

کدگذاری گفتار با استفاده از پیش‌بینی غیر خطی بر پایه بسط سری‌های ولترا

محمدحسن ساوجی و قاسم علی‌پور

۱- مقدمه

بیشتر روش‌های کدگذاری سیگنال گفتار بر پایه حذف افزونگی^۷‌های موجود در سیگنال گفتار و کدکردن آن در نرخ بیت پایین‌تر در کدکننده و بازسازی آن با کیفیت بالا در دکدکننده طراحی می‌شوند. یک روش ساده برای حذف این افزونگی‌ها انجام یک پیش‌بینی از مقدار هر نمونه سیگنال و چندی کردن سیگنال خطای پیش‌بینی^۸ یا سیگنال باقی‌مانده^۹ در یک نرخ بیت پایین‌تر است. پیش‌بینی می‌تواند با یک ترکیب خطی از تعداد محدودی از نمونه‌های پیشین سیگنال گفتار و سیگنال خطای پیش‌بینی (مدل کردن ساز و کار تولید سیگنال گفتار در انسان با یک فیلتر خطی صفر-قطب) و یا در حالت ویژه با یک ترکیب خطی از تعداد محدودی از نمونه‌های پیشین سیگنال گفتار (مدل کردن ساز و کار تولید سیگنال گفتار در انسان با یک فیلتر خطی تمام‌قطب) انجام گیرد [۱] و [۲]. ولی همچنانکه مشخص شده است ساز و کار تولید گفتار در انسان غیر خطی است و مدل‌های غیر خطی در عوض داشتن پیچیدگی‌های بیشتر می‌توانند همبستگی‌های غیر خطی موجود در سیگنال گفتار را به شکل مناسبی استخراج کنند [۳] و [۴].

کارهای فراوانی در زمینه استفاده از شبکه‌های عصبی برای مدل کردن فرایند تولید گفتار در جریان انجام عمل پیش‌بینی انجام گرفته است. گزارش شده است که استفاده از این تکنیک‌ها نسبت به روش‌های خطی باعث افزایش تا ۳dB در بهره^{۱۰} پیش‌بینی می‌شود [۵] تا [۹]. پیش‌بینی غیر خطی همچنین می‌تواند بر پایه بسط سری‌های ولترا باشد که در آن برای سادگی معمولاً تنها بخش‌های خطی و درجه دو در نظر گرفته می‌شوند. خطی بودن خروجی فیلترهای ولترا نسبت به عناصر هسته گسترش تئوری فیلتر خطی بهینه به فیلتر ولترای بهینه را مطابق با هر معیار متناسب با هنجار کمترین مربع^{۱۱} همچون معیار کمترین میانگین مربع خطا^{۱۲} (MMSE) و LS ممکن می‌سازد [۱۰]. همچنین این حقیقت که مفهوم فیلترهای ولترا یک گسترش از مفهوم فیلتر خطی است گسترش بیشتر، محاسبه سریع‌تر ضرایب و پیاده‌سازی آسان‌تر سخت‌افزاری و نرم‌افزاری این سیستم‌ها را ممکن می‌سازد. با این وجود تلاش‌های اندکی برای به کارگیری این نوع سیستم غیر خطی در کدگذاری سیگنال گفتار گزارش شده است [۱۱] تا [۱۲].

ورودی چندی کننده در یک کدکننده^{۱۳} ADPCM در هر زمان سیگنال

چکیده: در سال‌های اخیر به منظور کاهش بیشتر نرخ بیت و از آنجا پهنای باند توجه روزافزونی به استفاده از مدل‌ها و تکنیک‌های غیر خطی پیش‌بینی در کدگذاری گفتار شده است. معمولاً شبکه‌های عصبی برای این هدف به کار می‌روند که منجر به تا ۳dB کاهش بیشتر در انرژی سیگنال تحریک^۱ می‌شوند. پیش‌بینی غیر خطی همچنین می‌تواند بر پایه بسط سری‌های ولترا^۲ انجام گیرد که در آن برای سادگی معمولاً بسط به بخش‌های نخست و دوم محدود می‌شود (پیش‌بینی درجه دو^۳). مطالعات اولیه نشان دادند که در مقایسه با شبکه‌های عصبی استفاده از فیلترهای ولترا منجر به یک کاهش بسیار بیشتر در انرژی سیگنال تحریک می‌شود (۶ تا ۱۰dB). با این وجود به دلیل ناپایداری این کاهش نمی‌تواند منجر به کاهش نرخ بیت یا بهبود نسبت سیگنال به نویز شود. این ناپایداری در دکدکننده به دلیل وجود خطای محاسباتی (برای نمونه ناشی از چندی کردن سیگنال تحریک) و حساسیت بالای محاسبات به این خطا ایجاد می‌شود.

در کار اصلی که در اینجا ارائه می‌شود ناپایداری در کدک^۴ در هر دو نوع پیش‌بینی پیشرو و پسرو به ترتیب با استفاده از الگوریتم‌های کمترین مربع‌ها^۵ (LS) و کمترین میانگین مربع‌های (LMS) سیگنال خطا بررسی می‌شود. نشان داده می‌شود که پایداری در عوض فداکردن بخش عمده‌ای از صرفه‌جویی به دست آمده در انرژی سیگنال تحریک به دست می‌آید به گونه‌ای که سطح کاهش نهایی اغلب همانند شبکه‌های عصبی می‌باشد. در پیش‌بینی پیشرو پس از پایدارسازی و با وجود اندکی افزایش در پیچیدگی عملیاتی در ۲۰ تا ۴۵٪ قالب‌ها افزودن بخش درجه دو سودمند خواهد بود. بر این اساس الگوریتمی توسعه می‌یابد که پیش‌بینی غیر خطی تنها بر روی این قالب‌ها انجام گیرد. این الگوریتم باعث بهبود تا ۴dB در نسبت سیگنال به نویز نهایی می‌شود. پیش‌بینی غیر خطی پسرو متوالی با وجودی که از نقطه نظر پیاده‌سازی بسیار مناسب‌تر است در مقایسه با پیش‌بینی خطی کارایی بهتری را نتیجه نمی‌دهد.

کلید واژه: بسط سری‌های ولترا، پیش‌بینی غیر خطی، پیش‌بینی وقتی پسرو، پیش‌بینی وقتی پیشرو، کدگذاری گفتار، کمترین مربع‌ها، کمترین میانگین مربع‌ها.

این مقاله در تاریخ ۱۰ مهر ماه ۱۳۸۴ دریافت و در تاریخ ۱۲ خرداد ماه ۱۳۸۵ بازنگری شد.

محمدحسن ساوجی، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران، کد پستی ۱۹۸۳۹۶۳۱۱۳ (email: mh_savoji@yahoo.com).

قاسم علی‌پور، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران، کد پستی ۱۹۸۳۹۶۳۱۱۳ (email: alipoor_g@yahoo.com).

7. Redundancy
8. Prediction Error
9. Residual
10. Gain
11. Minimum Squared Norm
12. Minimum Mean Squared Error
13. Adaptive Differential Pulse Code Modulation

1. Excitation
2. Volterra
3. Quadratic Prediction
4. Codec
5. Least Squares
6. Least Mean Squares

۲- پیش بینی و فقی پیشرو (APF)

در این روش با فرض شبه ایستادن^۳ بودن سیگنال گفتار بر روی قالب^۴ هایی کوچک (بخش هایی با طول حدود ۲۰ تا ۳۰ msec)، برای هر قالب ضرایب فیلتر پیش بینی بهینه از دید LS سیگنال خطای پیش بینی از روش خودهمبستگی^۵ یا کوواریانس^۶ به دست می آید [۱۳] و [۱۴]. بر اساس معیار LS هدف کمینه کردن مجموع مربع نمونه های سیگنال خطا (به صورت $\sum d^*(n)$) بر روی M نمونه سیگنال درون قالب گفتار است. در آغاز سیگنال^۷ گفتار در یک پنجره با طول ثابت ضرب می شود. سپس فیلتر به دست آمده برای انجام عمل پیش بینی و از آنجا به دست آوردن سیگنال خطای پیش بینی مطابق با شکل ۱-الف بر روی سیگنال گفتار اعمال می شود.

اگر فیلتر پیش بینی را به شکل یک فیلتر غیر خطی شامل یک بخش درجه دو با طول حافظه N_f و با ضرایب $\{h_f(i, j)\}$ به همراه یک بخش خطی با طول حافظه N_s و با ضرایب $\{h_s(i)\}$ در نظر بگیریم مقدار پیش بینی شده $\tilde{s}(n)$ (خروجی پیش بینی کننده) در هر زمان به صورت زیر خواهد بود [۱۰] و [۱۵]

$$\tilde{s}(n) = \sum_{i=1}^{N_s} h_s(i) s(n-i) + \sum_{k=1}^{N_f} \sum_{j=k}^{N_f} h_f(k, j) s(n-k) s(n-j) \quad (۳)$$

برای سادگی ماتریس ضرایب بخش درجه دو به صورت بالامثلثی در نظر گرفته شده است. هدف کمینه کردن مجموع مربع نمونه های سیگنال خطا بر روی M نمونه از هر قالب به صورت زیر است

$$\sum_n d^*(n) = \sum_n (s(n) - \sum_{i=1}^{N_s} h_s(i) s(n-i) - \sum_{k=1}^{N_f} \sum_{j=k}^{N_f} h_f(k, j) s(n-k) s(n-j))^2 \quad (۴)$$

با در نظر گرفتن تقارن هسته درجه دو یک دستگاه معادلات خطی با $N_s + N_f(N_f + 1)/2$ معادله و همان تعداد مجهول به صورت

$$\mathbf{R}_r(\cdot, l) = \sum_{i=1}^{N_s} h_s(i) \mathbf{R}_r(i, l) + \sum_{k=1}^{N_f} \sum_{j=k}^{N_f} h_f(k, j) \mathbf{R}_r(k, j, l) \quad (۵-الف)$$

$l = 1, 2, \dots, N_s$

$$\mathbf{R}_r(\cdot, p, q) = \sum_{i=1}^{N_s} h_s(i) \mathbf{R}_r(i, p, q) + \sum_{k=1}^{N_f} \sum_{j=k}^{N_f} h_f(k, j) \mathbf{R}_r(k, j, p, q) \quad (۵-ب)$$

$p = 1, 2, \dots, N_f, q = p, p+1, \dots, N_f$

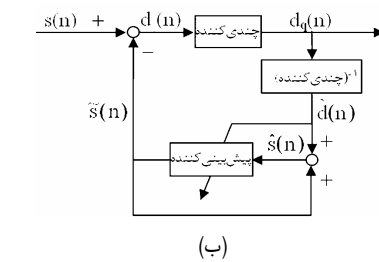
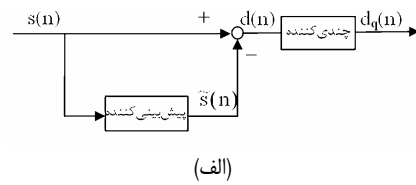
به دست می آید [۱۱] که

$$\mathbf{R}_r(i, j) = \sum_n s(n-i) s(n-j) \quad (۵-ج)$$

$$\mathbf{R}_r(i, j, k) = \sum_n s(n-i) s(n-j) s(n-k) \quad (۵-د)$$

$$\mathbf{R}_r(i, j, k, l) = \sum_n s(n-i) s(n-j) s(n-k) s(n-l) \quad (۵-ه)$$

3. Quasi-Stationary
4. Frame
5. Autocorrelation
6. Covariance



شکل ۱: ساختار کلی الگوریتم کدگذاری سیگنال گفتار با الگوریتم (الف) APF و (ب) APB.

باقی مانده یا خطای پیش بینی به صورت سیگنال تفاضلی زیر است

$$d(n) = s(n) - \tilde{s}(n) \quad (۱)$$

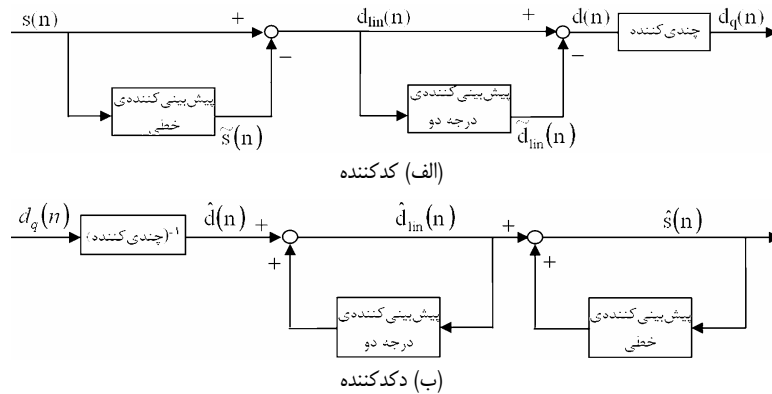
که در آن $\tilde{s}(n)$ یک پیش بینی از سیگنال گفتار در لحظه $n - s(n)$ می باشد. بهره پیش بینی متناسب با این پیش بینی به صورت

$$G = g = \sigma_s^2 / \sigma_d^2 \quad (۲)$$

تعریف می شود که σ_s^2 و σ_d^2 به ترتیب واریانس سیگنال های گفتار و باقی مانده می باشند. به دلیل نایستادن بودن سیگنال گفتار الگوریتم های استفاده شده برای این منظور همواره و فقی خواهند بود. فرستادن یا نفرستادن ضرایب فیلتر پیش بینی وابسته به چگونگی انجام عمل پیش بینی است. بر این اساس دو گونه کلی پیش بینی و فقی عبارتند از پیش بینی و فقی پیشرو^۱ (APF) و پیش بینی و فقی پسرو^۲ (APB) [۱] و [۲].

در اینجا به بررسی نتایج حاصل از به کارگیری فیلتر غیر خطی ولترا در حالت خاص درجه دو در کدگذاری سیگنال گفتار بر پایه پیش بینی می پردازیم. در هر دو روش پیش بینی پیشرو و پسرو پس از بررسی اثر این فیلتر بر روی فشردسازی سیگنال گفتار تلاش می کنیم که الگوریتم پایداری را برای بازسازی سیگنال اصلی گفتار از روی سیگنال خطای کد شده ارائه دهیم. سیگنال خطای پیش بینی پیش از ارسال چندی می شود. در هر حال یک چندی کننده و فقی متناسب با ساختار پیش بینی کننده به کار می رود. بر این اساس در بخش های ۲ و ۳ به ترتیب به معرفی ساختار کدکننده های ADPCM بر پایه بسط سری های ولترا با پیش بینی و فقی پیشرو و پسرو و نیز الگوریتم های پیشنهادی و نتایج به دست آمده می پردازیم. در پایان در بخش ۴ یک جمع بندی و نتیجه گیری ارائه می شود. برای آزمایش الگوریتم ها و ساختارهای معرفی شده در این مقاله همواره چهار سیگنال گفتار شامل دو نمونه ادا شده توسط زن و دو نمونه ادا شده توسط مرد هر کدام شامل یک جمله کامل در زبان انگلیسی را به کار می بریم. سیگنال ها در فرکانس ۸ kHz نمونه برداری شده اند و دامنه هر کدام به صورت غیر فشرده با ۸ بیت چندی شده است. همچنین پیش از انجام آزمایش ها دامنه این سیگنال ها را به مقدار ۱ نرمالیزه می کنیم.

1. Adaptive Prediction-Forward
2. Adaptive Prediction-Backward



شکل ۲: ساختار الگوریتم پیشنهادی در کدگذاری غیرخطی APF برای از میان بردن تزویج میان بخش‌های خطی و درجه دو.

برای هر قالب گفتار با حل این دستگاه ضرایب فیلتر غیر خطی $h_l(i)$ و $h_l(i, j)$ - بهینه مطابق با معیار LS به دست می‌آیند.

۱-۲ بدحالتی و از میان بردن تزویج میان بخش‌های خطی و درجه دوی فیلتر پیش‌بینی

روابط (۵) بیان‌کننده یک دستگاه معادلات خطی از $(N_1 + N_2(N_2 + 1))/2$ معادله با همان تعداد مجهول است که از نظر تئوری مقادیر دقیق این مجهول‌ها با حل دستگاه به دست می‌آیند. ولی یکی از مشکلات عملی حل این دستگاه حساسیت این حل به خطا در داده‌ها به هنگام بدحالت بودن ماتریس ضرایب است. یک دستگاه معادلات خطی را بدحالت گویند اگر یک تغییر نسبتاً کوچک در عناصر ماتریس ضرایب منجر به یک تغییر نسبتاً بزرگ در پاسخ دستگاه گردد. همچنین یک دستگاه معادلات خطی را خوش‌حالت گویند اگر یک تغییر نسبتاً کوچک در عناصر ماتریس ضرایب منجر به یک تغییر نسبتاً کوچک در پاسخ دستگاه گردد [۱۶]. یک معیار مرسوم برای سنجش میزان بدحالتی یک دستگاه معادلات خطی به نام عدد حالت^۳ به صورت نسبت مقادیر تکین^۴ بیشینه و کمینه ماتریس ضرایب تعریف می‌شود. هر اندازه این مقدار به مقدار ۱ نزدیک‌تر باشد ماتریس خوش‌حالت‌تر خواهد بود و مقادیر بزرگ‌تر عدد حالت نشان‌دهنده بدحالت‌تر بودن ماتریس است. به دلیل وجود تزویج میان عناصر خطی و درجه دوی فیلتر پیش‌بینی و نیز بزرگ بودن ابعاد ماتریس ضرایب معمولاً دستگاه (۵) بدحالت است.

برای رفع این مشکل ساختار شکل ۲-الف برای از میان بردن تزویج میان ضرایب بخش‌های خطی و درجه دوی فیلتر پیش‌بینی به هنگام به‌روز کردن آنها و از آنجا خوش‌حالت‌تر کردن ماتریس ضرایب پیشنهاد می‌شود. در این ساختار در آغاز یک فیلتر پیش‌بینی کننده خطی بهینه برای حذف افزونگی‌های خطی سیگنال گفتار بر روی این سیگنال اعمال می‌شود. سپس یک فیلتر درجه دوی بهینه تنها برای حذف افزونگی‌های غیر خطی باقی‌مانده از پیش‌بینی خطی بر روی سیگنال خطای پیش‌بینی خطی اعمال می‌شود. برای این منظور در آغاز باید طول حافظه فیلتر خطی را به اندازه‌ای افزایش دهیم که افزایش بیشتر آن منجر به بهبود چندانی در بهره پیش‌بینی نشود. سپس با ثابت نگه‌داشتن طول حافظه فیلتر خطی به بررسی اثر افزودن بخش درجه دوی فیلتر بر روی

نتایج پردازیم.

سیگنال خطای پیش‌بینی خطی به صورت

$$d_{lin}(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{i=1}^{N_1} h_l(i)s(n-i) \quad (الف-۶)$$

خواهد بود و برای هر قالب، فیلتر پیش‌بینی خطی بهینه می‌کوشد که مجموع مربع M نمونه این سیگنال (به صورت $\sum d_{lin}^2(n)$) را کمینه کند. همچنین سیگنال خطای پیش‌بینی غیر خطی به صورت^۵

$$d(n) = d_{lin}(n) - \tilde{d}_{lin}(n) \\ = d_{lin}(n) - \sum_{k=1}^{N_2} \sum_{j=k}^{N_2} h_l(k, j)d_{lin}(n-k)d_{lin}(n-j) \quad (ب-۶)$$

خواهد بود و برای هر قالب، فیلتر پیش‌بینی درجه دوی بهینه می‌کوشد که مجموع مربع M نمونه این سیگنال را (به صورت $\sum d^2(n)$) کمینه کند.

مطابق این ساختار (۵) را می‌توان به شکل زیر برای به دست آوردن ضرایب بهینه بخش‌های خطی و درجه دوی فیلتر پیش‌بینی، مطابق با معیار کمترین مربع‌های سیگنال خطا (LSE) تغییر داد

$$\mathbf{R}_r(\cdot, l) = \sum_{i=1}^{N_1} h_l(i)\mathbf{R}_r(i, l), \quad l = 1, 2, \dots, N_1 \quad (الف-۷)$$

$$\mathbf{R}_r(\cdot, p, q) = \sum_{k=1}^{N_2} \sum_{j=k}^{N_2} h_l(k, j)\mathbf{R}_r(k, j, p, q) \\ p = 1, 2, \dots, N_2, \quad q = p, p+1, \dots, N_2 \quad (ب-۷)$$

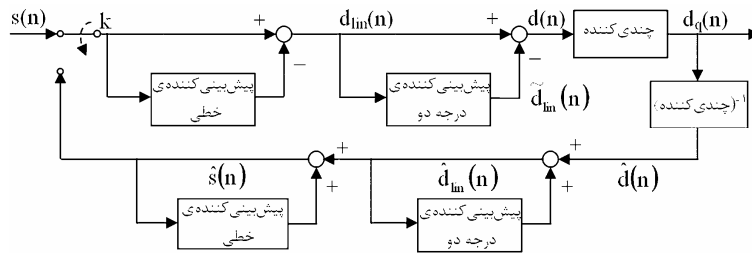
که

$$\mathbf{R}_r(i, j, k) = \sum_n d_{lin}(n-i)d_{lin}(n-j)d_{lin}(n-k) \quad (ج-۷)$$

$$\mathbf{R}_r(i, j, k, l) = \sum_n d_{lin}(n-i)d_{lin}(n-j) \\ \times d_{lin}(n-k)d_{lin}(n-l) \quad (د-۷)$$

حال به جای یک دستگاه معادلات خطی از $(N_1 + N_2(N_2 + 1))/2$ معادله باید دو دستگاه معادلات خطی به ترتیب با N_1 و $N_2(N_2 + 1)/2$ معادله را حل کرد که این به نوبه خود علاوه بر از میان بردن تزویج مذکور باعث کاهش ابعاد ماتریس ضرایب و از آنجا کاهش عدد حالت آن می‌شود. برای نمونه برای $N_1 = 10$ و چند مقدار N_2 مقدار عدد حالت (به صورت نسبت بزرگ‌ترین به کوچک‌ترین مقدار تکین ماتریس ضرایب فیلتر غیر خطی) در دو حالت با تزویج و بدون تزویج و با میانگین‌گیری بر روی ۸۰ قالب نخست از هر کدام از چهار سیگنال گفتار نمونه در جدول ۱ آورده شده است. از نتایج آورده‌شده در جدول پیداست که با استفاده از این

1. Ill-Condition
2. Well-Condition
3. Condition Number
4. Singular Value



شکل ۳: ساختار به کاررفته در کدکننده برای اطمینان از پایداری فیلتر بازسازی با پیش‌بینی غیرخطی پیشرو.

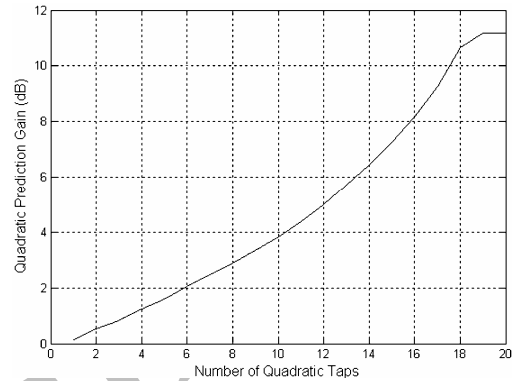
جدول ۱: مقدار میانگین عدد حالت ماتریس ضرایب فیلتر غیرخطی برای مقادیر مختلف طول حافظه

طول حافظه فیلتر غیر خطی	حالت بدون تزویج	حالت با تزویج
۱	۱	۳,۶۶۴۴e+۳
۲	۱,۰۱۱۹۸	۱,۱۰۴۵e+۵
۳	۶۰,۲۷۴۸	۱,۰۸۹۰e+۶
۴	۴۴۷,۰۹۳۱	۱,۳۳۵۳e+۷
۵	۵,۳۸۳۹e+۳	۸,۷۰۷۹e+۱۶
۶	۳,۹۱۴۳e+۶	۱,۰۴۵۹e+۱۸
۷	۳,۲۴۰۳e+۱۵	۳,۹۰۹۸e+۱۸
۸	۱,۶۲۸۴e+۱۶	۱,۵۴۷۷e+۱۸
۹	۱,۹۸۹۹e+۱۶	۵,۱۰۰۷e+۱۸
۱۰	۱,۳۷۴۵e+۱۶	۱,۰۹۵۰e+۲۱

در این ساختار برای هر قالب پس از چندین کردن سیگنال خطا و پیش از ارسال آن (با شبیه‌سازی عمل بازسازی در کدکننده) سیگنال‌های $s(n)$ و $d_{lin}(n)$ را بازسازی می‌کنیم تا از پایداری فیلتر بازسازی در کدکننده اطمینان حاصل کنیم. در صورت پایداری سیگنال خطای چندین شده به طرف کدکننده ارسال می‌شود در غیر این صورت با جایگزینی نمونه‌های بازسازی شده سیگنال گفتار به جای نمونه‌های اولیه آن (با تغییر وضعیت کلید k) محاسبه ضرایب فیلترهای خطی و درجه دو در یک حلقه تا رسیدن به یک فیلتر بازسازی پایدار تکرار می‌شود. با توجه به محدودبودن دامنه سیگنال گفتار اولیه به مقدار ۱ معیار پایداری که در اینجا در نظر گرفته شده است محدودبودن مقادیر سیگنال‌های بازسازی شده به مقدار ۱ می‌باشد.

۲-۳ نتایج با APF

با استفاده از یک فیلتر پیش‌بینی خطی با طول حافظه ثابت برابر با ۱۰ به بررسی اثر افزودن بخش درجه دو بر روی بهره پیش‌بینی می‌پردازیم. پنجره به کار رفته در این آزمایش یک پنجره مستطیل شکل با طول ۲۰ ms است. منحنی شکل ۴ نشان‌دهنده بهبود در میانگین بهره پیش‌بینی بخش درجه دو بر روی چهار نمونه سیگنال با پیش‌بینی پیشرو مطابق با ساختار شکل ۲-الف و با افزایش مقدار N_p از ۱ تا ۲۰ است. همچنانکه می‌بینیم بخش درجه دو بهبود بسیار خوبی (تا بیش از ۱۰dB) را در بهره پیش‌بینی نتیجه می‌دهد، هر چند با افزایش طول حافظه فیلتر خطی از مقدار ۱۰ بهبود در بهره پیش‌بینی بسیار ناچیز است. این نتیجه بیان‌کننده وجود ویژگی‌های غیر خطی در سیگنال گفتار و مناسب‌بودن فیلترهای غیر خطی برای پردازش سیگنال گفتار است. برای نمونه شکل موج یک سیگنال گفتار نمونه و سیگنال باقی‌مانده از پیش‌بینی خطی و غیر خطی با $N_1 = 10$ و $N_p = 10$ در شکل ۵ نشان داده شده است. آزمایش‌های انجام‌گرفته نشان دادند که در صورت استفاده از ساختار شکل ۲-ب برای بازسازی سیگنال گفتار فیلتر بازسازی بر روی برخی از



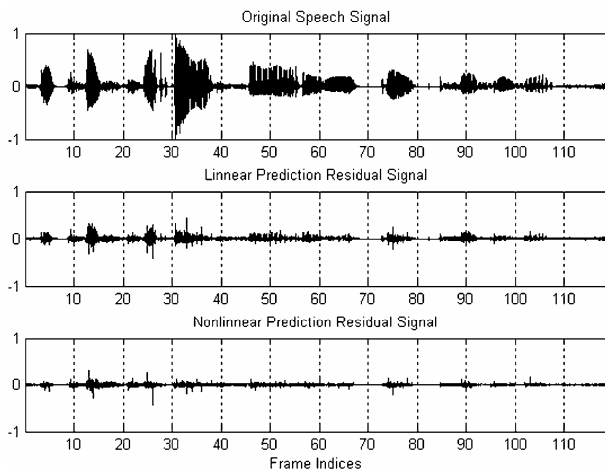
شکل ۴: میانگین بهره بخش درجه دو در الگوریتم APF.

ساختار عدد حالت ماتریس ضرایب بسیار کاهش پیدا می‌کند و در واقع ماتریس ضرایب بسیار خوش‌حالت‌تر خواهد شد. علاوه بر این سودمندی با توجه به افزایش نمایی تعداد عملیات مورد نیاز برای حل یک دستگاه معادلات خطی با افزایش ابعاد ماتریس ضرایب، در ساختار پیشنهادی شکل ۲ تعداد عملیات مورد نیاز و از آنجا پیچیدگی الگوریتم نیز بسیار کاهش می‌یابد. بر این اساس در ادامه این کار همواره این ساختار را برای بهینه‌سازی ضرایب فیلترهای پیش‌بینی‌کننده خطی و درجه دو به کار می‌بریم. می‌توان ساختار شکل ۲-ب را برای دزدکردن سیگنال کدشده مطابق با ساختار شکل ۲-الف به کار برد.

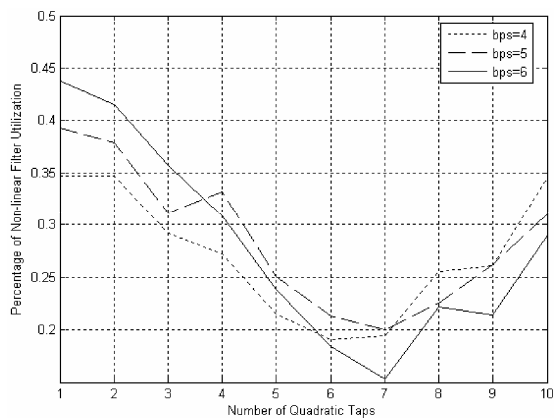
۲-۲ ناپایداری فیلتر بازسازی

در روش خودهمبستگی، ماتریس ضرایب (۷-الف) (برای یافتن ضرایب بهینه بخش خطی فیلتر پیش‌بینی) یک ماتریس توپلیتز^۱ است در نتیجه می‌توان دستگاه (۷-الف) را با روش لوینسن-دوربین^۲ حل کرد. این ویژگی باعث حداقل فزاید فیلتر تحلیل در کدکننده و از آنجا تضمین پایداری بخش خطی فیلتر بازسازی در کدکننده در شرایط دقت نامحدود می‌شود. همچنین ماتریس ضرایب (۷-ب) (برای یافتن ضرایب بهینه بخش درجه دو) همواره یک ماتریس مطلقاً مثبت^۳ است و می‌توان دستگاه (۷-ب) را با استفاده از روش تجزیه‌ی چلسکی^۴ حل کرد [۱۱]. با این وجود تجزیه چلسکی پایداری فیلتر بازسازی یا تجزیه^۵ را تضمین نمی‌کند. این ناپایداری زمانی روی می‌دهد که در کدکننده مقادیر چندین شده سیگنال باقی‌مانده را به جای سیگنال تحریک به کار ببریم. می‌توان با استفاده از ساختار شکل ۳ در کدکننده و در عوض از دست‌دادن بخشی از بهره پیش‌بینی پایداری فیلتر بازسازی را تضمین کرد.

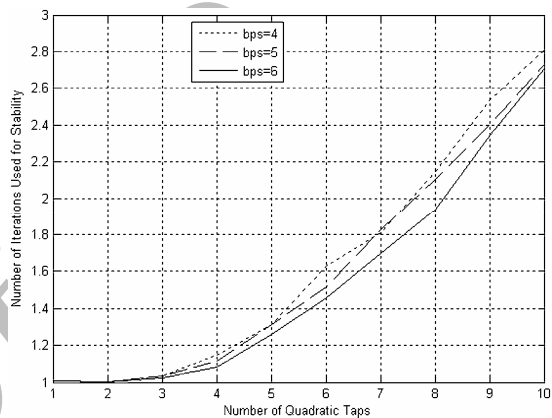
1. Toeplitz
2. Levinson-Durbin
3. Positive Definite
4. Cholesky
5. Synthesis



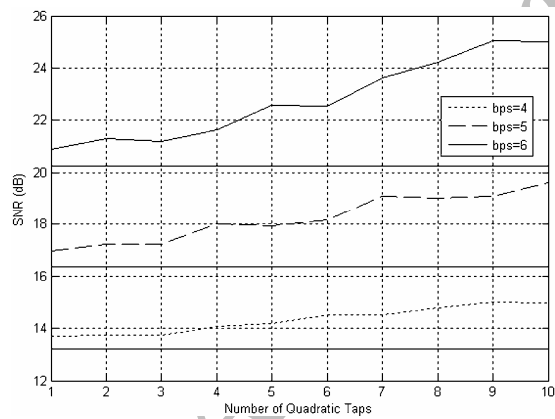
شکل ۵: یک سیگنال گفتار نمونه و سیگنال باقی‌مانده از پیش‌بینی خطی و غیر خطی.



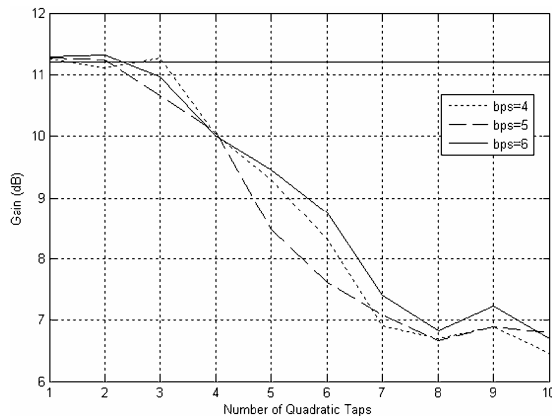
(ب)



(الف)



(د)



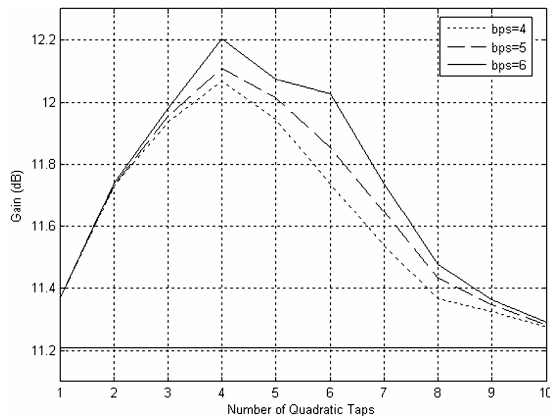
(ج)

شکل ۶: کارایی الگوریتم نهایی با APF با پیش‌بینی غیر خطی با انتخاب نسبت سیگنال به نویز نهایی به عنوان معیار سودمند بودن افزودن بخش درجه دو، میانگین (الف) تعداد تکرارهای مورد نیاز برای رسیدن به فیلتر بازسازی پایدار، (ب) نسبت تعداد قالب‌هایی که فیلتر غیر خطی بر روی آنها اعمال شده است به کل قالب‌ها، (ج) بهره و (د) SNR.

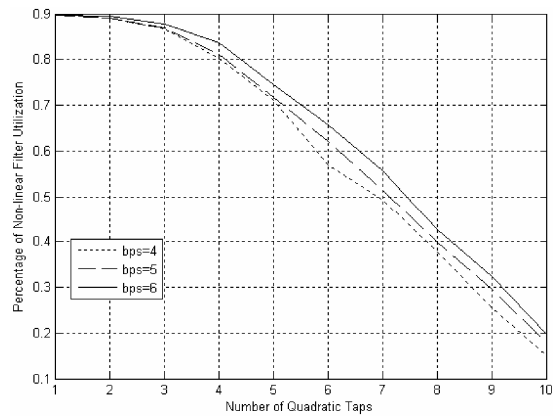
برای پیش‌بینی اکتفا می‌کنیم.

بهبود در نسبت سیگنال به نویز نهایی می‌تواند به عنوان یک معیار در تعیین سودمند بودن افزودن بخش درجه دوی فیلتر پیش‌بینی در نظر گرفته شود. بر این اساس شکل ۶-الف نشان‌دهنده میانگین تعداد تکرارهای لازم برای پایدارسازی فیلتر بازسازی بر روی همه قالب‌های سیگنال گفتار است که خود معیاری از پیچیدگی عملیاتی الگوریتم می‌باشد. باز این نتایج با میانگین‌گیری بر روی چهار نمونه سیگنال و به ازای سه مقدار ۴، ۵ و ۶ برای نسبت تعداد بیت بر نمونه (bps) و $N_1 = 10$ و نیز چند مقدار N_2 به دست آمده است. همچنانکه از این شکل پیداست تنها با اندکی افزایش در پیچیدگی می‌توان به یک فیلتر

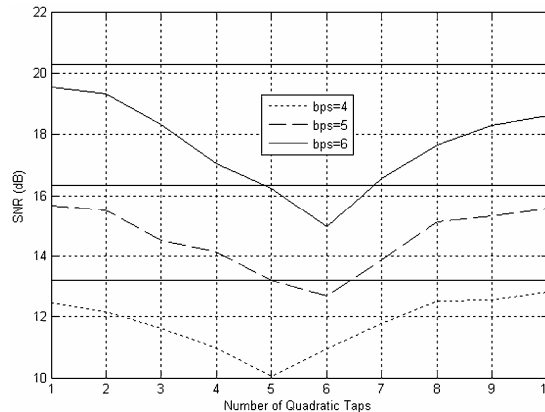
قالب‌ها ناپایدار می‌شود. همچنین در صورت استفاده از ساختار تضمین‌کننده پایداری شکل ۳ بر روی برخی از قالب‌ها (پس از پایدارسازی) افزودن بخش درجه دو به فیلتر خطی تنها باعث بدتر شدن نتایج نهایی (برای نمونه از دید SNR) می‌شود. به همین دلیل برای استفاده بهینه از این ساختار الگوریتم را می‌توان به گونه‌ای تغییر داد که برای هر قالب پس از پایدارسازی و پیش از ارسال مشخص شود که آیا افزودن بخش درجه دو به فیلتر پیش‌بینی باعث بهبود در نتایج نهایی خواهد شد یا نه؟ در صورت سودمند بودن افزودن بخش درجه دو یک فیلتر غیر خطی شامل بخش‌های خطی و درجه دو برای پیش‌بینی سیگنال گفتار به کار می‌رود و در غیر این صورت تنها به اعمال بخش خطی فیلتر



(ب)



(ف)



(ج)

شکل ۷: کارایی الگوریتم نهایی APF با پیش‌بینی غیر خطی با انتخاب مقدار بهره پیش‌بینی نهایی به عنوان معیار سودمندبودن افزودن بخش درجه دو، میانگین (الف) نسبت تعداد قالب‌هایی که فیلتر غیر خطی بر روی آنها اعمال شده است به کل قالب‌ها، (ب) بهره و (ج) SNR.

بیش از ۲۶٫۵dB می‌رسد.

پنجره به کار رفته در این آزمایش یک پنجره مستطیل شکل با طول ۲۰ ms بود. این آزمایش‌ها برای چند مقدار دیگر طول قالب از ۱۰ ms تا ۲۰ ms تکرار شدند که نتایج کمابیش یکسان بودند. در این آزمایش یک چندی‌کننده یکنواخت و قفی با برآورد پیش‌رو (AQF) برای چندی‌کردن سیگنال باقی‌مانده بر روی این سیگنال اعمال شد. آنچنانکه دیده می‌شود با به کار بردن این الگوریتم علاوه بر پایداری به بهبود تا ۴dB نسبت سیگنال به نویز نهایی نیز می‌رسیم. البته برای رسیدن به این هدف مجبور شدیم که از توانایی پیش‌بینی غیر خطی در فشرده‌سازی بیشتر سیگنال گفتار تنها بر روی بخش‌هایی از سیگنال بهره بگیریم، از این رو مقدار بهره پیش‌بینی به دست آمده بسیار کمتر از حالت بهینه خواهد بود.

آزمایش قبل را با در نظر گرفتن بهبود در بهره پیش‌بینی نهایی به عنوان معیار تعیین سودمندبودن پیش‌بینی غیر خطی تکرار کردیم که نتایج آن در شکل ۷ آورده شده است. باز بسته به مقدار N_p در ۱۵٪ تا بیش از ۹۰٪ قالب‌ها افزودن بخش درجه دو باعث بهبود نسبی نتایج (از نقطه‌نظر مقدار بهره نهایی) می‌شود. برای نمونه در $N_p = 5$ این نسبت بیش از ۷۰٪ است. باز این نسبت با افزایش طول حافظه فیلتر درجه دو کاهش می‌یابد. در این حالت همواره بهره نهایی بیشتر از بهره پیش‌بینی خطی تنها می‌باشد که بسته به مقدار N_p این بهبود از ۰٫۲ تا ۱dB است.

۲-۴ مقایسه شبکه‌های عصبی با بسط سری‌های ولترای

به کار رفته در APF

همچنانکه پیش از این گفته شد حل LSE برای ضرایب بهینه فیلتر ولترا می‌تواند به صورت تحلیلی بیان شود و این حل دارای یک مقدار

بازسازی پایدار رسید. برای نمونه این افزایش پیچیدگی برای $N_p = 5$ تنها ۲۵ تا ۳۰٪ است. با افزایش طول حافظه فیلتر درجه دو تعداد تکرارهای لازم برای پایداری فیلتر بازسازی افزایش می‌یابد که این خود به دلیل تأثیرپذیری مقدار بازسازی شده سیگنال در هر زمان از تعداد بیشتر نمونه‌های خطا و همچنین بزرگ‌تر شدن ابعاد ماتریس ضرایب و بد حالت‌تر شدن آن می‌باشد.

همچنین شکل ۶-ب نشان‌دهنده میانگین نسبت تعداد قالب‌هایی که افزودن بخش درجه دو باعث بهبود نسبت سیگنال به نویز در سیگنال بازسازی شده می‌شود به تعداد کل قالب‌ها است. بسته به مقدار N_p در ۱۵٪ تا بیش از ۴۵٪ قالب‌ها افزودن بخش درجه دو باعث بهبود نسبی نتیجه (از نقطه‌نظر نسبت سیگنال به نویز نهایی) می‌شود. برای نمونه در $N_p = 5$ این نسبت بیش از ۲۰٪ است. باز بنا به دلایلی که در مورد نتایج شکل ۶-الف بیان شد این نسبت با افزایش طول حافظه فیلتر درجه دو کاهش می‌یابد. میانگین مقدار بهره پیش‌بینی بخش درجه دو بر روی چهار نمونه سیگنال در شکل ۶-ج نشان داده شده است. همچنین میانگین مقدار SNR نهایی بر روی چهار نمونه سیگنال و برای چند مقدار N_p و سه مقدار bps در مقایسه با SNR به دست آمده از بخش خطی تنها نیز در شکل ۶-د آورده شده است. در این شکل برای هر مقدار بیت بر نمونه SNR به دست آمده از بخش خطی تنها (که با تغییر مقدار N_p ثابت می‌ماند) با خط پر و SNR به دست آمده از فیلتر غیر خطی با خط چین نشان داده شده است. با استفاده از این الگوریتم (در مقایسه با الگوریتم‌های خطی) می‌توان به یک بهبود تا ۴dB در SNR سیگنال بازسازی شده رسید. البته مقدار SNR بر روی برخی از قالب‌ها بیشتر از مقدار میانگین آن است به گونه‌ای که برای $\text{bps}=6$ مقدار این پارامتر به

$$\mathbf{X}_{\nu k} = [\hat{s}(k-1) \quad \hat{s}(k-2) \quad \dots \quad \hat{s}(k-N_{\nu})]^T \quad (ج-۸)$$

$$\mathbf{X}_{\tau k} = [\hat{s}^{\tau}(k-1) \quad \hat{s}(k-1)\hat{s}(k-2) \quad \dots \quad \hat{s}(k-1)\hat{s}(k-N_{\tau}) \quad \hat{s}^{\tau}(k-2) \quad \dots \quad \hat{s}^{\tau}(k-N_{\tau})]^T \quad (د-۸)$$

و مقادیر سیگنال خطای پیش‌بینی در زمان k به صورت زیر خواهد بود

$$d(k) = s(k) - \tilde{s}(k) = s(k) - \mathbf{H}_{\nu k}^T \mathbf{X}_{\nu k} - \mathbf{H}_{\tau k}^T \mathbf{X}_{\tau k} \quad (الف-۹)$$

می‌توان الگوریتم تکراری LMS نرمالیزه زیر را برای به روزکردن ضرایب فیلتر در هر تکرار k به کار برد [۱۰] و [۱۵]

$$\mathbf{H}_{\nu k+1} = \mathbf{H}_{\nu k} + \frac{\mu_{\nu} \mathbf{X}_{\nu k} \hat{d}(k)}{\sigma_k^{\nu}} \quad (ب-۹)$$

$$\mathbf{H}_{\tau k+1} = \mathbf{H}_{\tau k} + \frac{\mu_{\tau} \mathbf{X}_{\tau k} \hat{d}(k)}{\sigma_k^{\tau}} \quad (ج-۹)$$

در این روابط μ_{ν} و μ_{τ} پارامترهای همگرایی برای کنترل سرعت همگرایی ضرایب فیلتر به ضرایب فیلتر بهینه می‌باشند و

$$\sigma_k^{\nu} = \alpha \hat{s}^{\nu}(k) + (1-\alpha) \sigma_{k-1}^{\nu} \quad (د-۹)$$

برآوردی از توان سیگنال ورودی به فیلتر است که در آن α نیز یک عامل فراموشی می‌باشد.

اینک با پیاده‌سازی فیلتر غیر خطی مطابق ساختار پیشنهادی شکل ۸ برای از میان بردن تزویج میان عناصر خطی و درجه‌دوی فیلتر پیش‌بینی مقادیر سیگنال خطای پیش‌بینی خطی و غیر خطی در زمان k به ترتیب به صورت زیر خواهند بود

$$d_{lin}(k) = s(k) - \tilde{s}(k) = s(k) - \mathbf{H}_{\nu k}^T \mathbf{X}_{\nu k} \quad (الف-۱۰)$$

$$d(k) = d_{lin}(k) - \tilde{d}_{lin}(k) = d_{lin}(k) - \mathbf{H}_{\tau k}^T \mathbf{X}_{\tau k} \quad (ب-۱۰)$$

در این حالت \mathbf{X}_{ν} بردار شامل حاصل ضربهای نمونه‌های پیشین سیگنال $\hat{d}_{lin}(n)$ متناظر با بردار ضرایب \mathbf{H}_{τ} و به شکل زیر می‌باشد

$$\mathbf{X}_{\tau k} = [\hat{d}_{lin}^{\tau}(k-1) \quad \hat{d}_{lin}(k-1)\hat{d}_{lin}(k-2) \quad \dots \quad \hat{d}_{lin}(k-1)\hat{d}_{lin}(k-N_{\tau}) \quad \hat{d}_{lin}^{\tau}(k-2) \quad \dots \quad \hat{d}_{lin}^{\tau}(k-N_{\tau})]^T \quad (۱۱)$$

در اینصورت در هر تکرار می‌توان الگوریتم زیر را برای به روزکردن ضرایب فیلتر خطی و درجه دو به کار برد

$$\mathbf{H}_{\nu k+1} = \mathbf{H}_{\nu k} + \frac{\mu_{\nu} \mathbf{X}_{\nu k} \hat{d}_{lin}(k)}{\sigma_k^{\nu}} \quad (الف-۱۲)$$

$$\mathbf{H}_{\tau k+1} = \mathbf{H}_{\tau k} + \frac{\mu_{\tau} \mathbf{X}_{\tau k} \hat{d}(k)}{\sigma_k^{\tau}} \quad (ب-۱۲)$$

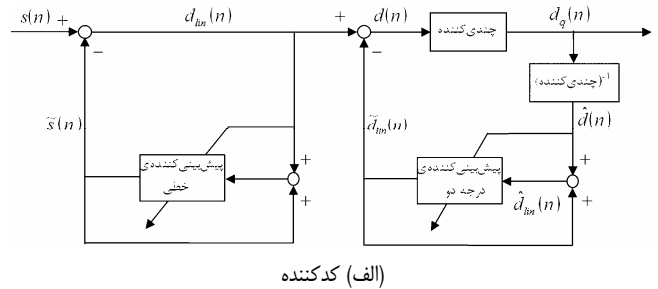
که

$$\sigma_k^{\nu} = \alpha \hat{s}^{\nu}(k) + (1-\alpha) \sigma_{k-1}^{\nu} \quad (ج-۱۲)$$

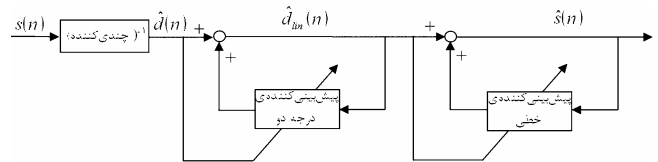
$$\sigma_k^{\tau} = \alpha \hat{d}_{lin}^{\tau}(k) + (1-\alpha) \sigma_{k-1}^{\tau} \quad (د-۱۲)$$

۳-۱ نتایج با APB

با ثابت نگه‌داشتن طول حافظه بخش خطی برابر با ۱۰ و با میانگین‌گیری بر روی چهار نمونه سیگنال، شکل ۹ نشان‌دهنده میانگین افزایش بهره بخش درجه دوی فیلتر پیش‌بینی با افزایش طول حافظه این



(الف) کدکننده



(ب) دکدکننده

شکل ۸: ساختار الگوریتم پیشنهادی در کدگذاری غیرخطی APB برای از میان بردن تزویج میان بخش‌های خطی و درجه دو.

بهینه سراسری است. همچنین با استفاده از اینگونه فیلترها ذاتاً می‌توان به یک بهره پیش‌بینی بسیار بالاتر (تا بیش از ۱۰dB) رسید در حالی که بیشترین بهره به دست آمده با استفاده از شبکه‌های عصبی از ۳dB فراتر نمی‌رود. به علاوه وجود الگوریتم‌های مناسب برای محاسبه فیلتر تحلیل، پیاده‌سازی فیلترهای ولترا را آسان‌تر کرده است در حالی که فیلتر تحلیل در شبکه‌های عصبی تنها می‌تواند به صورت بازگشتی و با الگوریتم تکثیر به عقب محاسبه شود. با این وجود تعداد ضرایب مورد نیاز برای توصیف فیلتر پیش‌بینی ولترا به صورت غیر خطی با اندازه فیلتر پیش‌بینی افزایش می‌یابد در حالی که این افزایش در مورد شبکه‌های عصبی بسیار آهسته‌تر است [۱۲]. همچنین در APF بر پایه بسط سری‌های ولترا فیلتر بازسازی ذاتاً ناپایدار است و پایدارسازی آن ممکن است در عوض کاهش عمده بهره پیش‌بینی به دست آید.

۳- پیش‌بینی وقتی پسرو

معایب اصلی APF تأخیر در کدگذاری و نیاز به بافرکردن داده و نیز ظرفیت کانال اضافی مورد نیاز برای ارسال اطلاعات جانبی می‌باشند. اگر ضرایب پیش‌بینی‌کننده بهینه را بتوان بر پایه داده چندجی شده و ارسال شده برآورد کرد می‌توان آنها را در دکدکننده بازسازی کرده تا دیگر نیازی به ارسال این ضرایب نباشد. بر این اساس کدکردن با پیش‌بینی وقتی پسرو مطابق با ساختار شکل ۱-ب معرفی شده است. در اینجا هدف رسیدن به کمینه میانگین مربع خطای (MMSE) پیش‌بینی است. در بیشتر موارد الگوریتم بر پایه روش تندترین شیب^۲ و با تقریب LMS پیاده می‌شود [۱۰].

در زمان k با به کار بردن یک فیلتر غیر خطی درجه دو با ماتریس ضرایب بخش‌های خطی و درجه دو به ترتیب به صورت

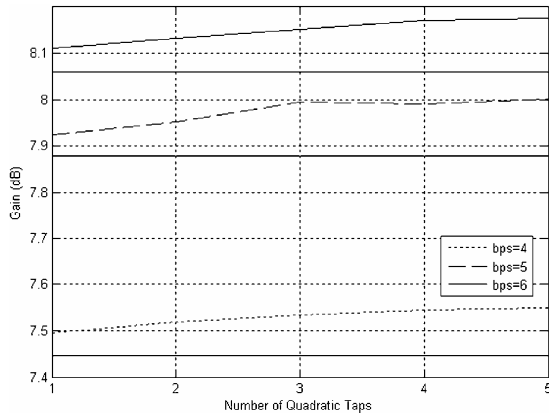
$$\mathbf{H}_{\nu k} = [h_{\nu k}(1) \quad h_{\nu k}(2) \quad \dots \quad h_{\nu k}(N_{\nu})]^T \quad (الف-۸)$$

$$\mathbf{H}_{\tau k} = [h_{\tau k}(1,1) \quad h_{\tau k}(1,2) \quad \dots \quad h_{\tau k}(1,N_{\tau}) \quad h_{\tau k}(2,2) \quad h_{\tau k}(2,3) \quad \dots \quad h_{\tau k}(N_{\tau},N_{\tau})]^T \quad (ب-۸)$$

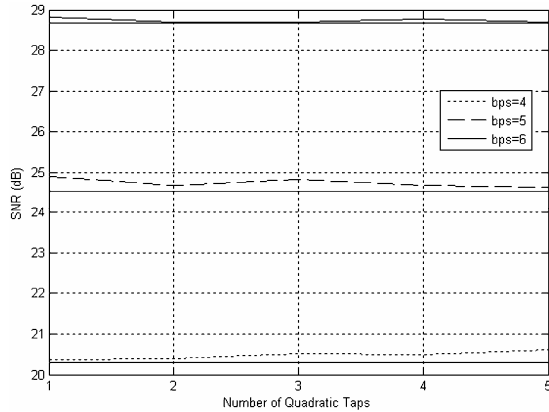
و بردارهای متناظر شامل نمونه‌های ورودی به ترتیب به صورت

1. Back Propagation

2. Steepest Descent



(الف)



(ب)

شکل ۱۱: کارایی الگوریتم نهایی APB با پیش‌بینی غیر خطی، میانگین (الف) بهره پیش‌بینی بخش درجه دو و (ب) SNR.

می‌باشد. این نتایج در مقایسه با نتایج به‌دست آمده از حالت خطی تنها با $N_1 = 10$ (که در همه موارد با خط پر نشان داده شده‌اند) آورده شده‌اند. آنچنانکه دیده می‌شود افزودن بخش درجه دو باعث مقداری بهبود در بهره و نیز مقدار کمی بهبود در SNR نهایی می‌شود.

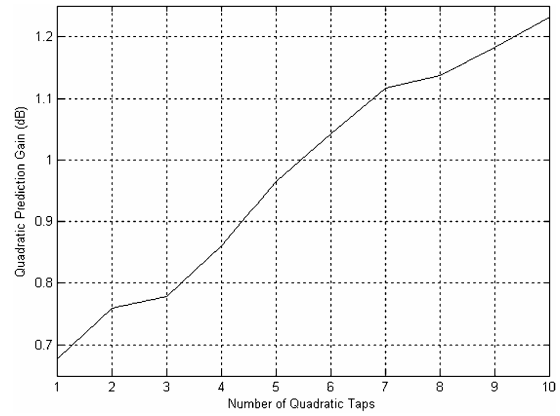
ناپایداری در ساختار شکل ۸-الف به دلیل نامحدود بودن طول حافظه بخش خطی فیلتر بازسازی و از آنجا منتشر شدن نویز وارد شده در جریان بازسازی سیگنال $d_{lin}(n)$ در بخش خطی روی می‌دهد. یک روش دیگر برای مقابله با این ناپایداری محدود کردن طول حافظه بخش خطی فیلتر بازسازی با استفاده از تکنیک لیکجینگ^۱ و با تغییر الگوریتم ۱۲-الف به صورت زیر است [۱]

$$\mathbf{H}_{1k+1} = a\mathbf{H}_{1k} + \frac{2(1-a)\mu_{1k}\mathbf{X}_{1k}d_{lin}(k)}{\sigma_{1k}^2} \quad (13)$$

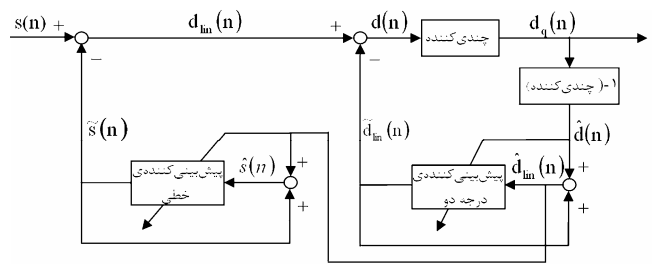
در اینصورت $0 < a < 1$ را عدد لیکجینگ گویند که طول حافظه فیلتر بازسازی را کنترل می‌کند. تغییر الگوریتم ۱۲-الف به الگوریتم ۱۳ باعث اندکی دور شدن از حالت بهینه فیلتر پیش‌بینی خطی و از آنجا کاهش بهره پیش‌بینی خطی می‌شود به گونه‌ای که در کل بهره پیش‌بینی از حالت خطی تنها هم کمتر خواهد شد.

۴- نتیجه‌گیری

در اینجا در آغاز ساختاری برای از میان بردن ترویج میان بخش‌های



شکل ۹: میانگین بهره بخش درجه دو در الگوریتم APB.



شکل ۱۰: ساختار الگوریتم پیشنهادی برای جلوگیری از انتشار نویز در APB.

بخش و با پیش‌بینی پسرو - مطابق ساختار شکل ۸-الف - است. پارامترهای همگرایی به صورت $\mu_1 = 0.05$ و $\mu_2 = 0.05$ انتخاب شده‌اند. البته در این آزمایش به منظور بررسی توان فشرده‌سازی این الگوریتم و در مقایسه با پیش‌بینی پیشرو سیگنال خطا چندی نشده است. این بار هم افزودن بخش درجه دو باعث مقداری بهبود در بهره پیش‌بینی می‌شود. البته بهبود بهره در این حالت بسیار کمتر از حالت پیش‌بینی پیشرو می‌باشد که بیشتر به دلیل تقریب LMS به کار رفته در این الگوریتم و خطای اضافی ناشی از آن [۱۴] و نیز این حقیقت که الگوریتم بازگشتی LMS حتی با یک فیلتر پیش‌بینی خطی هم ذاتاً غیر خطی است [۱۷] می‌باشد.

به دلیل وجود نویز چندی کردن سیگنال ورودی به فیلتر پیش‌بینی خطی در کدکننده و دکدکننده متفاوت خواهد بود که این خود باعث ناپایدار شدن بخش خطی فیلتر بازسازی می‌شود. بر همین اساس و به منظور جلوگیری از تأثیر نویز چندی کردن بر روی عملیات بازسازی در بخش خطی فیلتر بازسازی و در واقع شبیه‌سازی دقیق عملیات انجام گرفته برای بازسازی سیگنال در کدکننده ساختار کدکننده شکل ۸-الف را به صورت ساختار شکل ۱۰ تغییر می‌دهیم. در واقع در این حالت به جای مقادیر سیگنال $d_{lin}(n)$ مقادیر بازسازی شده آن $\hat{d}_{lin}(n)$ را که در کدکننده نیز موجود می‌باشد برای بازسازی سیگنال گفتار در بخش خطی به کار می‌بریم. به کارگیری سیگنال نویزی شده $\hat{d}_{lin}(n)$ به جای سیگنال $d_{lin}(n)$ در پیش‌بینی خطی منجر به اندکی کاهش در بهره پیش‌بینی این بخش می‌شود به گونه‌ای که افزودن بخش درجه دو در کل منجر به بهبود محسوسی در کارایی کدکننده (از دید بهره پیش‌بینی و سطح نویز وارد شده به سیگنال بازسازی شده) نمی‌شود.

شکل ۱۱ کارایی این الگوریتم را از دید بهره پیش‌بینی و نیز SNR سیگنال بازسازی شده برای $N_1 = 10$ و مقدار N_p از ۱ تا ۵ و برای سه نرخ بیت نشان می‌دهد. باز هم پارامترهای همگرایی به صورت $\mu_1 = 0.05$ و $\mu_2 = 0.05$ انتخاب شده‌اند. همچنین چندی کننده به کار رفته در این حالت یک چندی کننده یکنواخت و فقی با برآورد پسرو

- [5] M. Faúndez, F. Vallverdú, and E. Monte, "Nonlinear prediction with neural nets in ADPCM," in *Proc. IEEE ICASSP*, vol. 1, pp. 345-348, Seattle, US, 1998.
- [6] M. Faúndez-Zanuy and O. Oliva, "ADPCM with nonlinear prediction," in *Signal Process. IX: Theories and Applications, EUSIPCO, Rodas, Greece*, pp. 1205-1208, 1998.
- [7] N. Tishby, "A dynamical systems approach to speech processing," in *Proc. ICASSP*, pp. 365-368, 1990.
- [8] B. Townshend, "Non-linear prediction of speech," in *Proc. ICASSP*, pp. 425-428, 1991.
- [9] G. D'Alessandro, M. Faundez Zanuy, and F. Piazza, "A new sub-band non-linear prediction coding algorithm for narrowband speech signal -the NADPCMB-MLT coding scheme," in *Proc. ICASSP*, vol. 1, pp. 1025-1028, 2002.
- [10] V. J. Mathews and G. L. Sicuranza, *Polynomial signal processing*, Wiley, New York, 2000.
- [11] E. Mumolo and D. Francescato, "Adaptive predictive coding of speech by means of Volterra predictors," in *Proc. IEEE Winter Workshop on Nonlinear Digital Signal Process., Tampere, Finland*, pp. 2.1.4.1-2.1.4.4, Jul. 1993.
- [12] J. Thyssen, H. Nielsen, and S. D. Hansen, "Non-linear short-term prediction in speech coding," in *IEEE Proc. ICASSP, Aurlia*, pp. 185-188, Apr. 1994.
- [13] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.
- [14] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 1991.
- [15] G. L. Sicuranza, "Quadratic filters for signal processing," in *Proceeding of IEEE*, vol. 80, no. 8, pp. 1263-1285, Aug. 1992.
- [16] G. Lindfield and J. Penny, *Numerical Methods Using MATLAB®*, 2nd ed. Prentice-Hall, New Jersey, 1994.
- [17] M. Reuter, K. Quirk, J. Zeidler, and L. Milstein, "Non linear effects in LMS adaptive filters," in *Proc. of Symposium 20S00 on Adaptive Systems for Signal Process., Comm. and Control Symp.*, pp. 141-146, Lake Louise, Canada, Oct. 2000.

محمدحسن ساوچی تحصیلات خود را در مقطع کارشناسی و کارشناسی ارشد مهندسی برق در دانشگاه صنعتی شریف در سالهای ۱۳۵۱ و ۱۳۵۳ به اتمام رساند. وی در سال ۱۹۷۹ از انستیتو پلی‌تکنیک Grenoble فرانسه مدرک دکتر مهندسی خود را در رشته الکترونیک و مخابرات دریافت نمود و در سال ۱۹۸۱ در دانشگاه آکسفورد انگلستان دوره فوق دکترا گذراند. ایشان بین سالهای ۱۹۸۱ تا ۱۹۹۵ در دانشگاهها و مراکز تحقیقاتی مختلف اروپا مشغول تدریس و تحقیق بود. آخرین مسئولیت ایشان خارج از کشور استاد و مدیر گروه پردازش سیگنال در دانشگاه Santander کشور اسپانیا بوده است. دکتر ساوچی از سال ۱۳۸۰ تاکنون با سمت استاد الکترونیک و مخابرات در دانشگاه شهید بهشتی به فعالیت خود ادامه می‌دهد. برخی از زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش سیگنال، پردازش تصویر و ویدئو، پردازش سیگنال صحبت و فیلترهای وقتی و غیرخطی.

قاسم علی پور در سال ۱۳۸۱ مدرک کارشناسی مهندسی مخابرات خود را از دانشگاه تبریز و در سال ۱۳۸۴ مدرک کارشناسی ارشد مهندسی الکترونیک خود را از دانشگاه شهید بهشتی تهران دریافت نمود. وی از سال ۱۳۸۴ تاکنون به عنوان کارشناس ارشد ابزار دقیق در مجتمع صنایع پتروشیمی کرمانشاه به کار مشغول است. زمینه‌های تحقیقاتی مورد علاقه نامبرده عبارتند از: استفاده از روش‌های آماری و ایده‌های نوین پردازش سیگنال دیجیتال در پردازش سیگنال‌های صحبت و تصویر به ویژه برای مقاصد فشرده‌سازی، کدگذاری و نوین‌سازی.

خطی و درجه دوی فیلتر پیش‌بینی پیشنهاد شد. با مقایسه بهره به دست آمده از پیش‌بینی خطی و درجه دو مشخص شد که سیگنال گفتار دارای برخی ویژگی‌های غیر خطی است. با وجود توانایی بسیار فیلتر ولترا در فشرده‌سازی سیگنال گفتار عملاً پایدار کردن فیلتر بازسازی در ددکننده یک مسئله اساسی است. بر این اساس در اینجا و در هر دو روش پیش‌بینی پیشرو و پسرو هدف اصلی رسیدن به یک الگوریتم پایدار کدگذاری سیگنال گفتار با استفاده از پیش‌بینی بر پایه فیلترهای ولترا بود. در هر دو حالت الگوریتم‌ها و ساختارهای گوناگونی برای این هدف معرفی شدند. در پایان رسیدن به یک الگوریتم پایدار با وجود افزایش اندک در پیچیدگی عملیاتی مستلزم فداکردن مقدار زیادی از توان فشرده‌سازی فیلتر پیش‌بینی بود. آزمایش‌ها نشان دادند که در پیش‌بینی پیشرو برای ۱۵٪ تا ۴۵٪ قالب‌های سیگنال گفتار (پس از پایدارسازی) افزودن پیش‌بینی درجه دو باعث بهبود در نتایج (از نقطه‌نظر نسبت سیگنال به نویز نهایی) می‌شود. با اعمال بخش درجه‌دوی پیش‌بینی تنها بر روی این قالب‌ها می‌توان با وجود پایداری به یک بهبود تا ۴dB در نسبت سیگنال به نویز نهایی رسید. ولی پیش‌بینی غیر خطی پسرو متوالی با وجودی که از نقطه‌نظر پیاده‌سازی بسیار مناسب‌تر است در مقایسه با پیش‌بینی خطی کارایی بهتری را نتیجه نمی‌دهد.

در هر حال با توجه به اینکه کارهای انجام‌گرفته در زمینه استفاده از فیلترهای ولترا در کدگذاری سیگنال گفتار کم است جا دارد که بسیاری از جنبه‌های گوناگون موضوع بیشتر مورد بررسی قرار گیرد. برای نمونه بررسی روش‌های جلوگیری از انتشار خطا در جریان بازسازی سیگنال گفتار، بررسی الگوریتم‌های پیشنهادی بر روی قالب‌های مشخص از گفتار و تعیین کارایی این الگوریتم‌ها بر روی سیگنال‌های گفتار با ویژگی‌های مشخص، استفاده از مدل‌های صفر و قطب به جای مدل‌های تمام قطب در مدل‌کردن فرایند تولید گفتار، ترکیب مناسب کدکننده‌های PCM و DPCM و نیز ترکیب مناسب فیلترهای خطی و غیر خطی در مدل‌کردن فرایند تولید سیگنال گفتار و گسترش الگوریتم‌های وقتی که با وجود همگرایی سریع از نظر عددی پایدار باشند از جمله پیشنهاد‌های ما برای بررسی‌ها و پژوهش‌های آینده می‌باشند.

مراجع

- [1] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall Inc., 1984.
- [2] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publication Company, New York, 1993.
- [3] A. A. Beex and J. R. Zeidler, "Non-linear effects in adaptive linear prediction," in *4th IASTED Int. Conf. on Signal and Image Process. (SIP), Kaua'i, Hawaii*, pp. 21-26, Aug. 2002.
- [4] H. M. Teager, "Some observations on oral flow vocalization," *IEEE Trans. ASSP*, vol. 28, no. 5, pp. 559-601, 1980.