# بهبود روش تطبیق گوینده MLLR در بازشناسی گفتار غیر وابسته با کمک خطای کمینه کلاس‌بندی

رضا صحرائیان[۱]، بهزاد زمانی[۲]، احمد اکبری[۲] ، احمد آیت الهی[۱] ، بابک ناصرشریف[۳]

[۱]دانشکده مهندسی برق - دانشگاه علم و صنعت ایران

[۲]دانشکده مهندسی کامپیوتر - دانشگاه علم و صنعت ایران

[۳]دانشکده فنی و مهندسی - دانشگاه گیلان

**چکیده:** روش بیشینه همسانی با رگرسیون خطی (MLLR) به عنوان یکی از شناخته شده‌ترین روشهای تطبیق گوینده است که در دادگان تطبیقی کم با خوشه بندی پارامترهای مدل و تخمین ماتریس تبدیل برای هر خوشه باعث بهبود عملکرد تطبیق می‌شود در این مقاله سعی شده است که با استفاده از روش متمایز ساز خطای کمینه رده‌بندی با متمایز کردن ویژگی‌های سیگنال گفتار این مشکل را برطرف کرده، عملکرد تطبیق را بهبود دهیم. بکارگیری متمایز ساز خطای کمینه رده‌بندی ضعف روش MLLR در خوشه بندی را برطرف کرده و یک خوشه‌بندی یکنواخت ایجاد می‌کند. بر اساس آزمایشات انجام گرفته بر روی دادگان TIMIT و WSJ0 و نتایج بدست آمده، روش پیشنهادی یعنی بکارگیری خطای کمینه رده‌بندی نسبت به روشهای معادل بدون خطای کمینه رده‌بندی باعث افزایش ۰.۴۹ تا ۰.۶۲ درصد در نرخ بازشناسی واج برای TIMIT و ۰.۱۲ تا ۰.۷۲ درصد افزایش نسبی در نرخ بازشناسی واج برای دادگان WSJ0 شده است.

**واژه های کلیدی:** تطبیق گوینده،کمترین خطای رده‌بندی، بیشینه همسانی با رگرسیون خطی.

## ۱- مقدمه

تطبیق بر اساس بیشینه احتمال درستنمایی رگرسیون خطی ((Maximum Likelihood Linear Regression (MLLR)) به عنوان یکی از روش‌های مؤثر تطبیق گوینده با داده‌های تطبیقی کم شناخته شده است[۱]. در این روش برای پارامترهای مخلوط‌های گوسی مدلها (میانگین‌ها و واریانس‌ها) در سیستم مدل مخفی مارکوف (HMM) مجموعه‌ای از تبدیل‌های خطی در نظر گرفته می‌شود. توابع تبدیل به گونه‌ای تخمین زده می‌شوند که احتمال همسانی داده‌های تطبیقی بیشینه شود. از آن‌جایی که تعداد پارامترهای مدل زیاد بوده و معمولاً داده‌های تطبیقی کم می‌باشند، لذا تخمین دقیقی از پارمترهای مدل صورت نمی‌گیرد. از اینرو پارامترهای مدل را خوشه‌بندی کرده و برای هر کلاس رگرسیون تابع تبدیل تخمین زده می‌شود. فرض بر این است که تمام اجزای موجود در یک کلاس رگرسیون به شیوه یکسانی تبدیل می‌گردند. سؤال مهم این است که کدام پارامترهای مدل با یکدیگر در یک کلاس رگرسیون قرار گیرند. دو رویکرد در طراحی درخت‌های رگرسیون مورد توجه است: [۲]

۱. دانش آوایی: در این روش برای تصمیم‌گیری اینکه کدام اجزا با یکدیگر تشکیل یک کلاس دهند، از دانش آوایی استفاده می‌شود. اجزا بر اساس کلاس‌های آوایی متنوع (مثل nasalها، glideها) یا در سطح پایین‌تر بر اساس واج‌ها از هم جدا می‌شوند.
۲. فضای آکوستیکی: در این روش اجزا بدون در نظر گرفتن اینکه به کدام واج تعلق دارند، بر اساس نزدیکی‌شان در فضای آکوستیکی خوشه‌بندی می‌شوند. مزیت این روش عدم نیاز به دانش آوایی و وابستگی به داده‌های اعمالی می‌باشد، با این حال هیچ اطلاعاتی در مورد هویت آواها در کلاس‌ها وجود ندارد.

اعمال ماتریسهای تبدیل می‌تواند به صورت‌های مختلفی انجام گیرد. Gales در سال ۱۹۹۸ مستقیماً ماتریس‌های تبدیل را به بردارهای ویژگی آکوستیکی اعمال کرد [۳]. در آموزش تطبیقی گوینده ((speaker adaptive training (SAT))، تبدیلات بدست آمده از MLLR به هر دو دسته گوینده‌های آزمایش و آموزش اعمال می‌شوند، این کار جهت مدل کردن تغییرات بین گوینده‌ها و همچنین یافتن مدلی استاندارد برای تغییرات گفتار یک گوینده می‌باشد[۴]. علاوه بر این، تبدیل‌های MLLR به عنوان ویژگی‌هایی برای مدل کردن گوینده در بازشناسی گوینده [۵] استفاده می‌شوند.

برای بهبود عملکرد تطبیق MLLR تلاش‌هایی در زمینه بهبود درخت رگرسیون خطی انجام شده است. Gales الگوریتم تکرار شونده-ای از یک خوشه‌بندی اولیه برای رسیدن به درخت کلاس رگرسیونی که به راه‌حل درستنمایی بیشینه نزدیک‌تر باشند، ارائه داد. این روش به بهبود بازشناسی گفتار کمکی نکرد [۲].

استفاده از ((Minimum Classification Error (MCE)) برای تطبیق برای نخستین بار توسط Lin و همکارانش در [۶] گزارش شده است. در سایر مطالعاتی که در این زمینه انجام شده است از MCE برای بهبود آموزش مدلهای مدل مخفی مارکوف در مرحله تطبیق استفاده شده است. و همچنین به عنوان راه‌حلی در مقایسه با بیشینه همسانی برای بهبود یادگیری پارامترهای مدلهای مخفی مارکوف ارایه

شده است.[۷، ۸] در این مقاله برای بهبود عملکرد تطبیق گوینده از خطای کلاس‌بند کمینه در سیستم بازشناسی گفتار استفاده شده است.

همانطور که گفته شد برای تخمین تابع تبدیل MLLR ابتدا اجزای مدل مرجع را کلاس‌بندی می‌کنیم، همبستگی موجود بین بردارهای ویژگی می‌تواند اثر نامطلوبی روی کلاس‌بندی بگذارد. این اثر به صورت تجمع اجزای مدل در یک کلاس و کاهش این اجزا در کلاس‌های دیگر می‌باشد. تجمع گوسی‌ها در یک کلاس باعث می‌شود تطبیق MLLR چند کلاسی به تطبیق MLLR تک کلاسی شبیه شود. در نتیجه انتظار داریم بهبود تطبیق ناشی از چند کلاسی کردن، تحت تأثیر قرار گیرد. استفاده از روش متمایز ساز MCE به صورتی که در بخش ۳ و ۴ توضیح داده خواهد شد به ما کمک می‌کند اجزای گوسی مدل مرجع را به صورت یکنواخت‌تری کلاس‌بندی کنیم. نتایج حاصل نشان می‌دهد این کار باعث بهبود نسبی تطبیق MLLR در بازشناسی می‌شود.

ساختار ادامه مقاله به این صورت خواهد بود که در بخش دوم بیشینه احتمال همسانی با رگرسیون خطی گفته می‌شود. سپس در بخش سوم خطای کمینه کلاسبندی آورده شده است. بخش چهارم به ارائه روش پیشنهادی می‌پردازد. در ادامه نتایج آزمایشات در بخش پنجم بیان و در انتها در بخش ششم نتیجه‌گیری گفته شده است.

## ۲- بیشینه همسانی با رگرسیون خطی

روش MLLR مجموعه‌ای از تبدیلات مبتنی بر رگرسیون بدست آمده از داده‌های تطبیقی را برای تنظیم مدلهای HMM به کار می‌گیرد. روش MLLR با بکارگیری تبدیل اشتراکی قادر است تبدیل تطبیقی مقاومی را حتی برای آن دسته از پارامترهای مدل که در داده‌های تطبیقی نیستند فراهم سازد. این موضوع مشکل محدود بودن اطلاعات تطبیقی را از بین می‌برد. برای داده‌های تطبیقی کم، یک تبدیل مشترک برای تمام پارامترهای مدل‌ها استفاده می‌شود. معمولاً تطبیق تنها برای میانگین صورت می‌گیرد، چون فرض بر این است که تفاوت میان گوینده‌ها ناشی از متوسط موقعیت آواها در فضای آکوستیکی می‌باشد. از تطبیق کوواریانس کمتر استفاده می‌شود و اثر آن نیز غالباً کمتر از تطبیق میانگین است[۱].

برای تخمین پارامترهای مدل مربوط به گوینده جدید از یک تابع شباهت استفاده می‌شود. برای سادگی محاسبات احتمال به عنوان تابع شباهت بکار می‌رود. بنابراین هدف تخمین پارامترهای مدل است به گونه‌ای که پارامترهای جدید باعث افزایش تابع احتمال درستنمایی شوند. رابطه (۱) این مساله را نشان می‌دهد.

$$f_l(\hat{\lambda}) \geq f_l(\lambda) \qquad (۱)$$

که $\lambda$ نشان دهنده پارامترهای مدل و $f_l$ تابع احتمال درستنمایی می‌باشد. فرض کنید $\mu_s$ میانگین اجزای ترکیبی یک بردار n-بعدی باشد، می‌توان نوشت:

$$\hat{\mu} = W_k \xi_k \qquad (۲)$$

که $W_s$ ماتریس تبدیل با ابعاد $n \times (n+1)$ و $\xi_s = [w, \mu_{s1}, ..., \mu_{sn}]^t$ بردار میانگین گسترش یافته است. $w$ نشانگر افست است، اگر یک باشد، افست وجود دارد و اگر صفر باشد، افست وجود ندارد.

فرض کنید داده‌های تطبیقی مجموعه‌ای از T مشاهده باشد، $O=o_1,o_2,\ldots,o_T$ و $W$ برای R حالت $\{1,2,\ldots,r,\ldots,R\}$ تسهیم شده باشد، شکل عمومی رابطه تخمین تابع تبدیل MLLR برای کلاس kام به صورت رابطه (۳) است:

$$\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{kr}(t)C_{kr}^{-1}o_t\xi_{kr}^{t} = \sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{kr}(t)C_{kr}^{-1}W_k\xi_{kr}\xi_{kr}^{t} \qquad (۳)$$

که C ماتریس کوواریانس و $\gamma_r(t)$ احتمال پیشین گوسی rام در زمان t می‌باشد.

## ۳- خطای کمینه کلاس بندی

روش کمینه نمودن خطای کلاس‌بند (MCE)، یکی از شناخته‌شده‌ترین روش‌های متمایزسازی است که هم در گروه تبدیل ویژگی و هم در گروه آموزش کلاس‌بند بکار گرفته می‌شود[۹]. هنگامی‌که روش MCE برای آموزش HMM (به عنوان یک کلاس‌بند) استفاده می‌شود، MCE پارامترهای مدل را بگونه‌ای تغییر می‌دهد که خطای کل کلاس‌بند کاهش یابد[۱۰]. در روش آموزش به کمک MCE، ابتدا تابع هدف (که برای یافتن پارامترهای HMM بکار می‌رود) با استفاده از یک تابع پیوسته مدل می‌شود. سپس، کمینه این تابع با یک روش جستجوی گرادیان، مانند Gradient Probabilistic Descent، بدست می‌آید [۱۱].

ایده اصلی الگوریتم MCE، بهینه‌سازی نرخ خطای تجربی با استفاده از یک مجموعه داده آموزشی، جهت بهبود نرخ خطای بازشناسی می‌باشد. پس از آنکه نرخ خطای تجربی، بوسیله یک بازشناس یا کلاس‌بند، کمینه شد، تخمینی بایاس شده از نرخ خطای واقعی بدست می‌آید. یک راه مؤثر جهت کاهش این نرخ و بهبود کارآیی کلی سیستم، افزایش مرز میان کلاس‌ها در داده‌های آموزشی است. به این منظور از گرادیان کاهشی [۱۱] یا ترکیب امتیازهای حاشیه و نرخ خطای تجربی استفاده شده استروش MCE برای بدست آوردن تبدیل فضای ویژگی نیز استفاده می‌شود. در [۱۲]، MCE برای یافتن ماتریس تبدیل ویژگی‌ها، که بتواند خطای کلاس‌بند را کمینه کند، استفاده شده است. در ادامه و پس از انتشار مرجع [۱۲] ، Wang و Paliwal این روش را برای بازشناسی واکه‌ها بهبود بخشیدند[۱۳]. در همه این رویکردها از تابع هزینه‌ای بصورت رابطه (۴) استفاده شده است.

$$d_{k,Add}(O,F) = -g_k(O,F) + \frac{1}{\eta}\log\left[\frac{1}{M-1}\sum_{i=1,i\neq k}^{M}\exp\left(g_i(O,F)\eta\right)\right] \quad (۴)$$

که در آن $M$ تعداد کلاس‌ها، $\eta$ مقداری مثبت و $g_i(O,\lambda_i)$ لگاریتم احتمال تولید شده برای مشاهده $O$، به شرط مدل $\lambda_i$، بوسیله HMM می‌باشد و بصورت رابطه (۵) تعریف می‌شود:

$$g_i(O,\lambda_i) = \log p(O|\lambda_i) \quad (۵)$$

در روش MCE ، هدف کمینه نمودن تابع هزینه (۴) است. اما از آنجا که این توابع مشتق‌پذیر نیستند، از یک تابع پیوسته نرم مانند تابع سیگموید بعنوان تابع هزینه استفاده می‌کنیم. بنابراین خواهیم داشت:

$$l_k(O,F) = \frac{1}{1+\exp(-\alpha d_k(O,F))} \quad (۶)$$

که $d_k(O,F)$ تابع هزینه (۴) می‌باشد و $\alpha$ پارامتر تنظیم بزرگتر از یک است. بنابراین اگر $d_k(O,F)$ خیلی کوچکتر از صفر باشد، درواقع یک کلاس‌بند درست رخ داده و $l_k(O,F)$ به صفر نزدیک می‌شود. از طرفی مقدار مثبت برای $d_k(O,F)$ نشان دهنده جریمه برای کلاس-بندی نادرست است و در اینصورت $l_k(O,F)$ به ۱ متمایل می‌شود. کل خطای کلاس‌بندی برای مشاهده $O$ از رابطه (۷) بدست می‌آید:

$$L = \sum_{k=1}^{M} l_k(O,F) \quad (۷)$$

با استفاده از ماتریس تبدیل، بردارهای ویژگی از فضای ویژگی اصلی به فضای جدید برده می‌شود تا تمایز کلاس‌ها بیشتر گردد. تمایز بیشتر باعث کاهش خطا می‌شود. از اینرو با استفاده از معیار تابع هزینه سعی می‌شود ماتریس تبدیل بهینه بدست آید. مقادیر درایه‌های ماتریس تبدیل به روش تکراری با استفاده از گرادیان کاهشی برای تابع $L$، با رابطه (۸) قابل محاسبه خواهد بود.

$$w_{n,iter} = w_{n,iter-1} - \beta\frac{\partial L}{\partial w_n} \quad (۸)$$

رابطه فوق را می‌توان بصورت ماتریسی بیان کرد:

$$\mathbf{W}_{iter} = \mathbf{W}_{iter-1} - \beta\frac{\partial L}{\partial \mathbf{W}} \quad (۹)$$

که $\boldsymbol{W}$ ماتریس تبدیل؛ $r$ شاخص $r$امین درایه ماتریس تبـدیل $\boldsymbol{W}$؛ $iter$ اندیس تکرار الگوریتم و $\beta$ پارامتر یادگیری می‌باشـد. رابطـه (۹) نشان دهنده یک رویه تکراری است. این رویه زمانی متوقف می‌شود کـه مقدار تابع هزینه از یک حد آستانه کمتر شود. در این مقاله از این روش برای بهبود عملکرد تطبیق دهنده گوینده MLLR استفاده شده است.

## ۴- روش پیشنهادی

در تطبیق بر مبنای مدل، هدف نگاشت مدل مربوط به سیستم غیروابسته به گوینده به سمت مدل جدیدی است که ویژگی‌های گوینده جدید را به خوبی نشان دهد. همانطور که گفته شد در MLLR این نگاشت با تخمین ماتریس تبدیل و اعمال آن بر روی میانگین‌ها و واریانس‌های مربوط به مخلوط‌های گوسی مدل غیر وابسته به گوینده انجام می‌شود. در حالت ایده‌آل هدف داشتن ماتریس‌های تبدیلی است که هر یک از گوسی‌های مدل مرجع را به گوسی‌های مدل گوینده جدید نگاشت کند. از آن‌جا که داده‌های تطبیقی مربوط به گوینده جدید معمولاً کم است مجبور به کلاس‌بندی این پارامترها و تخمین ماتریس تبدیل برای هر کلاس از پارامترها هستیم[۱]. هر چه داده‌های تطبیقی بیشتری داشته باشیم می‌توان تعداد کلاس‌ها را افزایش و به حالت ایده‌آل نزدیکتر شد. از طرفی بنا به رابطه (۳) علاوه بر میزان داده‌های تطبیقی هر چه تعداد گوسی‌های هر کلاس بیشتر باشد تخمین مقاوم‌تری برای آن کلاس انجام می‌شود[۱۴]. بنابراین یکی از عواملی که می‌تواند باعث کاهش دقت تخمین ماتریس MLLR برای یک کلاس شود، کم بودن تعداد گوسی‌های آن کلاس است. به بیان دیگر با ثابت بودن داده‌های تطبیقی بین دو کلاس، برای آن کلاسی که تعداد گوسی‌های بیشتری دارد تخمین بهتری زده می‌شود. با توجه به ثابت بودن تعداد کل گوسی‌ها، زیاد شدن تعداد گوسی‌ها در یک کلاس معادل کم شدن آن‌ها در کلاس‌های دیگر می‌شود.

فرض کنید برای یک سیستم با تعداد ۱۰۰ مخلوط گوسی تطبیق ۴ کلاسی MLLR انجام می‌گیرد. اگر بعد از کلاس‌بندی، در کلاس یک ۹۰ مخلوط گوسی، کلاس دوم ۵ مخلوط گوسی، کلاس سوم ۲ مخلوط گوسی و کلاس چهارم ۳ مخلوط گوسی قرار گیرد و داده‌های تطبیقی به اندازه کافی موجود باشد. ماتریسی که برای کلاس اول تخمین زده می‌شود بسیار مقاوم است، ولی ۹۰٪ گوسی تحت این تبدیل قرار می-گیرند و تبدیل چند کلاسی ما از حالت ایده‌آل که قرار بود هر گوسی با یک تابع تبدیل به گوسی مربوط به مدل جدید نگاشت یابد، فاصله می-گیرد. بنابراین بهبود ناشی از تطبیق چند کلاسی عملاً معادل تطبیق یک کلاسی می‌باشد.

به طور کلی تجمع بیش از حدگوسی‌ها در یک کلاس نه تنها باعث کاهش دقت تخمین برای کلاس‌های دیگر می‌شود، بلکه باعث شبیه شدن تطبیق چند کلاسی به تطبیق تک کلاسی نیز می‌گردد.

برای جلوگیری از تجمع مخلوطهای گوسی در یک کلاس باید به معیار خوشه‌بندی گوسی‌ها مراجعه کنیم. همانطور که اشاره شد، این خوشه‌بندی معمولاً به دو صورت انجام می‌شود: بر اساس دانش آوایی، نزدیکی در فضای آکوستیکی. در MLLR استاندارد مربوط به HTK، کلاس‌بندی بر اساس نزدیکی توزیع ویژگی‌ها انجام می‌شود. این کلاس-بند با در نظر گرفتن نزدیکی میانگین گوسی‌ها آن‌ها را کلاس‌بندی می‌کند. بنابراین برای این‌که گوسی‌ها در یک کلاس متمرکز نشوند می-بایست فاصله بین میانگین‌های گوسی‌ها افزایش داده شود.

ساده‌ترین راه برای افزایش اختلاف میان میانگین دو کلاس داده، ضرب داده‌ها در یک عدد بزرگتر از یک می‌باشد. یک دسته n بعدی از ویژگی‌های سیگنال گفتار را بفرم زیر در نظر بگیرید.

$$fe = [fe_1 \ fe_2 \ ... \ fe_n]^T \qquad (۱۰)$$

اگر T ماتریس تبدیل باشد، بردار ویژگیها بصورت رابطه (۱۱) نگاشت می‌یابند.

$$fe' = T \times fe = T \times [fe_1 \ fe_2 \ ... \ fe_n]^T \qquad (۱۱)$$

چنانچه این تابع تبدیل خطی باشد(ماتریس قطری)، داریم:

$$fe = [a_1 fe_1 \ a_2 fe_2 \ ... a_n fe_n]^T \qquad (۱۲)$$

که $a_i$ها درایه‌های روی قطر اصلی ماتریس تبدیل می‌باشند. این تبدیل خطی باعث تغییر در توزیع احتمال هر یک از ویژگی‌ها می‌شود. اگر Fe نشان دهنده مجموعه بردارهای ویژگی $fe$ باشد و توزیع احتمال آن را گوسی در نظر بگیریم، داریم:

$$\begin{aligned} Fe &: N(\mu;\sigma) \\ Fe' &: N(a\mu;a\sigma) \end{aligned} \qquad (۱۳)$$

از آن‌جا که در HMM توزیع احتمال مربوط به هر یک از ویژگی‌ها به صورت ترکیب خطی وزن‌دار است، داریم:

$$\begin{aligned} P(Fe) &= \sum_i k_i N(\mu_i;\sigma_i) \\ P(Fe') &= \sum_i k_i N(a\mu_i;a\sigma_i) \end{aligned} \qquad (۱۴)$$

بنابران فاصله میان اجزای گوسی حتی در یک مدل نیز با ضریب $a$ تغییر می‌کند، بنابراین اختلاف میان میانگین‌های توزیع‌ها بیشتر نمود پیدا می‌کند.

الگوریتم خطای کمینه کلاس‌بندی به ما کمک می‌کند تا بتوانیم توزیع‌های ویژگی‌ها را در یک سطح مشخص از هم تفکیک‌پذیرتر کنیم. این الگوریتم ماتریس تبدیلی تخمین می‌زند که با اعمال بر بردارهای ویژگی، آن‌ها را به فضای جدیدی با قابلیت تفکیک‌پذیری بیشتر نگاشت می‌دهد[۱۲]. در این مقاله بازشناسی گفتار بر مبنای واج می‌باشد، لذا الگوریتم MCE نیز برای مجزاسازی واجها استفاده گردیده است. بنابراین در مدل جدید توزیع ویژگی‌های مربوط به واج‌ها از هم مجزاتر شده است، چنانچه المان‌های روی قطر اصلی ماتریس MCE به عنوان ضرایبی که در هر یک از ابعاد ویژگی‌ها ضرب می‌شوند بزرگتر از یک باشند طبق توضیح داده شده باعث می‌شود گوسی‌های مربوط به یک HMM از هم مجزاتر شوند.

در روش پیشنهادی ابتدا MCE به عنوان یک مرحله پیش پردازش قبل از خوشه‌بندی MLLR بکار می‌رود. این عمل باعث کلاس بندی بهتر پارامترهای مدل می‌گردد. انتظار می‌رود که با بکارگیری یک روش متمایزساز قبل از هر کلاس‌بندی کارائی آنرا بالا برد.

روش پیشنهادی را به همراه روشهای بدون MCE پیاده سازی شده و روی دادگان TIMIT و WSJ0 ارزیابی شده‌اند. در بخش بعد نتایج این آزمایشات آورده شده است.

## ۵- نتایج آزمایشات

به منظور بررسی عملکرد روش پیشنهادی، نرخ بازشناسی در سیستم بازشناسی غیروابسته با گوینده و سیستم تطبیق یافته با MLLR در دو حالت با MCE و بدون MCE مقایسه شده است. برای انجام آزمایشات از دادگان TIMIT و WSJ0 استفاده شده است.

در این بخش نتایج آزمایشات در سه قسمت بیان می‌شود. در ابتدا اثر MCE در خوشه‌بندی مخلوط‌های گوسی روی دادگان TIMIT بررسی می‌شود. سپس این اثر در نرخ بازشناسی روی دادگان TIMIT و بعد از آن روی دادگان WSJ0 گزارش می‌شود.

در ابتدا اثر MCE در خوشه بندی مخلوط‌های گوسی روی دادگان TIMIT بررسی می‌شود. به همین منظور سیستم بازشناسی غیروابسته به گوینده با ۳۹ مدل برای واجها آموزش داده شده است. در این سیستم هر مدل مخفی مارکوف دارای ۳ حالت و هر حالت دارای ۱۶ گوسی می‌باشد. همچنین خوشه بندی مخلوطهای گوسی را برای ۲، ۴، ۸ و ۱۶ کلاس انجام داده شد.

برای نمایش نتایج، روشها بفرم جدول (۱) نام‌گذاری شده‌اند.

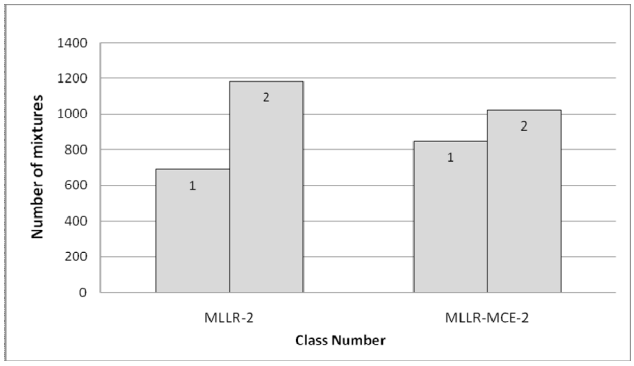
**جدول (۱) – نام‌گذاری روشها**

| نام روش | توضیح |
|---|---|
| SI | سیستم بازشناسی غیروابسته با گوینده بدون MCE |
| SI-MCE | سیستم بازشناسی غیروابسته با گوینده با MCE |
| MLLR-n | تطبیق گوینده سیستم SI بروش MLLR با n کلاس |
| MLLR-MCE-n | روش پیشنهادی - تطبیق گوینده سیستم SIMCE بروش MLLR با n کلاس |

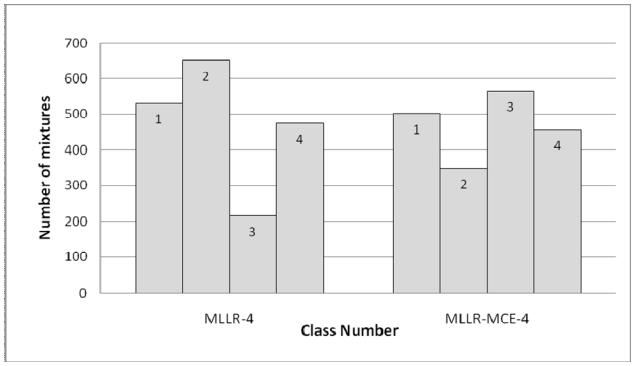
از آنجائی که تعداد مدل‌های در نظر گرفته شده برای واج‌های TIMIT، ۳۹ مدل ۳ حالته و هر حالت ۱۶ مخلوط گوسی است لذا تعداد کل مخلوطهای گوسی ۱۸۷۲ می‌باشد که تعداد مخلوط گوسی هر کلاس در شکل (۱) نشان داده شده است. در این شکل محور افقی شماره خوشه مخلوطهای گوسی و محور عمودی تعداد مخلوط گوسی در هر خوشه را بیان می کند. درضمن نمودار برای دو حالت MLLR بدون MCE (MLLR-n) و MLLR با MCE (MLLR-MCE-n) رسم شده است.

همان گونه که گفته شد مطلوب این است که توزیع مخلوطهای گوسی در کلاسها بطور یکنواخت انجام گیرد. چنانچه در شکل (۱) دیده میشود، روش پیشنهادی یعنی بکارگیری MCE قبل از MLLR باعث می‌شود که مخلوطهای گوسی بطور یکنواختی بین کلاسها توزیع شوند. چنانچه در این شکل دیده می شود با افزایش تعداد کلاسها عملکرد

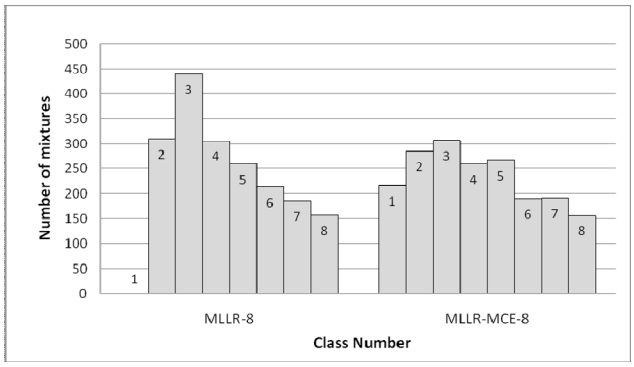
روش پیشنهادی بهتر می‌شود و این امر به این دلیل است که با افزایش تعداد کلاسها احتمال تداخل نمونه‌های کلاسها و پیچیدگی مساله بیشتر می‌شود بنابراین اثر MCE به عنوان متمایزساز بیشتر دیده می‌شود و توزیع مخلوط‌های گوسی بعد از اعمال MCE یکنواخت‌تر شده است. در ادامه نتایج بازشناسی واج برای روشهای مختلف روی دادگان TIMIT آورده شده است. در ادامه برای اینکه بتوان اثر روش پیشنهادی را بر روی تطبیق ۸-کلاسی و ۱۶-کلاسی دید، از دادگان WSJ0 برای آزمایش روی مدل آموزش داده شده استفاده شد. لازم به ذکر است برای هر گوینده در دادگان WSJ0 داده بیشتری وجود دارد.



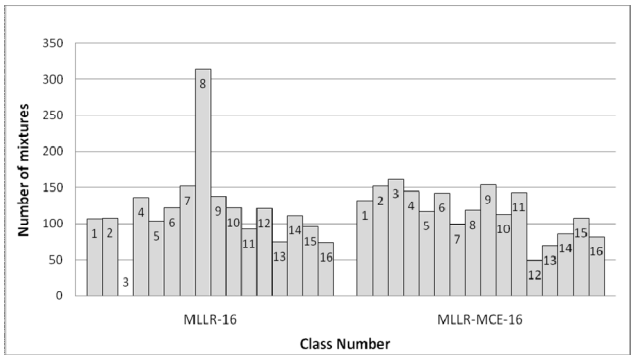

**(الف) ۲ کلاس مخلوط گوسی**




**(ب) ۴ کلاس مخلوط گوسی**




**(ج) ۸ کلاس مخلوط گوسی**




**(د) ۱۶ کلاس مخلوط گوسی**

**شکل (۱) – توزیع مخلوطهای گوسی در خوشه بندی MLLR برای دو روش با/بدون MCE**

سیستم بازشناسی با استفاده از دادگان بخش آموزش TIMIT آموزش داده شده است و برای تطبیق هر گوینده از ۱۰ جمله در بخش تست استفاده شده است. نتایج بازشناسی روی ۷۰ گوینده میانگین گرفته شده است. با توجه به اینکه دادگان تطبیق برای خوشه بندی MLLR به ۸ و ۱۶ کلاس کم می‌باشد، نتایج بازشناسی فقط برای خوشه‌بندی MLLR با ۲ و ۴ کلاس بیان شده است.

جدول (۲) نتایج نرخ بازشناسی واج را برای روشهای روی دادگان TIMIT نشان می‌دهد. با توجه به نتایج مندرج در جدول (۲) بکارگیری MCE در سیستم بازشناسی گفتار غیر وابسته به گوینده (SI-MCE) نسبت به سیستم بازشناسی گفتار غیر وابسته به گوینده (SI) باعث افزایش نرخ بازشناسی واج می شود. همچنین روش پیشنهادی یعنی بکارگیری MCE در MLLR باعث بهبود عملکرد سیستم می شود. همانطور که می توان دید با افزایش تعداد کلاس‌ها در MLLR عمل تطبیق بهتر انجام می‌گیرد. با توجه به نمودارهای شکل (۱)، دیده شد که با افزایش تعداد خوشه‌های مخلوط گوسی، MCE اثر بیشتری از خود نشان می‌دهد و مخلوطها را بطور یکنواخت‌تری بین کلاسها توزیع می‌کند. این امر را نیز در جدول (۲) می توان دنبال نمود، بگونه‌ای که نرخ بازشناسی واج در حالت ۲ کلاس حدود ۰.۴۹ درصد افزایش نسبی و در حالت ۴ کلاس حدود ۰.۶۲ درصد افزایش نسبی یافته است.

**جدول (۲) – نتایج نرخ بازشناسی واج روی دادگان TIMIT برای روشهای مختلف**

| روش | نرخ بازشناسی واج |
|---|---|
| SI | 80.77 |
| SI-MCE | 80.91 |
| MLLR-2 | 87.57 |
| MLLR-MCE-2 | **88.06** |
| MLLR-4 | 88.89 |
| MLLR-MCE-4 | **89.51** |

برای داشتن کلاسهای بیشتر در خوشه‌بندی MLLR نیاز به دادگان تطبیقی بیشتری برای هرگوینده می‌باشد. به همین دلیل

آزمایشات را بر روی دادگان WSJ0 نیز انجام داده که نتایج آن در ادامه آورده شده است.

در این آزمایش از مدل‌های آموزش داده شده با دادگان آموزشی TIMIT به عنوان مدل‌های مرجع استفاده شده است. سیستم بازشناسی گفتار غیر وابسته به گوینده دارای ۳۹ مدل برای واجها بوده که هر مدل دارای ۳ حالت و هر حالت دارای ۱۶ مخلوط گوسی می-باشد. نتایج مرحله تست بر اساس میانگین‌گیری روی ۲۰ گوینده از دادگان WSJ0 گزارش شده است، که برای هر گوینده ۴۰ جمله برای تطبیق و ۷۰ جمله برای تست استفاده شده است.

جدول (۳) نرخ بازشناسی واج روی دادگان WSJ0 برای روشهای مختلف نشان می‌دهد. چنانچه دیده می‌شود با افزایش تعداد کلاس در خوشه‌بندی MLLR نرخ بازشناسی واج افزایش می‌یابد. مطابق مقادیر نشان داده شده در جدول (۳) با بکارگیری MCE در روش MLLR در خوشه‌بندی ۲ کلاسی ۰.۱۲ درصد و با افزایش تعداد کلاسها به ۱۶ کلاس میزان بهبود به ۰.۷۴ درصد می‌رسد.

با افزایش تعداد کلاسهای خوشه بندی عملکرد روش MCE بهتر شده و تاثیر بیشتری در تطبیق گوینده و افزایش نرخ بازشناسی واج دارد.

**جدول (۳) - نتایج نرخ بازشناسی واج روی دادگان WSJ0 برای روشهای مختلف**

| روش | نرخ بازشناسی واج |
|---|---|
| SI | 45.29 |
| SI-MCE | 45.35 |
| MLLR-2 | 50.18 |
| MLLR-MCE-2 | **50.3** |
| MLLR-4 | 51.24 |
| MLLR-MCE-4 | **51.6** |
| MLLR-8 | 52.89 |
| MLLR-MCE-8 | **53.5** |
| MLLR-8 | 55.02 |
| MLLR-MCE-8 | **55.76** |

همانطور که نتایج آزمایشات انجام گرفته روی دادگان TIMIT و WSJ0 نشان دادند، روش پیشنهادی باعث افزایش نرخ بازشناسی در سیستم تطبیق گوینده شده است.

## ۶- نتیجه گیری

در سالهای اخیر تطبیق گوینده در سیستم‌های بازشناسی گفتار مستقل از گوینده مورد توجه محققان و پژوهشگران بوده است. روشهای متعددی در زمینه تطبیق ارائه شده‌اند که MLLR یکی از متداول‌ترین روشهای تطبیق گوینده است. این روش در دادگان تطبیقی کم با تخمین تابع تبدیل برای کلاس‌های مجزا از اجزای گوسی عملکرد نسبتاً خوبی دارد. تجمع گوسی‌ها در یک کلاس باعث نزدیک شدن تطبیق چند کلاسی به تطبیق تک کلاسی می‌شود که این امر می‌تواند به کاهش دقت منجر گردد.

روش پیشنهادی به این صورت است که از روش متمایز ساز خطای کمینه کلاس‌بندی برای تفکیک ویژگی‌های سیگنال گفتار بر اساس HMMهای مدل مرجع استفاده می‌شود. این امر ضعف روش MLLR در خوشه بندی را برطرف کرده و موجب خوشه‌بندی یکنواخت می-گردد. روش پیشنهادی با روش‌های معادل آن یعنی سیستم تطبیق MLLR بدون MCE بر روی دادگان TIMIT و WSJ0 مقایسه و ارزیابی شده‌اند. نتایج بدست آمده از آزمایشات نشان می‌دهد که روش پیشنهادی یعنی بکارگیری خطای کمینه کلاس‌بندی نسبت به روشهای معادل بدون خطای کمینه کلاس‌بندی باعث افزایش ۰.۴۹ تا ۰.۶۲ درصد در نرخ بازشناسی واج در TIMIT شده است. همچنین در دادگان WSJ0 نیز باعث افزایش ۰.۱۲ تا ۰.۷۴ درصدی در نرخ بازشناسی گفتار شده است.

با توجه به اینکه خطای کمینه کلاس‌بندی یکی از انواع روشهای متمایزساز می‌باشد، لذا بررسی بکارگیری روشهای دیگر متمایزساز می-تواند به عنوان ادامه کار پیشنهاد گردد.

## مراجع


[1] C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech & Language*, vol. 9, pp. 171–185, 1995.

[2] M. J. F. Gales, "The generation and use of regression class trees for MLLR adaptation," Cambridge Univ., Cambridge, U.K., Tech. Rep. CUED/F-INFENG/TR263, 1996.

[3] M. Gales, "Maximum likelihood linear transformations for HMM-based speech recognition". *Computer Speech & Language* 12, 75–98, 1998.

[4] T. Anastasakos, J. McDonough, J., Makhoul, "Speaker adaptive training: a maximum likelihood approach to speaker normalization". *In: Proc. of ICASSP*, vol. 2, pp. 1043–1046, 1997.

[5] A. Stolcke, L. Ferrer, S. Kajarekar, E. Shriberg, A. Venkataraman, "MLLR transforms as features in speaker recognition". *In:Proc. of Eurospeech*, pp. 2425–2428, 2005.

[6] C.-H. Lin, C.-H. Wu, and P.-C. Chang, "A study on speaker adaptation for Mandarin syllable recognition with minimum error discriminative training", IEICE Transaction Information & System, Vol. E78-D, No.6, pp.712-718, 1995.

[7] K. Laurila, M. Vasilache and O.Viikki, "A combination of discriminative and maximum likelihood techniques for noise robust speech recognition", *In: Proc. ICASSP*, 1998, pp.85-88.

[8] J. Wu, Q. Huo, "Supervised adaptation of MCE-trained CDHMMs using minimum classification error linear regression", *In: Proc. of ICASSP*, vol. 1, pp. 605–608, 2002.

[9] B. Zhang, and S., Matsoukas, Minimum phoneme error based heteroscedastic linear discriminant analysis for

[12] J. Hung and L. S. Lee, Data-driven temporal filters for robust features in speech recognition obtained via minimum classification error (MCE). *In: Proc. of ICASSP*, pp. 373-376, 2002.

[13] X. Wang and K. K., Paliwal, Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition*, vol. 36, pp. 2429-2439, 2003.

[14] Jonathan E. Hamaker, MLLR: A Speaker Adaptation Technique for LVCSR, Lecture for a course at ISIP - Institute for Signal and Information Processing, Department of Electrical and Computer Engineering, Mississippi State University, 1999.

speech recognition. *In: Proc. of ICASSP*, vol. 1, pp. 925-928, 2005.

[10] B. H., Juang, W. Chou and C.H., Lee, 1997. Minimum classification error rate methods for speech recognition. *IEEE Transaction on Speech and Audio Processing*, Vol. 5, No. 3, pp. 257-265, 1997.

[11] M., Loog, R. P.W., Duin, Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 732-739, 2004.