

مقاوم سازی سیستمهای بازشناسی چندباندی گفتار با استفاده از اصلاح آنتروپی مدل مخفی مارکف و تصویر وزن دار زیر باندی

احمد اکبری

بابک ناصرشریف

دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران - ایران

چکیده

مدل موسوم به تصویر وزن دار بهبود می یابد. با این روش، کارایی شیوه های وزن دهی برای ترکیب احتمالات در حضور انواع مختلف نویز، افزایش می یابد. در سیستم ترکیب ویژگیها، نیز تعمیمی از شیوه تصویر وزن دار، موسوم به تصویر وزن دار زیرباندی، برای بردارهای ویژگی مرکب از ویژگیهای زیرباندها ارائه می گردد. آزمایشها نشان می دهند که دقت بازشناسی سیستم چند باندی نسبت به سیستم تمام باند به ویژه در نسبتهای سیگنال به نویز متوسط و پایین بالاتر است. به علاوه، نتایج بیانگر آن است که دو شیوه اصلاح آنتروپی مدل مخفی مارکف و تصویر وزن دار زیرباندی سبب بهبود دقت بازشناسی سیستم چند باندی در حضور انواع مختلف نویز و در تمامی نسبتهای سیگنال به نویز می گردد. **واژه های کلیدی:** بازشناسی، مدل منحنی مارکف و شیوه های وزن دهی

سیستمهای بازشناسی چندباندی گفتار که بر اساس مکانیزم شنوایی انسان عمل می کنند، نرخ بازشناسی گفتار را در حضور نویزهای باند محدود نسبت به سیستم تمام باند بهبود می دهند. در بازشناسی چندباندی گفتار، سیگنال گفتار ابتدا به چند زیرباند فرکانسی تقسیم می شود. پس از استخراج بردارهای ویژگی از هر زیرباند، به دو شیوه می توان عمل کرد. در یک روش این بردارهای ویژگی با یکدیگر ترکیب می شوند. در روش دیگر احتمالات تولید شده توسط سیستمهای بازشناس متناظر با هر زیرباند برای این بردارها، با یکدیگر ترکیب می گردند. در مقاله حاضر، روشهایی برای برای بالا بردن دقت بازشناسی هر یک از دو نوع سیستم بازشناسی چند باندی گفتار (ترکیب احتمالات و ترکیب ویژگیها) در حضور نویز مورد استفاده قرار می گیرد. در سیستم ترکیب احتمالات، آنتروپی مدل مخفی مارکف به کمک یک روش مبتنی بر

Robust Multi-Band Speech Recognition Using Improved HMM Entropy and Sub-Band Weighted Projection Measure

Babak Nasersharif

Ahmad Akbari

Computer Engineering Department Iran University of Science and Technology

Abstract: Multi-band speech recognition systems, based on human hearing mechanism, improve speech recognition rate in presence of band-limited noises. In multi-band speech recognition, full band speech signal is divided into several frequency sub-bands. Then, a feature vector is extracted from each sub-band. There are two solutions for using these sub-band feature vectors.

The first solution, called feature combination, is to concatenate them and obtain a larger feature vector. In the second solution, called probability combination, separate sub-band recognizers process sub-band feature vectors independently and generate a probability estimate for them. These probability estimates are recombined in order to obtain final recognition result. In this paper, a

few methods are proposed to increase the correct recognition rate of above mentioned multi-band speech recognition systems in presence of background noise. In probability combination system, HMM entropy is improved using a model based method named as weighted projection measure (WPM). Results show that applying WPM to sub-band feature vectors increases the performance of different used weighting criteria for probability combination in presence of different noise types. In feature

combination system, a generalized form of WPM, called sub-band WPM, is applied to concatenated feature vector. Results show that multi-band speech recognition rate is higher than full-band system in low and medium SNRs. Furthermore, they indicate that improved HMM entropy and sub-band WPM increase multi-band system recognition rate in presence of different noise types with different SNRs.

۱- مقدمه

روش‌هایی هستند. در [۷] مروری کامل در این زمینه انجام شده است و روش‌های مختلف با هم مقایسه شده‌اند. در طریق دیگر، ویژگی‌های جدیدی استخراج می‌شوند که نسبت به تاثیرات نویز مقاوم باشند، همانند ویژگی‌های خودهمبستگی فاز^۵ [۵] یا ویژگی‌های کپسترال که از مشتق طیف استخراج می‌شوند [۸]. روش‌های مبتنی بر مدل، مدل آماری محیط را به نحوی اصلاح می‌کنند که با شرایط جدید محیطی همانند شرایط نویزی تطبیق یابد. در این تطبیق هیچ نوع فرض یا دانش خاصی درباره خود سیگنال گفتار لازم نیست. این روش‌ها معمولاً نیازمند آموزش برون خط^۶ بر روی دادگان گفتار نویزی هستند. به عنوان نمونه ای از این روش‌ها می‌توان به ترکیب موازی مدلها^۷ (PMC) [۱۹]، ترکیب موازی مدلها به صورت انتخابی^۸ (OPMC) [۶] و بازگشت خطی با بیشترین درستمایی^۹ (MLLR) اشاره کرد [۲۰،۹].

در بازشناسی چند باندهای گفتار، سیگنال گفتار به چندین باند فرکانسی تقسیم می‌شود و هر باند فرکانسی به صورت مستقل پردازش می‌شود. این روش بر اساس فرض استقلال زیرباندها از یکدیگر استوار است و بنابراین باندها باید به شیوه ای انتخاب شوند که همپوشانی میان زیرباندهای فرکانسی کمینه باشد. پس از انتخاب زیرباندها، یک بردار ویژگی از هر زیرباند استخراج می‌شود که زیربردار ویژگی نامیده می‌شود. پس از متعامد سازی زیربردارهای ویژگی و انجام دیگر پیش پردازشهای لازم در روش می‌تواند مد نظر قرار گیرد. در روش اول این زیربردارها در کنار یکدیگر

مسئله مقاوم سازی سیستمهای بازشناسی گفتار در برابر نویز می‌تواند به عنوان مسئله کاهش میزان عدم تطبیق میان شرایط آموزش و آزمون سیستم مطرح شود. روش‌هایی که برای کاهش این عدم تطبیق بکار می‌روند، به سه گروه اصلی تقسیم می‌شوند: روش‌های مبتنی بر داده، روش‌های مبتنی بر مدل و شیوه‌های پردازش چند باندهای^۱. روش‌های مبتنی بر داده از تاثیرات نویز بر سیگنال گفتار یا ویژگی‌های آن می‌کاهند. در مقابل روش‌های مبتنی بر مدل بجای خود سیگنال گفتار یا ویژگی‌های آن مدل آکوستیک گفتار را اصلاح می‌نمایند. شیوه پردازش چندباندی نیز معمولاً در مورد نویزهایی بکار گرفته می‌شود که سبب تخریب بخشی از طیف فرکانسی سیگنال گفتار می‌گردند.

روش‌های مبتنی بر داده به دو گروه عمده تقسیم می‌شوند: روش‌های مبتنی بر بهبود گفتار^۲ و روش‌های جبران ویژگی^۳. شیوه‌های بهبود گفتار مستقیماً با سیگنال نویزی گفتار سروکار دارند و با تخمین سیگنال تمیز از سیگنال نویزی درجهت کاهش میزان عدم تطبیق بکار می‌روند. روش تفاضل طیف و آستانه گذاری ضرایب تبدیل موجک^۴ [۲۲] نمونه‌هایی از این دسته هستند. روش‌های جبران ویژگی معمولاً عدم تطبیق را به دو طریق کاهش می‌دهند. در طریق اول، یک تبدیل به ویژگی‌ها اعمال می‌شود تا اثر نویز از آنها حذف گردد. تفاضل میانگین ضرایب کپسترال (CMS) [۱۲] و RASTA PLP [۲۵] از جمله چنین

1- Multi-band processing

2- Speech enhancement

3- Feature compensation

4- Wavelet Transform

5- Phase autocorrelation features

6- Offline

7- Parallel model combination

8- Optional parallel model combination

9- Maximum likelihood linear regression



می تواند کلمات مختلف را تشخیص دهد و لذا بازشناسی توسط سیستم شنوایی عملکرد مقاوم و خوبی از خود نشان می دهد. توجه به این نحوه عملکرد و مدلسازی آن، می تواند در مقاوم سازی بازشناسی گفتار موثر باشد. از جمله این شیوه ها می توان به مدلسازی نحوه پوشش نویز در گوش و بازشناسی گفتار در باندهای فرکانسی مختلف اشاره کرد. کار تحقیقاتی فلچر^{۱۴} و دانشجویانش نشان داده است که مکانیزم شنوایی انسان پیغام بیان شده به یک زبان خاص را، در چندین زیرباند فرکانسی پردازش و رمزگشایی می کند و بازشناسی نهایی را بر اساس ادغام اطلاعات بدست آمده از این زیرباندها انجام می دهد. به این ترتیب خطای بازشناسی گفتار حاصل ضرب خطاها در هریک از زیرباندها است و طبعاً از هر یک از این خطاها کوچکتر است. این مطلب توسط فلچر برای دو زیرباند نشان داده شده است و تعمیم آن به بیش از دو زیر باند توسط آلن^{۱۵} (۱۹۹۴) بیان گردیده است و تحت عنوان اصل فلچر-آلن یا قانون ضرب خطاها^{۱۶} (POE) شناخته می شود [۲۴]. این قانون اساس و انگیزه اصلی گسترش شیوه بازشناسی چند باندهای گفتار است. علاوه بر اینکه اخیراً نیز استدلالهای ریاضی بیشتری برای اثبات کارایی شیوه های بازشناسی چند باندهای ارائه شده است [۲]. چنانکه گفته شد، روش های بازشناسی چندباندی گفتار به دو دسته کلی تقسیم می شوند: ترکیب ویژگی ها و ترکیب احتمالات. ترکیب ویژگی زیرباندها از طریق قرار دادن زیربردارهای ویژگی در کنار یکدیگر، ابتدا توسط Okawa در سال ۱۹۹۸ پیشنهاد گردید [۱۶]. الحاق زیربردارهای ویژگی به یکدیگر، یک بردار ویژگی را ایجاد می کند که می توان آن را با شیوه های استاندارد پردازش تمام باند مدل کرد. به این ترتیب همبستگی ممکن میان زیربردارهای ویژگی در مدل آکوستیک در نظر گرفته می شود که معمولاً مدل را نیرومندتر و قابل اعتمادتر می کند [۱۶]. روند کلی شیوه ترکیب ویژگی ها در شکل (۱)-الف نشان داده شده است. اگر چه این شیوه ترکیب ساده است، لیکن امکان وزن دهی به زیرباندها بر اساس قابلیت اعتماد و میزان اطلاعات آنها را دارا نیست. این امر نقطه ضعف این شیوه محسوب می گردد.

قرار داده می شوند و به عنوان جایگزینی برای ویژگی های اصلی استفاده می گردند. به این شیوه ترکیب ویژگی ها^{۱۰} اطلاق می شود. در روش دیگر زیربردارهای ویژگی بوسیله بازشناس مجزای متناظر با هر زیرباند، مورد پردازش قرار می گیرند و احتمالی برای آنها تخمین زده می شود و این احتمالات به شیوه خطی یا غیر خطی با یکدیگر ترکیب می شوند. این شیوه ترکیب احتمالات^{۱۱} یا ترکیب مدلها نامیده می شود. هنگامی که بازشناس مورد استفاده در زیرباندها مدل مخفی مارکف باشد، این روش معمولاً تحت عنوان ترکیب درستنماییها^{۱۲} شناخته می گردد. مزیت مشترک هر دوی این روشها، این است که پردازش جداگانه بر روی زیربردارهای ویژگی سبب می شود تاثیر نویز از یک مجموعه ویژگی تخریب شده به دیگری سرایت نکند.

در کار حاضر، ترکیبی از روش های مبتنی بر مدل و بازشناسی چند باندهای گفتار ارائه شده است. در سیستم ترکیب درستنماییها از یک روش مبتنی بر مدل موسوم به معیار تصویروزن دار^{۱۳} (WPM)، برای اصلاح آنتروپی مدل مخفی مارکف استفاده گردیده است. این روش کارایی شیوه های مختلف وزن دهی را بهبود می بخشد. به علاوه، در مقاله حاضر تعمیمی از تصویر وزن دار موسوم به «تصویر وزن دار زیرباندی» برای سیستم ترکیب ویژگی ها ارائه گردیده است که بر بردار ویژگی مرکب از زیربردارهای ویژگی اعمال می گردد.

در ادامه مقاله، بخش دوم به بررسی اصول بازشناسی چندباندی گفتار و ترکیب ویژگی ها می پردازد. بخش سوم به بررسی شیوه های وزن دهی برای ترکیب درستنماییها می پردازد. بخش چهارم روش تصویر وزن دار را بررسی می کند. بخش پنجم به شرح انگیزه ها و اصول روش پیشنهادی اختصاص دارد. در بخش ششم نتایج آزمایشها ذکر می شود. بخش هفتم نیز حاوی جمع بندی و نتیجه گیری کلی است.

۲- بازشناسی چندباندی گفتار

در محیطهای دارای سر و صدای اضافی و نویز زمینه، انسان به خوبی

10- Feature Combination

11- Probability combination

12- Likelihood Combination

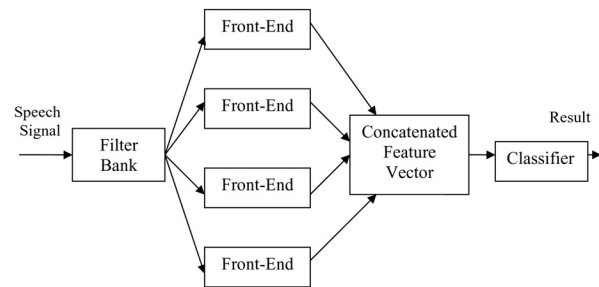
13- Weighted Projection Measure

14- Fletcher

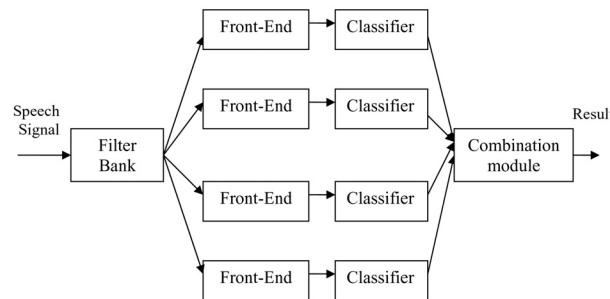
15- Allen

16- Product of Error Rule

در روش ترکیب احتمالات با هر زیرباند فرکانسی همانند یک منبع مجزای اطلاعاتی رفتار می شود. پس از عملیات پیش پردازش و استخراج زیربردارهای ویژگی از هر زیر باند، خروجیهای احتمال طبقه بندی کننده های^{۱۷} مربوط به هر زیرباند، با یکدیگر ترکیب می شوند. این ترکیب می تواند در سطح واج، در سطح هجا، در سطح کلمه یا در سطح جمله انجام شود. روند کلی شیوه ترکیب احتمالات در شکل (۱) -ب نمایش داده شده است. نحوه ترکیب خروجی طبقه بندی کننده های هر زیرباند، نقش موثری در کارایی یک سیستم بازشناسی چند باندی گفتار دارد. نحوه ترکیب بسته به نوع طبقه بندی کننده دارد. برخی طبقه بندی کننده ها همانند HMM و GMM در دستمای^{۱۸} را تخمین می زنند



(الف) سیستم ترکیب ویژگی ها



(ب) سیستم ترکیب احتمالات

شکل (۱) - معماری عمومی سیستم بازشناسی چند باندی گفتار

در حالیکه برخی دیگر احتمال پسین را در اختیار می گذارند (همانند سیستم مرکب HMM-ANN) [۱۱،۳]. در مجموع اکثر شیوه های ترکیب از یک تابع وزن دهی برای تعیین قابلیت اعتماد

نسبی و میزان اطلاعات موجود در هر زیر باند استفاده می نمایند تا وزنه های بیشتر به به زیرباند مورد اعتمادتر منتسب گردد. معیارهای شناخته شده وزن دهی در شیوه چند باندی، تخمین نسبت سیگنال به نویز در هر زیر باند [۱۱،۱۷،۱۸] و معکوس آنتروپی شرطی هر زیرباند [۱۷] هستند. استفاده از شبکه عصبی MLP به عنوان یک شیوه غیرخطی ترکیب احتمالات طبقه بندی کننده ها، نیز از دیگر شیوه های شناخته شده ترکیب است [۱۳،۱۸].

سیستم های چند باندی کلاسیک با محدودیت هایی مواجه هستند که سبب کاهش کارایی مورد انتظار آنها می گردند. از جمله کاهش داده های ورودی در نتیجه تقسیم به چند زیرباند که امکان ایجاد مدل های کوچکتر را فراهم می کند، گاه ممکن است سبب کمتر شدن نرخ بازشناسی در هر زیر باند شود. به عبارت دیگر از دست رفتن همبستگی میان زیرباندها سبب کاهش کارایی بازشناسی ها در هر زیرباند می شود [۱۱،۱۴]. روش های متعددی برای دخالت دادن همبستگی زیرباندها در سیستم بازشناسی چند باندی ارائه شده است که از آن جمله می توان به شیوه اجتماع احتمالات^{۱۹} [۴،۱۰] و شیوه ترکیب کامل^{۲۰} [۱۱،۳] اشاره نمود. مروری کامل بر شیوه های اخیر با فرض جبران عدم همبستگی بین زیرباندها در [۳] ارائه گردیده است. به منظور در نظر گرفتن همبستگی میان زیرباندها و جلوگیری از کاهش نرخ بازشناسی در هر زیرباند، روش های مبتنی بر مدل می توانند موثر واقع شوند. در کار حاضر نیز از این روش ها استفاده شده است.

۳- وزن دهی در سیستم ترکیب در دستمای ها

چنانکه گفته شد، در شیوه ترکیب احتمالات خروجیهای احتمال طبقه بندی کننده های مربوط به هر زیر باند، باید در سطحی از تقسیم بندی زمانی همانند سطح واج یا کلمه، با یکدیگر ترکیب شوند. این ترکیب می تواند به صورت خطی یا غیر خطی صورت گیرد. در کارهای پیشین نشان داده شده است که ترکیب خطی احتمالات بیش از ترکیب غیر خطی سبب مقاوم سازی سیستم بازشناسی چند باندی گفتار نسبت به نویز می شود [۶]. از این رو در کار حاضر، ترکیب خطی احتمالات انتخاب شده است. علاوه بر

17- Classifiers

18- Likelihood

19- Probabilistic union models

20- Full Combination

اینکه از مدل مخفی مارکف با چگالی پیوسته (CDHMM) به عنوان طبقه بندی کننده استفاده شده است. در این حالت احتمالات تولیدی بوسیله CDHMM درستنمایی نامیده می شوند. رابطه ترکیب خطی برای درستنمایی های تولید شده بوسیله مدل مخفی مارکف (HMM) به صورت زیر است:

$$S(x, M) = \sum_{b=1}^B \alpha_{b,M} P(x | M, b) \quad (1)$$

در این رابطه S بیانگر امتیاز گویش x بوسیله مدل M است. درستنمایی تولید شده بوسیله HMM متناظر با مدل M در زیر باند b است و B تعداد کل زیرباند هاست. $\alpha_{b,M}$ نیز ضریب وزن دهی است.

یک مسئله مهم در ترکیب درستنمایی ها تخمین ضریب $\alpha_{b,M}$ است. چنانکه گفتیم، این ضریب می تواند بر حسب قابلیت اعتماد نسبی و میزان اطلاعات موجود در هر زیر باند تعیین گردد. تخمین نسبت سیگنال به نویز در هر زیر باند [۱۸، ۱۷، ۱۱] و معکوس آنتروپی شرطی هر زیرباند [۱۶] از جمله مواردی هستند که برای تعیین قابلیت اعتماد هر زیر باند و به عنوان ضریب $\alpha_{b,M}$ قابل استفاده هستند. لازم به ذکر است که معمولاً نرمالیزاسیون برای ضرایب $\alpha_{b,M}$ انجام می شود به طوری که شرط $\sum_{b=1}^B \alpha_{b,M} = 1$ رعایت گردد. ما در کار حاضر یک معیار وزن دهی دیگر موسوم به آنتروپی مدل مخفی مارکف را بکار گرفته ایم که می تواند به عنوان ضریب $\alpha_{b,M}$ بکار رود و در ادامه معرفی می شود.

۱-۳- وزن دهی بر اساس آنتروپی مدل مخفی مارکف

توزیع درستنمایی تخمینی بوسیله مدل مخفی مارکف، حاوی اطلاعاتی درباره قابلیت اعتماد بردار مشاهدات ورودی مدل مخفی مارکف است. اگر مدل مخفی مارکف متناظر با یک کلاس دارای درستنمایی بالایی برای یک مشاهده باشد و سایر مدل های مخفی مارکف متناظر با سایر کلاسها درستنمایی پایینی برای آن مشاهده تولید کنند، این مطلب بیانگر قابلیت اعتماد آن مشاهده است. در حالیکه اگر مقدار درستنماییهای تخمینی همه مدل های مخفی

مارکف برای مشاهده ورودی به هم نزدیک باشد، مشاهده ورودی کمتر قابل اعتماد است. این اطلاعات را می توان با استفاده از محاسبه آنتروپی برای یک بردار مشاهده بوسیله تمامی مدل های مخفی مارکف به صورت زیر محاسبه کرد:

$$H(O) = - \sum_{i=1}^K P(O | M_i) \log_2 P(O | M_i) \quad (2)$$

در این رابطه O بردار مشاهدات، H(O) آنتروپی بردار مشاهدات بر روی مجموعه مدل های مخفی مارکف متناظر با کلاسها، $P(O | M_i)$ درستنمایی تخمینی برای مشاهده O بوسیله i-امین مدل مخفی مارکف و K تعداد کل کلاسها یا به عبارتی تعداد کل مدل های مخفی مارکف است. قابل ذکر است که بسته به اینکه بردار مشاهده O در سطح قاب یا سطح کلمه در نظر گرفته شود، محاسبه آنتروپی می تواند در سطح قاب یا کلمه انجام شود.

معیاری مبتنی بر آنتروپی در [۱۵] معرفی گردیده است که در سطح قاب و حالات HMM ابتدا درستنمایی را تبدیل به احتمال پسین می کند و سپس با استفاده از این احتمال پسین آنتروپی در سطح قاب و کلمه و سپس اطلاعات متقابل را در سطح حالات محاسبه می نماید و از اطلاعات متقابل ۲۱ موجود در مجموعه حالات نسبت به بردار مشاهده، به صورت یک ضریب به عنوان توان درستنمایی استفاده می کند. حال آنکه آنتروپی معرفی شده در رابطه (۲) در سطح کلمه و مدل با انگیزه ای کاملاً متفاوت تعریف می شود و در وزن دهی خطی مورد استفاده قرار می گیرد. این آنتروپی با احتمالات تولید شده در سطح حالات HMM مرتبط نیست و تنها احتمال نهایی تولید را مد نظر قرار می دهد که به این ترتیب معیاری برای نشان دادن وضعیت عملکرد HMM به هنگام بازشناسی نهایی است.

با توجه به آنکه آنتروپی فوق بیانگر قابلیت اعتماد یک مشاهده است، با محاسبه این آنتروپی برای هر زیر باند می توان رابطه ای میان آن و ضریب $\alpha_{b,M}$ در رابطه (۱) یافت. با توجه به آنچه گفته شد، مقادیر بزرگ آنتروپی بیانگر قابل اعتماد نبودن مشاهده و مقادیر کوچکتر آنتروپی بیانگر قابلیت اعتماد بیشتر مشاهدات خواهد بود. با توجه به این مطلب، می توان از عکس آنتروپی مشاهده در هر زیرباند، به عنوان $\alpha_{b,M}$ و به عبارتی وزن احتمال آن

مشاهده در آن زیرباند استفاده نمود. بنابراین با بازنویسی رابطه (۲) در زیرباند b به صورت زیر:

$$H_b(O) = -\sum_{i=1}^K P(O|M_i, b) \log_2 P(O|M_i, b) \quad (3)$$

را $\alpha_{b,M}$ (با انجام نرمالیزاسیون) می توان اینگونه محاسبه نمود:

$$\alpha_{b,M} = \frac{1}{H_b(O)} \quad (4)$$

۴- معیار تصویر وزن دار

تئوری معیار تصویر وزن دار (WPM) بر اساس این مشاهده Mansour و Juang استوار است که اندازه بردارهای ویژگی کپسترال در حضور نویز جمع پذیر سفید کاهش می یابد [۲۶]. طبق این خصوصیت در [۲۶] یک معیار محاسباتی بر پایه عمل تصویر کردن معرفی شد که کارایی بازشناسی گفتار بوسیله روش DTW را در حضور نویز به طرز قابل توجهی بهبود بخشید. Carlson و Celement این معیار را گسترش دادند و آن را در یک سیستم بازشناسی مبتنی بر مدل مخفی مارکف با چگالی پیوسته (CDHMM) بکار گرفتند [۲۳]. آنها یک عامل مقیاس را در توزیع احتمالات مشاهده در حالات CDHMM و به عبارت دیگر در توزیع گاوسی در دستنمایها دخالت دادند تا به این وسیله کاهش اندازه بردار کپسترال جبران گردد. ویژگی های مورد استفاده آنان ضرایب مل کپستروم (MFCC) بود. نتایج کار آنان نشان داد که بکار بردن این معیار تصویر وزن دار در یک سیستم وابسته به گوینده، نرخ بازشناسی کلمات مجزا را به طرز قابل توجهی در حضور انواع مختلف نویز از جمله نویز سفید و نویز رنگی بهبود بخشیده است [۲۳]. رابطه جبران تعریف شده آنها برای توزیع گاوسی را می توان به صورت زیر بیان کرد:

$$b_{j,i}(c_t) = \hat{I}(c_t, \lambda_{j,i,t}, \mu_{j,i}, C_{j,i}) = \frac{\exp\left(-\frac{1}{2}(c_t - \lambda_{j,i,t} \mu_{j,i})^T C_{j,i}^{-1} (c_t - \lambda_{j,i,t} \mu_{j,i})\right)}{(2\pi)^{\frac{n}{2}} |C_{j,i}|^{\frac{1}{2}}} \quad (5)$$

که پارامترهای موجود در آن این گونه تعریف می شوند:

C_t : بردار مشاهده کپسترال برای قاب t

n : بعد بردار مشاهده

$\mu_{j,i}$: بردار میانگین عنصر j ام مخلوط گاوسی در حالت i

$C_{j,i}$: ماتریس کوواریانس عنصر j ام مخلوط گاوسی در حالت i
عامل مقیاس برای قاب t در عنصر j ام مخلوط گاوسی در

حالت i

$b_{j,i}(c_t)$: احتمال تولید شده برای بردار مشاهده C_t توسط عنصر

j ام مخلوط گاوسی در حالت i

با گرفتن لگاریتم از رابطه (۵) و محاسبه لگاریتم درستنمایی، می

توان به رابطه ای برای الگوریتم ویتربی دست یافت که بیانگر معیار

تطبیق میان بردار مشاهده و بردارهای میانگین مخلوطهای

گاوسی است. این رابطه را می توان اینگونه نوشت:

$$\log b_{j,i}(c_t) = (c_t - \lambda_{j,i,t} \mu_{j,i})^T C_{j,i}^{-1} (c_t - \lambda_{j,i,t} \mu_{j,i}) + \log |C_{j,i}| + N \log(2\pi) \quad (6)$$

با گرفتن مشتق از رابطه (۶) نسبت به $\lambda_{j,i,t}$ می توان مقدار بهینه

ای برای بدست آورد که به واسطه آن احتمال مشاهده $b_{j,i}(c_t)$

بیشینه شود. این مقدار بهینه به صورت زیر بدست می آید:

$$\lambda_{j,i,t} = \frac{c_t^T C_{j,i}^{-1} \mu_{j,i}}{\mu_{j,i}^T C_{j,i}^{-1} \mu_{j,i}} \quad (7)$$

با جایگذاری این مقدار برای $\lambda_{j,i,t}$ در رابطه (۶)، مقداری برای

لگاریتم درستنمایی بدست می آید که WPM نامیده می شود.

رابطه (۶) و (۷) برای تمامی مخلوطهای گاوسی موجود در یک

حالت مدل مخفی مارکف بکار می روند و درستنمایی نهایی با

استفاده از مخلوطهای گاوسی تطبیق یافته با معیار تصویر وزن دار

محاسبه می شود.

۴-۱- مقایسه WPM و MLLR

در معیار تصویر وزن دار، چنانچه در روابط (۵) و (۶) مشاهده می

شود، میانگین مخلوطهای گاوسی در عامل مقیاس $\lambda_{j,i,t}$ ضرب

می شوند. این امر را می توان با روش MLLR پایه مقایسه کرد. در

روش MLLR پایه، میانگین مخلوطهای گاوسی تحت تاثیر تبدیلی

همانند زیر قرار می گیرند [۲۰]:

$$\hat{\mu} = A \mu + b \quad (8)$$

که در آن A یک ماتریس $n \times n$ و b یک بردار n بعدی است که n

بعد بردار مشاهدات است. ماتریس تبدیل A و بردار b با استفاده از

الگوریتم EM به تحوی تخمین زده می شوند که احتمال مشاهده داده های تطبیقی بیشینه شود. می توان WPM را فرم ساده ای از رابطه (۸) در نظر گرفت که در آن $A = \lambda I, b = 0$ که I ماتریس واحد $n \times n$ و λ یک مقدار عددی است. اما باید به این تفاوت نیز توجه کرد که در MLLR ماتریس A و بردار b با استفاده از داده های تطبیقی و در مرحله آموزش تعیین می شوند. لیکن در WPM ماتریس A و به عبارتی λ با استفاده از بردار مشاهده فعلی و در مرحله آزمایش تعیین می شوند که این امر تطبیق پذیری بیشتری با داده های ورودی را موجب می شود. علاوه بر اینکه محاسبات لازم در مرحله تطبیق نیز در WPM از پیچیدگی کمتری نسبت به MLLR برخوردار خواهد بود.

۵- روش های پیشنهادی

در کار حاضر، دوروش جداگانه برای بهبود هر یک از دو نوع سیستم پردازش چند بانندی یعنی سیستمهای ترکیب درستنمایی ها و سیستم ترکیب ویژگی ها پیشنهاد شده است. در سیستم ترکیب درستنمایی ها، اصلاح آنتروپی مدل مخفی مارکف در زیرباندها به وسیله تصویر وزن دار پیشنهاد شده است تا عملکرد شیوه های وزن دهی و به ویژه ضریب وزن دهی بر اساس عکس آنتروپی مدل مخفی مارکف در زیرباندها (رابطه ۴) بهبود یابد. در سیستم ترکیب ویژگی ها، تعمیمی از روش تصویر وزن دار برای اعمال بر بردارهای ویژگی تشکیل شده از زیربردارهای ویژگی استخراجی از زیرباندها ارائه شده است. ما این روش تعمیم یافته را تصویر وزن دار زیربانندی می نامیم. این دو روش در ادامه همین بخش به طرز دقیق تر مورد بررسی قرار می گیرند.

۵-۱- اصلاح آنتروپی مدل مخفی مارکف

با بکار گرفتن WPM در CDHMM برای محیط های نویزی، درستنمایی تولید شده بوسیله CDHMM برای یک بردار مشاهده نویزی، بیشینه می شود. این امر سبب بهبود قابل توجه نتایج بازشناسی در سیستم تمام باند می گردد. این بهبود نتیجه منطقی افزایش تفاوت میان درستنمایی تولیدی تمامی مدل های مخفی مارکف دخیل در بازشناسی برای بردار مشاهده نویزی است.

افزایش این میزان تفاوت، سبب کاهش میزان آنتروپی مدل مخفی مارکف تعریف شده در رابطه (۲) برای بردار مشاهده نویزی می گردد. بنابراین می توان گفت که بکارگیری WPM سبب کاهش آنتروپی مدل مخفی مارکف و در نتیجه عملکرد بهتر بازشناسی در محیط نویزی می گردد. بنابراین هنگامی که ما WPM را برای CDHMM در زیرباند b بکار ببریم، مقدار H_b در رابطه (۳) کاهش خواهد یافت. این امر نشانگر قابل اعتمادتر بودن نتایج بازشناسی در زیرباند b است. بنابراین ضریب وزن دهی $\alpha_{b,M}$ برای این زیرباند باید افزایش یابد که این امر با توجه به تعریف ارائه شده در رابطه (۴) برای $\alpha_{b,M}$ ، صورت خواهد گرفت.

۵-۲- تصویر وزن دار زیربانندی

چنانکه گفته شد، در شیوه ترکیب ویژگی ها پس از استخراج زیربردارهای ویژگی از هر زیرباند و کنار هم قرار دادن آنها، یک بردار ویژگی تشکیل می شود که در مورد سیگنال تمیز مدل با استفاده از آن آموزش می یابد و در مورد سیگنال نویزی بردار مشاهده ورودی برای CDHMM محسوب می گردد. روش تصویر وزن دار باید برای تطبیق این بردار ویژگی مرکب از زیربردارهای ویژگی با میانگین توزیعهای گاوسی CDHMM بکار رود. با توجه به اینکه بردار ویژگی در اینجا برداری مرکب از زیربردارهای ویژگی است، بکار بردن WPM در مورد آن نیازمند تغییراتی است که ما در اینجا این تغییرات را تحت عنوان تصویر وزن دار زیر بانندی (SWPM) مطرح می کنیم. چنانچه در بخش (۴-۱) گفته شد، می توان WPM را فرم ساده ای از رابطه (۸) در نظر گرفت که در آن $A = \lambda I, b = 0$. ما ماتریس A را با یک ماتریس قطری جایگزین می کنیم. به این ترتیب، برای اجزای مختلف بردار مشاهده مقادیر متفاوتی از λ قابل تعریف است، در حالیکه در WPM از یک مقدار یکسان برای تمامی المانهای بردار مشاهده استفاده می شود. A را به صورت زیر تعریف می کنیم تا آن را جایگزین $\lambda_{j,i,t}$ در رابطه (۶) نماییم:

با وارد کردن رابطه (۱۲) در رابطه (۱۰) و مشتق گیری از رابطه (۱۰) نسبت به λ_r ($r=1, 2, \dots, R$) و بازنویسی رابطه به فرم ماتریسی، λ_r می تواند به صورت زیر محاسبه شود:

$$\lambda_r = \frac{c_{r,t}'^T C_r'^{-1} \mu_r'}{\mu_r'^T C_r'^{-1} \mu_r'} \quad r=1, 2, \dots, R \quad (13)$$

که پارامترهای موجود در آن عبارتند از:

$c_{r,t}'$: زیربردار ویژگی r ام برای قاب t

μ_r' : زیر بردار میانگین r ام متناظر با زیربردار ویژگی r ام

C_r' : زیر ماتریس کوواریانس r ام متناظر با زیربردار ویژگی r ام

البته این محاسبه مستلزم مستقل فرض کردن زیربردارهای ویژگی از یکدیگر است. به این ترتیب برای هر زیربردار ویژگی، یک عامل مقیاس متفاوت تعریف می شود. با وارد کردن عاملهای مقیاس λ_r در رابطه (۱۲) و سپس بکارگیری رابطه (۱۲) در رابطه (۱۰)، لگاریتم احتمال مشاهده در رابطه (۱۰) تبدیل به معیاری می شود که ما آن را SWPM می نامیم.

۶- آزمایش ها و نتایج

نتایج در کار حاضر بر روی دادگان TIMIT برای بازنمایی کلمات مجزا گزارش شده اند. دادگان TIMIT یک بانک اطلاعاتی از گفتار پیوسته انگلیسی است که توسط شرکت TI^{۲۳} و دانشگاه MIT^{۲۴} تهیه شده است و اداره استاندارد آمریکا (NIST)^{۲۵} آن را تایید کرده است. این دادگان حاوی ۶۳۰۰ جمله است که توسط ۶۳۰ گوینده و با ۸ لهجه معمول آمریکای شمالی بیان شده اند. ۷۰٪ گویندگان مرد و ۳۰٪ آنها زن هستند. هر گوینده ۱۰ جمله را ادا کرده است که ۲ جمله از این ۱۰ جمله توسط سایر گویندگان نیز ادا شده است. در مجموع ۲۴۳۲ جمله متمایز در TIMIT موجود است که شامل ۲ جمله مشترک میان تمامی گویندگان، ۴۵۰ جمله مشترک میان گروههای ۷ نفری گویندگان و ۱۸۹۰ جمله تک گوینده است. کلیه کلمات و واجهای موجود در جملات این بانک اطلاعاتی دارای برچسب زمانی هستند. بانک اطلاعاتی TIMIT به دو بخش آموزش و آزمون تقسیم شده است که بخش آموزش شامل ۴۶۲ گوینده و بخش آزمون شامل ۱۶۸ گوینده است و گویندگان و

$$A_{j,i,t} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \quad (9)$$

با فرض قطری بودن ماتریس کوواریانس مخلوطهای گاوسی، با جایگزین کردن $\lambda_{j,i,t}$ با $A_{j,i,t}$ در رابطه (۶) و بازنویسی این رابطه به فرم جمع، رابطه زیر حاصل می شود:

$$\log b_{j,i}(c_t) = \log |C_{j,i}| + N \log(2\pi) + \sum_{k=1}^n \frac{(c_{t,k} - \lambda_k \mu_{j,i,k})^2}{\sigma_k^2} \quad (10)$$

که در آن $c_{t,k}$ ، $\mu_{j,i,k}$ به ترتیب عناصر k ام $C_{j,i}$ ، $\mu_{j,i}$ هستند و λ_k ، σ_k^2 به ترتیب عناصر k ام قطر اصلی ماتریسهای $C_{j,i}$ ، $A_{j,i,t}$ می باشند.

حاصل جمع موجود در سمت راست رابطه (۱۰) را می توان به صورت زیر نوشت:

$$\sum_{k=1}^n \frac{(c_{t,k} - \lambda_k \mu_{j,i,k})^2}{\sigma_k^2} = \sum_{k=1}^{k_1} \frac{(c_{t,k} - \lambda_k \mu_{j,i,k})^2}{\sigma_k^2} + \sum_{k=k_1+1}^{k_2} \frac{(c_{t,k} - \lambda_k \mu_{j,i,k})^2}{\sigma_k^2} + \dots + \sum_{k=k_{R-1}+1}^n \frac{(c_{t,k} - \lambda_k \mu_{j,i,k})^2}{\sigma_k^2} \quad (11)$$

بنا بر آنچه گفته شد، در روش ترکیب ویژگیها، بردار ویژگی مرکب از الحاق زیربردارهای ویژگی به یکدیگر بدست می آید. با توجه به رابطه (۱۱)، می توانیم فرض نماییم که بردار ویژگی مرکب در حالت کلی از الحاق R زیربردار ویژگی تشکیل شده است. به همین طریق، بردار میانگین و ماتریس کوواریانس مخلوطهای گاوسی نیز به R زیربردار میانگین و R زیر ماتریس کوواریانس متناظر با R زیربردار ویژگی تقسیم می شود. متناظر با هر یک از R زیربردار ویژگی باید یک مقدار متفاوت λ محاسبه شود. بنابراین معادله (۱۱) به صورت زیر نوشته می شود:

$$\sum_{k=1}^n \frac{(c_{t,k} - \lambda_k \mu_{j,i,k})^2}{\sigma_k^2} = \sum_{k=1}^{k_1} \frac{(c_{t,k} - \lambda_1 \mu_{j,i,k})^2}{\sigma_k^2} + \sum_{k=k_1+1}^{k_2} \frac{(c_{t,k} - \lambda_2 \mu_{j,i,k})^2}{\sigma_k^2} + \dots + \sum_{k=k_{R-1}+1}^n \frac{(c_{t,k} - \lambda_R \mu_{j,i,k})^2}{\sigma_k^2} \quad (12)$$

23- Texas Instruments

24- Massachusetts Institute of Technology

25- National Institute for Standards and Technologies



جملات ادا شده در هر یک از این دو بخش با یکدیگر متفاوتند. در کار حاضر، ۲ جمله مشترک ادا شده بین گویندگان در دادگان TIMIT با دو لهجه (از میان ۸ لهجه) انتخاب شده و با استفاده از برچسب زمانی جملات به کلمات تجزیه گردیده اند. به این ترتیب، ۲۱ کلمه بدست آمده است که بوسیله ۱۵۱ گوینده شامل ۱۰۲ گوینده مرد و ۴۹ گوینده زن ادا شده اند. از این میان، ۲۴۳۹ گویش از ۱۱۴ گوینده برای مجموعه داده آموزش در نظر گرفته شده است. مجموعه داده آزمون نیز حاوی ۷۷۷ گویش از ۳۷ گوینده است. نرخ نمونه برداری هر نمونه گفتاری نیز ۱۶ کیلوهرتز است. سیستم بازشناسی، مدل مخفی مارکف با چگالی پیوسته (CDHMM) است که دارای ۶ حالت است و در هر حالت نیز ۸ مخلوط گاوسی موجود است. آموزش مدل نیز با استفاده از گفتار تمیز (موجود در مجموعه داده آموزش) برای ۲۱ کلمه مذکور صورت می گیرد. سه نویز جمع پذیر در کار حاضر مورد استفاده قرار گرفته اند: نویز صورتی (Pink)، نویز محیط کارخانه (factory) و نویز سفید که از دادگان نویز NOISEX۹۲ انتخاب گردیده اند. این سه نوع نویز به هر دو مجموعه داده آموزش و آزمون اضافه شده اند.

گفتار ورودی با استفاده از تبدیل موجک گسسته [۲۷،۲۱] با تابع پایه Daubechi مرتبه پنجم (۵db) به چهار زیرباند ۱-۰ کیلوهرتز، ۲-۱ کیلوهرتز، ۴-۲ کیلوهرتز و ۸-۴ کیلوهرتز تقسیم می شود. نحوه انتخاب تعداد و محدوده زیرباندها براساس کارهای پیشین مولفین در [۲۹،۲۸،۱] و با توجه به کارهای دیگر محققین [۱۸،۱۴،۱۳] صورت گرفته است. سپس با توجه به توزیع بانک فیلتر MEL در روی زیرباندها، از هر زیر باند ضرایب مل کپستروم (MFCC) و مشتق آنها استخراج می شود. در کار حاضر از ۶ فیلتر MEL در هر زیرباند استفاده شده است و از هر زیرباند ۳ ضریب MFCC و ۳ ضریب مشتق مرتبه اول آن (در مجموع ۶ ضریب از هر زیرباند) استخراج گردیده است. این ضرایب از قایمهای ۳۰ میلی ثانیه با همپوشانی ۲۰ میلی ثانیه استخراج گردیده اند. طول هر قاب در هر زیرباند با توجه به میزان کاهش فرکانس نمونه برداری در آن زیرباند تعیین گردیده است. در سیستم تمام باند نیز ۱۲ ضریب MFCC و ۱۲ ضریب مشتق اول آن (در مجموع ۲۴ ضریب) از هر

قاب استخراج شده است.

۱-۶- اصلاح آنتروپی مدل مخفی مارکف در سیستم ترکیب درستنماییها

در این بخش نتایج اولین روش پیشنهادی، یعنی تاثیر اصلاح آنتروپی مدل مخفی مارکف بر شیوه های وزن دهی در سیستم ترکیب درستنماییها گزارش شده و مورد بررسی قرار گرفته است. نتایج به صورت متوسط خطای بازشناسی کلمه، در حضور سه نوع نویز سفید، صوتی و کارخانه در نسبتهای سیگنال به نویز مختلف ارائه گردیده است. این نتایج برای ۳۲۱۶ گویش موجود در دادگان نویزی آموزش و آزمایش حاصل شده اند. در شکل، واژه Full نشانگر نتیجه بازشناسی تمام باند و LC نشانگر نتایج بازشناسی سیستم ترکیب درستنمایی است. واژه Equal نشان دهنده وزن دهی مساوی به زیرباندهاست که در آن $\alpha_{b,M}$ در رابطه (۱) برای تمامی باندها برابر ۰/۲۵ در نظر گرفته می شود. واژه های SNR و E-HMM^{۲۶} نیز به ترتیب نشانگر وزن دهی بر اساس نسبت سیگنال به نویز در هر زیرباند و وزن دهی بر اساس آنتروپی مدل مخفی مارکف (رابطه ۴) در سطح کلمه هستند.

در روش وزن دهی SNR، نسبت سیگنال به نویز در هر زیرباند در شرایط شبیه سازی شده و با حضور سیگنال تمیز و نویز به صورت جداگانه محاسبه گردیده است که این شرایط ایده آل تخمین خوبی از سیگنال به نویز فراهم می کند و امکان مقایسه این ضریب را با دیگر ضرایب وزن دهی در بهترین حالتش فراهم می کند. به این ترتیب خطای تخمین نسبت سیگنال به نویز در ارزیابی آن به عنوان یک ضریب وزن دهی دخالت نخواهد کرد. از این رو که تخمین نسبت سیگنال به نویز در شرایط واقعی قطعاً با خطای تخمین همراه است.

LC+WPM ترکیب درستنمایی‌ها پس از اصلاح آنتروپی مدل

مخفی مارکف

Full: سیستم تمام باند

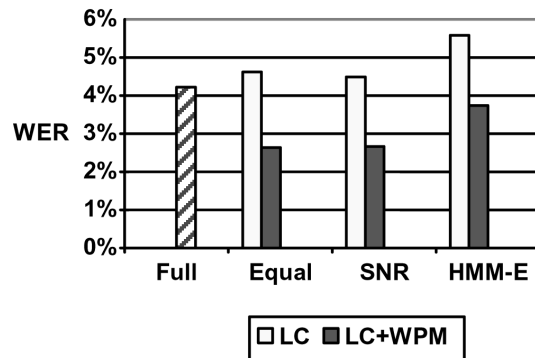
HMM-E: وزن دهی با عکس آنتروپی مدل مخفی مارکف در هر

زیرباند

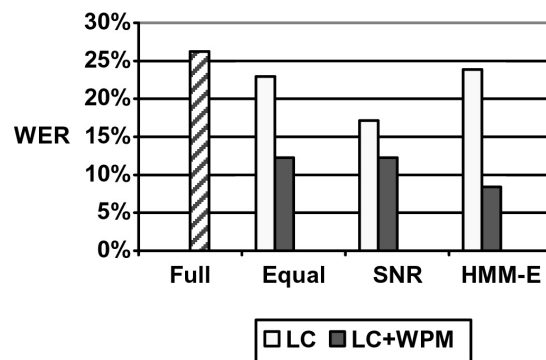
SNR: وزن دهی با نسبت سیگنال به نویز در هر زیرباند

Equal: وزن دهی مساوی

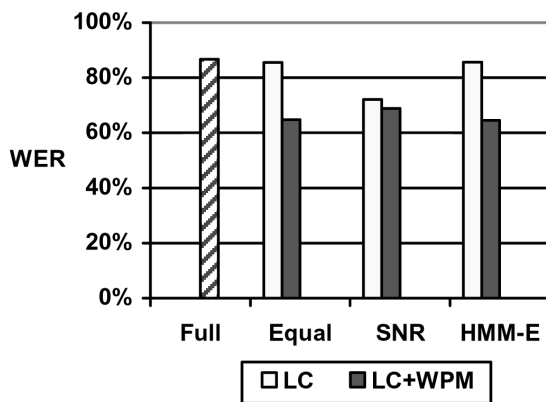
شکل (۲) متوسط خطای بازشناسی کلمه را برای سیستم بازشناسی چند باندهای به شیوه ترکیب درستنماییها و تاثیر اصلاح آنتروپی مدل مخفی مارکف با تصویر وزن دار، در حضور سه نوع نویز سفید، صورتی و کارخانه نشان می‌دهد. طبق این شکل در نسبت سیگنال به نویز ۲۰dB، دقت بازشناسی سیستم ترکیب درستنماییها به میزان ۱٪ تا ۴٪ کمتر از سیستم تمام باند است. از آنجا که در این نسبت سیگنال به نویز، سیگنال تقریبا تمیز محسوب می‌شود و کارایی سیستمهای چندباندی برای گفتار تمیز به علت در نظر نگرفتن همبستگی میان زیرباندها کمتر از سیستم تمام باند است [۱۴، ۱۱، ۴]، این نتیجه چندان دور از انتظار نیست. علاوه بر این، چنانکه مشاهده می‌شود اصلاح آنتروپی مدل مخفی مارکف به کمک WPM، سبب بهبود دقت بازشناسی سیستم ترکیب درستنماییها به میزان ۱٪ تا ۲٪ درصد می‌گردد و عملکرد آن را نسبت به سیستم تمام باند بهتر می‌سازد. این امر نشان می‌دهد که در این نسبت سیگنال به نویز، اصلاح آنتروپی مدل مخفی مارکف و درستنماییها و به عبارت دیگر تنظیم عملکرد مدل (طبقه بندی کننده) می‌تواند موثرتر از تنظیم وزنها باشد. در نسبتهای سیگنال به نویز ۰dB و ۱۰dB، خطای بازشناسی سیستم چند باندهای بر اساس ترکیب درستنماییها کمتر از خطای بازشناسی سیستم تمام باند است. به علاوه، اصلاح آنتروپی مدل مخفی مارکف به کمک WPM نیز سبب بهبود دقت بازشناسی در تمامی شیوه‌های وزن دهی گردیده است. از سوی دیگر، می‌توان مشاهده کرد که تاثیر اصلاح آنتروپی به کمک WPM، تاثیر بیشتری بر افزایش دقت بازشناسی در روش‌های وزن دهی HMM-E و Equal دارد که مستقیما از اطلاعات احتمال خروجی CDHMM استفاده می‌نمایند. با توجه به شکل می‌توان



الف) SNR = 20 dB



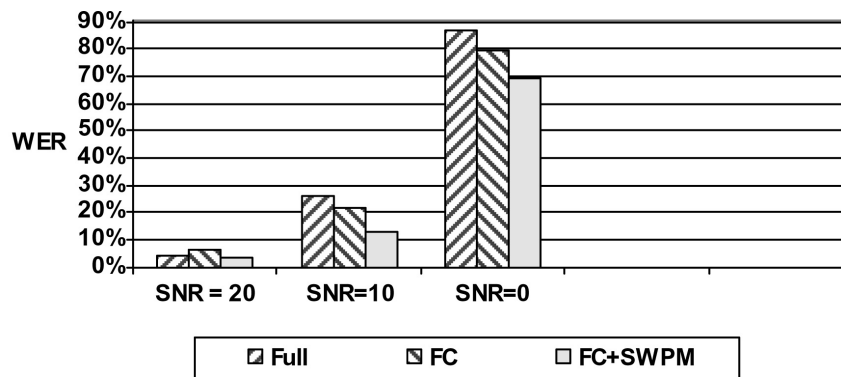
ب) SNR = 10 dB



ج) SNR = 0 dB

شکل (۲) - متوسط خطای بازشناسی کلمه در حضور سه نوع نویز سفید، کارخانه و صورتی برای نمایش تاثیر اصلاح آنتروپی مدل مخفی مارکف با WPM در سیستم ترکیب درستنماییها در نسبتهای مختلف سیگنال به نویز

LC: ترکیب درستنماییها



شکل (۳) - متوسط خطای بازشناسی کلمه در حضور سه نوع نویز سفید، کارخانه و صورتی برای نمایش تاثیر تصویر وزن دار زیرباندی (SWPM) در سیستم ترکیب ویژگی‌ها در نسبت‌های مختلف سیگنال به نویز

برای ۳۲۱۶ گویش موجود در دادگان نویزی آموزش و آزمایش گزارش شده اند. در شکل واژه Full نشانگر نتیجه بازشناسی تمام باند است. واژه FC بیانگر شیوه ترکیب ویژگی‌ها و FC+SWPM بیانگر بکارگیری شیوه تصویر وزن دار زیرباندی در روش ترکیب ویژگی‌ها است.

مطابق شکل، در نسبت سیگنال به نویز ۲۰ dB، دقت بازشناسی شیوه ترکیب ویژگی‌ها در حضور هر سه نوع نویز در حدود ۱٪ تا ۲٪ کمتر از سیستم تمام باند است. دلیل این امر را مشابه با سیستم ترکیب در دستنمایی‌ها می توان نزدیکی سیگنال تمیز به سیگنال نویزی و عدم در نظر گرفتن وابستگی زیرباندها دانست. چنانکه مشاهده می گردد،

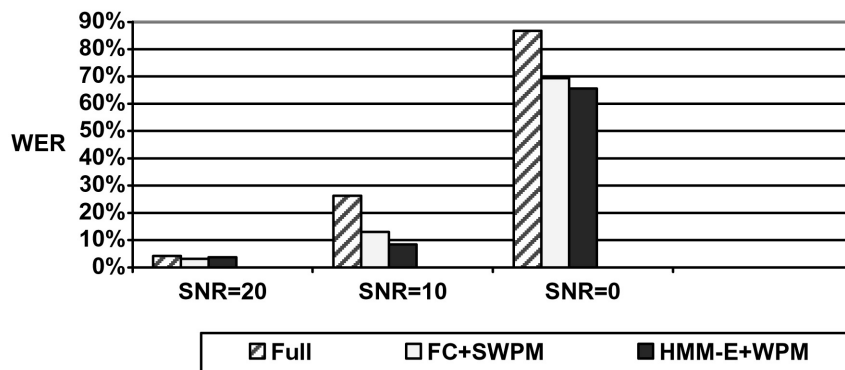
با بکارگیری SWPM، خطای بازشناسی سیستم ترکیب ویژگی‌ها در این نسبت سیگنال به نویز در حدود ۳٪ کاهش می یابد، به نحوی که دقت بازشناسی آن در حدود ۱/۵٪ بیشتر از سیستم تمام باند می گردد.

در نسبت‌های سیگنال به نویز ۱۰,۰ dB نیز دقت بازشناسی سیستم ترکیب ویژگی‌ها بهتر از سیستم تمام باند است. علاوه بر اینکه، استفاده از SWPM نیز سبب افزایش دقت بازشناسی سیستم ترکیب ویژگی‌ها می گردد به نحویکه خطای بازشناسی آن در نسبت‌های سیگنال به نویز ۱۰ dB و ۰ dB به ترتیب در حدود ۹٪ و ۱۰٪ کاهش می یابد.

گفت که در این دو نسبت سیگنال به نویز، در مجموع وزن دهی بر اساس آنتروپی اصلاح شده مدل مخفی مارکف (روش HMM-E به همراه WPM در نتایج LC+WPM) بهترین عملکرد را داراست. روش وزن دهی SNR پیش از اصلاح آنتروپی مدل مخفی مارکف (در نتایج LC) در هر سه نسبت سیگنال به نویز ۰, ۱۰, ۲۰ dB، نسبت به دو روش وزن دهی دیگر بهتر عمل می کند. اما پس از اصلاح آنتروپی مدل مخفی مارکف (نتایج LC+WPM)، غالباً دو روش دیگر نتایج بهتری دارند. این امر به دلیل آن است که روش‌های وزن دهی HMM-E و Equal بر اساس احتمالات خروجی CDHMM تعریف شده اند که اصلاح آنتروپی به کمک WPM مستقیماً بر این روش‌ها تاثیر می گذارد. حال آنکه، روش وزن دهی SNR بر اساس نسبت سیگنال به نویز در هر زیرباند عمل می کند که مرتبط با اطلاعات زیرباند است و مستقیماً ارتباطی به اطلاعات مدل ندارد. از این رو، تاثیر اصلاح آنتروپی بر دقت بازشناسی آن کمتر از دو روش دیگر است.

۲-۶- تصویر وزن دار زیرباندی در سیستم ترکیب ویژگی‌ها

در این قسمت، نتایج دومین روش پیشنهادی یعنی بکارگیری تصویر وزن دار زیرباندی در سیستم ترکیب ویژگی‌ها، مورد بررسی قرار گرفته است. شکل (۳) نشانگر متوسط خطای بازشناسی کلمه در حضور سه نویز جمع پذیر صورتی، محیط کارخانه و سفید برای سه نسبت مختلف سیگنال به نویز ۲۰, ۱۰, ۰ dB است. این نتایج نیز



ترکیب درست‌نمایی با وزن دهی بر پایه عکس آنتروپی مدل مخفی مارکف پس از اصلاح آنتروپی مدل مخفی مارکف HMM-E+WPM: ترکیب درست‌نمایی با وزن دهی بر پایه عکس آنتروپی مدل مخفی مارکف پس از اصلاح آنتروپی مدل مخفی مارکف
 FC+SWPM ترکیب ویژگی‌ها با بکارگیری تصویر وزن دار زیرباندی Full سیستم تمام باند
 شکل (۴) - قیاس متوسط خطای بازشناسی کلمه سیستم‌های بهبود یافته چند باندی (ترکیب ویژگی‌ها و ترکیب درست‌نماییها) در حضور سه نوع نویز سفید، کارخانه و صورتی در نسبت‌های مختلف سیگنال به نویز

دار زیرباندی در روش ترکیب ویژگی‌ها است. HMM-E+WPM نیز بیانگر اصلاح آنتروپی به کمک WPM برای روش وزن دهی بوسیله عکس آنتروپی مدل مخفی مارکف است. چنانکه در شکل دیده می‌شود، در نسبت سیگنال به نویز ۲۰ dB، دقت بازشناسی سیستم ترکیب ویژگی‌ها به همراه SWPM نسبت به سیستم تمام باند و آنتروپی اصلاح شده مدل مخفی مارکف بهتر است. البته در این نسبت سیگنال به نویز، با رجوع به شکل (۲)-الف می‌توان دریافت که شیوه وزن دهی Equal به همراه WPM دقت بازشناسی بهتر از FC+SWPM ارائه می‌کند. مطابق شکل (۴) در نسبت‌های سیگنال به نویز ۱۰، ۰ dB، شیوه HMM-E+WPM بهتر از FC+SWPM عمل نماید. به این ترتیب و با توجه به نکته فوق در باب شیوه وزن دهی Equal، اصلاح آنتروپی مدل مخفی مارکف در بهبود عملکرد سیستم ترکیب درست‌نماییها نسبت به کاربرد SWPM برای بهبود سیستم ترکیب ویژگی‌ها موفق‌تر عمل کرده است. هر چند که سیستم ترکیب درست‌نماییها به محاسبات بیشتری از سیستم ترکیب ویژگی‌ها نیاز دارد.

۷- جمع بندی و نتیجه گیری

در کار حاضر، بهبود عملکرد سیستم‌های بازشناسی چند باندی گفتار در حضور نویز مد نظر قرار گرفت برای هر کدام از انواع سیستم‌های بازشناسی چند باندی گفتار یک روش بهبود دقت

۳-۶ مقایسه روش‌های پیشنهادی: اصلاح آنتروپی مدل مخفی

مارکف و تصویر وزن دار زیرباندی

در این بخش، با جمع بندی نتایج بخش‌های (۱-۶) و (۲-۶) نتایج بازشناسی دو نوع سیستم پردازش چند باندی معرفی شده با یکدیگر مقایسه می‌شوند: سیستم ترکیب درست‌نماییها و سیستم ترکیب ویژگی‌ها. چنانکه در بخش‌های پیشین گفته شد، برای افزایش دقت بازشناسی سیستم ترکیب درست‌نماییها، از اصلاح آنتروپی مدل مخفی مارکف بوسیله WPM استفاده شده است و برای بهبود نرخ بازشناسی سیستم ترکیب ویژگی‌ها، تصویر وزن دار زیرباندی یا همان SWPM بکار گرفته شده است. لازم به ذکر است که با توجه به تعریف SWPM که برای بردارهای ویژگی مرکب از زیربردارهای ویژگی استخراجی از زیرباندها در نظر گرفته شده است، این روش تنها برای سیستم ترکیب ویژگی‌ها و یک مدل مخفی مارکف واحد قابل بکار بردن است و در سیستم ترکیب درست‌نماییها که چندین مدل مخفی مارکف (به تعداد زیرباندها) مورد نیاز است، قابل استفاده نیست. لذا قیاس حاضر، تنها از این جهت صورت می‌گیرد که دقت بازشناسی دو سیستم چندباندی بهبود یافته در کنار هم مشاهده گردند.

شکل (۴) مقایسه متوسط خطای بازشناسی روش‌های پیشنهادی را نشان می‌دهد. در شکل واژه Full نشانگر نتیجه بازشناسی تمام باند است. واژه FC+SWPM بیانگر بکارگیری شیوه تصویر وزن

ing, vol. 2, pp. 133-136, 2003.

[6] C.Cerisara, I.Illina, "Robust speech recognition to non-stationary noise based on model driven approach", European Conference on Speech Communication and Technology, pp.3053-3056, 2003.

[7] S.Molau, F.Hilger, H.Ney, "Feature Space Normalization in Adverse Acoustic Conditions", International Conference on Acoustics, Speech, and Signal processing, vol. 1, pp. 656-659, 2003.

[8] J.Chen, K.K.Paliwal, S.Nakamura, "Cepstrum derived from differential power spectrum for robust speech recognition", Speech Communication, Vol. 41, Issues 2-3, pp. 469-484, October 2003.

[9] X. Cui, A.Bernard, A. Alwan, "A noise-robust ASR back-end technique based on weighted Viterbi recognition", European Conference on Speech Communication and Technology, pp. 2169-2172, 2003.

[10] J. Ming, P. Jancovic, F.J.Smith, "Robust speech recognition using probabilistic union models", IEEE Trans. on Speech and Audio Processing, vol. 10, No. 6, pp. 403-414, September 2002.

[11] A. Hagen, Robust speech recognition based on multi-stream processing, Ph.D. dissertation, Lab. d'Intell. Artif. Perceptive, École Polytechnique Fédérale, Lausanne, Switzerland, 2001.

[12] X. Huang, A.Acero, H. Hon, Spoken Language processing, Prentice Hall, 2001.

[13] C. Cerisara, D. Fohr, "Multi-band automatic speech recognition", Computer Speech and Language, vol.15, Issue. 2, pp. 151-174, April 2001.

[14] N. Mirghafoori, A multi-band approach to automatic speech recognition, PhD thesis, ICSI, Berkeley, 1999.

[15] S. Okawa, T. Nakajima, K.Shirai, "A recombination strategy for multi-band speech recognition based on mutual information criterion", European Conference on Speech Communication and Technology, vol. 2, pp. 603-606, 1999.

[16] S. Okawa, E. Boochieri, A.Potamianos "Multi-band speech recognition in noisy environment", IEEE International Conference on Acoustics, Speech, and Signal processing, vol. 2, pp. 641-644, 1998.

[17] F. Berthommier, H. Glotin, E. Tessier, H. Bourlard, "Interfacing of CASA and partial recognition based on a multi-stream technique", International Conference on Spoken Language Processing, Australia, 1998.

[18] S. Tibrewala, H. Hermansky "Sub-band based recognition of noisy speech", IEEE International Conference on Acoustics, Speech, and Signal processing, pp.1255-1258, 1997.

[19] M.J.F.Gales, S.J.Young, "Robust continuous speech recognition using parallel model combination",

پیشنهاد گردید: اصلاح آنتروپی مدل مخفی مارکف برای سیستم ترکیب درستنماییها و تصویر وزن دار زیرباندی برای سیستم ترکیب ویژگیها.

در سیستم ترکیب درستنماییها، از روش تصویر وزن دار برای بهبود آنتروپی مدل مخفی مارکف استفاده گردید. نتایج آزمایشها بیانگر آن است که اصلاح آنتروپی مدل مخفی مارکف سبب بهبود عملکرد سیستم ترکیب درستنماییها و روش های وزن دهی در حضور انواع مختلف نویز با نسبت های سیگنال به نویز متفاوت می گردد. به علاوه، اصلاح آنتروپی عملکرد روش وزن دهی مبتنی بر آنتروپی مدل مخفی مارکف را در نسبت های سیگنال به نویز متوسط و پایین، در قیاس با دیگر روش های وزن دهی بیشتر تحت تاثیر قرار می دهد.

در سیستم ترکیب ویژگیها، تعمیمی از شیوه تصویر وزن دار بنام تصویر وزن دار زیرباندی ارائه گردید که بر بردارهای ویژگی مرکب از زیربردارهای ویژگی استخراجی از زیرباندها اعمال گردید. نتایج بیانگر این است که استفاده از این شیوه، دقت بازشناسی سیستم ترکیب ویژگیها را در حضور انواع متفاوت نویز با نسبت های سیگنال به نویز مختلف بهبود می بخشد.

مراجع

[1] B. Nasersharif, A. Akbari, "Application of wavelet transform and wavelet thresholding in robust sub-band speech recognition", European signal processing Conference, vol.1, pp. 345-348, 2004.

[2] N.Poh, S.Bengio, "Why do multi-stream, multi-band and multi-modal approaches work on biometric user authentication tasks?", International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, pp. 893-896, 2004.

[3] A.Hagen, A.Morris, "Recent advances in the multi-stream HMM/ANN hybrid approach to noise robust ASR", Computer Speech and Language, Vol. 18, pp. 1-28, 2004.

[4] J. McAuley, J. Ming, P. Hanna, D.Stewart, "Modeling sub-band correlation for noise robust speech recognition", International Conference on Acoustics, Speech, and Signal processing, Vol. 1, pp. 1017-1020, 2004.

[5] S. Ikbal, H. Misra, H. Bourlard, "Phase autocorrelation derived robust speech features", IEEE International Conference on Acoustics, Speech, and Signal process-