تنظیم تفریق طیفی به منظور بهبود کارایی سیستم‌های بازشناسی گفتار

مهران صفایانی[×]، حسین ثامتی، باقر باباعلی، محمدتقی منظوری شلمانی

چکیده

در این مقاله روشی جدید برای تنظیم یک فیلتر تفریق طیفی ارائه می‌گردد، به گونه‌ای که این فیلتر بیشترین تأثیر را بر بهبود نتایج یک سیستم بازشناسی گفتار داشته باشد. در حال حاضر واحدهای بازشناسی و بهسازی گفتار به صورت دو واحد مستقل عمل می‌کنند، بدین صورت که در ابتدا الگوریتم‌های بهسازی گفتار بروی سیگنال گفتار اعمال می‌شوند و سپس سیگنال بهسازی شده به واحد بازشناسی گفتار وارد می‌شود. در این سیستم‌ها فرض می‌شود که بهترین الگوریتم بهسازی گفتار مطمئناً منجر به بهترین نتیجه بازشناسی خواهد شد، ولی بازشناسی گفتار یک مسأله دسته بندی الگوهاست که از بردارهای ویژگی استخراج شده از سیگنال گفتار برای دسته بندی استفاده می‌کند. بنابراین تنها در صورتی نتایج بازشناسی افزایش خواهد یافت که این ویژگی‌ها درست‌نمایی[†] دنبالهٔ آوایی صحیح را نسبت به سایر دنباله‌های آوایی رقیب افزایش دهد. در روش جدید ارائه شده، فیلتر تفریق طیفی به نحوی تنظیم می‌شود که درست نمایی ویژگی‌های استخراج شده از سیگنال خروجی این فیلتر بیشینه شود. با بکارگیری این روش دقت سیستم‌های بازشناسی گفتار بروی دادگان فارس‌دات نویزی شده به میزان ۱۸٪ افزایش می‌یابد.

کلمات کلیدی

بهسازی گفتار، بازشناسی گفتار، تفریق طیفی، بیشینه کردن درست‌نمایی.

Calibration of Spectral Subtraction for Improving Performance of Speech Recognition Systems

Mehran Safayani, H. Sameti, B. Babaali, M.T. Manzuri Shalmani
Computer Engineering Dept. Sharif University of technology

Abstract

In this paper, we present a novel approach for adjusting a spectral subtraction filter which has the most effect on improving speech recognition performance. Currently speech enhancement and recognition units work as two independent sets. At first speech enhancement stage process the speech data then enhanced speech signal is processed by speech recognition stage. These systems make the assumption that the best speech enhancement algorithm, will result in the best recognition performance. However speech recognition is a pattern classification problem that uses the extracted feature vectors from the speech signal, so if these feature vectors increase the likelihood of the correct transcription with respect to other competing incorrect hypothesis, speech recognition performance will increase surely. In our new method, parameters of the spectral subtraction are optimized to maximize the likelihood of the recognition features extracted from the resultant output signal. By incorporating the speech recognition system into adjusting filter parameter, speech recognition performance increase up to 18% on FARSDAT noisy database.

Keywords

Speech Enhancement, Speech Recognition, Spectral Subtraction, Maximizing Likelihood.

---

[*] دانشگاه صنعتی شریف، دانشکده مهندسی کامپیوتر, {Safayani , Babaali}@ce.sharif.edu , {Sameti, Manzuri}@sharif.edu .

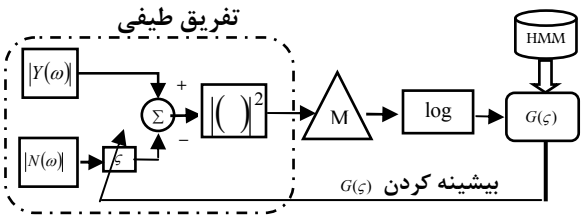
[†] Likelihood

۱- مقدمه

دقت سیستم‌های بازشناسی گفتار در حضور نویزهای جمع‌شونده و انعکاسی به شدت افت می‌کند. برای حذف اثر نویز روشهای بسیاری پیشنهاد شده است که می‌توان از تفریق طیفی، فیلتر وینر، فیلتر وفقی، فیلتر کالمن و سایر روشها نام برد [1]. تفریق طیفی یکی از عمومی-ترین روشها، در بهسازی گفتار است [3]. ایدهٔ اصلی آن بسیار ساده بوده و به راحتی قابل پیاده‌سازی است. این روش طیف سیگنال اصلی را توسط تفریق کردن طیف نویز از طیف سیگنال نویزی بدست می-آورد. در این روش فرض می‌شود که نویز ناهمبسته بوده و با سیگنال جمع می‌شود. به منظور افزایش کارایی این روش مطالعات بسیاری صورت گرفته است که بیشتر آنها کیفیت سیگنال گفتار را بر اساس معیارهایی نظیر افزایش توان سیگنال به نویز یا تجربیات شنیداری انسان افزایش می‌دهند. هنگامی‌که این روشها قبل از سیستم‌های بازشناسی بکار گرفته شود، فرض می‌شود که با افزایش کیفیت سیگنال صحبت، دقت سیستم‌های بازشناسی گفتار نیز افزایش خواهد یافت. البته اگر این روشها دنباله‌ای از ویژگی‌هایی تولید کنند که درست-نمایی دنبالهٔ آوایی صحیح را نسبت به سایر فرضیه‌ها افزایش دهد، نتیجه بازشناسی قطعاً بهبود خواهد یافت ولی معیار افزایش توان سیگنال به نویز یا سایر معیارهای مبتنی بر شکل موج لزوماً باعث بهبود دقت بازشناسی نمی‌شوند زیرا مسأله بازشناسی گفتار یک مسأله دسته‌بندی الگو است در حالیکه طراحی فیلتر یک مسأله پردازش سیگنال است. برای مثال یکی از مشکلات رایج تفریق طیفی نویز موزیکال[1] است. این نویز آزاردهنده باعث کاهش دقت بازشناسی گفتار می‌شود در حالی‌که ممکن است توان سیگنال به نویز افزایش یافته باشد. سلتزر[2] در [4] از این فرضیه استفاده کرده و نشان می‌دهد که نتیجه بازشناسی گفتار با تنظیم پارامترهای فیلتر بر اساس نتایج بازشناسی افزایش چشم‌گیری می‌یابد. در کار ایشان که بازشناسی گفتار از راه دور به کمک آرایه میکروفن است، یک فیلتر-و-جمعی[3] ارائه می‌شود که سیستم بازشناسی گفتار در تنظیم پارامترهای آن دخیل است و ضرایب شکل‌دادن پرتو[4] بر اساس بیشینه کردن درست‌نمایی دنبالهٔ آوایی صحیح تطبیق داده می‌شود. در این روش فرض می‌شود که با بیشینه کردن و یا حتی افزایش درست‌نمایی دنبالهٔ آوایی صحیح، دقت بازشناسی افزایش خواهد یافت. در [5] یک نسخهٔ زیرباند از این الگوریتم ارائه شده است و ضرایب فیلتر در حوزهٔ فرکانس بهینه می‌شود.

در این مقاله یک روش جدید به منظور تنظیم یک فیلتر تفریق طیفی بر اساس نتایج بازشناسی گفتار ارائه می‌شود. در این روش واحد بازشناسی گفتار در تخمین پارامتر فیلتر تخمین طیفی دخیل می‌شود و نشان داده می‌شود با انجام این کار نتایج بازشناسی گفتار افزایش خواهد یافت. این روش شامل دو مرحله است. مرحلهٔ اول، بهینه‌سازی است که در این مرحله از گفتاری که دنبالهٔ آوایی‌اش از قبل معلوم است، استفاده می‌شود. در این مرحله ضریب فیلتر تفریق طیفی به گونه‌ای پیدا می‌شود که فاصلهٔ مابین درست‌نمایی دنبالهٔ آوایی صحیح و درست‌نمائی فرضیه نادرست با بیشترین امتیاز، بیشینه شود. در مرحلهٔ دوم با استفاده از فیلتر استخراج شده در مرحلهٔ اول بازشناسی بروی همهٔ مجموعهٔ آزمون انجام می‌شود. در شکل (۱) نمودار بلوکی این روش نشان داده شده است. سیستم بازشناسی بکاررفته بر مبنای مدل مخفی مارکف بوده و بردارهای ویژگی بر اساس MFCC استخراج شده‌اند. این روش را می‌توان به صورت تفریق طیفی چند کاناله نیز تعمیم داد. بطوریکه یک ضریب مجزا بروی هر باند فرکانسی اعمال گردد.

باقی‌مانده این مقاله در ۵ بخش ارائه می‌شود. در بخش دوم روش‌های تفریق طیفی مرور می‌شود. در بخش سوم روش جدید ارائه شده به منظور تنظیم فیلتر تفریق طیفی بر مبنای نتایج واحد بازشناسی گفتار بیان می‌شود. در بخش چهارم پایگاه داده و سیستم بازشناسی گفتار بکار رفته توضیح داده می‌شود و آزمایش‌های انجام شده و تحلیل نتایج در بخش پنجم توضیح داده می‌شود و سرانجام در بخش ششم نتیجه مقاله ارائه شده است.



شکل (۱): نمودار بلوکی الگوریتم تنظیم کردن تفریق طیفی بر اساس نتایج بازشناسی

۲- تفریق طیفی

یکی از معروف‌ترین روش‌های بهسازی گفتار، روش تفریق طیفی است. در این روش اندازهٔ طیف سیگنال اصلی توسط کم کردن اندازهٔ طیف نویز از اندازهٔ طیف سیگنال نویزی بدست می‌آید. در این روش فرض می‌شود که نویز ناهمبسته با سیگنال صحبت بوده و با سیگنال صحبت جمع می‌شود. اگر این فرض برقرار باشد که در مورد نویز پس‌زمینه برقرار است. می‌توان رابطهٔ سیگنال با نویز را به صورت (۱) نوشت و تبدیل فوریه این رابطه به صورت (۲) بدست می‌آید. همچنین توان طیف سیگنال نویزی به صورت (۳) تخمین زده می‌شود.

$$y(k) = s(k) + n(k) \quad (۱)$$

$$Y(w) = S(w) + N(w) \quad (۲)$$

$$|Y(w)|^2 \approx |S(w)|^2 + |N(w)|^2 \quad (۳)$$

که در آن $|Y(w)|$, $|S(w)|$ به ترتیب اندازهٔ طیف سیگنال نویزی و تمیز است. از آنجائیکه طیف نویز نمی‌تواند به طور مستقیم

تخمین زده شود، در این رابطه نویز توسط میانگیری بروی M فریمی که تصور می‌شود که نویز است به صورت (۴) بدست می‌آید:

$$\left|\hat{N}(w)\right|^2 = \frac{1}{M}\sum_{i=0}^{M}\left|Y_i(w)\right|^2 \qquad (۴)$$

یکی از مشکلات تفریق طیفی وجود نویز موزیکال است که باعث افت کیفیت سیگنال گفتار می‌شود. بروتی[5] [6] با انجام تغییراتی در نسخه اصلی این الگوریتم روشی برای کاهش اثر این نویز ارائه داد که در آن تخمین بهتری از نویز بدست می‌آورد و از اینکه سیگنال حاصل از یک حد خاصی کمتر شود ممانعت می‌کند. این الگوریتم به صورت (۵) ارائه شده است.

(۵)

$$\left|\hat{Y}(w)\right|^2 = \begin{cases} \left|\hat{S}(\omega)\right|^2 - \alpha\left|\hat{N}(w)\right|^2 & \left|\hat{S}(\omega)\right|^2 > \beta\left|\hat{N}(w)\right|^2 \\ \beta\left|\hat{N}(w)\right|^2 & else \end{cases}\Bigg\}_{for\ \ all\ \ w}$$

که در آن α پارامتر تفریق طیفی است و β حداقل اندازه طیف را تعیین می‌کند. مقادیر زیاد α ممکن است باعث اعوجاج شود. برای ممانعت از این مشکل مقدار α بصورت فریم به فریم تطبیق داده می‌شود. تفریق طیفی‌ای که در این مقاله استفاده می‌شود در رابطهٔ (۶) و (۷) نشان داده شده است. در این فرمول‌ها ضریب ς یا بصورت تجربی بدست می‌آید و یا با توجه به توان سیگنال به نویز شکل موج خروجی تنظیم می‌شود که اگر مشخصات محیط تغییر کند این ضریب بایستی دوباره تنظیم شود. در این مقاله روشی برای تنظیم این پارامتر ارائه شده است بطوریکه بیشترین تأثیر را بر نتایج بازشناسی داشته باشد. البته روش ارائه شده قابل تعمیم برای سایر فرمول‌های تفریق طیفی نیز هست.

$$\left|\hat{S}(k)\right| = \left|Y(k)\right| - \varsigma\left|\hat{N}(k)\right| \qquad (۶)$$

$$\left|\hat{S}(k)\right| = \begin{cases} \left|\hat{S}(k)\right| & \left|Y(k)\right| > \varsigma\left|\overline{N}(k)\right| \\ \left|\overline{N}(k)\right| & otherwise \end{cases} \qquad (۷)$$

۳- بهینه سازی تفریق طیفی به منظور بازشناسی مقاوم گفتار

در این قسمت نحوهٔ بدست آوردن یک فیلتر تفریق طیفی مناسب بر مبنای افزایش فاصله بین درست‌نمایی دنبالهٔ آوایی صحیح نسبت به سایر دنباله‌های آوایی بیان می‌شود. فرض کنید که $L(k,\varsigma)$ امتیاز مدل واحد k ام (یا بطور خلاصه مدل k ام) نسبت به پارامتر فیلتر تفریق طیفی ς باشد که به صورت (۸) بیان می‌گردد.

$$L(k,\varsigma) = P\left(O(\varsigma)\,|\,\lambda_k\right) \qquad (۸)$$

در این رابطه $O(\varsigma)$ بیانگر دنبالهٔ مشاهدات است که تابعی از پارامتر فیلتر تفریقی طیفی است و λ_k نشان‌دهندهٔ مدل k ام است. از روی $L(k,\varsigma)$ ها، می‌توان متغیر مفید $H(i,\varsigma)$ را بصورت (۹) تعریف نمود:

$$H(i,\varsigma) = \log\frac{L(i,\varsigma)}{\sum_{j=1}^{K}L(j,\varsigma)} \qquad (۹)$$

که در آن K تعداد همهٔ مدل‌هاست. اگر c نشان دهندهٔ اندیس مدل صحیح باشد، می‌توان ضریب فیلتر تفریق طیفی را به صورتی بهینه کرد که درست نمایی $H(c,\hat{\varsigma})$ بیشینه گردد و در نتیجه دقت سیستم بازشناسی افزایش یابد که به صورت (۱۰) بیان می‌گردد.

$$H(c,\hat{\varsigma}) = \arg\max_{\varsigma}\left\{\log\frac{L(c,\varsigma)}{\sum_{j=1}^{K}L(j,\varsigma)}\right\} \qquad (۱۰)$$

در این مقاله به منظور کاهش پیچیدگی محاسباتی، در مخرج عبارت (۱۰) فقط مدل برنده (مدل با بیشترین امتیاز) قرار می‌گیرد و تابع هدفی به صورت (۱۱) بدست می‌آید.

(۱۱)

$$H(c,\hat{\varsigma}) = \arg\max_{\varsigma}\left\{\log\left(L(c,\varsigma)\right) - \log\left(L(r,\varsigma)\right)\right\}$$

که در آن c اندیس مدلی است که واقعاً بایستی بازشناسی شود و r اندیس مدلی است که هم‌اکنون بازشناسی شده است و ممکن است که مدل صحیح نباشد. فرض می‌شود که بیشینه کردن (۱۱) نسبت به ς منجر به افزایش فاصله، بین درست‌نمایی مدل صحیح و سایر مدلهای رقیب می‌شود. در این فرض یک پارادوکس وجود دارد. اگر دنبالهٔ واجی صحیح و به عبارتی دنبالهٔ مدل‌های صحیح معلوم باشند، دیگر نیازی به بازشناسی گفتار نیست. این مسأله بدین صورت توجیه می‌گردد که در الگوریتم جدید یک مرحلهٔ بهینه‌سازی وجود دارد و کاربر بایستی یک جمله‌ای که دنبالهٔ آوایی آن از قبل معلوم است بیان کند. سپس سیستم با استفاده از این دانش، بهینه‌سازی را انجام می‌دهد و یک فیلتر مناسب تنظیم می‌کند و بعد از آن همهٔ مجموعهٔ آزمون از این فیلتر به منظور بهسازی استفاده می‌کنند. $L(c,\varsigma)$ و $L(r,\varsigma)$ به ترتیب به صورت (۱۲) و (۱۳) بیان می‌شود.

(۱۲)

$$L(c,\varsigma) = \sum_{i=1}^{T}\log P\left(X_i(\varsigma)\,|\,s_i\right) + \log\left(s_1, s_2, ..., s_T\right)$$

(۱۳)

$$L(r,\varsigma) = \sum_{i=1}^{T}\log P\left(X_i(\varsigma)\,|\,\hat{s}_i\right) + \log\left(\hat{s}_1, \hat{s}_2, ..., \hat{s}_T\right)$$

که در آن‌ها $X_i(\varsigma)$ بردار ویژگی i امین فریم است و فرض می‌شود که درست‌نمایی دنبالهٔ واجی صحیح و همچنین دنبالهٔ واجی غیر صحیح با بیشترین امتیاز، با یک دنباله حالت که بیشترین امتیاز را دارد بیان می‌گردد که به ترتیب به صورت $S_C = \{s_1, s_2, ..., s_T\}$ و $\hat{S} = \{\hat{s}_1, \hat{s}_2, ..., \hat{s}_T\}$ بیان می‌گردند. همچنین بردارهای ویژگی X_i به روش MFCC استخراج می‌گردد. که می‌توان این ویژگی‌ها را به صورت رابطهٔ (۱۴) بیان نمود.

$$X_i = DCT\left(\log\left(M\,|\,S_i\,|^2\right)\right) \qquad (۱۴)$$

که در آن X_i بیانگر ضرایب MFCC فریم iام گفتار، S_i خروجی فیلتر تفریق طیفی و M بیانگر ماتریس ضرایب وزن دار فیلترهای مل هستند. توزیع‌های درون حالت‌های مدل مخفی مارکف توسط چند تلفیق گوسی مدل می‌شوند، بنابراین می‌توان رابطهٔ (۱۱) را به صورت (۱۵) بدست آورد.

(۱۵)

$$H(c,\hat{\varsigma}) = \arg\max_{\varsigma}\left\{\left(\sum_{t=1}^{|S_C|}\log\sum_{k=1}^{K}\mathrm{N}\left(X(t,\varsigma),\mu_{S_c(t),k},\Sigma_{S_c(t),k}\right)\right) - \left(\sum_{t=1}^{|S_r|}\log\sum_{k=1}^{K}\mathrm{N}\left(X(t,\varsigma),\mu_{S_r(t),k},\Sigma_{S_r(t),k}\right)\right)\right\}$$

که در آن N توزیع نرمال، $|S_C|$ تعداد کل فریم‌ها، K تعداد تلفیق‌های گوسی و $\mu_{S_c(t),k}$، $\Sigma_{S_c(t),k}$ به ترتیب بردار میانگین و ماتریس کوواریانس توزیع نرمال حالت k ام هستند. برای بیشینه کردن عبارت (۱۵) تابع $G(\varsigma)$ به صورت (۱۶) تعریف می‌شود.

(۱۶)

$$G(\varsigma) = \left[\left(\sum_{t=1}^{|S_c|}\log\sum_{k=1}^{K}\exp\left(-\frac{1}{2}\vartheta_C{}^H\Sigma^{-1}_{S_c(t),k}\vartheta\right)\right) - \left(\sum_{t=1}^{|S_r|}\log\sum_{k=1}^{K}\exp\left(-\frac{1}{2}\vartheta_r{}^H\Sigma^{-1}_{S_r(t),k}\vartheta_r\right)\right)\right]$$

که در آن ϑ_c به صورت (۱۷) و ϑ_r به صورت (۱۸) تعریف می‌شوند.

$$\vartheta_c = X(t,\varsigma) - \mu_{S_c(t),k} \qquad (۱۷)$$

$$\vartheta_r = X(t,\varsigma) - \mu_{S_r(t),k} \qquad (۱۸)$$

که در آنها $X(t,\varsigma)$ بردار ویژگی در فریم t ام نسبت به پارامتر فیلتر ς است، $\mu_{S_c(t),k}$ و $\mu_{S_r(t),k}$ به ترتیب برابر بردار میانگین مربوط به مدل صحیح و مدل بازشناسی شده است. گرادیان (۱۶) نسبت به ς به صورت (۱۹) بدست می‌آید. با استفاده از (۱۶) و گرادیان آن می‌توان (۱۵) را با استفاده از روش گرادیان مزدوج [2] بیشینه کرد و ضریب بهینه فیلتر تفریق طیفی را بدست آورد.

(۱۹)

$$\nabla_{\varsigma}G(X) = \left[\left(-\sum_{t}\sum_{k=1}^{K}\left(\frac{\exp(\rho_c)}{\sum_{j=1}^{K}\exp(\rho_c)}\right)\Sigma^{-1}_{S_c(t),k}\vartheta_c\frac{\partial X_t^c(\varsigma)}{\partial\varsigma}\right) + \left(\sum_{t}\sum_{k=1}^{K}\left(\frac{\exp(\rho_r)}{\sum_{j=1}^{K}\exp(\rho_r)}\right)\Sigma^{-1}_{S_r(t),k}\vartheta_r\frac{\partial X_t^c(\varsigma)}{\partial\varsigma}\right)\right]$$

که در آن ρ_c به صورت (۲۰) و ρ_r به صورت (۲۱) بیان می‌شوند و $\frac{\partial X_t^c(\varsigma)}{\partial\varsigma}$ یک ماتریس ژاکوبین است که از مشتق جزئی هر مؤلفهٔ بردار ویژگی نسبت به ضریب فیلتر تفریق طیفی بدست می‌آید.

$$\rho_c = -\frac{1}{2}\vartheta_c{}^H\Sigma^{-1}_{S_c(t),k}\vartheta_c \qquad (۲۰)$$

$$\rho_r = -\frac{1}{2}\vartheta_r{}^H\Sigma^{-1}_{S_r(t),k}\vartheta_r \qquad (۲۱)$$

به طور کلی الگوریتم این روش در ادامه بیان می‌شود:

۱. یک مقدار پیش‌فرض برای ضریب فیلتر در نظر گرفته می‌شود برای مثال ضریب برابر با صفر.
۲. کاربر یک جمله‌ای را که دنبالهٔ آوایی آن از قبل معلوم است بیان می‌کند.
۳. سیستم بازشناسی بهترین دنبالهٔ حالت را با استفاده از الگوریتم ویتربی، دنبالهٔ آوایی معلوم و ضریب فیلتر تفریق طیفی بدست می‌آورد.
۴. با استفاده از روش گرادیان، (۱۶) نسبت به ضریب فیلتر تفریق طیفی بیشینه می‌شود
۵. اگر فیلتر تولید شده همگرا شد به مرحلهٔ ۶ برو وگرنه به مرحلهٔ ۳ برگرد.
۶. ضریب بهینه بدست آمده به عنوان پارامترهای فیلتر نهایی بکار می‌رود.

۴- پایگاه دادگان و سیستم بازشناسی گفتار

به منظور آموزش و آزمایش این سیستم از دادگان فارس‌دات[7] استفاده گردیده است. این پایگاه داده شامل ۶۰۸۰ گفتار فارسی است که توسط ۳۰۴ گوینده بیان شده است. این گویندگان از ۱۰ منطقهٔ جغرافیایی مختلف در ایران انتخاب شده‌اند بنابراین ۱۰ لهجهٔ رایج فارسی در این پایگاه وجود دارد. نسبت مرد به زن در این پایگاه دو به یک است و ۴۰۵ جمله در آن وجود دارد که ۲۰ جمله بازای هر

گوینده است. هر گوینده ۱۸ جمله را به صورت تصادفی بیان کرده است باضافهٔ دو جمله که در بین همهٔ گویندگان مشترک است. این جمله‌ها بر اساس ۱۰۰۰ کلمهٔ فارسی شکل می‌گیرند. پایگاه داده در محیط با نویز کم ضبط شده است و توان سیگنال به نویز آن به طور متوسط ۳۱ دسی‌بل است. به منظور آزمودن سیستم از ۱۰۰ جملهٔ آن استفاده شده است و باقی جمله‌ها برای آموزش استفاده گردیده است.

به منظور نشان دادن اثر الگوریتم جدید بر نتایج بازشناسی گفتار از یک سیستم بازشناسی واج‌های گسسته مبتنی بر مدل مخفی مارکف استفاده می‌شود. HMMهای به کار برده شده از نوع چگالی پیوسته و با تلفیق‌های گوسی بوده و پرش مجاز بین حالت‌های آن‌ها فقط به صورت چپ به راست است. در آزمایش‌های ما مدل سازی گفتار براساس واج و با استفاده از دادگان فارس‌دات انجام گردید و سعی شد از بهترین پارامترهای ممکن استفاده شود. بردارهای ویژگی استخراج شده از فریم‌های گفتار، شامل ۱۲ ضریب مل-کپستروم (C0-C11) همراه با مشتقات زمانی اول و دوم آن‌ها می‌باشد. مدل‌های مخفی مارکوف با استفاده از الگوریتمsegmental k-means آموزش داده شده‌اند و توپولوژی آن‌ها برای تمام واج‌ها یکسان در نظر گرفته شده است. لازم به ذکر است که در دادگان فارس‌دات کل آواهای زبان فارسی به ۴۳ واج تقسیم‌بندی شده است. از این ۴۳ واج، ۱۲ واج مربوط به قسمت بستار واج‌های انسدادی و بقیه واج‌های استاندارد زبان فارسی می‌باشند. قسمت‌های غیر گفتاری سیگنال (مانند سکوت) نیز به عنوان یک واج در نظر گرفته شده‌اند؛ بنابراین در مجموع ۴۴ واج می‌تواند مدل گردد. ولی در این پروژه با توجه به عملکرد بهتر بازشناسی، با ادغام قسمت بستار واج‌های انسدادی با قسمت رهش و همچنین ترکیب بعضی از واج‌های شبیه با یکدیگر، تعداد مدل‌ها به ۳۰ مدل واج کاهش یافته است.

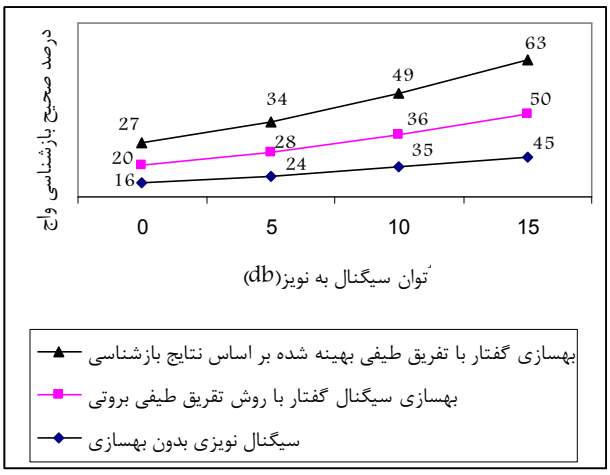
۵- آزمایش‌ها

در ابتدا دادگان آزمون شامل ۱۰۰ جملهٔ فارس‌دات با نویز سفید جمع گردید و در توان سیگنال به نویزهای مختلف، دادگان آزمون آماده شدند. سپس یک جملهٔ آن به منظور بهینه‌سازی استفاده گردید و با استفاده از روش توضیح داده شده در بخش (۳) فیلتر مناسب تنظیم گردید و از آن در مرحلهٔ بازشناسی بر روی ۹۹ جمله استفاده شد. درصد بازشناسی واج‌های صحیح در جدول (۱) نشان داده شده است. همانگونه که در این جدول مشاهده می‌شود نتایج بازشناسی در توان سیگنال به نویزهای مختلف افزایش داشته است که برای نمونه می‌توان به افزایش ۱۸ درصدی دقت بازشناسی واج در توان سیگنال به نویز ۱۵ دسی‌بل اشاره کرد.

جدول (۱): درصد صحیح بازشناسی واج بروی ۹۹ جملهٔ فارس‌دات

توان سیگنال به نویز (db)	بدون بهسازی	بهینه‌سازی به روش بیشینه کردن درست نمائی
۰	۱۶	۲۷
۵	۲۴	۳۴
۱۰	۳۵	۴۹
۱۵	۴۵	۶۳

در این قسمت الگوریتم بهینه سازی تفریق طیفی ارائه شده با یک روش تفریق طیفی مبتنی بر توان سیگنال به نویز مقایسه می‌شود. یکی از متداول‌ترین الگوریتم‌های تفریق طیفی، الگوریتمی است که توسط بروتی [7] ارائه گردید که رابطهٔ این روش در (۶) ارائه شده است. در این قست در ابتدا با استفاده از الگوریتم بروتی دادگان نویزی که شامل ۱۰۰ جملهٔ فارس‌دات است، بهسازی می‌شود و سپس این سیگنال‌های بهسازی شده به سیستم بازشناسی وارد می‌گردد. نتایج حاصل از اعمال این روش در شکل (۲) با روش بهینه سازی تفریق طیفی به روش بیشینه کردن درست نمایی مقایسه شده است. همانگونه که از شکل (۲) مشاهده می‌شود، روش تفریق طیفی با استفاده از بیشینه کردن درست نمائی در همهٔ موارد نسبت به تفریق طیفی بروتی بهبود داشته است و این بهبود در دادگان با توان سیگنال به نویز ۱۵ دسی‌بل به طور وضوح مشاهده می‌گردد که روش بروتی تنها ۵ درصد نتایج بازشناسی را افزایش داده است درحالیکه روش جدید ۱۸ درصد دقت بازشناسی را افزایش داده است . با بررسی سیگنال‌های خروجی روش تفریق طیفی بروتی مشاهده شد، که این روش مقدار زیادی نویز موزیکال به سیگنال اضافه می‌کند که باعث می‌شود که بسیاری از واج‌ها به درستی بازشناسی نشود و در نتیجه اعمال این روش در ورودی سیستم بازشناسی گفتار نتایج بازشناسی را به میزان قابل توجهی افزایش نمی‌دهد.



شکل (۲): مقایسه الگوریتم تفریق طیفی بهینه شده با تفریق طیفی بروتی

۶- نتیجه

در این مقاله روشی جدید به منظور بهینه‌سازی یک فیلتر تفریق طیفی به منظور استفاده در ورودی سیستم‌های بازشناسی گفتار ارائه گردید و نشان داده شد که بکارگیری این روش نتایج بازشناسی را نسبت به روش‌های متداول تفریق طیفی افزایش می‌دهد. با بکارگیری روش جدید ارائه شده مطمئن هستیم که این فیلتر ویژگی‌هایی از سیگنال صحبت را که برای سیستم بازشناسی مهم تر است تقویت می‌کند و در نتیجه نتایج سیستم‌های بازشناسی گفتار افزایش می‌یابد. این روش را می‌توان به صورت تفریق طیفی چند کاناله نیز گسترش داد و همچنین بروی سایر الگوریتم‌های بهسازی گفتار نیز اعمال کرد.

مراجع

[1] J. R. Deller, J. G. Proakis, H. L. Hansen, “Discrete Time Processing of Speech Signals”, Macmillan, New York, NY, 1993.
[2] E. Polak, "Computational methods in optimizations, New York: Academic Press. 1971.
[3] S.F. Boll, “Suppression of Acoustic Noise in Speech using Spectral Subtraction”, IEEE Trans Acoustics, Speech & Signal Processing, April 1979.
[4] M. L. Seltzer, B. Raj, R. M. Stern, “Likelihood-Maximizing Beamforming for Robust Hands-Free Speech Recognition”, IEEE Transaction on Speech and audio processing, VOL. 12, NO. 5, September 2004.
[5] M. L. Seltzer, R. M. Stern, "Subband parameter optimization of microphone arrays for speech recognition in reverberant environments", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, July 2003.
[6] M. Berouti, R. Schwartz, J. Makhoul, “Enhancementof speech corrupted by acoustic noise”, Proceedings of the IEEE International Conference on Acoust, Speech and Signal Processing, pp. 208-211, Apr. 1979.
[7] M. Bijankhan, J. Sheikhzadegan, “FARSDAT-The Farsi Spoken Language Database”, proceeding of the 5th International conference of speech science and technology, vol. 2, pp. 826-831, 1994.

زیرنویس‌ها

[1] Musical
[2] Seltzer
[3] Filter-and-Sum
[4] Beamforming
[5] Berouti