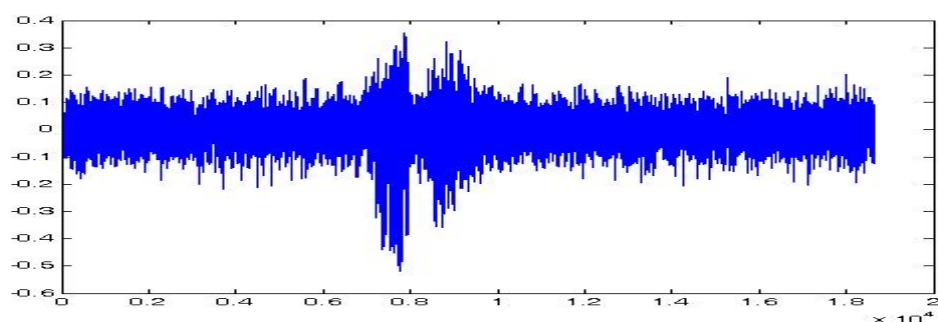
شکل زمانی سیگنال اصلی و سیگنال ساخته شده در کد کننده سلپ عمومی و سلپ بهبود یافته و شکل فرکانسی سیگنال اصلی و سیگنال ساخته شده با سلپ عمومی و سلپ بهبود یافته برای بازه [240-480] به ترتیب در شکل های ۷-۶، ۷-۷، ۸-۷ آورده شده است. همانطور که در شکل زمانی سیگنال ها در شکل ۷-۶ دیده می شود شکل کلی سیگنال اولیه در هر دو کد کننده به طور کلی حفظ می گردد ولی در کد کننده بهبود یافته شکل سیگنال ساخته شده شباهت بیشتری به شکل سیگنال اصلی دارد. در این شکل ها نیز برای ساخت سیگنال خروجی دامنه نمونه ها برای شنیده شدن راحت تر دو برابر شده است. در این شکل ها محور افقی تعداد نمونه ها و محور عمودی دامنه نمونه ها را نمایش می دهد. با توجه به شکل ۷-۷ و ۸-۷ مشاهده می شود که به ازای فرمنت اول

فصل هفتم – کد کننده پیشنهادی و نتایج عملی

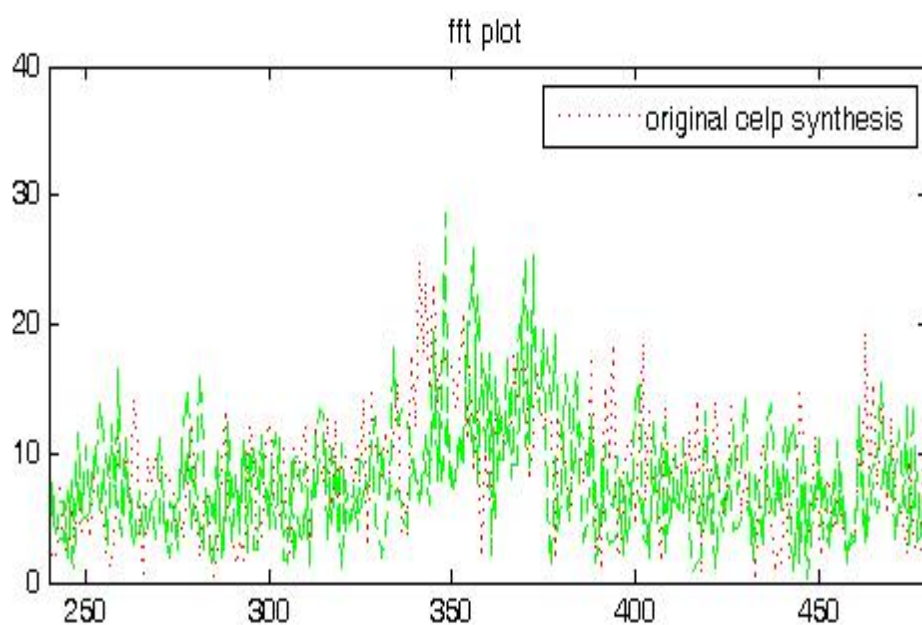
(نقطه ماکزیمم فرکانسی اول) سیگنال ساخته شده با کد کننده بهبود یافته شباهت بیشتری با سیگنال اصلی در حوزه فرکانس دارد.

جدول ۲-۷. نتایج سیگنال با سیگنال به نویز ۱۷ دسی بل و بدون موسیقی

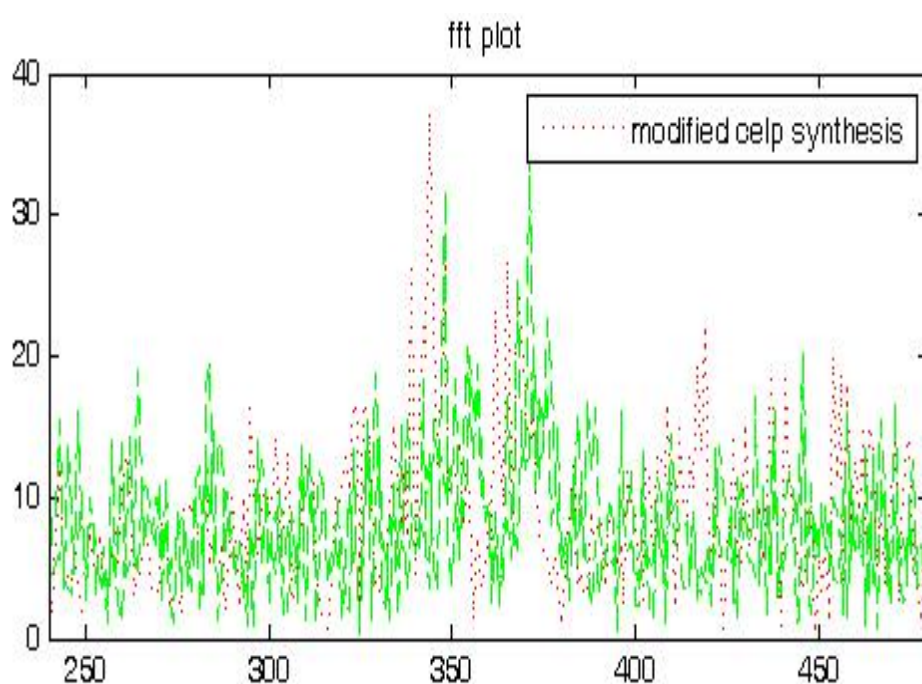
میانگین نمره های DMOS	کد کننده سلف عمومی	کد کننده سلف بهبود یافته
	۴/۰۱	۴/۳۸



شکل ۶-۷. سیگنال اصلی (شکل اول) سیگنال ساخته شده با کد کننده عمومی سلف (شکل دوم) و کد کننده بهبود یافته (شکل سوم) با سیگنال به نویز ۱۷ دسی بل



شکل ۷-۷. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلپ (نقطه چین)



شکل ۷-۸. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلپ (نقطه چین)

۷-۵-۳. نتایج برای سیگنال صحبت همراه با موسیقی بدون نویز

در این تست سیگنال های صدای همراه با موسیقی مورد آزمایش قرار گرفت. سیگنال های مورد تست قرار گرفته شامل فایل های ذخیره شده صوتی از کلیپ های فارسی و انگلیسی می باشند که با فرمت فایل های صوتی (wav) ذخیره شده اند و طول های مختلفی دارند. در برخی از این فایل ها از گوینده هایی درخواست شده است که به صورت جداگانه همراه با پخش موسیقی نیز صحبت نمایند تا کیفیت کد کردن صدای آنها نیز مورد تست قرار گیرد. همچنین در انتخاب کلیپ های صوتی، کلیپ هایی با موسیقی زمینه زیاد و کم برای ذخیره شدن انتخاب گردید و تست ها برای موسیقی زمینه با شدت های مختلف انجام گرفت.

در این تست ها نیز از پارامتر های آزمایش قبل استفاده گردید. ($\gamma=0.8$ برای فیلتر وزن دار ادراکی و $\alpha=0.8$ و $\beta=0.5$ برای فیلتر نهایی در دیکدر و استفاده از ۱۰ ضریب LPC برای هر فریم) در تست های انجام شده از ۴۰ نفر درخواست گردید تا کیفیت سیگنال های ساخته شده را مورد ارزیابی قرار دهند. و طبق اظهار نظر افراد تست شونده سیگنال ساخته شده با کد کننده بهبود یافته دارای مقدار کمی نویز اما کیفیت تفکیک مناسبی داشت اما سیگنال ساخته شده با کد کننده عمومی سلپ نویز کمتری داشت اما حالت صدای گوینده همراه با ناهمواری هایی بود و تشخیص گوینده در کد کننده بهبود یافته بهتر بود.

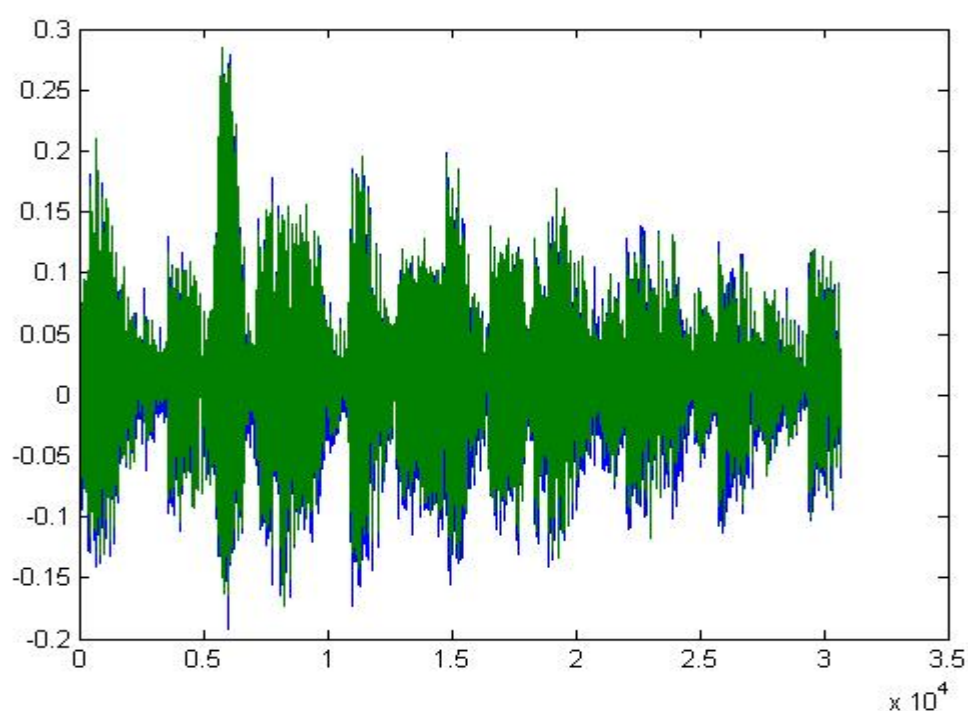
نتایج تست مذکور برای ۴۰ نفر بصورت میانگین در جدول ۷-۳ آورده شده است. مطابق جدول و مشاهده مقدار میانگین نمره های داده شده کارکرد دو کد کننده تقریباً برابر است البته در این تست افراد مختلف نمره های مختلف زیادی به کد کننده ها دادند و برخی از افراد نمره بیشتر به کد کننده عمومی و برخی به کد کننده بهبود یافته دادند که میانگین نمره ها در جدول ذکر شده است.

شکل زمانی سیگنال اصلی و سیگنال ساخته شده در کد کننده سلپ عمومی و سلپ بهبود یافته و شکل فرکانسی سیگنال اصلی و سیگنال ساخته شده با سلپ عمومی و سلپ بهبود یافته برای بازه [240-480] برای یکی از فایل های مورد آزمایش قرار گرفته، به ترتیب در شکل های ۷-۹، ۷-۱۰، ۷-۱۱ و ۷-۱۲ آورده شده است. همانطور که در شکل زمانی سیگنال ها در شکل ۷-۹ و ۷-۱۰ دیده می شود شکل کلی سیگنال اولیه در هر دو کد کننده به طور کلی حفظ می گردد ولی در کد کننده بهبود یافته شکل سیگنال ساخته شده شباهت بیشتری به شکل سیگنال اصلی دارد. در این شکل ها

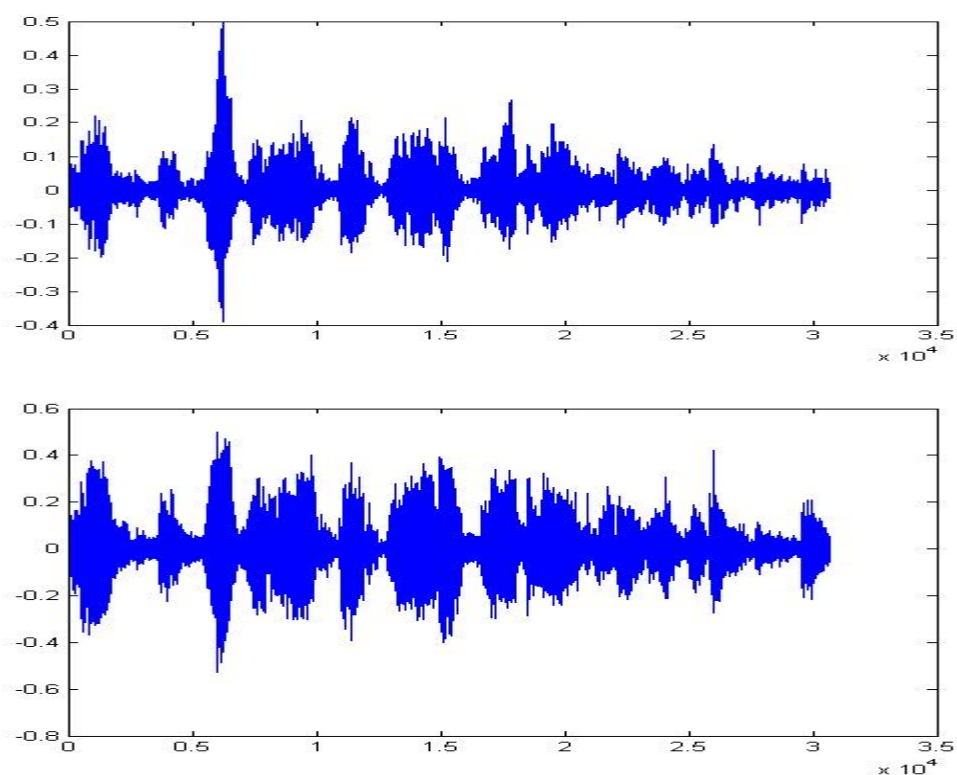
نیز برای ساخت سیگنال خروجی دامنه نمونه ها برای شنیده شدن راحت تر دو برابر شده است. در این شکل ها محور افقی تعداد نمونه ها و محور عمودی دامنه نمونه ها را نمایش می دهد. با مشاهده طیف فرکانسی سیگنال های ساخته شده با دو کد کننده عمومی و بهبود یافته در شکل های ۷-۱۱ و ۷-۱۲ مشاهده می شود که در مقایسه با طیف سیگنال اصلی نقاط ماکزیمم سیگنال های ساخته شده تقریباً مانند هم هستند.

جدول ۷-۳. نتایج سیگنال صحبت به همراه موسیقی

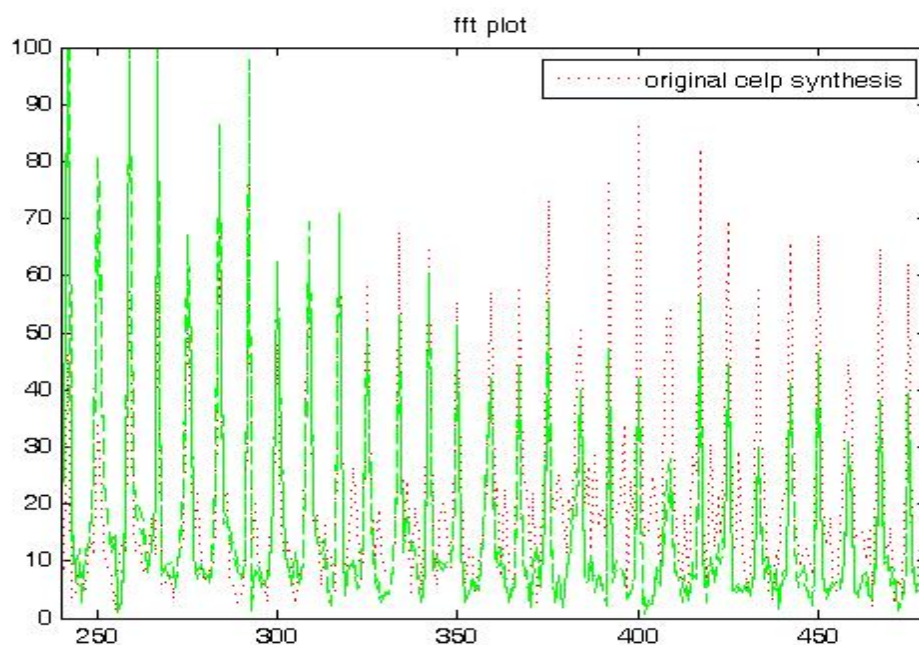
میانگین نمره های DMOS	کد کننده سلپ عمومی	کد کننده سلپ بهبود یافته
	۳/۹۵	۴/۰۷



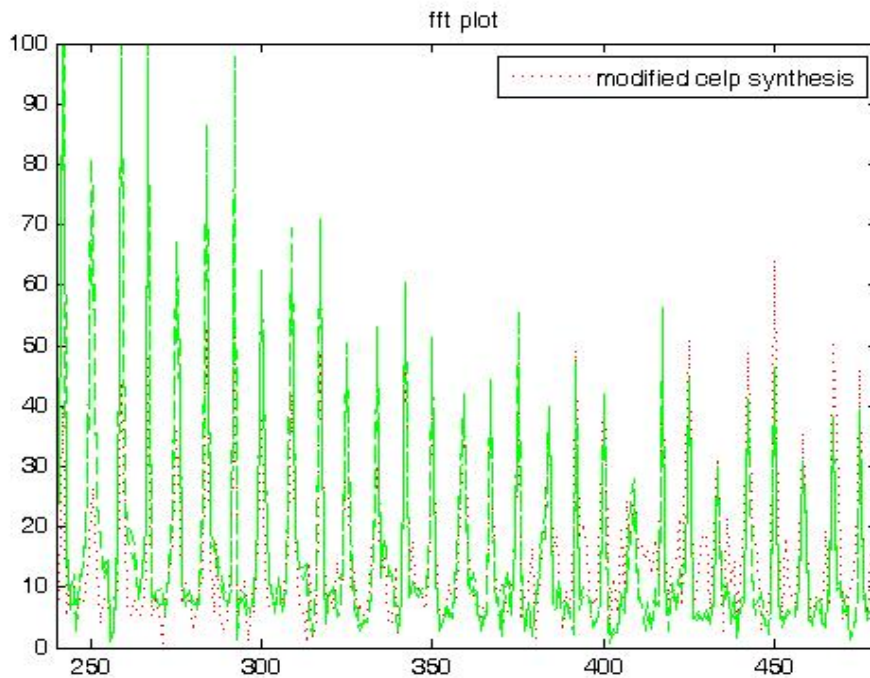
شکل ۷-۹. سیگنال اصلی صدای همراه با موسیقی



شکل ۷-۱۰. سیگنال ساخته شده سیگنال همراه با موسیقی کد کننده عمومی سلپ (شکل بالا) کد کننده بهبود یافته (شکل پایین)



شکل ۷-۱۱. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلپ (نقطه چین) سیگنال همراه با موسیقی



شکل ۷-۱۲. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلپ (نقطه چین) سیگنال همراه با موسیقی

۷-۵-۴. نتایج برای سیگنال صحبت همراه با موسیقی و نویز

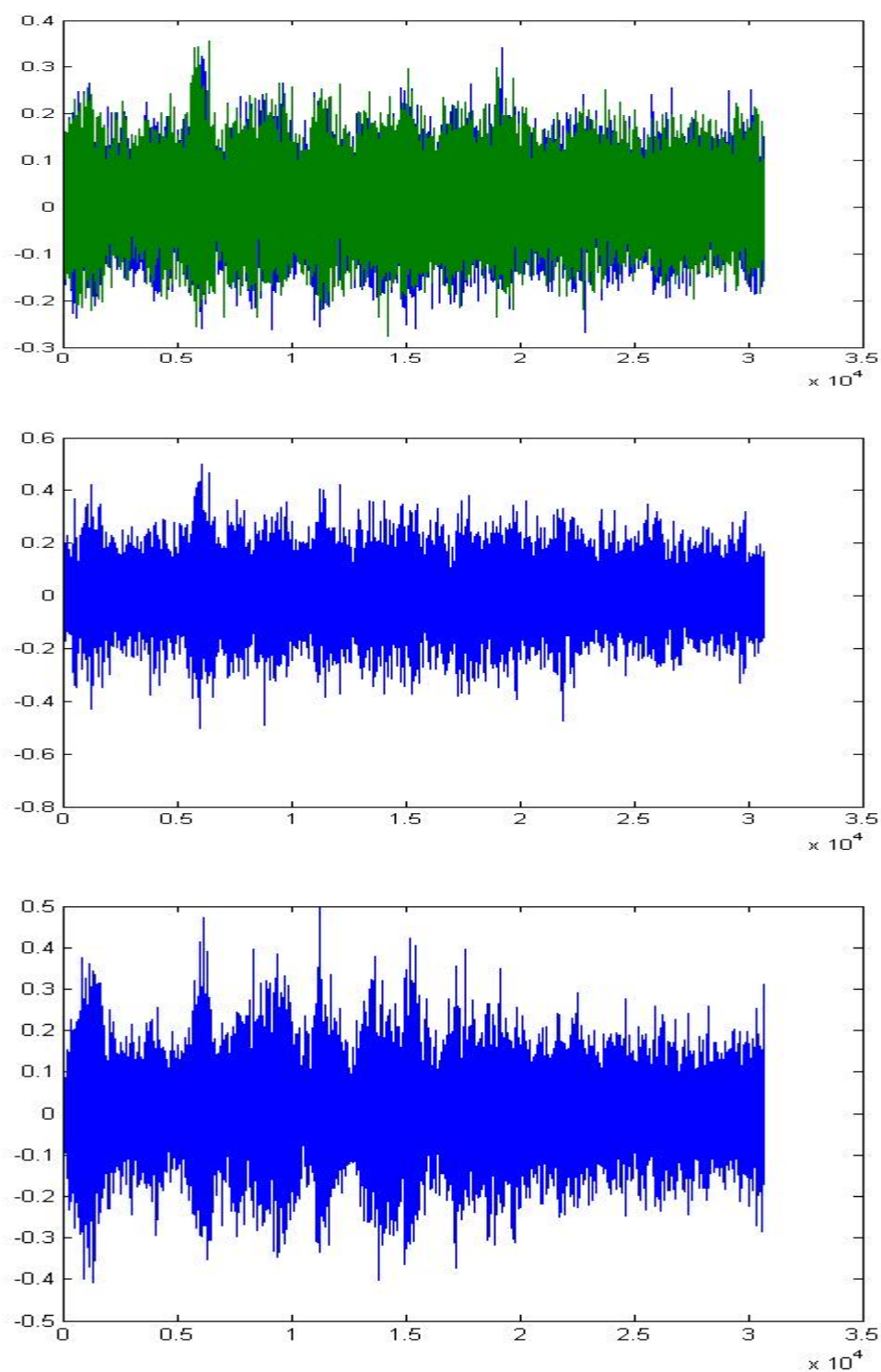
در این تست فایل های صوتی استفاده شده در قسمت ۷-۴-۳ برای بررسی کارکرد کد کننده بهبود یافته و کد کننده عمومی سلپ مورد آزمایش قرار گرفت. تست انجام شده در این مرحله به گونه ای بود که برای فایل های صوتی از طریق دستور افزودن نویز سفید گوسی جمع شونده با نرم افزار MATLAB (دستور AWGN) با درصد سیگنال به نویز مختلف نویز به سیگنال اصلی که حاوی موسیقی زمینه نیز بود افزوده شد. در این تست ها نیز از پارامتر های تست قبل برای تست استفاده گردید. ($\gamma=0.8$ برای فیلتر وزن دار ادراکی و $\alpha=0.8$ و $\beta=0.5$ برای فیلتر نهایی در دیکدر و استفاده از ۱۰ ضریب LPC برای هر فریم) همچنین در تست انجام شده، مقادیر سیگنال به نویز از ۱۴ دسی بل تا ۲۰ دسی بل مورد تست قرار گرفت و دلیل این مقادیر این بود که در سیگنال به نویز های پائین تر سیگنال صحبت همراه با موسیقی خیلی به سختی در سیگنال حاوی نویز قابل تشخیص بود و یا هیچ صدای برای بررسی کیفیت کار کد کننده وجود نداشت. در این تست سیگنال های خروجی ساخته شده با سیگنال به نویز ۱۷ دسی بل مورد تست ۴۰ نفر قرار گرفت و میانگین نمره های داده شده برای این ۵۰ فایل صوتی مطابق جدول ۷-۴ می باشد. در این تست نیز طبق اظهار نظر افرادی

که در تست شرکت کرده بودند سیگنال ساخته شده توسط کد کننده بهبود یافته، نسبت سیگنال ساخته شده توسط کد کننده سلپ استاندارد دارای وضوح بیشتری بود و در نواحی که سیگنال صحبت همراه با موسیقی وجود داشت کد کننده سلپ بهبود یافته سیگنال مناسب تری را تولید می نمود. طبق جدول نتایج، کارکرد کد کننده بهبود یافته در شرایط افزودن نویز سفید جمع شونده و وجود موسیقی زمینه، به نسبت کد کننده عمومی سلپ نتایج بهتری داشت و اختلاف قابل تشخیصی در نتایج شنیده می شد و کیفیت شنیداری بهتر و همچنین شناسایی بهتر گوینده در نتایج دیده می شد.

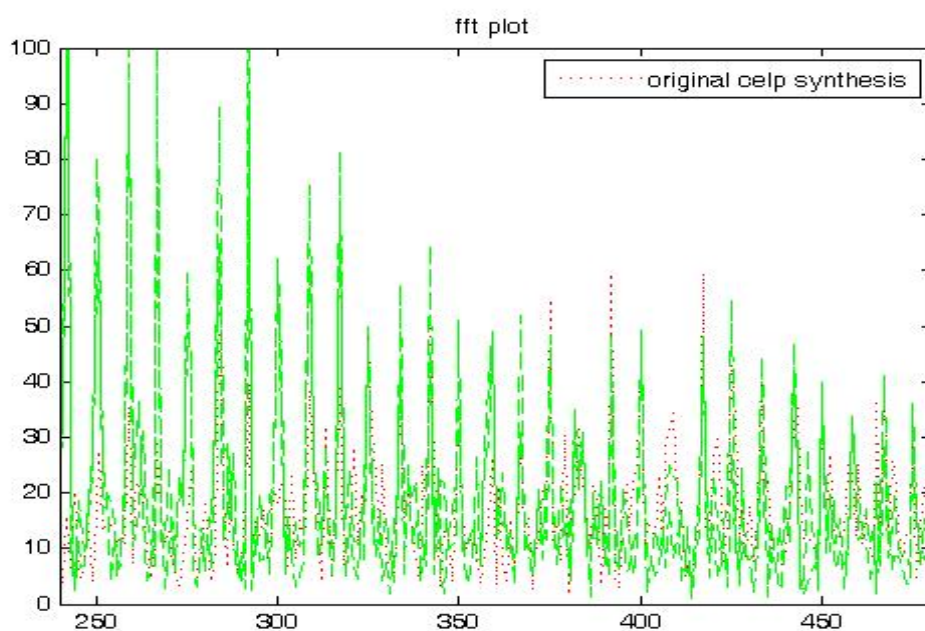
جدول ۷-۴. نتایج سیگنال صحبت به همراه موسیقی و نویز

میانگین نمره های DMOS	کد کننده سلپ عمومی	کد کننده سلپ بهبود یافته
	۳/۵۱	۳/۸۷

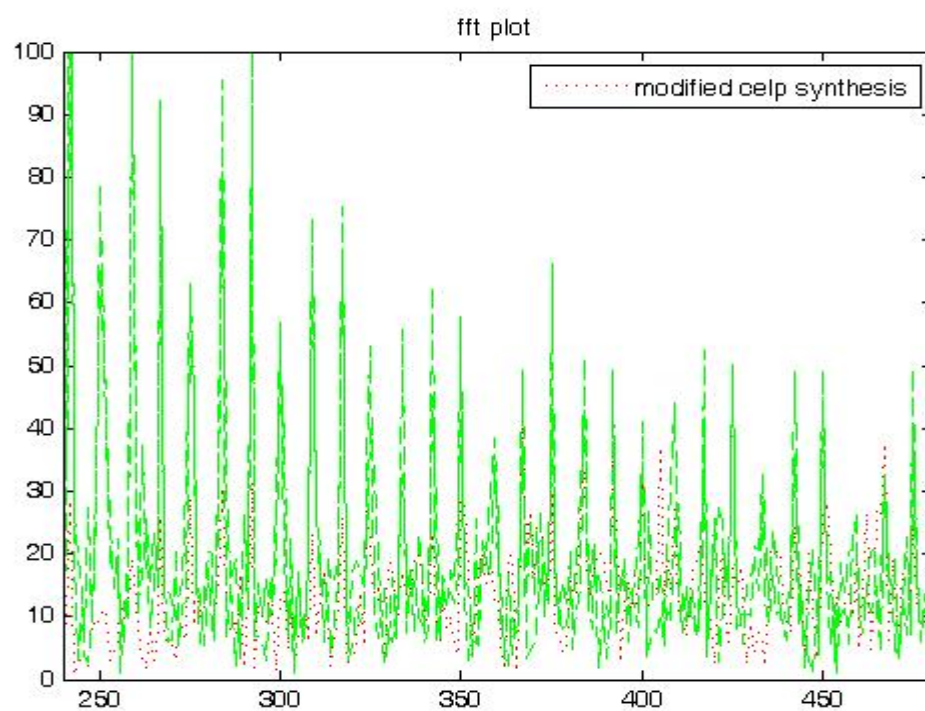
شکل زمانی سیگنال اصلی و سیگنال ساخته شده در کد کننده سلپ عمومی و سلپ بهبود یافته و شکل فرکانسی سیگنال اصلی و سیگنال ساخته شده با سلپ عمومی و سلپ بهبود یافته برای بازه [240-480] برای یکی از فایل های مورد آزمایش قرار گرفته، به ترتیب در شکل های ۷-۱۳، ۷-۱۴ و ۷-۱۵ آورده شده است. بررسی شکل زمانی و فرکانسی سیگنال های فوق همانند بخش ۷-۴-۲ می باشد و شکل زمانی سیگنال های ساخته شده کلیات شکل زمانی سیگنال اصلی را دارد و طیف فرکانسی در نقاط ماکزیمم تقریباً یکی است.



شکل ۷-۱۳. سیگنال اصلی سیگنال همراه با موسیقی و نویز ۱۷ دسی بل (شکل بالا) و سیگنال ساخته شده با کد کننده عمومی سلپ (شکل دوم) و کد کننده بهبود یافته (شکل سوم)



شکل ۷-۱۴. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلف (نقطه چین) سیگنال همراه با موسیقی و نویز ۱۷ دسی بل



شکل ۷-۱۵. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلف (نقطه چین) سیگنال همراه با موسیقی و نویز ۱۷ دسی بل

۶-۷. نرخ بیت مورد استفاده

از آنجا که در کد کننده بهبود یافته معرفی شده در این پایان نامه از پارامترهای کد کننده استاندارد FS1016 استفاده شده است و درجه تخمین خطی همانند این استاندارد ۱۰ انتخاب شده است، نرخ بیت استفاده شده برای ساخت سیگنال مطابق نرخ بیت استاندارد فوق و جدول تخصیص بیت برای کد کننده بهبود یافته مطابق جدول ۶-۳ می باشد. مطابق جدول فوق برای ضرایب LPC که در هر فریم ۱۰ ضریب می باشد بنابه شماره ضریب، تعداد ۳ یا ۴ بیت تخصیص داده شده است که مبنای این تخصیص میزان دقت و اهمیت ضریب مورد نظر می باشد.

برای انکد کردن رشته بیت سلف برای ارسال برای دیکدر، ۴ عدد در هر فریم برای ارسال دوره تناوب کتاب تطبیقی با دقت ۸-۶-۸-۶ بیت، ۴ عدد بهره کتاب کد تطبیقی با دقت ۵ بیت برای هر عدد، ۴ عدد کد تحریک منبع تحریک تصادفی با دقت ۹ بیت برای هر کد تحریک، ۴ عدد بهره کد تحریک با دقت ۴ بیت برای بهره، یک بیت همزمانی و ۴ بیت تصحیح خطا برای این کد کننده و یک بیت رزرو در نظر گرفته شده است. البته در تست های انجام شده برای کاهش نرخ بیت استفاده شده بیت های تصحیح خطا و بیت رزرو در برخی از تست ها حذف گردید و کیفیت سیگنال تغییری نکرد. با حذف ۵ بیت از فریم ارسالی برای دیکدر میزان نرخ بیت به اندازه $5 \times 800 / 240$ کاهش یافت یعنی نرخ بیت مورد استفاده در کد کننده بهبود یافته ۴۶۳۳ بیت بر ثانیه گردید.

۷-۷. بررسی اثر تغییر برخی پارامترها بر روی کیفیت و نرخ بیت

برای بررسی اثر تغییر پارامترهای کد کننده بهبود یافته، بر روی کیفیت سیگنال ساخته شده خروجی و نرخ بیت مورد نیاز برای ساخت سیگنال با پارامترهای جدید، اثر تغییر برخی از پارامترهای قابل تغییر کد کننده بهبود یافته مورد آزمایش قرار گرفت که نتایج آن در این بخش ارائه می گردد.

۷-۱. بررسی اثر تغییر پارامترهای γ, β, α

در آزمایش های انجام گرفته برای بررسی اثر تغییر پارامترهای γ, β, α بر روی کیفیت سیگنال تولید شده توسط کد کننده بهبود یافته، مقادیر مختلف این پارامتر ها برای دیتا بیس موجود مورد استفاده قرار گرفت تا با استفاده از گوش کردن به صدای تولید شده و تغییر پارامتر ها بهینه ترین پارامترها انتخاب گردد. در تست های مذکور ۵۰ فایل صدای ضبط شده و ۲۰ فایل همراه با موسیقی مورد تست قرار گرفت. همانطور که قبلا بیان گردید در تست های مقایسه ای با کد کننده عمومی پارامتر های فوق به ترتیب $\alpha=0.8, \beta=0.5, \gamma=0.8$ انتخاب شد اما در تست های انجام شده بر روی دیتا بیس با توجه به حجم زیاد تست های مورد نیاز برای بررسی اثر تغییر این پارامتر ها، محدوده تغییرات این پارامتر ها به ترتیب $0.5 < \gamma < 0.8, 0.5 < \alpha < 0.8, 0.3 < \alpha < 0.3$ انتخاب گردید که هر پارامتر با دقت ۰.۲ تغییر می کرد.

با توجه به آزمایش های انجام شده مقادیر بهینه ای که برای تمامی سیگنال های ورودی جواب قابل قبولی داشته باشد بدست نیامد چرا که مثلا برای سیگنال های ضبط شده با صدای بلند در دیتا بیس مقادیر بهینه $\alpha=0.7, \beta=0.48, \gamma=0.72$ بود و برای سیگنال های ضبط شده همراه با موسیقی با میزان مختلف موسیقی موجود در زمینه این مقادیر متفاوت بود. به همین منظور در تمامی تست های انجام شده برای این کد کننده مقادیر همان مقادیر ذکر شده در استاندارد که قبلا بیان گردید انتخاب شدند.

۲-۷-۷. بررسی اثر کاهش ضرایب LPC

در تست های انجام شده برای بررسی اثر کاهش تعداد ضرایب LPC بر روی کیفیت سیگنال ساخته شده و کاهش نرخ بیت، تعداد این ضرایب از ۷ تا ۱۲ تغییر نمود و سیگنال ساخته شده خروجی برای تست کیفیت سیگنال مورد تست DMOS قرار گرفت. در تست های انجام شده بدیهی بود که با افزایش تعداد ضرایب LPC کیفیت سیگنال ساخته شده، بهبود می یافت و این در نتایج تست DMOS نیز مشهود بود. در این تست که برای ۳۰ نمونه از صدا های ضبط شده و ۲۰ نمونه از صدا های همراه با موسیقی انجام گرفت اثر کاهش ضرایب LPC که باعث کاهش نرخ بیت ارسالی می شود مورد تست قرار گرفت. که نتایج تست های انجام شده برای ضرایب فوق در جدول ۵-۷ که میانگین نمره ۴۰ نفر می باشد آورده شده است. مطابق جدول با کاهش تعداد ضرایب LPC کیفیت سیگنال تولید شده نیز کاهش می یابد در عوض نرخ بیت مورد استفاده نیز کاهش می یابد. میزان نرخ بیت مورد نیاز نیز برای ارسال در هر مرحله در جدول ۵-۷ آورده شده است.

جدول ۵-۷. تغییرات کیفیت و نرخ بیت بر اساس تغییر ضرایب LPC

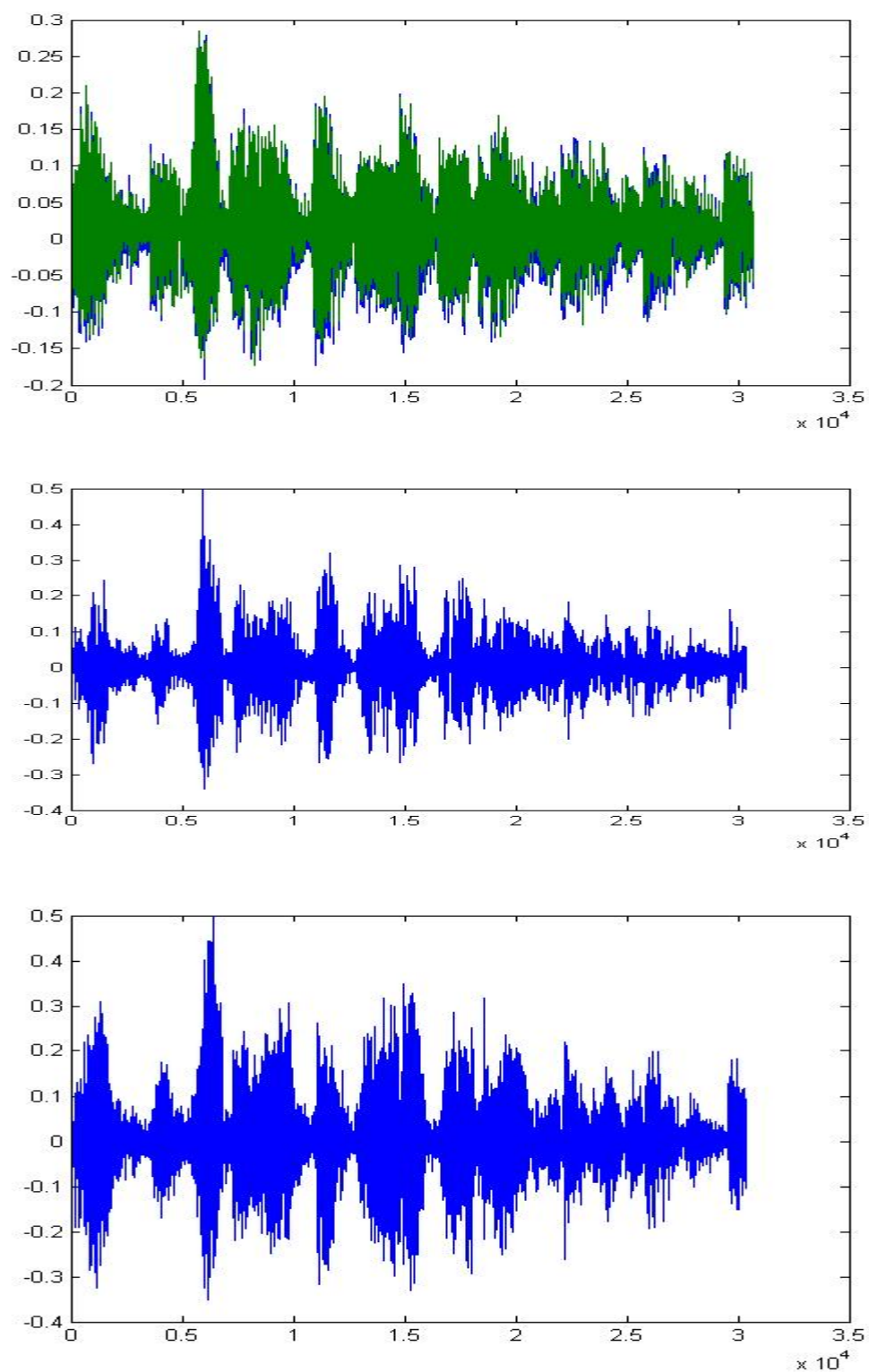
تعداد ضرایب LPC تخصیص داده شده	نرخ بیت مورد نیاز (بیت بر ثانیه)	نمرات میانگین DMOS (کد کننده عمومی)	نمرات میانگین DMOS (کد کننده بهبود یافته)
۷	۴۵۰۰	۳/۹۷	۳/۹۱
۸	۴۶۰۰	۴/۲۳	۴/۱۶
۹	۴۷۰۰	۴/۳۸	۴/۳۰
۱۰	۴۸۰۰	۴/۵۱	۴/۴۵

۷-۳-۷. اثر تغییر طول پنجره

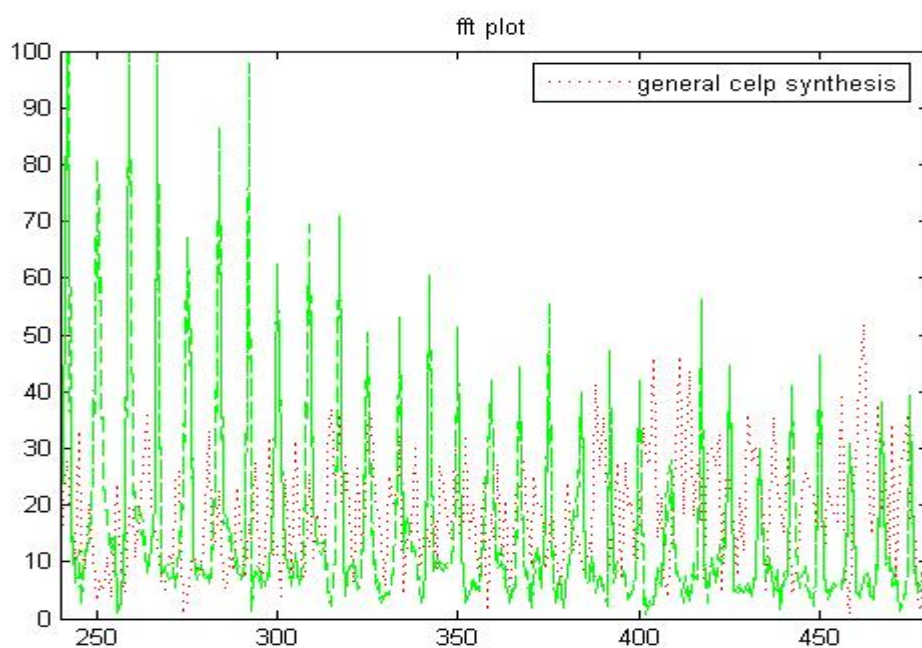
در تست انجام شده برای بررسی اثر افزایش طول فریم، بر روی کیفیت سیگنال ساخته شده و نرخ بیت ارسالی، ۳۰ فایل صوتی ضبط شده و ۲۰ نمونه ضبط شده همراه با موسیقی از دیتا بیس مورد آزمایش قرار گرفت و نتایج حاصل از تست DMOS که میانگین نمره های ۴۰ نفر می باشد همراه با نرخ بیت مورد نیاز برای ارسال داده ها در هر مورد در جدول ۷-۶ آورده شده است. تست های فوق برای طول فریم های ۲۴۰-۲۵۰-۲۶۰-۲۷۰-۲۸۰-۲۹۰ انجام گرفت. و شکل سیگنال های خروجی حاصل برای طول فریم ۲۹۰ در شکل ۷-۱۵ و ۷-۱۶ آورده شده است.

جدول ۷-۶. تغییرات کیفیت و نرخ بیت بر اساس تغییر طول فریم

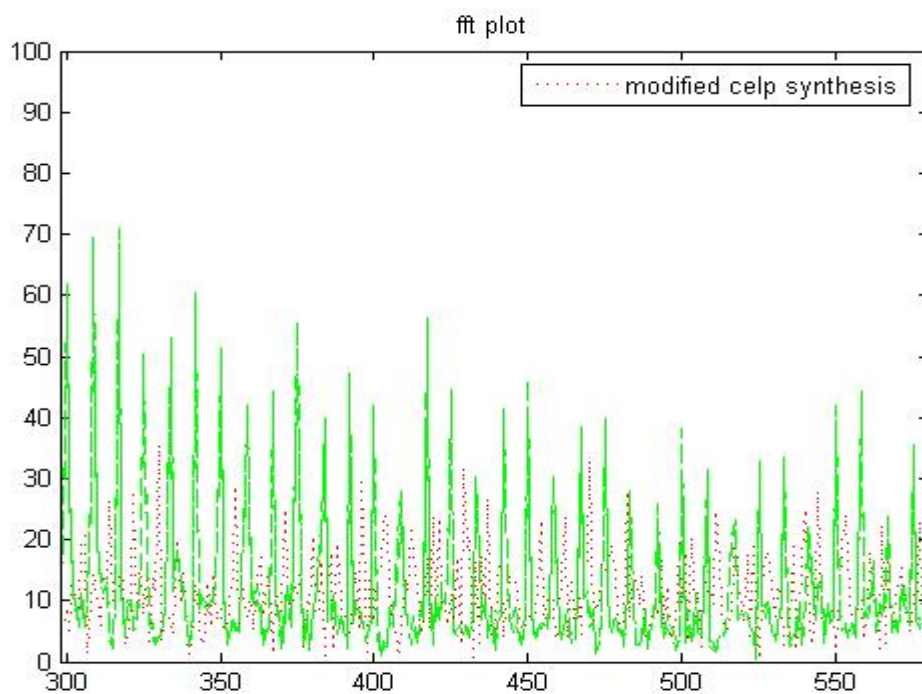
نمرات میانگین DMOS (کد کننده بهبود یافته)	نمرات میانگین DMOS (کد کننده عمومی)	نرخ بیت مورد نیاز (بیت بر ثانیه)	تعداد نمونه های موجود در فریم
۴/۴۵	۴/۵۱	۴۸۰۰	۲۴۰
۴/۳۶	۴/۴۱	۴۶۰۸	۲۵۰
۴/۲۷	۴/۳۰	۴۴۳۰	۲۶۰
۴/۱۱	۴/۱۵	۴۲۶۶	۲۷۰
۳/۹۲	۳/۹۵	۴۱۱۴	۲۸۰
۳/۷۶	۳/۷۷	۳۹۷۲	۲۹۰



شکل ۷-۱۶. سیگنال اصلی (شکل بالا) سیگنال ساخته شده با کد کننده سلف عمومی (شکل دوم) و کد کننده بهبود یافته (شکل سوم) با طول فریم ۲۹۰



شکل ۷-۱۷. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده عمومی سلپ (نقطه چین) با طول فریم ۲۹۰



شکل ۷-۱۸. FFT سیگنال اصلی (شکل پیوسته) و سیگنال ساخته شده با کد کننده بهبود یافته سلپ (نقطه چین) با طول فریم ۲۹۰

۷-۸. نتیجه گیری و پیشنهادات

در این پایان نامه کد کننده معروف CELP و انواع متداول آن معرفی گردید. سپس استاندارد FS1016 که یک استاندارد دولتی آمریکا در سالهای اخیر بوده است تشریح و به طور عملی پیاده سازی گردید. در مرحله بعد روش استخراج ضرایب PLP که روش بسیار متداول در تشخیص گفتار می باشد بعنوان مبنایی برای کد کننده CELP استفاده شد و بر اساس آن یک کد کننده بهبود یافته معرفی گردید.

کد کننده سلپ بهبود یافته معرفی شده با استفاده الگوریتم PLP، مورد آزمایش و تست های مختلف کیفیت و نرخ بیت قرار گرفت و در برخی موارد نتایج قابل قبولی داشت. در این تست ها که بر روی دیتا بیسی متشکل از ۵۰ نمونه ضبط شده فارسی، ۳۰ نمونه همراه با موسیقی فارسی و ۳۰ نمونه به زبان انگلیسی بود، کار کرد کد کننده بهبود یافته با کد کننده عمومی سلپ با استفاده از معیار DMOS مقایسه گردید. در این مقایسه کد کننده بهبود یافته برای سیگنال های گفتار بدون نویز و موسیقی نسبت به کد کننده سلپ عمومی کیفیت خوبی نداشت اما برای سیگنال های گفتار همراه با نویز و سیگنال های گفتار همراه با موسیقی و نویز کارکرد نسبتا قابل قبولی داشت. همچنین در این تست ها اثر تغییر پارامتر های فیلتر نهائی و فیلتر وزن دار بررسی گردید و پارامتر های استاندارد این مقادیر در کد کننده بهبود یافته مورد استفاده قرار گرفت. اثر کاهش تعداد ضرایب LPC و افزایش طول فریم های مورد استفاده از دیگر تست های انجام شده بر روی کد کننده بهبود یافته می باشد که نتایج کاهش میزان نرخ بیت و کیفیت متناظر در جداول مربوطه در داخل پایان نامه آورده شده است.

در ادامه پایان نامه پیشنهاد می گردد اثر استفاده از کوانتیزه کننده برداری و استفاده از جفت طیف خطی (LSP) در کد کننده بهبود یافته بررسی گردد. همچنین اثر استخراج ضرایب PLP و ارسال

مستقیم آنها برای دیکدر بررسی گردد. برای افزایش سرعت کار کد کننده می توان از الگوریتم های ژنتیک و سیستم های فازی در انتخاب کد تحریک استفاده نمود.

- [1] L.R. Rabiner and R.W. Schafer, "Digital Processing of Speech Signals" , Prentice Hall, Inc., Englewood Cliffs, New Jersey.
- [2] K.Sam Shanmugam,"Digital and Analog Communication Systems",1979.
- [3] WAI C. CHU. " Speech Coding Algorithms Foundation and Evolution of Standardized Coder". John-Wiley, pp. 299-325, 2003.
- [4] A.M.Kondoz,,"DIGITAL SPEECH CODING FOR LOW BIT RATE COMMUNICATION SYSTEMS", University of Survey, UK, JOHN WILEY & SONS, 1994.
- [5] Tamanna Islam , " Interpolation of Linear Prediction Coefficients for Speech Coding" , Department of Electrical Engineering, McGill University, Montreal, Canada, April 2000.
- [6] Randy Goldberg , Lance Riek, " A Practical Handbook of Speech Coding", CRC Press, Boca Raton London New York Washington, D.C, 2000.
- [7] Petr Motlicek, "Feature Extraction in Speech Coding and Recognition", , Report of PhD research internship in ASP Group, OGI-OHSU, March 19, 2003.
- [8] Vahid Abolghasemi, Hossein Marvi," MODIFIED CELP CODER USING CEPSTRAL ANALYSIS ", ICSV14, Cairns,Australia, july 2007.
- [9] H. Hermansky," Perceptual Linear Predictive(PLP) analysis for speech", J. Acoust. Soc. Am., Vol. 87, No. 4, pp. 1738– 1752, 1990.
- [10] P. Boda, J.M. Boite, A.E. Cetin, K. Elenius, Z. Ka_ci_c, J.P. Martens, J. Nouza, F, Drygajlo (editor)," Feature extraction ".
- [11] Jae Sam YOON, Gil Ho LEE," A MFCC-Based CELP Speech Coder for Server-Based Speech Recognition in Network Environments", IEICE TRANS. FUNDAMENTALS, VOL.E90–A, NO.3 MARCH 2007.
- [12] Shivali Srivastava, "FUNDAMENTALS OF LINEAR PREDICTION ", Department for Electrical and Computer Engineering Mississippi State University.
- [13] Andreas S. Spanias. " Speech Coding: A Tutorial Review, " Proceeding of the IEEE Vol. 82, No. 10, pp. 1541-1582, October 1994.
- [14] Chien-Kuang Lin, "Low-Complexity CodeBook Searching Algorithms for FS1016", For the Degree of Master of Science, Advisory by Prof. Po-Ning Chen, Department of

Communications Engineering National Chiao Tung University Hsinchu, Taiwan 300, R.O.C, June, 2002.

[15] Steven K Smith, "Digital Signal Processing", Newnes is an imprint of Elsevier Science, 2003.

[16] Sen M Kuo, Bob H Lee, "Real-Time Digital Signal Processing", JOHN WILEY & SONS, LTD, 2001.

[17] Joon-Hyuk Chang, "Perceptual weighting filter for robust speech modification", Department of Electronic Engineering, Inha University, Incheon, 402-751, Korea, 2005.

[18] J.-H. Chen, "Toll-quality 16 kb/s CELP speech coding with very low complexity," in proceedings IEEE International Conference on Acoustics, speech and signal processing , pp. 9-12, 1995.

[19] G.Madre, E.H. Baghious, S. Azou and G. Burel," FAST PITCH MODELLING FOR CS-ACELP CODER FERMAT NUMBER TRANSFORMS ", IEEE Int. symp. On Signal Processing and Information Technology (ISSPIT), 2003.

[20] Houman Zarrinkoub, Paul Mermelstein, " JOINT OPTIMIZATION OF SHORT-TERM AND LONG-TERM PREDICTIONS IN CELP SPEECH CODERS", ICASSP, 2003

[21] K.Abboud, P. Kabal , "Wideband CELP Speech coding at 12kbits/sec", proc. Biennial Symp. Commun. (Kingston, on), pp. 25-28, May 1992

[22] G.H.Lee, J.S. Yoon, Y.R. Oh and H.K. Kim, " Design of a speech coder utilizing speech recognition parameters for server-based wireless speech recognition," , proc. ISPACS, pp.159-163, seoul, korea, Nov, 2004

[23] Cheng, Y.M., and O'Shaughnessy, "Speech Enhancement Based Conceptually on Auditory Evidence," IEEE Trans. on Signal Processing, 39. 1943-1954, 1991

[24] Edward Painter, Student Member IEEE, and Andreas Spanias, Senior Member IEEE, "A MATLAB SOFTWARE TOOL FOR THE INTRODUCTION OF SPEECH CODING FUNDAMENTALS IN A DSP COURSE" , Department of Electrical Engineering, Telecommunications Research Center Arizona State University, Tempe, Arizona 85287-7206

[25] Thomas Millhouse & Frantz Clermont," ACOUSTIC DESCRIPTION OF A SOPRANO'S VOWELS BASED ON PERCEPTUAL LINEAR PREDICTION", ICPHS, Saarbrücken, 6-10 August 2007.

[26] Millhouse, T. J. & Clermont, "Perceptual characterization of the singer's formant region" A preliminary study.Proc.11th Australian / New Zealand International Conference on Speech Science and Technology.SST06 253-258, 2006.

[27] Wei Li, Anu Sridhar and Tina Teng, "Comparison of Speech Coding Algorithms: ADPCM, CELP and VSELP", University of Texas at Dallas, 1999

- [28] Y-H.Kao, "Low Complexity CELP Speech Coding at 4.8 kbps", University of Maryland, Master thesis 1992.
- [29] Andreas S. Spanias. "Speech Coding: A Tutorial Review, " Proceeding of the IEEE Vol.82, No. 10, pp. 1541-1582, October 1994.
- [30] K.Koishida, K.Tokuda, T. Kobayashi and S. Imai, "CELP coding system based on melgeneralized cepstral analysis," Proc. ICSLP'96, pp.318–321, 1996.
- [31] Hongwu YANG, Dezhi HUANG, and Lianhong CAI, "Perceptually Weighted Mel–Cepstrum Analysis of Speech based on PsychoacousticModel", IEICE TRANS.INF.& SYST., VOL.E89–D, NO.12 DECEMBER 2006.
- [32] Jerry D. Gibson, "Speech Coding Methods, Standards, and Applications", Department of Electrical & Computer Engineering University of California, Santa Barbara Santa Barbara, CA 93106-6065.
- [33] L. Hanzo, et al, "Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels", New York: Wiley-IEEE Press, 2001.
- [34] T.Painter and A. Spanias, "Perceptual Coding of Digital Audio," Proceedings of the IEEE, vol. 88, pp. 451-513, April 2000.
- [35] ITU-T Recommendation G.711, Appendix II: A comfort noise payload definition for ITU-T G.711 use in packet-based multimedia communication systems, Feb. 2000.
- [36] ITU-T, G.711, Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711," Sept. 1999.
- [37] Susanna Varho, "NEW LINEAR PREDICTIVE METHODS FOR DIGITAL SPEECH PROCESSING", Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing Espoo 2001.
- [38] Petr Motlicek and Hynek Hermansky," Speech Coding based on Spectral Dynamics", IDIAP–RR 06-30, MIA 2006.
- [39] Vahid Abolghasemi and Hossein Marvi, "Modified CELP Coder Using Root Cepstrum", 2007
- [40] Chan-woo Kim, Gyeoggi-do, "SPEECH CODING APPARATUS AND METHOD THEREFOR", United States Patent Application Publication, Pub. No: US2006/0149534 A1, JUL.6, 2006.
- [41] Richard C. Hendriks, Richard Heusdens and Jesper Jensen, "PERCEPTUAL LINEAR PREDICTIVE NOISE MODELLING FOR SINUSOID-PLUS-NOISE AUDIO CODING", IEEE, 2004.
- [42] Abeer Alwan, A. Bernard, X.Cui and Q. Zhu, "JOINT SPEECH COMPRESSION AND AUTOMATIC SPEECH RECOGNITION", University of California, Los Angeles, CA90095, 2000-2001.

- [43] Hongwu YANG, Dezhi HUANG and Lianhong CAI, "Perceptually Weighted Mel-Cepstrum Analysis of Speech Based on Psychoacoustic Model", IEICE TRANS. INF. & SYST., VOL.E89-D, NO.12 DECEMBER 2006.
- [44] Wira Gunawan and Mark Hasegawa-Johnson, "PLP COEFFICIENTS CAN BE QUANTIZED AT 400 BPS", Department of Electrical and Computer Engineering University of Illinois at Urbana-Champaign, USA.
- [45] DEN BRINKER AND RIERA-PALOU, " PURE LINEAR PREDICTION", Audio Engineering Society Convention Paper, Presented at the 115th Convention, New York, NY, USA, 2003 October 10.13.
- [46] Tom Bäckström, " LINEAR PREDICTIVE MODELLING OF SPEECH CONSTRAINTS AND LINE SPECTRUM PAIR DECOMPOSITION", Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing Espoo, 2004.
- [47] David S. Petruncio, Jr. and Mark A. Hasegawa-Johnson, " Evaluation of Various Features for Music Genre Classification with Hidden Markov Models", March , 2002.
- [48] Nachiappan and PM Abdul Manan Ahmad, " SPEECH CODING EFFECTS ON RECOGNITION ACCURACY FOR TIMIT", Postgraduate Annual Research Seminar 2007.
- [49] Alexis Bernard and Abeer Alwan, "SOURCE AND CHANNEL CODING FOR REMOTE SPEECH RECOGNITION OVER ERROR-PRONE CHANNELS", Dept. of Electrical Engineering, UCLA Los Angeles, CA 90095