کدگذاری سیگنال صحبت تحت محیط‌های نویزی مبتنی بر مدل سیستم شنوایی انسان

سید ایمان ابطحی[(۱)] - محمد رضا آشوری[(۲)] - رسول امیر فتاحی[(۳)]

(۱) مربی - گروه مهندسی برق، دانشگاه آزاد اسلامی، واحد میمه

(۲) استادیار - آزمایشگاه تحقیقاتی پردازش سیگنال‌های دیجیتال، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی اصفهان

(۳) دانشیار - آزمایشگاه تحقیقاتی پردازش سیگنال‌های دیجیتال، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی اصفهان

تاریخ دریافت: بهار ۱۳۹۰ تاریخ پذیرش: بهار ۱۳۹۱

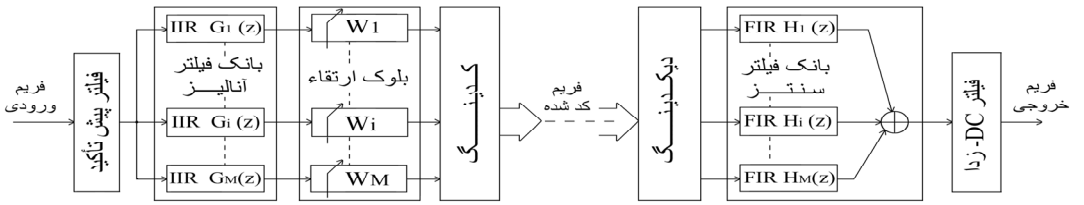
خلاصه: در این مقاله یک سیستم آنالیز/ سنتز، بر اساس مدل طبیعی حلزونی گوش و ویژگی‌های درک شنوایی انسان ارائه شده که قادر به کد کردن سیگنال گفتار در شرایط دشوار آکوستیکی است. بدین منظور، سیگنال نویزی توسط یک بانک فیلترگاماتن مختلط به تعدادی زیرباند شنوایی تجزیه شده و سیگنالِ هر زیرباند به طور مستقل و وفقی، از جهت حذف نویز پردازش می‌شود. استخراج پارامترها و فشرده‌سازی نیز از طریق ماسک گذاری کوتاه مدت، یک روش کوانتیزاسیون غیریکنواخت جدید و الگوریتم‌های کدینگ بدون تلفات صورت می‌گیرد. ارزیابی کیفیت از طریق آزمون‌های استاندارد کمی و کیفی، نشان می‌دهد که علیرغم کاهش قابل توجه نرخ بیت تا حدود 14.6 Kbps، کیفیت سیگنال‌های سنتزشده بهبود معناداری یافته، و عملکرد سیستم در برابر انواع نویزهای سفید، رنگی و پریودیک، باثبات و مؤثر است. همچنین کیفیت سیگنال‌های خروجی در مقایسه با نتایج چند نمونه کدینگ استاندارد، قابل رقابت ارزیابی شده است.

کلمات کلیدی: آزمون‌های استاندارد کمی و کیفی، بانک فیلتر گاماتن مختلط، کدینگ بدون تلفات، مدل طبیعی حلزونی گوش، ویژگی‌های درک شنوایی.

۱- مقدمه

ضبط سیگنال گفتار با استفاده از سنسورهای آکوستیک و در حضور نویزهای پس زمینه و محیطی، عملکرد سیستم کدینگ در تخمین و استخراج پارامترها را مختل نموده و باعث افت شدید کیفیت در سیگنال سنتز شده می‌شود. از این رو به همراه سیستم کدینگ، اعمال تکنیک‌های مؤثر ارتقاء، اهمیت زیادی دارد. الگوریتم‌های بسیاری نیز در این زمینه ارائه شده‌اند. به عنوان مثال در [۱] یک تکنیک ارتقاء کیفیت دو شاخه قبل از کدینگ LPC[۱]، در [۲] یک کدکننده در شرایط نویزی به صورت کوانتیزاسیون برداری چند لایه، و در [۳] یک سیستم با نرخ بیت بسیار پایین بر اساس مدل مارکف مخفی ارائه شده است.

اخیراً توانایی عملکرد سیستم شنوایی انسان در محیط‌های نویزی، محققین را به استفاده از این ویژگی در سیستم‌هایشان تشویق کرده است [۴، ۵]. بر این مبنا در این مقاله، یک سیستم کدینگ/ دیکدینگ ارائه شده که قادر به بهبود کیفیت سیگنال‌های نویزی، علیرغم فشرده‌سازی و کاهش نرخ بیت آن است. در این سیستم، ابتدا سیگنال ورودی از طریق مدل سازی حلزونی گوش به زیرباندهای شنوایی تجزیه می‌شود. بدین منظور اکثر مدل‌های شنوایی از فیلترهای گاماتن حقیقی مرتبه چهار در مقیاس ERB[۲]، که شباهت زیادی با شکل فیلترهای شنوایی دارند، استفاده می‌کنند [۵، ۶، ۷]. به عنوان نمونه در [۸] با استفاده از این فیلترها و بر پایه تکنیک بهینه‌سازی حداقل مربعات، یک بانک فیلتر شنوایی آنالیز/ سنتز بهینه با بار محاسباتی و تأخیر اندک، طراحی شده است. اما فیلترهای گاماتن با وجود مزایای زیاد، در نواحی فرکانس بالا دارای پاسخ فرکانسی با شیب هموار است که منجر به کاهش دقت فرکانسی مدل و عدم انطباق کامل با شیب تند فیلترهای شنوایی و در نتیجه ایجاد اعوجاج در سیگنال خروجی می‌شود. جهت کاهش این اثر، در [۹] از یک بانک فیلتر مبتنی بر منحنی‌های ماسکینگ شنوایی استفاده شده است. راه حل دیگر افزایش مرتبه فیلتر است.



شکل (۱): بلوک دیاگرام سیستم پیشنهادی
Fig. 1: Diagram of proposed system

اما ثابت‌های مدل حلزونی اغلب برای فیلترهای مرتبه چهارم توصیه شده‌اند [۶]. در این مقاله تبدیل فیلترهای حقیقی به فرم مختلط، با هدف دوبل شدن فیلترها درهر زیرباند و در نتیجه افزایش رزولوشن فرکانسی و کاهش اعوجاج، پیشنهاد می‌شود. مقایسه مدل آنالیز/ سنتز مختلط پیشنهادی و مدل حقیقی در [۸] در شرایط مشابه، برتری این مدل را از نظر اعوجاج به اثبات می‌رساند.

در الگوریتم پیشنهادی، بانک فیلتر آنالیز به همراه مدل IHC[۳] (سلول‌های مویی درونی)، پروسه تبدیل امواج آکوستیک صدا به سیگنال‌های آتش نورونی در حلزونی را شبیه‌سازی می‌کند. این سیگنال‌ها جهت حذف نویز به بلوک ارتقاء اعمال می‌شوند. در این بلوک، عمل حذف نویز از طریق وزن‌دهی سیگنالِ هر زیرباند به صورت وفقی و با بهره‌ای که از تخمین واریانس سیگنال و نویز به دست می‌آید، انجام می‌شود. استخراج پارامترها و فشرده‌سازی نیز بر اساس مدل‌های ادراکی، کوانتیزاسیون و کدکننده‌های بدون تلفات می‌باشد. الگوریتم‌های استفاده شده در هر مرحله، نسبت به مراجع مربوطه و متناسب با ساختار این سیستم، تغییر یافته و در اکثر موارد بهبود نیز یافته‌اند. بازسازی سیگنال در قسمت دیکدینگ نیز با روشی نوین صورت پذیرفته است.

با اعمال این مدل، نرخ بیت در فرکانس نمونه برداری 8 KHz، به طور متوسط 14.6 Kbps محاسبه شده که معادل با 1.82 بیت برای هر نمونه است. نتایج آزمون‌های استاندارد کمی و کیفی[۴] و همچنین مقایسه با چند نمونه کدینگ استاندارد، نشان دهنده عملکرد با کیفیت، مؤثر و با ثبات سیستم پیشنهادی در شرایط متنوع آکوستیک و در برابر انواع ورودی‌های تمیز و نویزی است.

در این مقاله، بخش (۲) به تشریح الگوریتم پیشنهادی شامل طراحی بانک فیلتر آنالیز/سنتز، مرحله ارتقاء و مرحله کدینگ/ دیکدینگ، و بخش (۳) به بررسی نتایج شبیه سازی و همچنین مقایسه با چند نمونه کدینگ استاندارد، اختصاص یافته است.

۲- الگوریتم پیشنهادی

شکل (۱) بلوک دیاگرام سیستم پیشنهادی را نشان می‌دهد. در این سیستم، فیلتر پیش تأکید جهت حذف مؤلفه‌های Subsonic و DC از سیگنال ورودی، که می‌توانند باعث بروز خطا در باندهای پایین شوند، اضافه شده است. ما آن را با تابع تبدیل $H(z) = 1 - 0.97\, z^{-1}$ پیاده سازی کردیم. سایر بلوک‌ها در ادامه شرح داده می‌شوند.

۳- بانک فیلتر آنالیز/ سنتز گاماتن مختلط

در دستگاه شنوایی وظیفه حلزونی، تبدیل ارتعاشات مکانیکی صدا به ایمپالس‌های عصبی است. حلزونی مانند یک اسپکتروم آنالایزر، صدا را به مؤلفه‌های فرکانسی آن تجزیه می‌کند و به همین دلیل به عنوان یک مدل فیلترینگ شنوایی در سیستم‌های پردازش صوت، به کار می‌رود. پروسه شبیه سازی این مدل، یک پروسه غیرخطی دو طبقه شامل: ۱- آنالیز طیفی و شبیه سازی ارتعاشات غشای قاعده‌ای (BMM[۵]) و ۲- تبدیل سیگنال‌های BMM به ایمپالس‌های عصبی توسط مدل سلول‌های مویی درونی (IHC[۶])، است [۶]. سیگنال‌های BMM را می‌توان با گروهی از فیلترهای همپوشانی کرده میان گذر خطی که موج آکوستیک ورودی را به زیرباندهای فرکانسی تجزیه می‌کنند، مدل کرد. در این زمینه استفاده از فیلترهای گاماتن مرتبه چهار بسیار کارآمد می‌باشد.

در این مقاله با هدف افزایش رزولوشن فرکانسی و کاهش اعوجاج، پاسخ ضربه فیلترهای گاماتن حقیقی معرفی شده در [۸] به فرم مختلط تعمیم یافته و یک بانک فیلتر آنالیز/ سنتز مختلط با ۲۵ زیرباند طراحی شده است. پاسخ ضربه i-امین فیلترگاماتن مختلط چنین است: ($j = \sqrt{-1}$)

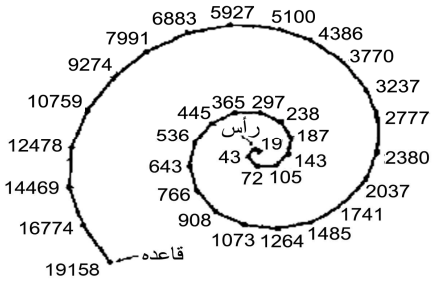
$$g_{ci}(t) = a_i t^{N-1} \exp\left(-2\pi b_i t + j\left(2\pi f_{ci} t + \varphi\right)\right), \qquad (۱)$$

در این رابطه، N مرتبه، φ فاز، ai ضریب نرمالیزکننده، و fci و bi به ترتیب فرکانس مرکزی و پهنای باند فیلتر در مقیاس ERB است. ERB فیلترها مشخص کننده مقادیر گزینندگی فرکانسی در مدل حلزونی می‌باشند. این پارامترها تابعی از مکان بر روی غشای قاعده‌ای حلزونی و مطابق با فیزیولوژی آن چنین به دست می‌آیند [۶]:

$$f_{ci} = 165.4(10^{2.1 x_i} - 1), \qquad (۲)$$

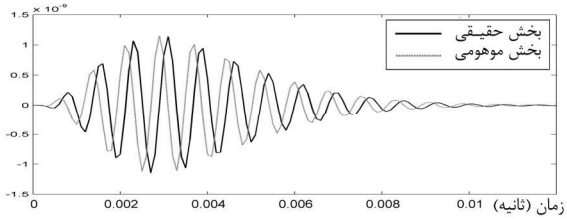
$$b_i = 1.019 \times 24.7\left(1 + 4.37\frac{f_{ci}}{1000}\right), \qquad (۳)$$

در اینجا، $x_i \in (0,1)$ فاصله نرمالیز شده از قاعده حلزونی و مقدار 24.7 برابر با حداقل پهنای باند است. شکل (۲) نقشه فرکانسی حلزونی و شکل (۳) پاسخ ضربه تابع یک فیلتر گاماتن مختلط نمونه را نشان می‌دهد.



شکل (۲): نقشه فرکانسی حلزونی [۶]

Fig. 2: Cochlear frequency mapping [6]



شکل (۳): پاسخ ضربه یک فیلتر گاماتن نمونه با فرکانس مرکزی 1293 هرتز و پهنای باند 167.5 هرتز

Fig. 3: A sample Gammatone filter impulse response with center frequency = 1293 Hz, bandwidth= 167.5Hz

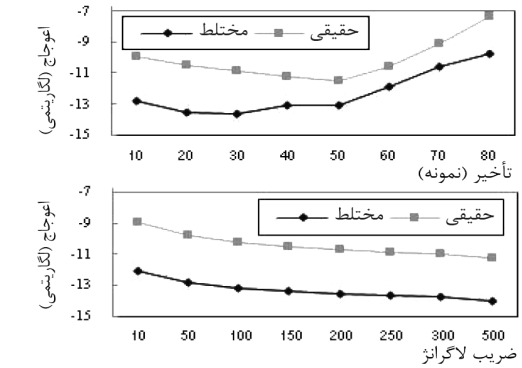
جهت تبدیل این فیلترهای آنالوگ به فرم فیلترهای دیجیتال تکنیک‌های گوناگونی وجود دارد. در [۷] یک مقایسه کلی بین این تکنیک‌ها صورت پذیرفته و تبدیل Impulse Invariant را از نظر مطابقت با پاسخ ضربه، اندازه و فاز فیلترهای گاماتن آنالوگ، مناسب معرفی کرده است. در این مقاله نیز از این روش استفاده شده است. هر فیلتر IIR طراحی شده دارای چهار قطب و سه صفر بوده و تابع تبدیل آن برای i-امین فیلتر چنین است:

$$G_i(z) = \frac{a_i e^{-p_i T} z^{-1} + 4a_i e^{-2p_i T} z^{-2} + a_i e^{-3p_i T} z^{-3}}{1 - 4e^{-p_i T} z^{-1} + 6e^{-2p_i T} z^{-2} - 4e^{-3p_i T} z^{-3} + e^{-4p_i T} z^{-4}}, \quad (۴)$$

T دوره تناوب نمونه برداری و p یک عدد مختلط است.

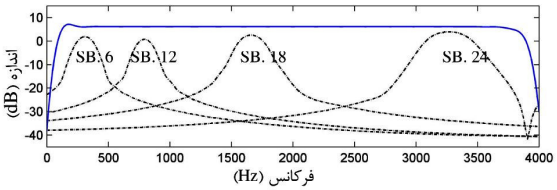
در شکل (۴) مقایسه‌ای از نظر اعوجاج بین مدل آنالیز/ سنتز مختلط در این مقاله و مدل حقیقی در مرجع [۸]، در شرایط مشابه، صورت پذیرفته است. همان طور که مشاهده می‌شود با تغییر پارامترهای طراحی بانک فیلتر سنتز شامل میزان تأخیر و ضریب لاگرانژ، همواره اعوجاج در مدل مختلط کمتر از حقیقی است، که این امر کیفیت بهتر سیگنال سنترشده را در پی دارد. تأخیر، جهت علّی شدن فیلترهای سنتز در نظر گرفته شده است.

مرتبه هر فیلتر FIR سنتز 80، تأخیر 5 میلی ثانیه و ضریب لاگرانژ 250 انتخاب شده است. مطابق با شکل (۵)، کانولوشن فیلترهای آنالیز و سنتز در هر زیرباند پاسخ میان گذر می‌دهد و جمع آثار این فیلترها، شبیه به یک فیلتر تمام گذر با اندازه صاف است.



شکل (۴): نمودار تغییرات اعوجاج نسبت به: (بالا): تأخیر، با ضریب لاگرانژ= 250؛ و (پایین): ضریب لاگرانژ، با تأخیر=5ms؛ برای دو بانک فیلتر آنالیز/ سنتز مدل مختلط و مدل حقیقی [۸] با مرتبه برابر و پارامترهای مشابه

Fig. 4: Curve of distortion as a function of: (up) delay (D) with L=250; & (down): Lagrange multiplier (L) with D=5ms; for two analysis/synthesis filter banks of proposed complex and real [8] models; with equal order and similar paratemeters



شکل (۵): نمودار توپر: پاسخ فرکانسی بانک فیلتر آنالیز/ سنتز گاماتن مختلط؛ نقطه‌چین: پاسخ فرکانسی فیلترهای آنالیز و سنتز در چهار زیرباند نمونه

Fig. 5: Solid: The overall complex GTF analysis/synthesis spectrum.
Dashed: Complex GTF analysis/synthesis individual channel spectrum (only 4 channel are shown)

۴- الگوریتم حذف نویز

هدف از این قسمت، علاوه برافزایش SNR و بالا بردن کیفیت، بهبود عملکرد کدینگ جهت استخراج پارامترهای مفقود شده در نویز و همچنین کاهش نرخ بیت است. در این مقاله از تکنیک معرفی شده در [۵]، با کمی اصلاح از جهت کاهش خطای تخمین و همچنین پس از بسط به فرم مختلط، استفاده شده است. در این تکنیک با فرض مستقل بودن سیگنال و نویز، سیگنال تمیز $\hat{V}_i(n)$ در هر زیرباند به صورت ضریبی از سیگنال نویزی $V_i(n)$ در آن زیرباند تخمین زده می‌شود:

$$\hat{V}_i(n) = w_i \cdot V_i(n), \quad (۵)$$

بهره w_i از طریق تعریف تابع خطا، چنین به دست می‌آید:

$$w_i = \left(\sigma^2 V_i - \sigma^2 N_i\right) / \sigma^2 V_i, \quad (۶)$$

$\sigma^2 N_i$ واریانس نویز و $\sigma^2 V_i$ واریانس سیگنال نویزی در زیرباند iام است. این رابطه شباهت زیادی به بهره فیلتر وینر دارد. از آنجایی که فقط سیگنال نویزی را در اختیار داریم، واریانس نویز را در فواصل زمانی سکوت تخمین می‌زنیم. از این رو الگوریتم نیازمند یک آشکارساز فواصل زمانی سکوت نیز هست. بدین منظور از روش [۱۰]

استفاده شده است. این روش بر اساس انرژی‌های کوتاه مدت و اندازه‌گیری نرخ عبور از صفر می‌باشد.

به منظور کاهش خطای تخمین، می‌توان با تعریف پارامتر α، واریانس نویز را، در فریم‌های مجاور اندازه‌گیری کرد. بین کاهش نویز و افزایش اعوجاج نیز رابطه معکوسی وجود دارد که با تعریف پارامتر β می‌توان آن را کنترل کرد. بنابراین واریانس نویز $d_{i,j}$ در زیرباند i -ام و فریم jام چنین اصلاح می‌شود:

$$d_{i,j} = \beta_{i,j}\left((1-\alpha)\ \sigma^2 N_{i,j-1} + \alpha\ \sigma^2 N_{i,j}\right) \qquad (۷)$$

به علاوه جهت افزایش اثر نویز زدایی: توان γ، جهت طبیعی‌تر شدن صدای خروجی: آفست w_0 و یک مقدار حداقلی نیز جهت اجتناب از منفی شدن بهره در نظر گرفته می‌شود. مقدار این پارامترها، به صورت تجربی و با ارزیابی کیفیت سیگنال خروجی به دست می‌آیند. با اعمال موارد مذکور بهره در زیرباند i-ام و فریم j-ام چنین است:

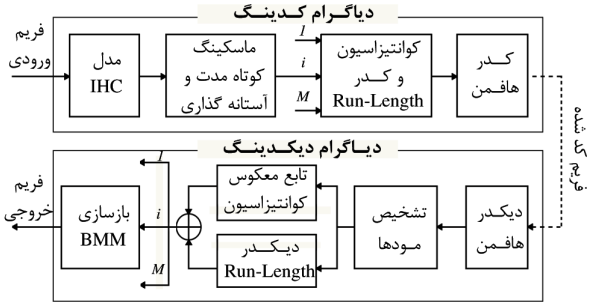
$$w_{i,j} = \begin{cases} w_0 + \left(\dfrac{\sigma^2 V_{i,j} - d_{i,j}}{\sigma^2 V_{i,j}}\right)^{\gamma}, & \text{if } (\sigma^2 V_{i,j} - d_{i,j}) \geq 0 \\ w_0 & \text{otherwise} \end{cases} \qquad (۸)$$

طبیعت زیرباند پروسه، آن را برای کاهش نویزهای سفید و رنگی مناسب ساخته است. ما یک بلوک پس پردازش نیز جهت کاهش نویزهای پریودیک یا باریک باند قرار دادیم. چنانچه ورودی آلوده به این نویز باشد، واریانس در زیرباند یا زیرباندهایی که طیف نویز در محدوده باندگذر آن قرار دارد، به طور غیر عادی زیاد می‌شود. با آشکارسازی این افزایش ناگهانی می‌توان این زیرباندها را به طور اتوماتیک تشخیص و با اصلاح بهره، آن قسمت از باند فرکانسی را تضعیف و نویز را کاهش داد.

۵- الگوریتم کدینگ

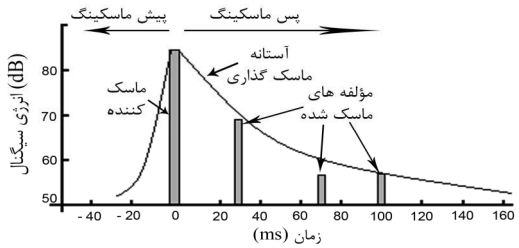
بلوک دیاگرام کدینگ/دیکدینگ در شکل (۶) رسم شده است. سیگنال‌های BMM پس از ارتقاء، توسط مدل IHC به ایمپالس‌های عصبی تبدیل می‌شوند. این مدل را می‌توان با یک یکسوساز نیم موج و یک تابع فشرده‌ساز قانون توان ساخت [۱۱]. ازنظرفیزیکی پروسه یکسوسازی به این علت است که امواج صدا همواره در حلزونی از قاعده به سمت رأس حرکت می‌کنند و هرگز به عقب برنمی‌گردند.

در ادامه، به منظور حذف مؤلفه‌های اضافه و نارسا از سیگنال به دست آمده، بر روی آن یک مدل ماسک گذاری موقتی و کوتاه مدت، اعمال می‌شود. این مدل بر طبق ویژگی‌های ادراکی، بیان می‌کند که دریک بازه زمانی کوچک، تُن‌های ضعیف توسط تُن‌های مجاور قوی‌تر صدا، غیر قابل شنیدن می‌شوند. در اینجا از روش [۹] استفاده شده است. در دو مرحله پس- ماسکینگ و پیش- ماسکینگ، آستانه‌ها به فرم نمایی نزولی بوده و ثابت زمانی آن‌ها در زیرباندهای مختلف به صورت تجربی مشخص می‌شوند. ما مقادیر آن را با فواصل مساوی بین 0.005 تا 0.03 برای زیرباندهای از 1 تا 25 انتخاب کردیم. شکل (۷) اثر این ماسک گذاری را نشان می‌دهد. علاوه بر این یک آستانه گذاری ساده نیز جهت حذف پالس‌های کوچک و نارسای باقی مانده در مکان‌های تصادفی، اعمال می‌شود. مقدار آستانه به صورت وفقی در هر فریم و در هر زیرباند از روی متوسط دامنه پالس‌ها محاسبه می‌شود. آنچه باقی می‌ماند، پارامترهایی است که باید کد شوند.



شکل (۶): بلوک دیاگرام: بالا: سیستم کدینگ؛ پایین: سیستم دیکدینگ

Fig. 6: Block diagram of the coding/decoding scheme



شکل (۷): اثر ماسک گذاری کوتاه مدت

Fig. 7: Short-term temporal masking threshold

۶- کوانتیزاسیون و کدگذاری

پس از بلوک IHC و حذف اضافات، اکثر مکان‌ها در بردار زیرباندها صفر شده‌اند. این بدان معناست که می‌توان به جای کد کردن تک تک آن‌ها، فقط رابطه بین محل پالس‌های همسایه، یا به عبارتی تعداد صفرهای بین آن‌ها را کد کرد. بدین منظور از تکنیک Run-Length استفاده شده است [۹].

در هر فریم، ابتدا بردار همه زیرباندها درون یک بردار بزرگتر به هم متصل شده و تعداد صفرهای بین هر دو پالس با 6 بیت کد می‌شود (مود صفر). دامنه پالس‌های غیرصفر نیز با استفاده از یک روش کوانتیزاسیون غیر یکنواخت جدید، کوانتیزه می‌شوند. این روش قادر است رنج وسیعی از اعداد را با خطای اندک کوانتیزه، و در مقایسه با روش [۹]، با تعداد بیت برابر، سطوح کوانتیزاسیون بیشتر و در نتیجه نویز کمتری ایجاد کند. بر این اساس، به وسیله 6 ضریب نرمالیزاسیون nf، 6 مود مختلف تعریف شده و بسته به دامنه هر پالس x، یکی ازمودها به طور اتوماتیک انتخاب می‌شود. پالس کوانتیزه شده x_q برای k-امین مود چنین به دست می‌آید:

$$x_q = \text{Round}\left(x.nf_k^{\frac{3}{2}}/12\right) \qquad (۹)$$

Round عدد محاسبه شده را به نزدیکترین عدد صحیح گرد می‌کند. ضرایب نرمالیزاسیون طوری انتخاب شده‌اند که x_q حداکثر برابر 31

است. از اینرو هر پالس کوانتیزه شده با 5 بیت و شماره هر مود با 3 بیت مشخص می‌شود. قالب اطلاعات 8 بیتی و آرایش بیت‌ها مطابق با جدول (۱) است. این روش ضریب امنیت سیگنال کد شده را نیز بالا می‌برد. علاوه بر این به منظور فشرده‌سازی بیشتر و کاهش نرخ بیت، یک کدینگ بدون تلفات هافمن نیز بر روی دنباله بیت به کار رفته است [۱۲].

Table (1): Value of each coded data package in binary

جدول (۱): آرایش بیت‌های وضعیتی و اطلاعاتی، قبل از کدینگ هافمن

قالب 8 بیتی – شماره بیت		B7	B6	B5	B4	B3	B2	B1	B0
تعداد صفر	Mode 0	0	0	-	-	-	-	-	-
پالس‌های غیر صفر کوانتیزه شده	Mode 1	0	1	0	-	-	-	-	-
	Mode 2	0	1	1	-	-	-	-	-
	Mode 3	1	0	0	-	-	-	-	-
	Mode 4	1	0	1	-	-	-	-	-
	Mode 5	1	1	0	-	-	-	-	-
	Mode 6	1	1	1	-	-	-	-	-

۷- دیکدینگ

مطابق با شکل (۶)، هدف بازگشت به فرم نمایش شنوایی یا بازسازی BMM است. بدین منظور ابتدا دنباله بیت کد شده توسط یک دیکدر هافمن، کدگشایی و به صورت بسته‌های 8 بیتی در می‌آید. سپس، پس از استخراج بیت‌های وضعیتی و تشخیص مود به کار رفته، با استفاده از یک دیکدر Run-Length، پالس‌های صفر و با استفاده از تابع معکوس کوانتیزاسیون، پالس‌های غیرصفر مجدداً تولید می‌شوند. رابطه این تابع چنین است:

$$x = 12\ x_q \cdot nf_k^{-\frac{3}{2}} \qquad (۱۰)$$

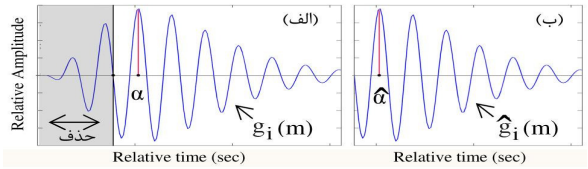
پارامترها مشابه با رابطه (۹) است. در ادامه، بردار حاصله به بردارهای کوچکتر مربوط به هر زیرباند تقسیم شده و ایمپالس‌های عصبی به دست می‌آیند. این ایمپالس‌ها باید مجدداً به فرم BMM تبدیل شوند. بدین منظور در این مقاله، یک روش جدید پیشنهاد می‌شود. در هر زیرباند پاسخ ضربه پنجره شده فیلتر گاماتن مختلط مربوط به آن زیرباند $g_i(m)$ در نظرگرفته می‌شود. سپس مطابق با شکل (۸)، قسمت سایه خورده آن حذف شده و با شیفت منحنی، $\hat{g}_i(m)$ به دست می‌آید. اگر α نقطه ماکزیمم مطلق تابع $g_i(m)$ و m_l ریشه‌های آن باشد (l=0,1,…,L) به نحوی که:

$$0 = m_0 < \ldots < m_{k-1} < m_k < \alpha \ ;\ g_i(\alpha) = \max\left(g_i(m)\right) \qquad (۱۱)$$

در این صورت ریشه m_{k-1} مرز ناحیه سایه خورده را مشخص می‌کند. سیگنال BMM خروجی $B_i(n)$ چنین محاسبه می‌شود:

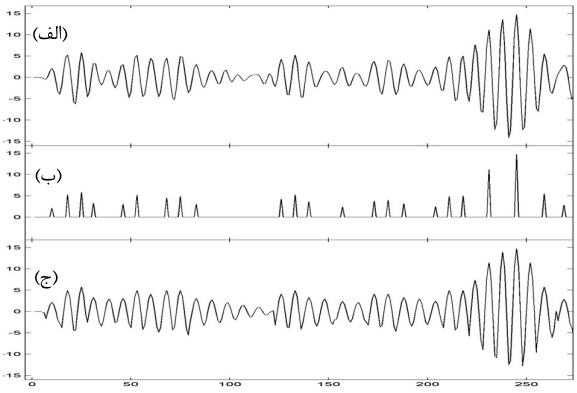
$$B_i(n) = \sum_n x_i(n) . \hat{g}_i(m-n+\hat{\alpha}) \qquad (۱۲)$$

$x_i(n)$ سیگنال دیکد شده در i-امین زیرباند و $\hat{\alpha}$ محل جدید ماکزیمم است. شکل (۹) سیگنال‌های BMM اصلی، ایمپالس‌های عصبی و BMM بازسازی شده را به طور نمونه نشان می‌دهد. مجموع سیگنال‌های خروجی از بانک فیلتر سنتز، سیگنال گفتار نهایی را تولید می‌کند. یک فیلتر DC-زدا نیز جهت افزایش کیفیت و حذف مؤلفه‌های DC ناشی از مرحله بازسازی، اضافه شده است.



شکل (۸): پردازش پاسخ ضربه گاماتن مختلط جهت بازسازی BMM در یک زیر باند نمونه؛ الف: پیش ب: پس از پردازش (فقط بخش حقیقی رسم شده است).

Fig. 8: Processing truncated complex GTF impulse response (only real part is shown); a: After; b: Before



شکل (۹): الف: BMM اصلی؛ ب: ایمپالس‌های عصبی پس ازحذف اضافات (پارامترها)؛ ج: BMM سنترشده؛ برای یک سیگنال نمونه در زیرباند شماره ۸

Fig. 9: Above: Original BMM signal; middle: Coded signal; below: Reconstructed BMM signal, in sub-band number 8

۸- نتایج آزمایش‌ها

مدل پیشنهادی در محیط Matlab شبیه سازی و عملکرد آن از جهت کیفیت و نرخ بیت بررسی شده است. فایل‌های ورودی شامل سیگنال‌های تمیز مربوط به چهار گوینده زن و سه گوینده مرد، سیگنال‌های آلوده به نویزهای سفید ($1/f^0$)، صورتی ($1/f^1$)، قهوه‌ای ($1/f^2$)، ماشین و babble، همگی با $SNR=10_{dB}$، و نویز پریودیک با دامنه نصف ماکزیمم ورودی و فرکانس 1000 Hz است. فایل‌ها به صورت ۱۶ بیت و مونو با پهنای باند 300-3400 Hz می‌باشند که با فرکانس 8 KHz در فرمت Microsoft Wav نمونه‌برداری شده و پس از نویز آلود شدن، فریم به فریم به سیستم وارد می‌شوند [۱۳]. طول هر فریم ورودی 50 میلی‌ثانیه است. شکل (۱۰) اسپکتروگرام سیگنال‌های اصلی، نویزی و سنترشده را به طور نمونه نشان می‌دهد.

۹- ارزیابی کیفیت از طریق آزمون کیفی (Subjective)

ارزیابی برپایه معیارMOS[7] و تحت توصیه‌نامه ITU-T P.835 صورت پذیرفته است [۱۴]. از شنوندگان خواسته شده با هدفون به فایل‌های گفتار ورودی و سنتز شده گوش دهند و برای هر کدام سه MOS از یک تا پنج، برای نرخ اعوجاج گفتار، نرخ اعوجاج پس زمینه و کیفیت کلی سیگنال بدهند. این تست بر روی جمعاً شانزده شنونده با جامعه آماری متنوع، در محیطی آرام و عاری از نویز انجام شده و معدل نتایج آن در جدول (۲) لیست شده است. نتایج حاصله کیفیت خوب سیگنال‌های سنتزشده را تأیید می‌کند.

۱۰- ارزیابی کیفیت از طریق آزمون های کمی (Objective)

در [۱۵] مقایسه‌ای بین آزمون‌های مختلف کمی صورت پذیرفته و سه آزمون LLR[8]، WSS[9] و PESQ[10] را دارای کرولیشن مناسب بـا MOSهای آزمون کیفی، معرفی کرده است. آزمون LLR مبتنی بر اندازه گیری LPC[11] بوده و از این نظر شباهت زیادی با روش‌های Itakura-Saito و Cepstrum دارد [۱۵]. آزمون WSS نیز برپایه یک مدل شنوایی، اختلاف وزن یافته بین شیب‌های طیفی را در هر باند فرکانسی برحسب دسی بل محاسبه می‌کند [۱۵]. نتایج این آزمون‌ها، هرچه کوچکتر باشد، به ایده‌آل نزدیکترند. آزمون پیچیده و استاندارد PESQ نیز تحت توصیه‌نامه ITU-T P.862 بوده و امتیاز آن در بهترین حالت 4.5 است [۱۶].

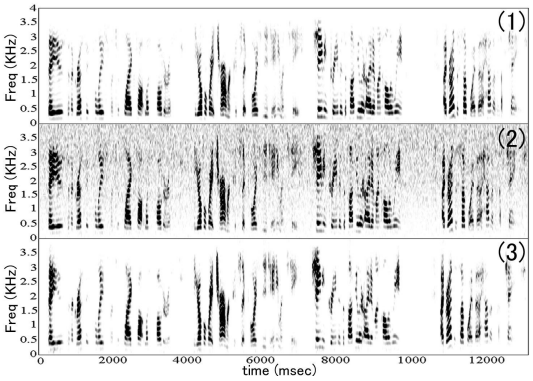
آزمون‌های مذکور بر سیستم پیشنهادی اعمال و نتایج آن درکنار آزمون کیفی در جدول (۲) آمده است. همان طور که مشاهده می‌شود، نتایج آزمون‌های مختلف، ضمن تأیید یکدیگر، عملکرد بسیار خوب سیستم برای ورودی‌های تمیز و عملکرد با ثبات و البته رضایت‌بخش آن را در برابر انواع نویز، نشان می‌دهد.

۱۱- محاسبه نرخ بیت:

در فرکانس نمونه برداری ۸ KHz، متوسط نرخ بیت نتیجه شده از این روش، در حدود 14.6 KHz محاسبه شده که معادل با 1.82 بیت برای هر نمونه است. این مقدار با توجه به نرخ بیت سیگنال ورودی که 128 KHz است، کاهش 88.6 درصد را نشان می‌دهد.

۱۲- مقایسه با دیگر سیستم‌ها

جهت مقایسه، ۶ نوع کدینگ استاندارد شامل CELP، G.723.1، G.726، G.728، G.729 وGSM-FR پیاده‌سازی شده و در شرایط مشابه از آن‌ها نیز آزمون‌های کمی اخذ شده است. سیگنال‌های ورودی برای همه سیستم‌ها یکسان و همگی تمیز می‌باشند. نتایج در جدول (۳) ثبت شده‌اند. همان طور که مشاهده می‌شود، نتایج سیستم پیشنهادی قابل رقابت و در مواردی بهتر نیز هست. البته با توجه به متفاوت بودن نرخ بیت، بار محاسباتی و حوزه عملکرد در الگوریتم‌های مختلف، رتبه‌بندی آن‌ها صحیح نیست و جدول فقط از جهت مقایسه تهیه شده است.



شکل (۱۰): اسپکتروگرام: ۱: اصلی؛ ۲: نویزی (سفید SNR=10dB)؛ ۳: سنتزشده

Fig. 10: Spectrogram of original utterance (1), After adding white noise (10 dB SNR) (2), After synthesis (3)

۱۳- نتیجه‌گیری

یک سیستم کدینگ/ دیکدینگ در حوزه شنوایی، با هدف قابلیت عملکرد در محیط‌های نویزی ارائه و عملکرد آن از طریق آزمون‌های MOS، LLR، WSS، و PESQ مورد ارزیابی قرار گرفت. نتایج حاصله، عملکرد مؤثر، با ثبات و قابل اعتماد سیستم در حضور انواع نویز پس زمینه و همچنین کیفیت رضایت‌بخش سیگنال کد شده در عین کاهش قابل توجه نرخ بیت، را نشان می‌دهد. علاوه بر این، سیستم با چند نمونه کدینگ استاندارد مقایسه شد. در این مدل، مراحل پردازش اگرچه به ظاهر زیادند، اما روندی ساده و در حوزه زمان دارند و درگیر فوریه و توابع پیچیده ریاضی نمی‌شوند. سیستم پیشنهادی به هیچ گونه پیش بینی از نویز و تنظیمات اولیه و یا حتی به هیچ دیتابیس آموزشی نیاز ندارد و برای همه انواع نویز یک سیستم واحد به کار می‌رود.

پی‌نوشت

1- Linear Predictive Coding
2- Equivalent Rectangular Bandwidth
3- Inner Hair Cells
4- Objective & Subjective quality measures
5- Basilar membrane motion
6- Inner Hair Cells
7- Mean Opinion Score
8- Log-likelihood ratio
9- Weighted-slope spectral distance
10- Perceptual evaluation of speech quality
11- Linear predictive coding

Table (2): The results of objective and subjective quality measurements for various input signals (SD=signal distortion; BD=background distortion; Q=overall distortion; STD=standard deviation

جدول (۲): نتایج آزمون‌های کمی و کیفی برای انواع ورودیSD=نرخ اعوجاج گفتار؛ BD=نرخ اعوجاج پس زمینه؛ Q=کیفیت کلی؛ STD=انحراف معیار)

	ایده آل	سیستم پیشنهادی		CELP		G.723.1 ACELP		G.726 ADPCM		G.728 LD-CELP		G.729 CS-ACELP		GSM-FR	
		مرد	زن	مرد	زن	مرد	زن	مرد	زن	مرد	زن	مرد	زن	مرد	زن
نرخ بیت	-	14.6 Kbps		4.8 Kbps		6.4 Kbps		32 Kbps		16 Kbps		8 Kbps		13 Kbps	
آزمون LLR	کوچک	0.28	0.35	0.28	0.26	0.32	0.28	0.24	0.20	0.30	0.27	0.42	0.43	0.15	0.16
آزمون WSS	کوچک	23.6	16.8	42.9	55.4	32.8	33.7	15.4	16.8	24.2	25.6	42.3	40.7	18.1	12.1
نرخ اعوجاج گفتار (SD)	5	4.71	4.57	4.06	3.79	4.71	4.67	4.69	4.80	4.90	4.81	3.96	3.97	4.56	4.47
نرخ اعوجاج پس زمینه (BD)	5	3.17	3.27	2.54	2.33	3.02	2.95	4.08	4.08	3.17	3.09	2.57	2.58	3.79	3.83
نرخ کیفیت کلی (Q)	5	4.38	4.22	3.35	3.03	4.19	4.12	4.02	4.13	4.19	4.05	3.33	3.34	4.38	4.23
آزمون PESQ	4.5	4.04	3.83	2.74	2.44	3.72	3.61	3.41	3.53	3.87	3.70	2.79	2.80	3.96	3.73

Table (3): Quality comparison between the proposed system and some standard codec under same conditions

جدول (۳): نتایج آزمون‌های کمی برای سیستم پیشنهادی و چند نمونه کدینگ استاندارد در شرایط یکسان و با ورودی‌های مشابه

	آزمون کیفی (Subjective)					آزمون‌های کمی (Objective)					
	SD	BD	Q	میانگین	STD	LLR	WSS	SD	BD	Q	PESQ
حالت ایده آل	5	5	5	5	کوچک	کوچک	کوچک	5	5	5	4.5
تمیز زن	4.346	4.115	4.308	4.256	0.484	0.350	16.84	4.573	3.275	4.215	3.834
تمیز مرد	4.423	4.192	4.356	4.321	0.448	0.285	23.63	4.712	3.166	4.379	4.037
نویز سفید	3.846	3.692	3.731	3.756	0.767	0.690	38.24	3.622	2.842	3.086	2.625
نویز 1/f	3.846	3.577	3.654	3.692	0.711	0.843	49.30	3.074	2.393	2.715	2.489
نویز قهوه ای	4.077	3.654	3.923	3.885	0.702	0.485	32.66	4.022	2.877	3.589	3.200
نویز ماشین	3.885	3.577	3.769	3.744	0.515	0.454	45.10	3.852	2.545	3.250	2.782
نویز babble	-	-	-	-	-	0.687	49.71	3.349	2.646	2.778	2.341
نویز پریودیک	3.154	3.769	3.808	3.577	0.698	0.950	73.85	2.266	2.476	2.673	2.227

مراجع

[1] T. Agarwal, P. Kabal, "Pre-processing of noisy speech for voice coders", Proc. IEEE Work. Spee. Cod., Tsukaba, Japan, pp.169-171, Oct. 2002.

[2] V. Krishnan, "A framework for low bit-rate speech coding in noisy environment", PHD Thesis, Georgia Institute of Technology, 2005.

[3] M. Padellini, F. Capman, G. Baudoin, "Very low bit-rate speech coding in noisy environments", Proc. 10th Int. Conf. on Spee. and Comp., SPECOM, 2005.

[4] S. Haque, R. Togneri, A. Zaknich, "Perceptual features for automatic speech recognition in noisy environments", Elsevier Science Publishers, SPECOM, Vol.51, pp.58-75, 2009.

[5] L. Lin, E. Ambikairajah, "Speech denoising based on an auditory filterbank", Proc. 6th Int. Conf. on Sig. Proc., Beijing, pp.552-555, 2002.

[6] R.D. Patterson, "Auditory images: How complex sounds are represented in the auditory system", J. Aco. Soc. Jpn. E., Japan, Vol.21, No.4, pp.183-190, 2000.

[7] L.V. Immerseel, S. Peeters, "Digital implementation of linear Gammatone filters: Comparison of design methods", Acoustical Society of America, ARLO 4(3), July 2003.

[8] L. Lin, W.H. Holmes, E. Ambikairajah, "Auditory filter bank inversion", Proc. ISCAS, Sydney, Vol.2, pp.537-540, 2001.

[9] L. Lin, E. Ambikairajah, W.H. Holmes, "Perceptual domain based speech and audio coder", Proc. 3th Int. Sym. DSPCS, Sydney, pp.6-11, 2002.

[10] L. Rabiner, M. Sambur, "An algorithm for determining the endpoints of isolated utterance", Bell Sys. Tech. J. (BSTJ), Vol.54, pp.297-315, 1975.

[11] G. Kubin, W.B. Kleijn, "On speech coding in a perceptual domain", Proc. ICASSP, Phoenix.USA, pp.205-208, 1999.
[12] D.A. Huffman, "A method for the construction of minimum-redundancy code", Proc. of the IRE, Vol.40, 1952.
[13] Online:http://www.signalogic.com/index.pl?page=codec_samples
[14] ITU-T P.835, "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm", Geneva: International Telecommunication Union.
[15] Y. Hu, P.C. Loizou, "Evaluation of objective quality measures for speech enhancement", IEEE Trans. on Aud. Spee. and Lang. Proc., Vol.16, No.1, 2008.
[16] ITU-T P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", Geneva: InternationalTelecommunication Union, 2001.