طراحی و شبیه‌سازی یک کدکنندهٔ صحبت LPC به صورت بلادرنگ

مقالاتی از کارنامهٔ پژوهشی «شریف» سال ۷۰

مجله شریف در ادامهٔ چاپ مقالاتی از کارنامهٔ پژوهشی شریف، در این شماره نیز اقدام به چاپ یکی دیگر از این مقالات نموده‌است.

تشخیص نوع تحریک در کدکنندهٔ صحبت LPC توسط شبکهٔ عصبی

همایون برهانی (مربی)
سید بهرام ظهیراعظمی
مرکز تحقیقات الکترونیک

چکیده

یکی از مسائل مهم در ساخت کدکننده‌های صحبت، تشخیص صحیح نوع تحریک است. در این گزارش ضمن توضیح مشکلات موجود در روشهای معمول تشخیص نوع تحریک، روش جدیدی که مبتنی بر شبکه‌های عصبی است ارائه می‌گردد. قدرت این روش تا حدی است که روی حدود ۲۰۰۰ قطعهٔ صحبت نمونه که برای آموزش مورد استفاده قرارگرفتند بدون خطا عمل کرد و میزان خطا برای قطعات انتخاب شده خارج از این مجموعه بسیار ناچیز بود.

مقدمه

در ساختمان کدکننده‌های صحبت، قسمتهای مختلفی وجود دارد که یکی از آنها مربوط به تعیین نوع تحریک ـ متناوب یا مغشوش ـ است.[۱]

ارتعاش تارهای صوتی (Vocal Cords) عامل متناوب شدن تحریک و واک‌داری (Voicing) یک قطعه (Frame) است. به این شکل که با قرار گرفتن این تارها در فاصلهٔ مناسبی از یکدیگر و دمیده شدن هوا با فشار معین از بین این تارها یک ارتعاش منظم ایجاد می‌گردد که به «واک» موسوم است.

این پدیده را می‌توان با تماس آرام انگشتان دست روی ناحیه‌ای از گلو (کنار غدهٔ تیروئید) که تارهای صوتی و حنجره در پشت آن قرار گرفته‌اند، حس نمود.

در موقع بیان بسیاری از واج‌ها (Phonemes) این ارتعاش در تارهای صوتی وجود دارد. به طوری که از ۳۱ واج زبان فارسی تنها ده واج بی‌واک هستند. بسامد وقوع واجهای بی‌واک نیز معمولاً کمتر از واجهای واک‌دار است. به طوری که می‌توان ادعا کرد در حدود ۹۰ درصد از زمان صحبت، تارهای صوتی در حال ارتعاش هستند. باوجود کم بودن احتمال وقوع واجهای بی‌واک تشخیص صحیح قطعه‌های واک‌دار و بی‌واک از هم، دارای اهمیت بسیار زیادی است. زیرا اولاً بسیاری از واجهای واک‌دار دارای جفت بی‌واک هستند. بدین معنی که دو واج در کلیهٔ علایم و مظاهر به یکدیگر شبیه هستند، به جز پارامتر واک‌داری که یکی واک‌دار است و دیگری بی‌واک. طبیعی است که بروز اشتباه در تشخیص واک داری در بین این قبیل واجها می‌تواند موجب ایجاد اشتباه بین آنها گردد. از جفتهای واک‌دار و بی‌واک می‌توان موارد زیر را نام برد:

(ب /b/ و پ /p/)، (د /d/ و ت /t/)؛ (گ /g/ و ک /k/)، (ژ /z/ و ش /s/)، (ز /z/ و س /s/)، (و /v/ و ف /f/) و (ج /j/ و چ /c/) که در هر جفت به ترتیب، اولی واک‌دار و دومی بی‌واک است.[۲]

به این صورت تشخیص غلط واک‌داری در یک کدکنندهٔ صحبت (Vocoder) ممکن است باعث گردد تا کلمهٔ «سیب» /sib/ به اشتباه «زیپ» /zip/ شنیده شود. این موضوع در درک واجها (Phoneme Recognition) نیز اهمیت پیدا می‌کند. البته نباید از نظر دور داشت که گرچه واک‌داری یکی از مشخصه‌های هر واج است ولی تا حد زیادی تحت تأثیر عوامل دیگر بویژه بافتی که واج در آن قرار می‌گیرد نیز هست.

به علاوه اشتباه در تخمین واک‌داری ـ بسته به روش تخمین گام (Pitch Detection) که استفاده می‌شود ـ می‌تواند موجب تغییرات غیرهماهنگ در گام صحبت شود که چنین پدیده‌ای کیفیت صدای حاصله را شدیداً خراب می‌کند.

روش تشخیص واک

با بررسی چند کدکنندهٔ صحبت نمونه [۵و۶و۷] ملاحظه شد که تشخیص واک در بسیاری از کدکننده‌های صحبت با انتخاب مقادیر آستانه روی کمیت‌هایی که در تخمین گام به کار می‌روند انجام می‌گردد. به عنوان مثال در روش کپستروم (Cepstrum)[۵]، ماکزیمم تابع کپستروم در فاصلهٔ معقول برای گام ـ که اغلب حدود ۲ تا ۲۰ میلی‌ثانیه در نظر گرفته می‌شود ـ با یک مقدار آستانه مقایسه می‌شود و درصورتی که از این مقدار کمتر بود قطعه بی‌واک معرفی می‌گردد و در غیر این صورت قطعه واک‌دار بوده و محل ماکزیمم به عنوان گام معرفی می‌شود و در نهایت اصلاحاتی نیز باتوجه به گام قطعهٔ قبل و مقدار تابع کپستروم در نصف محل ماکزیمم مطلق، صورت می‌گیرد.

این روش دارای خطای زیادی است، چراکه تجربه نشان داده است که با انتخاب تنها یک معیار نمی‌توان جداسازی قطعه‌های واک‌دار و بی‌واک را بدرستی انجام داد بلکه به اطلاعات بیشتری نیاز هست. حتی با افزودن اطلاعات دیگری از قبیل نرخ عبور از صفر (Zero Crossing Rate) و انرژی قطعه نیز امکان جداسازی صحیح برای مجموعه‌ای گسترده از افراد و واجهای بیان شده، وجود ندارد.[۴]

بنابراین لازم است از معیارهای بیشتری برای جداسازی استفاده شود. شکلهای ۱ و ۲ به ترتیب نمایشگر قطعه‌های واک‌دار و بی‌واک صحبت هستند. در هر شکل منحنی زمانی، طیف و پوش طیف تخمین زده شده رسم گردیده است. تفاوتهایی که بین قطعه‌های واک‌دار و بی‌واک مشهود است به قرار زیر می‌باشند[۴]:

۱ـ در منحنی زمانی، قطعه‌های واک‌دار حالت تناوبی و در قطعه‌های بی‌واک حالت مغشوش (نویزی) مشاهده می‌گردد.

۲ـ اغلب قطعه‌های واک‌دار نسبت به قطعه‌های بی‌واک دارای انرژی بیشتری هستند.

۳ـ معمولاً نرخ عبور از صفر در قطعه‌های بی‌واک بیش از قطعه‌های واک‌دار است.

۴ـ تابع خودبستگی قطعه‌های واک‌دار یکنواخت‌تر از قطعه‌های بی‌واک است.

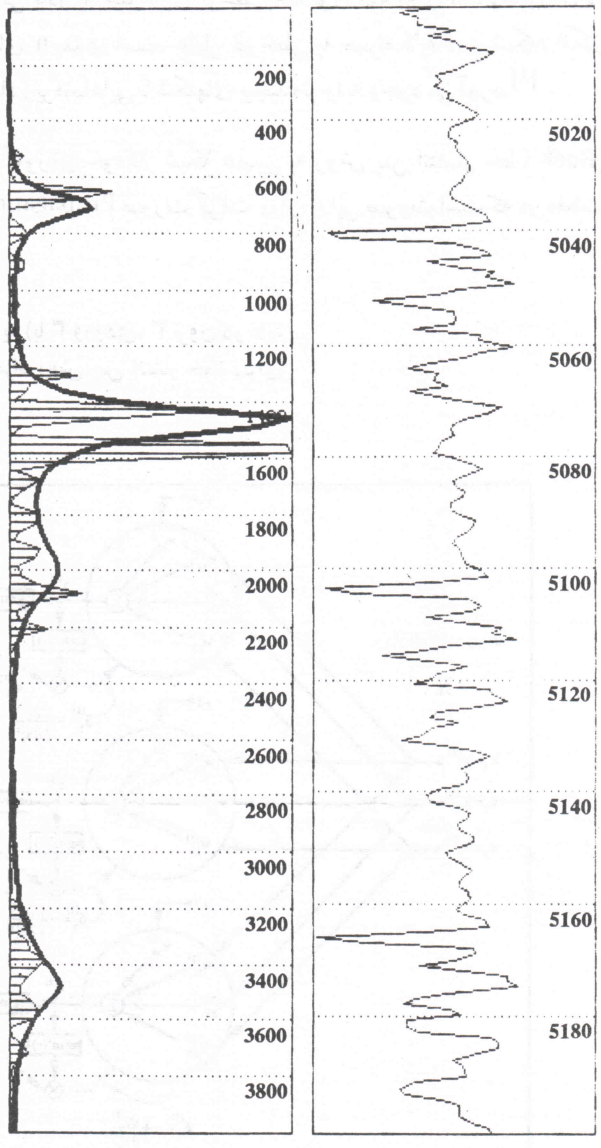
۵ـ معمولاً قطعه‌های واک‌دار دارای مولفه‌های فرکانسی قوی در فرکانسهای پایین‌تر از ۱۵۰۰ هرتز هستند، در حالی که قطعه‌های بی‌واک اغلب دارای مولفه‌های فرکانسی قوی در فرکانسهای بالاتر از ۱۵۰۰ هرتز هستند.

۶ـ در طیف قطعات واک‌دار گلبرگهای (Lobe) هم عرض وجود دارد.

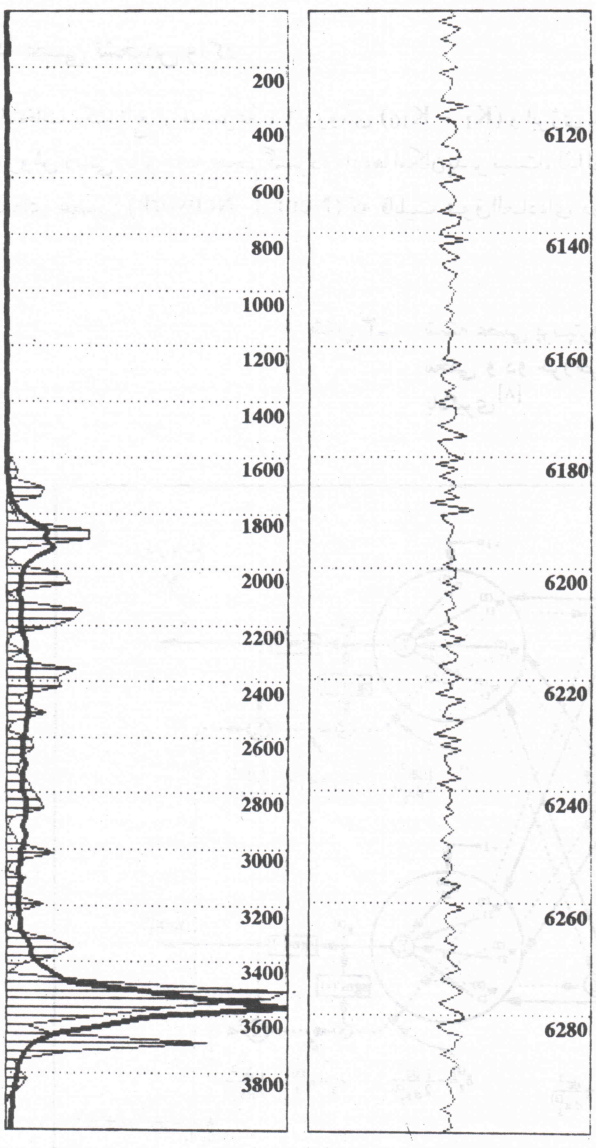
به عنوان مثالی از تفاوت اشاره شده در مورد ۳، میانگین نرخ عبور از صفر در قطعات واک‌دار، ۱۵۳۰ بار در ثانیه و برای قطعه‌های بی‌واک ۴۰۴۰ بار در ثانیه بوده است. علی‌رغم این مسئله چون قطعه‌های واک‌داری با حدود ۳۸۰۰ عبور از صفر در ثانیه وجود داشته‌اند، معیار نرخ عبور از صفر نمی‌تواند به تنهایی در جداسازی قطعات به‌کار رود.[۴]

علاوه بر اینها باید توجه داشت که تاثیر نویز ممکن است بر روی یکی از این معیارها بیش از حد زیاد باشد ـ مثلاً یک نوع نویز می‌تواند سبب افزایش فوق‌العاده نرخ عبور از صفر شود ـ مسئله دیگر تفاوتهایی است که در محدودهٔ تغییرات هریک از این معیارها برای افراد، جنسیت‌ها، لهجه‌ها

شکل ۱ـ بالا منحنی تغییرات زمانی و پایین طیف (خطوط نازک) و پوش طیف (خط کلفت) مربوط به یک قطعهٔ ۲۵ میلی ثانیه‌ای از صحبت واک‌دار



شکل ۲ـ بالا منحنی تغییرات زمانی و پایین طیف (خطوط نازک) و پوش طیف (خط کلفت) مربوط به یک قطعهٔ ۲۵ میلی‌ثانیه‌ای از صحبت بی‌واک



و سنین مختلف وجود دارد. باتوجه به بند ۵ تفاوتهای ذکر شده در بالا این فکر به وجود می‌آید که ممکن است بتوان با استفاده از تمام اطلاعات طیف سیگنال در هر قطعه، واک‌داری قطعه یا عدم آن را تعیین نمود. در مقالهٔ حاضر نحوهٔ انجام این عمل به نحوی که برای یک کدکنندهٔ صحبت LPC (Linear PredictionCoding) مناسب باشد ارائه شده است. چنین روشی برای تعیین واک‌داری باید تامین‌کنندهٔ خواسته‌های زیر باشد:

۱ـ دربرنداشتن محاسبات اضافی قابل توجه، به طوری که تحقق آن در یک کدکنندهٔ صحبت بلادرنگ امکان‌پذیر باشد.

۲ـ حجم گسترده‌ای از داده‌ها را ـ که از صحبت افراد مختلف و با جنسیت‌ها و سنین متفاوت که کلیهٔ واج‌ها را بیان کرده‌اند جمع‌آوری شده ـ در نظر گرفته باشد.

۳ـ درصد خطای ناچیزی داشته باشد.

از آنجاکه در یک کدکنندهٔ LPC، ضرایب $(K_1 - K_p)$ کلیهٔ اطلاعات پوش طیف تخمین زده شده را در بردارند[۳]، می‌توان باکمک این ضرایب و همچنین انرژی محاسبه شده، عمل تشخیص واک‌داری را به نحو مطلوب انجام داد. باتوجه به اینکه ضرایب $(K_1 - K_p)$ و انرژی

لزوماً باید محاسبه گردد، استفاده از آنها در تشخیص واک‌داری متضمن محاسبات اضافی نبوده و از این جهت برای یک سیستم بلادرنگ مناسب است.

شبکهٔ عصبی تشخیص واک

انتخاب یک تابع از مجموعهٔ ۱۱ ورودی (K_1 - K_{10}) و انرژی به یک روش دستی، باتوجه به حجم گسترده داده‌ها امکان‌پذیر نیست، لذا از شبکه‌های عصبی (Neural Network) که قابلیت فوق‌العاده‌ای در دسته‌بندی (Classification) خودکار از خود نشان داده‌اند، استفاده شد.

شبکهٔ عصبی از تعدادی پردازشگر کوچک موسوم به نرون (Neuron) تشکیل شده است. ورودی هر نرون مجموع وزن‌دار (Weighted) خروجی نرونهای شبکه و ورودی خارجی به شبکه است. خروجی هر نرون نیز از اعمال یک تابع غیرخطی مانند تانژانت هذلولی (Hyperbolic Tangent) بر روی ورودی به دست می‌آید.

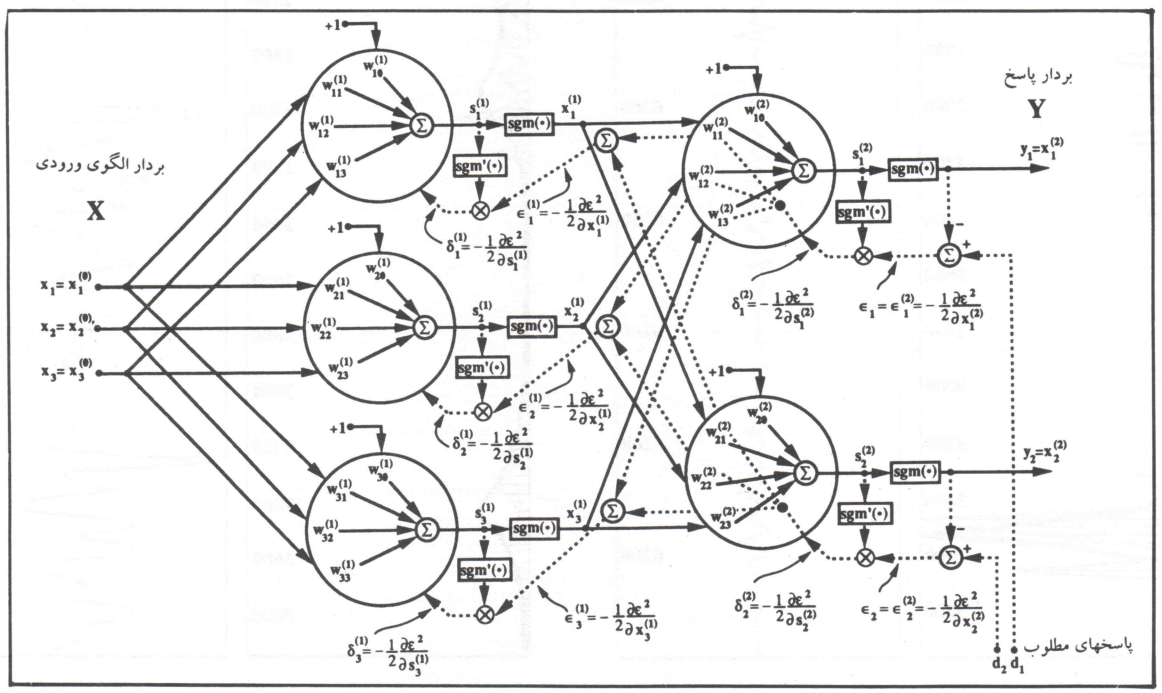
شبکهٔ پرسپترون (Perceptron) که برای تشخیص واک مورد استفاده قرارگرفت، یکی از انواع شبکه‌های عصبی است که رو به جلو (Feed Forward) و بدون حافظه (Memoryless) است. (شکل ۳)

هر یک از نرونهای لایهٔ نخست (لایهٔ مخفی) در شبکهٔ پرسپترون به تنهایی قادر به جداسازی توسط یک فوق صفحه (Hyperplane) در فضای n بعدی است. عامل غیرخطی به همراه لایهٔ دوم شبکه، امکان ایجاد زیرفضاهایی با شکلهای پیچیده‌تر را به وجود می‌آورد.[۸]

آموزش خودکار شبکهٔ عصبی به روش پس انتشار خطا (Back Propagation) صورت گرفت و آن به این صورت است که در دفعات متعدد هربار یک ورودی نمونه به شبکه اعمال می‌گردد، سپس اگر خروجی حاصل با خروجی مطلوب و موردنظر متفاوت بود کلیهٔ ضرایب شبکه به نحوی تغییر داده می‌شوند که خطا تقلیل داده شود. این روش به صورت مرحله به مرحله انجام می‌گردد تا به یک جواب نسبتاً بهینه برسد.

شبکهٔ عصبی با ۵ نرون در لایهٔ میانی ـ که به این صورت آموزش داده شد ـ توانست شناسایی قطعات واک‌دار از بی‌واک را در مجموعهٔ حدود ۲۰۰۰ قطعه که برای آموزش مورد استفاده قرار گرفته بود بدون هیچ مورد خطا انجام دهد و در خارج از محدوده داده‌های آموزشی نیز

شکل ۳ـ شبکه عصبی پرسپترون دو لایه (با ۳ ورودی، ۳ نرون در لایه مخفی و دو خروجی) به همراه روش پس انتشار خطا برای یادگیری[۸]



عملکرد بسیار خوبی ارائه نمود. علت این است که خاصیت تعمیم‌دهی (Generalization) شبکه‌های عصبی باعث می‌گردد که در پاسخگویی به قطعات دیگری که در مجموعهٔ یادگیری نبوده ولی خواص نسبتاً مشابهی با قطعات مجموعه داشته باشند، موفق باشند.

از آنجایی که برای افزایش سرعت همگرایی (covergence) شبکهٔ عصبی در مرحلهٔ عصبی در مرحلهٔ آموزش بهتر است ورودیها در محدودهٔ عددی یکسانی قرار داشته باشند، مقدار انرژی نیز با یک تابع غیرخطی به عددی در محدودهٔ $[-1,1]$ (مشابه ضرائب $(K_1 - K_{10})$ تبدیل گردید.[۴]

نتیجه‌گیری

توسط یک شبکهٔ عصبی دو لایه با ۵ نرون در لایهٔ مخفی می‌توان به درصد صحت بسیار بالایی در تشخیص واک‌داری قطعات صحبت دست یافت. تحقق نرم‌افزاری این شبکه با حدود ۶۰ عمل ضرب و ۶۰ عمل جمع اعشاری و ۵ عمل تانژانت هیپربولیک به ازای هر قطعهٔ ۲۵ میلی‌ثانیه‌ای امکان‌پذیر است. همچنین برنامه باید ۶۶ مقدار وزنه را ذخیره کند که آنها هم متغیرهای اعشاری هستند. در استفاده از این شبکه در کدکنندهٔ صحبت LPC نیاز به محاسبات اضافه بر آنچه در بالا گفته شد ـ به جز در مورد نرمالیزاسیون انرژی ـ نیست.

استفاده از شبکهٔ عصبی یک لایه، یعنی تنها یک نرون و بدون لایهٔ مخفی نیز بی‌آنکه خطای قابل توجهی ایجاد کند میزان محاسبات را تا حد فقط ۱۰ ضرب و ۱۰ جمع برای هر قطعه کاهش می‌دهد و بنابراین می‌تواند در ساخت کدکنندهٔ صحبت بلادرنگ مورد استفاده قرار گیرد.

منابع

۱- ب. ظهیر اعظمی، م. اکبر، ه. برهانی، «سیستم آنالیز و سنتز صحبت به روش پیشگویی خطی، کارنامهٔ پژوهشی شریف، ۱۳۶۸.

۲- ی. ثمره، «آواشناسی زبان فارسی، آواها و ساخت آوایی هجا»، نشر دانشگاهی، ۱۳۶٤.

۳- ب. ظهیر اعظمی، ه. برهانی، «روشهای LPC» گزارش ۶۹.۰۲، مرکز تحقیقات الکترونیک دانشگاه صنعتی شریف.

٤- ب. ظهیر اعظمی، ه. برهانی، «تشخیص واک به کمک شبکهٔ عصبی، گزارش ۶۹.۰۳، مرکز تحقیقات الکترونیک دانشگاه صنعتی شریف.

5- A. M. Null, "Cepstrum Pitch Determination" J. Acoust. Soc. Am., Vol. 41, Feb. 1967.

6- M.J.Ross, H.L. Shaffer, A. Cohem, R. Freudberg, H. J. Manley "Average Magnitude Difference Function Pitch Extraction" IEEE Trans. on ASSP, Vol. 22, Oct. 1974.

7- M.M. Sondhi, "New Methods of Pitch Extraction" IEEE Trans. on Audio Electroacoust., Vol. Au-16, June 1968.

8- B. Widrow, M.A.Lehr, "30 Years of Adaptive Neural Networks: Perceptron, Madaline & Back Propagation." IEEE Proc., Vol. 78, No. 9, Sep. 1990.