بکارگیری شبکه‌های عصبی ARTMAP فازی و KSOFM برای کاهش نرخ بیت کدکننده گفتار مبتنی‌بر استاندارد FS-1015

منصور شیخان[۱]، داود غرویان[۲]، علی اسلام‌زاده[۳]
۱- استادیار، گروه مهندسی مخابرات مرکز تحصیلات تکمیلی، دانشگاه آزاد اسلامی، واحد تهران‌جنوب، msheikhn@azad.ac.ir
۲- استادیار، گروه مهندسی الکترونیک، دانشگاه صنعت آب و برق، gharavian@pwut.ac.ir
۳- کارشناس ارشد مهندسی الکترونیک، دانشگاه آزاد اسلامی، واحد تهران‌جنوب، a_eslamzade@azad.ac.ir

چکیده

طیف سیگنال گفتار به پارامترهای پیش‌بینی خطی (LPC) حساسیت زیادی داشته و بروز خطاهای کوچک در چندی‌سازی منجر به ناپایداری در فیلتر سنتز می‌شود و این درحالی است که پارامترهای زوج طیفی (LSP) در این مورد به‌عنوان روش بازنمایی مؤثرتر مطرح هستند. از سوی دیگر شبکه‌های عصبی مصنوعی تاکنون به‌عنوان ابزاری موفق برای بهبود کیفیت و کاهش پیچیدگی محاسباتی کدکننده‌های گفتار بکارگرفته شده‌اند. براین اساس در این مقاله با ایجاد تغییراتی در ساختار کدکننده مبتنی‌بر استاندارد FS-1015، سعی در کاهش نرخ بیت ۲/۴ کیلوبیت برثانیه‌ای این الگوریتم، ضمن بهبود نتایج عملکردی آن شده است. در این راستا، با بکارگیری پارامترهای LSP به‌جای LPC و نیز چندی‌کننده‌های برداری مبتنی بر شبکه‌های عصبی KSOFM (نسخه با آموزش تحت نظارت) و ARTMAP فازی، نرخ بیت کدکننده به kbps ۱/۹ کاهش و کیفیت گفتار سنتزشده در ملاک MOS نیز به‌ترتیب به‌میزان ۰/۱۳ و ۰/۲۶ بهبود یافته است. همچنین زمان اجرای الگوریتم در صورت بکارگیری دو روش چندی‌سازی عصبی مذکور به‌ترتیب به‌میزان ۲۷٪ و ۴۳٪ کاهش پیدا می‌کند.

واژه‌های کلیدی

کدکننده گفتار، شبکه‌های عصبی، چندی‌سازی برداری، ARTMAP فازی

۱– مقدمه

بسیاری از کدکننده‌های گفتار که نرخ بیت پایینی دارند، مبتنی‌بر مدل پیش‌بینی خطی[۱](LPC) می‌باشند. در این مدل، طیف کوتاه مدت گفتار توسط یک فیلتر تمام-قطب بازنمایی می‌شود [۱]. اما از آنجا که پارامترهای LPC کراندار نیستند، تعریف ناحیه چندی‌سازی با مشکل مواجه است. از سوی دیگر، طیف گفتار به پارامترهایLPC حساسیت زیادی داشته و لذا خطاهای کوچک در چندی‌سازی منجر به ناپایداری فیلتر سنتز می‌شود. بر این اساس پارامترهای LPC کارآیی لازم را در چندی‌سازی ندارند [۲]. به همین دلیل در این مقاله از پارامترهای زوج طیفی[۲](LSP) به‌عنوان روش بازنمایی مؤثرتر استفاده شده است [۳].

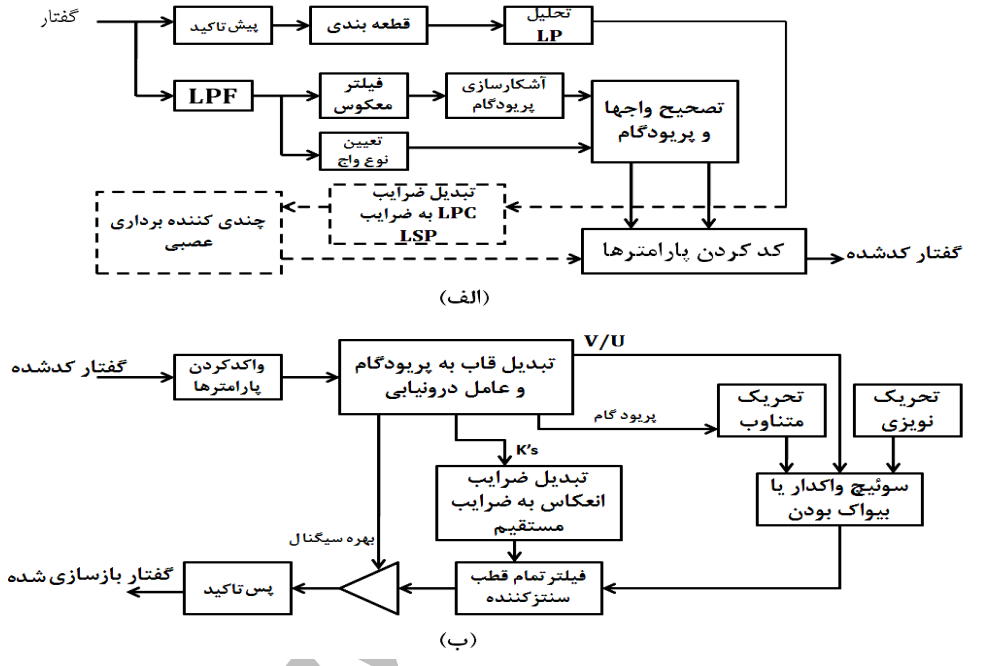
از سوی دیگر، شبکه‌های عصبی به‌عنوان ابزاری موفق تاکنون در کاربردهای گوناگونی از پردازش گفتار و زبان مورد استفاده قرار گرفته‌اند. در این راستا، کاربردهای بازشناسی خودکار گفتار[۳](ASR) [۴-۶]، سنتزگفتار طبیعی [۷-۹] و پردازش زبان طبیعی[۴](NLP) [۱۰-۱۲] به‌عنوان نمونه‌هایی که توسط مؤلف برای زبان فارسی تجربه شده‌اند، قابل ذکر است. برای کدکننده‌های گفتار نیز شبکه‌های عصبی در دو حوزه کاری مورد استفاده قرار گرفته‌اند: پیش‌بینی‌کننده‌های نرونی برای بهبود کیفیت [۱۳-۱۷] و کاهش

پیچیدگی محاسباتی در کدکننده‌ها [۲۳-۱۸].

در این مقاله، با هدف کاهش نرخ بیت کدکننده مبتنی‌بر استاندارد FS-1015، تغییراتی در ساختار آن پیشنهاد شده است تا نرخ بیت ۲/۴kbps این کدکننده تا ۱/۹kbps کاهش یابد. تغییرات اعمالی به این ساختار عبارتند از:

۱) استفاده از پارامترهای LSP به‌جای LPC

۲) بکارگیری شبکه‌های عصبی خودسازمانده (SOM)[۵] و ARTMAP فازی برای چندی‌سازی برداری و کاهش بار محاسباتی.



شکل ۱- کدکننده استاندارد FS-1015 به‌همراه تغییرات ساختاری، الف) کدکننده، ب) واکدکننده.

ساختار مقاله نیز بدین‌ترتیب است که در بخش دوم الگوریتم کدکننده مبتنی‌بر استاندارد FS-1015 مرور خواهد شد. در بخش سوم، چگونگی تبدیل پارامترهای LPC به LSP بیان می‌شود. در بخش چهارم، مبانی چندی‌سازی برداری به‌کمک شبکه‌های عصبی و با تمرکز بر روی شبکه ARTMAP فازی مرور خواهد شد. مدل پیشنهادی برای کدکننده با نرخ کاهش‌یافته در بخش پنجم معرفی و جزئیات شبیه‌سازی و نتایج تجربی در بخش ششم ارائه خواهد شد. بخش هفتم نیز به نتیجه‌گیری اختصاص یافته است.

۲- الگوریتم استاندارد FS-1015

این کدکننده گفتار که تحت عنوان LPC-10 نیز شناخته می‌شود، برای کدکردن مبتنی‌بر پیش‌بینی خطی گفتار با نرخ ۲/۴kbps ارائه شده است [۲۴]. در این الگوریتم روش AbS[۶] بکار گرفته شده و از یک فیلتر نردبانی[۷] مرتبه ۱۰ برای تولید پارامترهای پیش‌بینی استفاده می‌شود. طول قاب ۲۲/۵میلی‌ثانیه و ۵۴ بیت برای کدکردن پارامترهای هر قاب تخصیص می‌یابد. شمای بلوکی کدکننده و واکدکننده LPC-10 به‌همراه تغییراتی که در این مقاله در ساختار کدکننده آن ایجاد شده و با خط‌چین نشان داده شده‌اند، در شکل (۱) آورده شده است.

در این کدکننده برای کاهش اثرات ممیز ثابت در محاسبات ریاضی، سیگنال گفتار (با پهنای باند ۱۰۰ تا ۳۶۰۰ هرتز) از یک فیلتر پیش‌تاکید[۸] بالاگذر FIR[۹] مرتبه اول عبور داده می‌شود. تقطیع[۱۰] و پردازش قاب بستگی به واکدار یا بیواک بودن سیگنال گفتار دارد. اطلاعات مربوط به واکدار بودن و زمان تناوب گام[۱۱] با پردازش گفتار خروجی از یک فیلتر پایین‌گذر با فرکانس قطع ۸۰۰ هرتز، بدست می‌آید. یک فیلتر معکوس مرتبه ۲ جهت بهبود عمل تخمین زمان تناوب گام برای سیگنال‌های ورودی که شامل فرکانس‌های زیر ۳۰۰ هرتز می‌باشند، بکار می‌رود. استخراج زمان تناوب گام، از روی شکل موج فیلترشده و به‌روش تابع میانگین اندازه تفاضل (AMDF)[۱۲] بدست می‌آید [۲۵]. واکدار بودن با توجه

به‌میزان انرژی، نرخ عبور از صفر (ZCR)[۱۳] و نسبت بیشینه به کمینه AMDF تقریب زده می‌شود.

۳– استخراج پارامترهای LSP

یادآوری می‌شود که نظریه پیش‌بینی خطی مبتنی‌بر پیش‌بینی نمونه بعدی سیگنال گفتار از روی ترکیب خطی مقادیر قبلی بود:

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s_{n-k} \qquad (۱)$$

که در این رابطه a_k پارامترهای پیش‌بینی بودند. با فرض e_n به‌عنوان خطای پیش‌بینی می‌توان نوشت:

$$s_n = e_n + \sum_{k=1}^{p} a_k s_{n-k} \qquad (۲)$$

با گرفتن تبدیل z داشتیم:

$$S(z) = E(z) + \left[\sum_{k=1}^{p} a_k z^{-k}\right] S(z)$$

$$S(z) = \frac{E(z)}{A(z)} = E(z)H(z) \qquad (۳)$$

که $H(z)$ تابع انتقال فیلتر دیجیتال تمام-قطب بود. این درحالی است که یک تابع انتقال کلی لوله صوتی دارای صفر و قطب است. به این دلیل در روش LSP، دو چندجمله‌ای $P(z)$ و $Q(z)$ به‌این ترتیب از روی $A(z)$ تعریف می‌شوند:

$$P(z) = A(z) + z^{-(k+1)}A(z^{-1}) \qquad (۴)$$

$$Q(z) = A(z) - z^{-(k+1)}A(z^{-1}) \qquad (۵)$$

ریشه‌های $P(z)$، LSPهای فرد و ریشه‌های $Q(z)$، LSPهای زوج را ارائه می‌دهند. معنی فیزیکی ضرایب LSP همان فرکانس‌های تشدید طیفی لوله صوتی است. به‌همین دلیل ضرایبLSP با توجه به ویژگی‌های فیزیکی سیستم تولید گفتار انسان، دارای کران بوده و مقادیر حقیقی با ترتیب طبیعی هستند. براساس همین دو ویژگی، این ضرایب برای چندی‌سازی بسیار مفیدتر از ضرایبLPC می‌باشند [۲۶].

۴– چندی‌سازی برداری

چندی‌سازی برداری روشی برای نگاشت دنباله‌ای از سیگنال‌های پیوسته یا بردارهای گسسته به دسته‌ای از دنباله‌های دیجیتالی است. این روش در سیستم‌های کدکننده کاربرد زیادی دارد [۲۷]. در این مقاله از شبکه‌های عصبی خودسازمانده کوهونن (KSOFM) و ARTMAP فازی برای چندی‌سازی برداری و با هدف کاهش تعداد بیت‌های هر قاب از الگوریتم LPC-10 استفاده شده است.

۴–۱– شبکه KSOFM با آموزش تحت نظارت

در شبکه عصبی کوهونن، هر نورون i در نقشه خودسازمانده متناظر با یک بردار کتاب کد[۱۴] n بعدی است. نرون مذکور با نرون‌های مجاور خود ارتباط همسایگی N_i را دارد. شبکه KSOFM اغلب به‌صورت بدون نظارت[۱۵] آموزش می‌بینند. هرچند که آموزش تحت نظارت آن نیز پیشنهاد شده است [۲۸]. در آموزش تحت نظارت، اطلاعات مربوط به کلاس بردار ورودی نیز ضمیمه می‌شود.

در این تحقیق، از شبکه KSOFM با آموزش تحت نظارت به‌عنوان یکی از چندی‌کننده‌های برداری استفاده شده است. در این مورد، با استفاده از رابطه (۶)، نرون با بیشترین تطبیق (m_r) به‌ازای بردار ورودی **x** به‌عنوان برنده انتخاب می‌شود:

$$r = \arg\min_i \|(x - m_i)\Lambda\| \qquad (۶)$$

که در این رابطه، Λ ماتریس قطری $n\times n$ با عناصر قطر اصلی $\lambda_{11}....\lambda_{pp}$ برابر μ_1 و سایر عناصر قطری برابر μ_2 است. iامین عنصر از jامین بردار کتاب‌کد m_j براساس رابطه (۷) بهنگام‌سازی می‌شود:

$$\Delta m_{ij} = \begin{cases} -\beta\alpha(t) f(\Delta_{jr}) h(x_i, m_{ij}) & \text{اگر x و } m_r \text{ به کلاس‌های مختلفی تعلق دارند} \\ \alpha(t) f(\Delta_{jr})(x_i - m_{ij}) & \text{در غیر این‌صورت} \end{cases} \qquad (۷)$$

که در آن $f(\Delta_{jr})$ تابع همسایگی گوسی با رابطه (۸) است:

$$f(\Delta_{jr}) = exp\left(-\frac{\|l_j - l_r\|^2}{2\sigma(t)^2}\right) \qquad (۸)$$

که در آن $\sigma(t)$ تابع کاهشی با تعداد تکرارها، l_r موقعیت نرون برنده و l_j موقعیت نرون jام است. به‌همین ترتیب α نرخ یادگیری است که در طول آموزش به‌صورت خطی کاهش یافته تا به صفر برسد. β نیز نرخ تضعیف است که تأثیر جمله تضعیف h(۰) را مشخص می‌کند. هدف از بکارگیری جمله تضعیف، دورکردن m_r و همسایگانش از x است:

$$h(x_i, m_{ij}) = \text{sgn}(x_i - m_{ij})(\rho_i - |x_i - m_{ij}|) \qquad (۹)$$

که در آن $sgn(.)$ تابع علامت و ρ_i انحراف معیار استاندارد است و به‌صورت زیر تعریف می‌شود:

$$\rho_i = \sqrt{\frac{\sum_{l=1}^{N}(x_{li}-\bar{x}_i)^2}{N}} \qquad ;\bar{x}_i = \frac{\sum_{l=1}^{N} x_{li}.\rho_i}{N} \qquad (۱۰)$$

N در این رابطه، مبین تعداد گره‌ها در مجموعه آموزشی است.

مقادیر μ_1 و μ_2 به ابعاد و اندازه مؤلفه‌ها در l و c بستگی داشته و فاصله اقلیدسی ارائه‌شده در رابطه (۶) نیز چنین محاسبه می‌شود:

$$d = \sqrt{\mu_1\sum_{i=1}^{p}(l_i-m_i)^2 + \mu_2\sum_{j=1}^{p}(c_j-m_{n+j})^2} \qquad (۱۱)$$

راهی برای تعیین مقادیر μ_1 و μ_2 در مرجع [۲۹] براساس رابطه (۱۲) پیشنهاد شده است:

$$\frac{\mu_1}{\mu_2} = \frac{n}{m}\frac{\sum_{j=1}^{m}(\phi(|l_j|)-\sigma(|l_j|))}{\sum_{i=1}^{n}(\phi(c_i)-\sigma(c_i))} \qquad (۱۲)$$

که در آن $\varphi(|l_i|)$ میانگین قدر مطلق iامین عضو از برچسب‌های داده در دادگان و $\varphi(c_i)$ میانگین iامین عضو از تمام مختصات است. در این تحقیق فرض شده که $\mu_1+\mu_2=1$ باشد.

۴-۲- شبکه ARTMAP فازی

ساختار شبکه‌های مبتنی‌بر نظریه تشدید وفقی[۱۶](ART) با آموزش تحت نظارت، با نام ARTMAP شناخته شده‌اند [۳۰]. هر سیستم ARTMAP از دو مدول (ART_b, ART_a) تشکیل شده که طبقات بازشناسی پایداری را در پاسخ به دنباله‌های دلخواه از الگوهای ورودی ایجاد می‌کند. این دو مدول با یکدیگر از طریق یک مدول واسط به‌نام ناحیه نگاشت (F^{ab}) پیوند می‌یابند. ARTMAP باینری سیستم ART_1 را به‌عنوان مدول‌های ART_a و ART_b بکار می‌گیرد و این درحالی است که ARTMAP فازی از سیستم‌های ART فازی، که در ادامه به‌اختصار معرفی خواهند شد، بدین‌منظور بهره می‌گیرد. بدین‌ترتیب‌که مثلاً عملگر اشتراک $(\cap)$ با عملگر AND فازی [۳۱] جایگزین می‌شود. حال در ادامه قبل از بیان مختصر الگوریتم ARTMAP فازی، مبانی الگوریتم ART فازی مرور می‌شود:

تعریف ۱- بردارهای فعالیت: هر سیستم ART شامل سه لایه F_0، F_1 و F_2 است. گره‌های مبین بردار ورودی فعلی، لایه F_0 را تشکیل می‌دهند. لایه F_1 ورودی‌هایی از لایه پایینی خود (F_0) و نیز از لایه بالایی خود (F_2) دریافت می‌کند. بردار فعالیت F_0 با $I=(I_0,...,I_M)$ و مؤلفه‌های بهنجار I_i نمایش داده می‌شود. بردار فعالیت F_1 با $x=(x_1,...,x_M)$ و بردار فعالیت F_2 با $y=(y_1,...,y_M)$ نمایش داده می‌شوند.

تعریف ۲- بردار وزن: به‌ازای هر گره مشخص‌کننده دسته در لایه F_2، بردار وزن $w_j=(w_{j_1},...,w_{j_M})$ به‌عنوان حافظه بلندمدت[۱۷](LTM) وجود دارد.

تعریف ۳- پارامترها: در شبکه ART فازی پارامترهای انتخاب $(\alpha\rangle 0)$، نرخ یادگیری $(\beta\in[0,1])$ و مراقبت $(\rho\in[0,1])$ درنظر گرفته شده‌اند.

تعریف ۴- انتخاب طبقه: به‌ازای هر ورودی I و گره j در لایه F_2، تابع انتخاب T_j براساس رابطه (۱۳) تعریف می‌شود:

$$T_j = \frac{|I\Lambda w_j|}{\alpha+|w_j|} \qquad (۱۳)$$

که عملگرهای نرم $|.|$ و $\wedge$ چنین تعریف می‌شوند:

$$|p| \equiv \sum_{i=1}^{M}|p_i| \qquad (۱۴)$$

$$(p\Lambda q)_i \equiv \min(p_i, q_i) \qquad (۱۵)$$

طبقه انتخاب‌شده با J مشخص می‌شود:

$$T_J = \max\{T_j;\ j=1...N\} \qquad (۱۶)$$

در این شرایط $y_J=1$ و $y_j=0\ ;\ j\neq J$ بردار فعالیت F_1 نیز از رابطه (۱۷) تبعیت می‌کند:

$$x = \begin{cases} I & \text{اگر } F_2 \text{ غیرفعال باشد} \\ & ; \\ I\ \Lambda\ w_J & \text{اگر نرون J‌ام از لایه } F_2 \text{ انتخاب شود;} \end{cases} \qquad (۱۷)$$

تعریف ۵- تشدید یا بازنشانی: اگر شرط زیر برقرار باشد، آنگاه پدیده‌ی تشدید اتفاق می‌افتد:

$$\frac{|I\Lambda w_J|}{|I|} \geq \rho \qquad (۱۸)$$

بدیهی است که درصورت عدم برقراری شرط فوق، فرمان بازنشانی مبین عدم تطبیق صادر و نمایه جدیدی به‌جای J انتخاب و

جستجو برای یافتن دسته‌ای که شرط (۱۷) را برآورده کند ادامه می‌یابد.

تعریف ۶- یادگیری: پس از اتمام فرآیند جستجو، بردار وزن w_J براساس رابطه (۱۹) تجدید می‌شود:

$$w_J^{(new)} = \beta(I \wedge w_J^{(old)}) + (1-\beta) w_J^{(old)} \qquad (۱۹)$$

ورودی‌های به ART_a و ART_b در شبکه ARTMAP فازی به‌صورت کد مکمل $A=(a,a^c)$ و $B=(b,b^c)$ هستند. x^a و y^a به‌ترتیب مبین بردارهای خروجی F_1^a و F_2^a هستند. w_j^a نیز بردار وزن Jامین گره از ART_a است. مشابه این نمادگذاری برای ART_b نیز درنظر گرفته می‌شود. برای ناحیه نگاشت نیز x^{ab} مبین بردار خروجی F_{ab} و w_j^{ab} مبین بردار وزن از Jامین گره F_2^a به F^{ab} است.

تعریف ۷- فعالیت ناحیه نگاشت: اگر یکی از دسته‌های ART_a یا ART_b فعال شوند، آنگاه F^{ab} فعال می‌شود. اگر هم هر دوی ART_a و ART_b فعال باشند، آنگاه اگر ART_a همان دسته‌ای که ART_b پیش‌بینی کرده را ارائه نماید، درآن‌صورت F^{ab} فعال خواهد شد. بردار خروجی F^{ab} نیز از رابطه (۲۰) تعیین می‌شود:

$$x^{ab} = \begin{cases} y^b \wedge w_J^{ab}; & \text{Jامین گره از } F_2^a \text{ فعال و } F_2^b \text{ نیز فعال است} \\ w_J^{ab}; & \text{Jامین گره از } F_2^a \text{ فعال و } F_2^b \text{ غیرفعال است} \\ y^b; & F_2^a \text{ غیرفعال و } F_2^b \text{ فعال است} \\ 0; & F_2^a \text{ و } F_2^b \text{ غیرفعال هستند} \end{cases} \qquad (۲۰)$$

۵- مدل پیشنهادی

در این مقاله، در جریان دو پیشنهاد سعی در کاهش نرخ بیت کدکننده گفتار مبتنی بر استاندارد FS-1015 شده است. در جریان پیشنهاد اول، از پارامترهای LSP به‌جای LPC در ساختار سنتی این کدکننده استفاده شده است. پیشنهاد دوم نیز در راستای بکارگیری چندی‌کننده‌های برداری عصبی به‌منظور خوشه‌بندی و کاهش تعداد بیت‌ها در کدکردن پارامترها ارائه شده است. در این مورد شبکه ARTMAP فازی و نیز KSOFM با یادگیری تحت نظارت بکار گرفته شده و عملکرد آنها با یکدیگر مقایسه شده است.

در این راستا، تعداد بیت‌های مورد نیاز در هر قاب ۲۲/۵ میلی‌ثانیه‌ای کدکننده LPC-۱۰ و نیز تعداد مربوط در هر قاب از کدکننده پیشنهادی در جدول (۱) آورده شده است.

همان‌گونه که در جدول مشاهده می‌شود، در کدکننده پیشنهادی، ضرایب اول و دوم LSP قطعات واکدار و نیز بیواک با ۸ بیت (به‌جای ۱۰ بیت در کدکننده LPC-10) کد می‌شوند. ضرایب سوم و چهارم LSP نیز همین وضعیت را دارند. از سوی دیگر، برای کدکردن ضرایب نهم و دهم LSP، تعداد بیت بیشتری نسبت به کدکننده LPC-10، درنظرگرفته شده است. در کدکننده LPC-10، ۲۱ بیت تحت عنوان بیت‌های محافظت در برابر خطا (EPB)[۱۸] برای قاب‌های بیواک درنظر گرفته شده است. این تعداد در کدکننده پیشنهادی، با توجه به عملکرد مناسب چندی‌کننده‌های برداری عصبی به‌میزان ۷ بیت نیز قابل کاهش شده است.

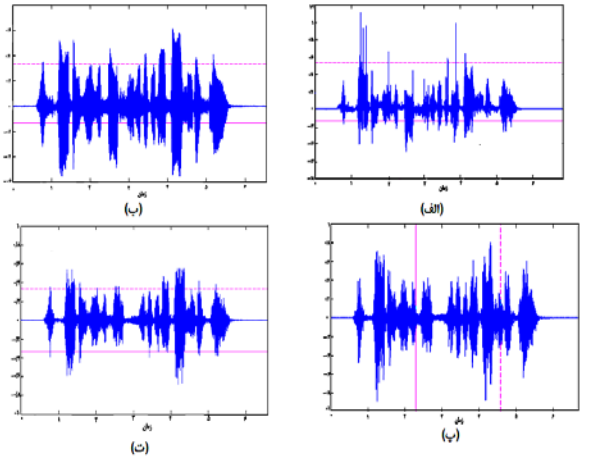
شبکه KSOFM بکارگرفته‌شده دارای ۲۵۶ نرون برای کدکردن پارامترهای [P_2 و P_1] یا [P_4 و P_3] در هر یک از قطعات واکدار یا بی‌واک است. همین روند برای دسته‌بندی به‌کمک شبکه ARTMAP فازی نیز بکارگرفته شده است.

جدول ۱- مقایسه تعداد بیت‌های مورد نیاز در هر قاب از کدکننده LPC-10 و کدکننده پیشنهادی

پارامترها	تعداد بیت‌ها به‌ازای هر قاب			
	کدکننده LPC-10 استاندارد		کدکننده پیشنهادی	
	واکدار	بی‌واک	واکدار	بی‌واک
دوره تناوب گام	۷	۷	۷	۷
بهره	۵	۵	۵	۵
P_1	۵	۵	۸	۸
P_2	۵	۵		
P_3	۵	۵	۸	۸
P_4	۵	۵		
P_5	۴	-	۴	-
P_6	۴	-	۴	-
P_7	۴	-	۴	-
P_8	۴	-	۴	-
P_9	۳	-	۴	-
P_{10}	۲	-	۴	-
همزمان‌سازی	۱	۱	۱	۱
محافظت در برابر خطا	-	۲۱	-	۷
مجموع	**۵۴**	**۵۴**	**۵۳**	**۳۶**

۶– شبیه‌سازی و نتایج تجربی

در این مقاله از ۱۶۰ جملهٔ دادگان FARSDAT [۳۱]، که توسط دو گوینده مرد و دو گوینده زن ادا شده‌اند، به‌عنوان دادگان گفتاری استفاده شده است. در چندی‌سازی برداری به کمک شبکه KSOFM با آموزش تحت نظارت، نرخ یادگیری $\alpha(0)=0/7$، نرخ تضعیف $\beta(0)=0/1$، گسترش همسایگی $\sigma=40$، $\mu_1=0/35$ و $\mu_2=0/65$ درنظر گرفته شده‌اند. تنظیم دقیق شبکه نیز با نرخ اولیه ۰/۰۵ انجام شده است. در شبیه‌سازی شبکه ARTMAP فازی نیز پارامتر انتخاب $\alpha=0/01$، نرخ یادگیری اولیه $\beta=0/01$ و پارامتر مراقبت $\rho=0/95$ درنظر گرفته شده است. به‌عنوان نتایج شهودی، نمونه گفتار مربوط به جمله‌ی "*مثل آدم‌های مسخ‌شده به‌نظر می‌رسد*" که با حالت هیجانی ادا شده در قسمت (الف) شکل (۲) و شکل موج سنتزشده با کدکننده LPC-10 در قسمت (ب) شکل (۲) و شکل موج‌های سنتزشده با کدکننده پیشنهادی و چندی‌سازی برداری با شبکه‌های عصبی KSOFM و ARTMAP فازی به‌ترتیب در قسمت‌های (پ) و (ت) شکل (۲) آورده شده‌اند.



شکل ۲– مقایسه شهودی گفتار سنتزشده با کدکننده‌های مختلف، الف) گفتار اصلی، ب) LPC-10، پ)مدل پیشنهادی با چندی‌سازی به‌کمک KSOFM، ت) مدل پیشنهادی با چندی‌سازی به‌کمک ARTMAP فازی

برای ارزیابی کیفیت گفتار سنتزشده از معیار MOS استفاده و ۴۰ جمله از دادگانFARSDAT توسط۳۰ شنونده امتیازدهی شدند. نتایج عملکرد مدل مبتنی‌بر استاندارد FS-1015 و دو روش پیشنهادی از لحاظ کیفیت گفتار سنتزشده در ملاک MOS و نیز زمان اجرای الگوریتم‌ها برای ۱۸۳۷۰ قاب گفتاری نمونه در جدول (۲) آورده شده است.

جدول۲- نتایج ارزیابی عملکرد روش‌های پیشنهادی با کدکننده مبتنی بر استاندارد FS-1015

روش	MOS	زمان‌اجرا (ثانیه)
استاندارد FS-1015	۲/۶۷	۴۳۸
مدل پیشنهادی با چندی‌سازی به‌کمک KSOFM	۲/۸۰	۳۳۱
مدل پیشنهادی با چندی‌سازی به‌کمک ARTMAP فازی	۲/۹۳	۲۴۹

۷– نتیجه‌گیری

در این مقاله، با اعمال تغییراتی در ساختار کدکننده مبتنی‌بر استاندارد FS-1015، نرخ بیت ۲/۴ kbps این الگوریتم با توجه به تغییر طول قاب از ۲۲/۵ میلی‌ثانیه به ۲۷/۷۸ میلی‌ثانیه و نیز چندی‌سازی برداری به‌کمک شبکه‌های عصبی KSOFM و ARTMAP فازی به ۱/۹ kbps کاهش پیدا کرد. این درحالی است که کیفیت گفتار سنتزشده برحسب ملاک MOS در روش پیشنهادی در هر دو حالت بکارگیری چندی‌کننده‌های برداری عصبی KSOFM و ARTMAP فازی نسبت به کدکننده‌ی استاندارد به‌ترتیب به میزان ۰/۱۳ و ۰/۲۶ بهبود یافت. به‌همین ترتیب زمان اجرای الگوریتم نیز در دو روش مذکور به‌ترتیب ۷۳٪ و ۵۷٪ زمان اجرای الگوریتم استاندارد بوده و با کاهش قابل ملاحظه‌ای مواجه شد.

۸– مراجع

[1] W. T. K. Wong, J. Joe, K. Joe, M. Joe; **"Low Rate Speech Coding for Telecommunications"**, BT Technol. J., 14, pp. 28 - 43, 1996.

[2] F. Itakula**; "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signal"**, J. Acoust. Soc. Amer., 57, pp. 535(A), 1975.

[3] M. Hasegawa-Johnson**; "Line Spectral Frequencies Are Poles and Zeros of the Glottal Driving-Point Impedance of a Discrete Matched-Impedance Vocal Tract Model"**, J. Acoust. Soc. Amer., 108, pp. 457-460, 2000.

[۴] م. شیخان، ش. پاینده، ف. رزاقیان؛ "**سیستم بازشناسی و درک گفتار فارسی**"، مجموعه مقالات دومین کنفرانس مهندسی برق ایران، صفحات ۳۵۲-۳۴۵، ۱۳۷۳.

[5] M. Sheikhan, M. Tebyani, M. Lotfizad;

[۸] م. شیخان، م. نصیرزاده، ع. دفتریان؛ **"طراحی و پیاده‌سازی سیستم تبدیل متن به‌گفتار طبیعی برای زبان فارسی"**، مجله علمی – پژوهشی دانشکده مهندسی دانشگاه فردوسی مشهد، سال ۱۷، شماره ۲، صفحات ۴۸ - ۳۱، ۱۳۸۴.

[۹] م. شیخان؛ **"تولید خودکار نوای گفتار به‌کمک مدل آمیختار عصبی- آماری با امکان انتخاب واحد در سنتز"**، مجلهٔ علمی- پژوهشی مهندسی پزشکی زیستی، دوره جدید، شماره اول، صفحات ۲۴۰ - ۲۲۷، ۱۳۸۶.

[۱۰] م. شیخان، ف. رزاقیان، ر. بوذرجمهری؛ **"تلفیق شبکه عصبی با هوش‌مصنوعی جهت تفکیک، تصحیح، بررسی معنایی گفتار فارسی و تبدیل گفتار به متن نوشتاری"**، مجموعه مقالات دومین کنفرانس کامپیوتر ایران، صفحات ۱۲۹ - ۱۲۱، ۱۳۷۲.

[۱۱] م. شیخان،م. طبیانی،م. لطفی‌زاد؛ **"دسته‌بندی مفهومی و رفع ابهام معنایی کلمات فارسی توسط شبکه‌های عصبی"**، مجموعه مقالات کنفرانس بین‌المللی سیستم‌های هوشمند و شناختی، صفحات ۳۹-۳۵، ۱۳۷۵.

[12] M. Sheikhan, M. Tebyani, M. Lotfizad; **"Using Symbolic and Connectionist Approaches to Automate Editing Persian Sentences Syntacticly",** Proc. Int. Conf. Intell. & Cogn. Syst, pp. 250-253, 1996.

[13] [13] M. Birgmeier; **"Nonlinear Prediction of Speech Signals Using Radial Basis Function Networks",** Proc. Europ. Signal Process. Conf, Vol. 1, pp. 459-462, 1996.

[14] A. Kumar, A. Gersho; **"LD-CELP Speech Coding with Nonlinear Prediction",** IEEE Signal Processing Letters, 4, pp. 89-91, 1997.

[15] N. Ma, G. Wei; **"Speech Coding with Nonlinear Local Prediction Model",** Proc. IEEE ICASSP, Vol. 2, pp. 1101-1104, 1998.

[16] M. Faúndez-Zanuy, S. McLaughlin, A. Esposito, A. Hussain, J. Schoentgen, G. Kubin, W. B. Kleijn, P. Maragos; **"Nonlinear Speech Processing: Overview and Applications",** Control and Intelligent Systems, 30, pp. 1 - 10, 2002.

[17] M. Faúndez-Zanuy; **"Nonlinear Speech Coding with MLP, RBF and Elman Based Prediction",** Lecture Notes in Computer Science, 2687, pp. 671-678, 2003.

[18] M. G. Easton, C. C. Goodyear; **"A CELP Codebook and Search Technique Using a Hopfield Net",** Proc. IEEE ICASSP, pp. 685 - 688, 1991.

[19] A. Indrayanto, A. Langi, W. Kinsner; **"A Neural Network Mapper for Stochastic Codebook Parameter Encoding in Code Excited Linear Predictive Speech Processing"**, Proc. IEEE West. Canada Conf. Comp., Power and Commun. Sys., pp. 221 - 224, 1991.

[20] L. A. Hernandez-Gomez, E. Lopez-Gonzalo; **"Phonetically-Driven CELP Coding Using Self-Organizing Maps",** Proc. IEEE ICASSP, Vol. 2, pp. 628-631, 1993.

[21] L. Wu, M. Niranjan, F. Fallside; **"Fully Vector-Quantized Neural Network-Based Code Excited Nonlinear Predictive Speech Coding",** IEEE Trans. Speech and Audio Processing, Vol. 2, pp. 482 - 489, 1994.

[22] S. Wu, G. Zhang, X. Zhang, Q. Zhao; **"A LD-CELP Speech Coding Algorithm Based on Modified SOFM Vector Quantizer",** Proc. Int. Symp. Intell. Inform. Technol. Appl., pp. 408 - 411, 2008.

[23] V. Huong, B. J. Min, D. C. Park, D. M. Woo; **"A New Vocoder Based on AMR 7.4 Kbit/S Mode in Speaker Dependent Coding System",** Proc. ACIS Int. Conf. Soft. Engng., Artif. Intell., Network., and Parallel/Distributed Comp., pp. 163 - 167, 2008.

[24] T. Tremain; **"The Government Standard Linear Predictive Coding Algorithm: LPC-10"**, Speech Technology, 1, pp. 40 - 49, 1982.

[25] A. Spanias; **"Speech Coding: A Tutorial Review",** Proc. IEEE, No. 82, pp. 1541-1582, 1994.

[26] O. Wiriyanuruknakorn, J. Srinonchat; **"A Finite State Vector Quantizer for New Bit Rate Speech Compression",** Proc. Int. Conf. Signal Processing, Commun. and Network., pp. 255-259, 2008.

[27] A. Gersho, R. M. Gray; **Vector Quantization and Signal Compression**, Kluwer Academic Publishers, 1992.

[28] T. Kohonen, **Self-Organizing Maps**, Springer Series in Information Sciences, 1995.

[29] M. Hagenbuchner, A. Sperduti, A. Tsoi; **"A Self-Organizing Map for Adaptive Processing of Structured Data",** IEEE Trans. on Neural Networks, 14, pp. 491 - 505, 2003.

[30] G. A. Carpenter, S. Grossberg, J. H. Reynolds;

"ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network", Neural Networks, Vol. 4, pp. 565-588, 1991.

[31] L. Zadeh; **"Fuzzy Sets",** Inform. Contr, Vol. 8, pp. 338-353, 1965.

[32] M. Bijankhan, J. Sheikhzadegan, M. R. Roohani, Y. Samareh, C. Lucas and M. Tebyani; **"FARSDAT- The Speech Database of Farsi Spoken Language",** Proc. Australian Conf. on Speech Science and Technology, Vol. 2, pp. 826 - 830, 1994.

۹– پی‌نوشت‌ها

1- Linear Predictive Coding
2- Line Spectral Pairs
3- Automatic Speech Recognition
4- Natural Language Processing
5- Self Organizing Maps
6- Analysis by Synthesis
7- Lattice
8- Pre-emphasized
9- Finite Impulse Response
10- Segmentation
11- Pitch
12- Average Magnitude Difference Function
13- Zero Crossing Rate
14- Codebook
15- Unsupervised
16- Adaptive Resonance Theory
17- Long Term Memory
18- Error Protection Bit