تاریخ دریافت مقاله: ۹۷/۰۳/۰۴
تاریخ پذیرش مقاله: ۹۷/۰۷/۱۵

زاویهٔ اَبعاد کاهش یافته: ویژگی جدید و مقاوم مبتنی بر اطلاعاتِ زاویهٔ سیگنالِ گفتار و کاربرد آن در شناسایی لهجه

علیرضا فلک رفعت
دانشجوی کارشناسی ارشد دانشکده مهندسی کامپیوتر و فناوری اطلاعات ـ دانشگاه آزاد اسلامی واحد قزوین ـ قزوین ـ ایران
پست الکترونیکی: alirafat2012@gmail.com

اعظم ربیعی*
استادیار گروه کامپیوتر ـ دانشگاه آزاد اسلامی واحد دولت آباد ـ اصفهان ـ ایران
پست الکترونیکی:azrabiee@gmail.com

چکیده

ایـن مقاله، ویژگـی جدیدی مبتنی بـر اطلاعات زاویه سـیگنال گفتار به نام زاویهٔ اَبعاد کاهش یافته(DRP)، ارائه کرده اسـت. ویژگی DRP طی دو مرحله نرمال‌سـازی و کاهش اَبعاد از زاویهٔ اسپکتروگرامِ سیگنال گفتار استخراج می‌گـردد. فرایند کاهش اَبعاد در این تحقیق با کمک تکنیک تحلیل اجزاء اصلی (PCA) انجام شده است. در این تحقیق، از ترکیب ویژگی پیشنهادی با ویژگی‌های مبتنی بر اندازه، برای شناسایی لهجهٔ گوینده استفاده شده است. شناسایی لهجهٔ گوینده با کمک شبکه‌های عصبی مصنوعی به‌عنوان دسته‌بندی کننده انجام شده است. مقایسه‌های آزمایش‌های انجام شده در شرایط بدون نوفه و با سیگنال به نوفه‌های صفـر، ۵ و ۱۰ دسـی بل، برتری این ویژگـی و مقاوم به نوفه بودن آن در شناسایی لهجه گوینده را نشان می‌دهد. بیشترین مقدار کارایی مربوط به ترکیب ویژگی پیشنهادی (زاویهٔ اَبعاد کاهش یافته) بـا ضرایب برداری Rasta PLP در محیط سـالم برابر با ۹۸/۱۴٪ و محیط آغشته به نوفه ۹۵/۵۵٪ نشان داده شده است.

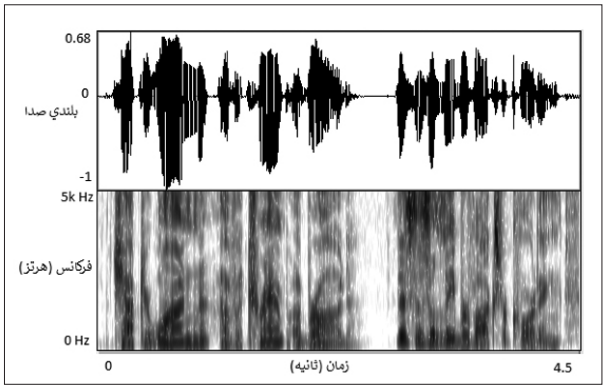
واژه‌هـای کلیـدی: ویژگی مبتنی بر زاویه، دسـته‌بندی لهجه، زاویهٔ اَبعاد کاهش یافته، تحلیل اجزای اصلی، ویژگی مقاوم به نوفه

۱- مقدمه

سیسـتم‌های مبتنی بر پردازش گفتار، طی سـال‌های متوالی، به‌عنوان دستیار افراد ناتوان، جهت یادگیری زبان و بسیاری از کاربردهای مربوط به ارتباط انسان و کامپیوتر مورد توجه بوده‌اند. شناسـایی گفتار، تولید گفتار، ارتقاء کیفیت گفتار نوفه‌دار، شناسـایی گوینده و لهجهٔ گوینده از جملهٔ سیستم‌های مبتنی بر پردازش گفتار هستند.

سـیگنال گفتار عـلاوه بـر اطلاعات کلامـی، حاوی ویژگی‌های مفیدی در مـورد گوینده و حالات درونی وی مانند سن، جنسیت، هیجان، لهجه و غیره می‌باشد. در علم پردازش سـیگنال، به منظور شناسـایی دقیق‌تر گوینده و حالات درونی او، ویژگی‌هایی از سـیگنال گفتار استخراج می‌شـود و همان‌طور که ذکر شـد عمدتاً در کاربردهای مختلف مربـوط به ارتباط انسـان و کامپیوتـر، از آن‌ها

* نویسندهٔ مسئول



شکل ۱: سیگنال گفتار در حوزه زمان (تصویر بالا) و اندازهٔ اسپکتروگرام یا نمایش زمان- فرکانس سیگنال (تصویر پایین)

بهره‌بـــرداری می‌گردد. این ویژگی‌ها هـــم در حوزه زمان و هم در حوزه فرکانس از ســـیگنال گفتار قابل اســـتخراج هســـتند. لازم به ذکر اســـت ســـیگنال گفتار با کمک تبدیل فوریه به حوزه فرکانس منتقل می‌شـــود. همچنین با کمک تبدیل فوریه زمان کوتاه[1]، نمایش زمان- فرکانســـی به نام «اسپکتروگرام» از این سیگنال به‌دست می‌آید که گاه حاوی اطلاعات دقیق‌تر و مفیدتری است.

شـــکل ۱، نمونه‌ای از یک سیگنال گفتار و اسپکتروگرام یـــا نمایـــش زمان-فرکانســـی آن را نشـــان می‌دهـــد. در اســـپکتروگرام، از آنجا که تبدیل فوریه یک تبدیل مختلط است، ســـیگنالِ منتقل شـــده در حوزهٔ فرکانس، شامل دو بخـــش حقیقـــی و موهومی مانند x=a+ib اســـت که در آن a بخـــش حقیقـــی و b بخش موهومی x اســـت. از این رو، معمـــولاً ویژگی‌ها از انـــدازهٔ این مقادیـــر موهومی (یعنی $|x| = \sqrt{a^2 + b^2}$) اســـتخراج می‌شـــوند و اطلاعـــات زاویه[2] سیگنال گفتار (یعنی $\theta = \arctan(\frac{b}{a})$) دور ریخته می‌شود.

محققان در گذشته بر این باور بودند که اطلاعات زاویه در ســـیگنال گفتار نقش مهمی در پردازش، بازشناســـی و درک گفتار ایفا نمی‌کند [۱]. اما با گذشت زمان و تحقیقات انجام شـــده، نتایجی حاصل شد که نشـــان داد می‌توان از اطلاعات زاویه نیز در پردازش گفتار بهره برد [۲]. همچنین، می‌توان ویژگی‌های استخراج شده مبتنی بر اطلاعات زاویه از سیگنال گفتار را با ویژگی‌های مستخرج از اندازه ترکیب

1- Short-Time Fourier Transform (STFT)
2- Phase Information

جدول۱: چند نمونه از ویژگی‌های مستخرج از سیگنال گفتار

	نام ویژگی	اختصار لاتین
مبتنی بر اندازه اسپکتروگرام	ضریب کپسترال فرکانسی مل	MFCC[1]
	ضریب پیشگویی خطی	LPC[2]
	ضریب کپسترال پیشگویی خطی	LPCC[3]
	پیشگویی خطی ادراکی	PLP[4]
	تبدیل طیفی نسبی پیشگویی خطی ادراکی	RastaPLP[5]
مبتنی بر زاویهٔ اسپکتروگرام	زاویه هارمونیک	HP[6]
	شیفت نسبی زاویه	RPS[7]
	زاویه باقیمانده	RP[8]
	زاویه خالص (زاویه اسپکتروگرام نرمال‌شده)	PP
	زاویهٔ اَبعاد کاهش یافته (پیشنهاد این مقاله)	DRP[9]

1- Mel Frequency Cepstral Coefficient
2- Linear Predictive Coefficient
3- Linear Predictive Cepstral Coefficient
4- Perceptual Linear Predictive
5- Relative Spectral Transform PLP
6- Harmonic Phase
7- Relative Phase Shift
8- Residual Phase
9- Dimention Reduced Phase

کـــرد و نتایج بهتری حتی در محیط‌های نوفه‌دار به‌دســـت آورد [۳،٤]. جـــدول ۱ تعدادی از ویژگی‌های مســـتخرج از ســـیگنال گفتار، مبتنی بـــر اطلاعات زاویه و انـــدازه را به تفکیک نشـــان می‌دهد. جزئیات مربوط به نحوهٔ اســـتخراج ویژگی‌های مبتنی بر اندازه در بخش ۱-۱ و مبتنی بر زاویه در بخش ۱-۲ شرح داده شده است.

تحقیقات اخیر، بیشتر به بررسی اطلاعات زاویهٔ سیگنال گفتار در کاربردهایی مانند شناسایی گوینده یا هیجان وی پرداخته‌اند. این تحقیقات در بخش ۲ مرور خواهند شد. در مورد تاثیر اطلاعات زاویهٔ گفتار در شناسایی لهجه تاکنون تحقیقات زیادی انجام نشـــده اســـت. از ایـــن رو، به کلیاتِ تحقیقاتِ مربوط به شناســـایی لهجه نیز در همان بخش ۲ اشاره‌ای خواهد شد.

از آنجا کـــه تحقیقی به بررســـی اهمیـــت ویژگی‌های مبتنی بر زاویه در شـــرایط نوفه‌دار برای شناسایی لهجه نپرداخته اســـت، مقاله حاضر، یـــک ویژگی جدید مبتنی بر اطلاعات زاویه، به نام «زاویهٔ اَبعاد کاهش یافته» برای این منظور ارائه می‌کند. مدل پیشنهادی برای شناسایی لهجه به

شکل ۲: مدل پیشنهادی شناسایی لهجهٔ گفتار

همراه ویژگی پیشنهادی در شکل ۲ نشان داده شده است. همان‌طور که در شکل نشان داده شده است، مدل پیشنهادی شامل سه مرحلهٔ استخراج اسپکتروگرام، استخراج ویژگی و دسته‌بندی با کمک شبکه‌های عصبی مصنوعی است. واژگان تخصصی به کار رفته در مراحل اول و دوم به‌طور عمده در بخش‌های ۱-۱ و ۱-۲ پوشش داده می‌شوند. مدل شناسایی لهجه به همراه ویژگی پیشنهادی مبتنی بر اطلاعات زاویه در بخش ۳ شرح داده خواهد شد.

در بخش ٤، آزمایش‌هایی جهت ارزیابی مدل، در محیط‌های سالم و آغشته به نوفه به ترتیب در زیربخش‌های ٤-۱ و ٤-۲ ارائه شده است. در انتها، بخش ٥ به جمع‌بندی و نتیجه‌گیری می‌پردازد.

۱-۱- ویژگی‌های مبتنی بر اندازه

ویژگی‌های مختلف و گوناگونی در سیگنال گفتار نهفته است. هر کدام از این ویژگی‌ها، مشخصه خاصی از گفتار را بهتر نمایش می‌دهند. جدول ۱ و شکل ۲ ویژگی‌های مستخرج از سیگنال گفتار را به دو دستهٔ ویژگی‌های مبتنی بر اندازه (ویژگی‌های برداری معمولی طیفی) و ویژگی‌های مبتنی بر اطلاعات زاویه، تقسیم کرده است.

هر دو دسته ویژگی‌های مدنظرِ این تحقیق برگرفته از اسپکتروگرام یا نمایش زمان-فرکانسی سیگنال گفتار است. از این رو، اشارهٔ مختصری به نحوه استخراج اسپکتروگرام خواهیم داشت. همچنین، به استخراج اسپکتروگرام به‌عنوان اولین مرحله از مدل پیشنهادی شناسایی لهجه در شکل ۲ اشاره شده است.

جهت استخراج نمایش زمان-فرکانسی، ابتدا فرکانس‌های بالا طی عملیات پیش‌تاکید، تقویت می‌شوند. منظور از مرحلهٔ پیش‌تاکید، استفاده از فیلتر بالاگذر $1-\alpha Z^{-1}$ است که در آن $\alpha=0.97$ در نظر گرفته شده است. این فیلتر برگرفته از مدل لب در تولید گفتار و معادل یک فیلتر بالاگذر است و باعث تقویت فرکانس‌های بالا می‌شود.

به دلیل ماهیت اتفاقی و پویای سیگنال گفتار، این سیگنال در حوزه زمان، طی فرایند پنجره‌زنی و قاب‌بندی به تعدادی قاب با زمان محدود و همپوشانی‌های زمانی تبدیل می‌گردد. سیگنال گفتار در قاب‌های با مدت زمان محدود، ایستایی بیشتری دارد. همچنین، برای پیشگیری از اثر ناصافی[3]، قاب‌های گفتار در پنجره‌ای ضرب می‌شوند تا اثر لبه‌ها از بین برود.

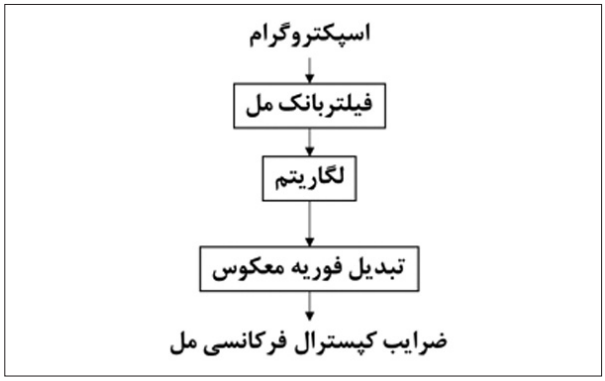
سپس، تبدیل فوریه گسسته از قاب گرفته می‌شود. این تبدیل فوریه به منظور استخراج اطلاعات فرکانسی سیگنال است. خروجی این تبدیل، به‌صورت عدد مختلط است. در اکثر سیستم‌های مبتنی بر پردازش گفتار، از ویژگی‌های مستخرج از اندازهٔ این عدد مختلط استفاده می‌شود. در ادامه برخی از ویژگی‌های برداری معمولی طیفی، که عمدتاً مبتنی بر اندازهٔ اسپکتروگرام هستند و در این تحقیق از آن‌ها استفاده شده، شرح داده می‌شود.

۱-۱-۱- ضریب کپسترال فرکانسی مل

ضرایب کپسترال فرکانسی مل از پرکاربردترین و معروف‌ترین ویژگی‌های مبتنی بر اندازه اسپکتروگرام هستند که از مدل کردن غیرخطی پاسخ سیستم شنوایی گوش انسان به‌دست می‌آید [٥]. نمودار بلوکی استخراج این ضرایب در شکل ۳ نشان داده شده است.

استخراج اسپکتروگرام از سیگنال گفتار، مشابه مراحل

3-aliasing



شکل ۳: نمودار بلوکی استخراج ضرایب کپسترال فرکانسی مل

شرح داده شده در ابتدای بخش ۱-۱ است. جهت استخراج ضرایب کپسترال فرکانسی مل، همان‌طور که شکل ۳ نشان می‌دهد، در مرحله اول بانک فیلتر مل که الهام گرفته از سیستم شنوایی انسان است، روی اسپکتروگرام اعمال می‌شود. سپس، لگاریتم روی خروجی فیلتر مل محاسبه شده است و درنهایت دامنه حاصل از تبدیل کسینوسی گسسته معکوس روی خروجی لگاریتم، به‌عنوان ویژگی‌های مورد نیاز در نظر گرفته می‌شود.

ضرایب ویژگی، شکل و ویژگی مجرای گفتار در هنگام بیان گفتار را نشان می‌دهند. در برخی موارد نظیر بازشناسی گفتار می‌توان از سرعت تغییر این ویژگی‌ها و نیز آهنگ تغییر آن‌ها استفاده کرد، که این امر مزایای خود را دارد. به همین دلیل، از مشتقات اول و دوم این ضرایب نیز در این تحقیق استفاده شد.

۱-۱-۲- ضریب پیشگویی خطی

ایدهٔ اصلی استخراج این ضرایب از اینجا گرفته شده است که هر نمونه از گفتار، توسط یک ترکیب خطی از تعدادی از نمونه‌های قبلی خود به‌دست می‌آید. یعنی تخمین نمونه nام گفتار یا $\hat{s}(n)$را به‌صورت زیر می‌توان محاسبه کرد [۶].

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

ضرایب پیشگویی خطی همان a_kها در رابطه (۱) هستند که با تکنیک حداقل میانگین مربعات خطا و با حداقل کردن فاصله$s(n)$ از$\hat{s}(n)$، از طریق یکی از روش‌های همبستگی یا کواریانس به‌دست می‌آیند. در بخش نتایج ضرایب LPC براساس کواریانس محاسبه شده‌اند. لازم به ذکر است، اگر این ویژگی از نمونه‌ها در حوزه زمان استخراج شوند، شامل هر دو اطلاعات اندازه و زاویه اسپکتروگرام خواهند بود. بنابراین، بسته به نحوه استخراج این ضرایب، می‌توان این ویژگی را به‌عنوان یک ویژگی برداری معمولی قلمداد کرد.

۱-۱-۳- ضریب کپسترال پیشگویی خطی

هر چند ضرایب پیشگویی خطی، ویژگی‌های مفیدی از سیگنال گفتار را ارائه می‌دهند، ولی بسیار حساس هستند و محاسبات آن‌ها باید دقیق باشد. محاسبه کپسترال، ضرایبی را تحت عنوان ضرایب کپسترال پیشگویی خطی ارائه می‌دهد که بسیار مقاوم‌تر و پایدارتر از آن است [۷]. در این تحقیق از ضرایب کپسترال پیشگویی خطی نیز استفاده شده است.

۱-۱-۴- پیشگویی خطی ادراکی

ضرایب پیشگویی خطی ادراکی ، ویژگی‌هایی هستند که محاسبه آن‌ها از مدل شنوایی انسان ایده گرفته است. محاسبه این ویژگی بسیار شبیه به ضریب پیشگویی خطی است [۸]. این ضرایب استخراج شده را با PLP_lpcas نیز می‌شناسیم. این ضرایب می‌توانند از اسپکتروم یا کپستروم استخراج شوند. از هر دسته از این ضرایب می‌توان به‌عنوان ویژگی‌های مناسب از گفتار استفاده کرد [۹].

۱-۱-۵- تبدیل طیفی نسبی پیشگویی خطی ادراکی

ضرایب پیشگویی خطی ادراکی، مثل بسیاری از دیگر ضرایب که مبتنی بر طیف زمان کوتاه هستند بسیار آسیب پذیرند. تبدیل طیفی نسبی آن‌ها با کمک گرفتن از طیف نسبی، باعث مقاوم شدن این ضرایب به نوفه‌های طیفی می‌شود [۹]. در استخراج این ضرایب از یک فیلتر باند میانی که به یک نمایش لگاریتم- طیفی اعمال می‌شود بهره گرفته می‌شود [۱۰].

۱-۲- ویژگی‌های مبتنی بر زاویه

هر چند اکثر سیستم‌های مبتنی بر پردازش گفتار،

از ویژگی‌های مستخرج از اندازهٔ اسپکتروگرام استفاده می‌کنند ولی زاویه اسپکتروگرام نیز حاوی اطلاعات مفیدی در مورد گفتار و گوینده است. این تحقیق، ویژگیِ جدیدی مبتنی بر اطلاعات زاویه ارائه می‌کند. جهت ارزیابی و نمایش برتری ویژگی پیشنهادی، تعدادی از ویژگی‌های مبتنی بر زاویه، در این تحقیق به کار گرفته شده‌اند. از این‌رو، در ادامه به مرور آن‌ها پرداخته می‌شود. ویژگی‌های زاویه خالص نرمال‌شده، زاویه هارمونیک، شیفت زاویه نسبی و زاویه باقی‌مانده در ادامه مرور شده‌اند.

۱-۲-۱- زاویه خالص نرمال شده

در این تحقیق، منظور از زاویه خالص، زاویه اسپکتروگرام نرمال شده است [۳]. برای استخراج زاویه خالص، ابتدا سیگنال از حوزه زمان به حوزهٔ زمان-فرکانس (اسپکتروگرام) منتقل می‌شود. سپس، زاویه اسپکتروگرام در فرکانس‌های مختلف در هر لحظه با استفاده از یک فرکانس مبنا نرمال می‌گردد. زاویه خالص نرمال‌شده، به عنوال یک ویژگی مبتنی بر زاویه در پردازش سیگنال استفاده می‌شود. مراحل استخراج این ویژگی به شرح زیر است:

مرحله اول : تبدیل سیگنال از حوزه زمان به زمان-فرکانس

مرحله دوم : استخراج زاویه اسپکتروگرام از دامنه آن

مرحله سوم : نرمال‌سازی زاویه اسپکتروگرام

فرض کنید $S(\omega,t)$، محصول تبدیل فوریه یک قاب در زمان یا همان نمایش اسپکتروگرام است. طبق رابطه (۲) اندازه و زاویه این مقدار موهومی، قابل تفکیک می‌باشد.

$$S(\omega,t)=X(\omega,t)+jY(\omega,t) = \sqrt{X^2(\omega,t)+Y^2(\omega,t)}*e^{j\theta(\omega,t)} \quad (۲)$$

در رابطه (۲)، $X(\omega,t)$ قسمت صحیح و $Y(\omega,t)$ قسمت موهومی را شامل می‌شود. همچنین، $\sqrt{X^2(\omega,t)+Y^2(\omega,t)}$ و $\theta(\omega,t)$، به ترتیب اندازه و زاویه (زاویه) این قاب هستند. اندازه زاویه‌ها در تمام قاب‌ها می‌تواند در محدوده $[0,2\pi]$ یا $[-\pi,\pi]$ تعریف گردد. با این حال، تغییرات زاویه $\theta(\omega,t)$ به موقعیت برش‌های سیگنال ورودی با فرکانس‌های زاویه‌ای ω مختلف بستگی دارد. عملیات نرمال‌سازی، به منظور کاهش تفاوت بین دو زاویه از دو پنجره متفاوت لازم است. توضیح فرایند نرمال‌سازی برای از بین بردن تاثیرات موقعیت‌های مختلف، به شرح زیر است [۳]:

برای نرمال کردن زاویه‌های هر قاب، از یک فرکانس مبنا (به‌عنوان مثال $\omega_b = 2\pi \times 1000\ hz$) در تمام قاب‌ها استفاده می‌شود. مابقی زاویه‌ها براساس این فرکانس مبنا بر حسب رادیان قابل تخمین است. برای نمونه، اگر زاویه فرکانس مبنا $\theta(\omega_b,t)=\pi/4$ در نظر گرفته شود، آنگاه،

$$S'(\omega_b,t)=\sqrt{x^2(\omega_b,t)+Y^2(\omega_b,t)} * e^{j\theta(\omega_b,t)} * e^{j\left(\frac{\pi}{4}-\theta(\omega_b,t)\right)} \quad (۳)$$

تفاوت رابطهٔ (۲) و (۳) مقدار زاویه $\left(\frac{\pi}{4}-\theta(\omega_b,t)\right)$ می‌باشد. با $\omega=2\pi f$ (در فرکانسی غیر از فرکانس مبنا)، تفاوت زاویه برابر با $\frac{\omega}{\omega_b}\left(\frac{\pi}{4}-\theta(\omega_b,t)\right)$ می‌شود. بنابراین :

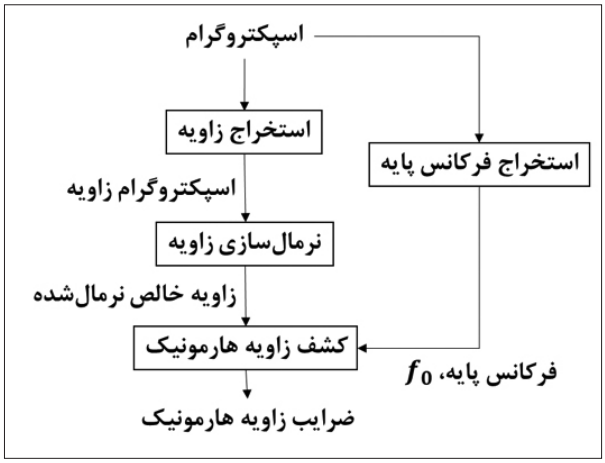
$$S'(\omega_b,t)=\sqrt{x^2(\omega,t)+Y^2(\omega,t)} * e^{j\theta(\omega,t)} * e^{j\frac{\omega}{\omega_b}\left(\frac{\pi}{4}-\theta(\omega_b,t)\right)} = \sqrt{X^2(\omega,t)+Y^2(\omega,t)} * e^{j\bar{\theta}(\omega,t)} \quad (٤)$$

که در آن $\bar{\theta}$ زاویه نرمال شده است. پس، زاویه نرمال شده طبق رابطه (۵) به‌دست می‌آید [۳].

$$\bar{\theta}(\omega,t)=\theta(\omega,t)+\frac{\omega}{\omega_b}\left(\frac{\pi}{4}-\theta(\omega_b,t)\right) \quad (۵)$$

۱-۲-۲- زاویهٔ هامونیک

منظور از زاویهٔ هارمونیک، در واقع مقدار زاویه در فرکانس‌های هارمونیک سیگنال می‌باشد. فرکانس پایه به‌عنوان اولین هارمونیک در نظر گرفته می‌شود و منظور از هارمونیک‌ها، ضرایب صحیح فرکانس پایه است [۱۱]. در واقع قله‌های سیگنال در حوزه فرکانس به‌عنوان هارمونیک‌های آن در نظر گرفته می‌شوند. مثلاً اگر اولین قله در فرکانس ۱۰۰ هرتز قرار گرفته باشد، سایر قله‌ها



شکل ۴: نمودار بلوکی استخراج ضرایب زاویه هارمونیک

در فرکانس‌های ۲۰۰، ۳۰۰، ۴۰۰ و غیره قرار دارند. رابطهٔ فرکانس‌های هارمونیک با فرکانس پایه در رابطه (٦) نشان داده شده است.

$$f_k = k * f_0 \qquad k = 1,2,... \qquad (٦)$$

که در آن f_0 ، k و f_k به ترتیب فرکانس پایه سیگنال، عدد صحیح مثبت و فرکانس‌های هارمونیک می‌باشد. نمودار بلوکی استخراج زاویه هارمونیک در شکل ٤ نشان داده شده است.

همان‌طور که در شکل ٤ نشان داده شده است، ابتدا فرکانس پایه (f_0) در هر قاب از اسپکتروگرام محاسبه می‌شود. باید این نکته را در نظر داشت که فرکانس پایه در هر دو حوزه زمان و اسپکتروگرام قابل محاسبه است. سپس زاویه اسپکتروگرام استخراج می‌گردد. در مرحلهٔ بعد، با استفاده از زاویه اسپکتروگرام و فرکانس پایه، زاویهٔ هارمونیک‌ها به‌دست آمده و هارمونیک‌های ابتدایی که از درجهٔ اهمیت بیشتری برخوردار هستند، به‌عنوان ضرایب زاویه هارمونیک استخراج می‌گردند.

هارمونیک‌های ابتدایی اطلاعات مفیدتری را در خود نگه می‌دارند و دارای اهمیت بیشتری هستند. از این رو در اکثر مطالعات و پژوهش‌ها از هارمونیک‌های بالاتر استفاده نمی‌شود. دلیل این امر این است که گوش انسان به فرکانس‌های پائین حساس‌تر می‌باشند [۱۲]. در آزمایش‌های این تحقیق، تعداد پنج ضریب اول هارمونیک در نظر گرفته شد (k=1,...,5).

۱-۲-۳- شیفت زاویه نسبی

مراحل استخراج ضرایب شیفت زاویه نسبی، بسیار شبیه به ضرایب زاویه هارمونیک است با این تفاوت که در مرحله آخر، اختلاف زاویه هارمونیک اول با دیگر هارمونیک‌ها طبق رابطه (٦) محاسبه می‌شود، [۱۳]

$$RPS_k = \theta(kf_0) - \theta(f_0) \qquad k = 2,3,...,n \qquad (٦)$$

که در رابطه فوق، منظور از θ(f) مقدار زاویه در فرکانس f بر حسب هرتز است. مقدار f_0 برابر با زاویه هارمونیک اول و kf_0 هارمونیک kام می‌باشد.

۱-۲-۴- زاویه باقی‌مانده

منظور از زاویه باقی‌مانده، مابقی زاویه هارمونیک‌های استخراج شده از هر قاب می‌باشد [۱٤]. در این تحقیق، ٥ ضریب زاویه هارمونیک اول به‌عنوان زاویه هارمونیک در نظر گرفته شد. مابقی زاویه‌ها (از ٦ تا ۱۱ امین زاویه هارمونیک) به‌عنوان زاویه باقی‌مانده استفاده می‌شود.

۲- تحقیقات مرتبط

همان‌طور که ذکر شد، اخیراً محققان به اهمیت اطلاعات زاویه سیگنال گفتار پی برده‌اند. برای اولین بار، در سال ۲۰۱۰ ساراکساگا و همکارانش، تحقیقی در مورد استفاده از اطلاعات زاویه هارمونیک به منظور بهبود بازشناسی گفتار انجام دادند [۱۳]. در آن تحقیق، از شیفت زاویه نسبی برای بهبود نرخ بازشناسی گفتار اتوماتیک استفاده شد. این افراد در سال ۲۰۱۱ نیز به بررسی استفاده از زاویه هارمونیک در بازشناسی گوینده پرداختند [۱٥]. در این تحقیق آن‌ها از مجموعه‌ای از ویژگی‌ها به منظور بازشناسی گوینده کمک گرفتند. نتایج نشان داد که ویژگی‌های مبتنی بر زاویه عملکرد بسیار عالی و بسیار نزدیک به عملکرد پارامترهای MFCC دارند. از آن زمان، نتایج تحقیقات مختلفی نشان داده است که از اطلاعات مبتنی بر زاویه می‌توان با موفقیت برای بازشناسی گفتار و گوینده استفاده کرد [۱٦و۱۷].

علاوه بر این، اهمیت اطلاعات زاویه در بازشناسی

هیجان گوینده در تحقیقات مختلف نیز نشان داده شده است. نالینی، پالانیول در سال ۲۰۱۳ [۱٤]، با استفاده از ویژگی زاویه باقی‌مانده و MFCC به بازشناسی هیجانات گوینده پرداختند. مدل آن تحقیق، هیجانات گفتار را به دسته‌های از پیش تعریف شده مانند خشم، ترس، شادی، خنثی یا غمگین دسته‌بندی می‌کرد. نتایج این تحقیق نشان داد که نرخ خطای سیستم با استفاده از ویژگی‌های MFCC 20% و با ترکیب هر دو، زاویه باقی‌مانده و ویژگی‌های MFCC، نرخ خطا تا ۱٦٪ کاهش می‌یابد.

در سال ۲۰۱٦، نیز در تحقیق دیگری از اطلاعات زاویه برای بازشناسی هیجانات در گفتار استفاده شده است [۱۸]. مولایی و جوزف‌کولمر در همین سال، یک روش تخمین زاویه هارمونیک با تکیه بر فرکانس پایه ارائه دادند [۷]. همچنین، مالی و مولایی اهمیت زاویه هارمونیک را در بهبود بازسازی سیگنال گفتار نشان دادند [۱۹].

همان‌طور که ذکر شد، تحقیقات اخیر، بیشتر به بررسی اثر اطلاعات زاویه سیگنال گفتار در کاربردهای مختلف پرداخته‌اند. در مورد تاثیر اطلاعات زاویه بر دسته‌بندی یا شناسایی لهجه تاکنون تحقیقات زیادی انجام نشده است. به تحقیقات انجام شده در زمینه دسته‌بندی لهجه، بدون در نظر گرفتن ویژگی‌های مبتنی بر زاویه، می‌توان به تحقیق گی در سال ۲۰۱۵ اشاره کرد [۲۰]. مدل این تحقیق از ویژگی‌های نرمال شده PLP استفاده کرد.

ربیعی و ستایشی در سال ۲۰۱۲ [۹]، بهترین و موثرترین ویژگی‌ها از نظر مقاومت در برابر نوفه برای دسته‌بندی لهجه را شناسایی کردند. نتایج این تحقیق، نشان داد که پیشگویی خطی ادراکی، تبدیل طیفی نسبی پیشگویی خطی ادراکی و ضرایب پیشگویی خطی، تحت هر دو شرایط تمیز و نوفه‌دار عملکرد خوبی دارند. همچنین، بررسی اهمیت ویژگی‌های مبتنی بر زاویه سیگنال گفتار در دسته‌بندی لهجه در شرایط بدون نوفه در تحقیقی در سال ۲۰۱٦ انجام شد [۲۱].

۳- مدل پیشنهادی شناسایی لهجه

هدف از این تحقیق، بررسی اهمیت ویژگی‌های مبتنی بر زاویه در دسته‌بندی لهجه است. همان‌طور که در شکل ۲ نشان داده شد، مدل پیشنهادی جهت شناسایی یا دسته‌بندی لهجه‌های گفتار در این تحقیق، شامل مراحل استخراج اسپکتروگرام، استخراج ویژگی و سپس دسته‌بندی لهجه با استفاده از شبکه عصبی مصنوعی می‌باشد.

نمایش زمان-فرکانس سیگنال شامل اطلاعات و ویژگی‌های مهم‌تری نسبت به نمایش حوزهٔ زمان آن است. از این رو، در اولین مرحله از روش پیشنهادی، بعد از فیلتر پیش‌تاکید، جهت تقویت فرکانس‌های بالا، قاب‌های ۲۵ میلی‌ثانیه از گفتار پیوسته با شیفت ۱۰ میلی‌ثانیه استخراج شده است. به عبارت دیگر هر قاب با قاب بعدی و قبلی ۱۵ میلی‌ثانیه همپوشانی دارد. سپس قاب‌ها در پنجره همینگ ضرب شده‌اند تا اثر لبه‌های قاب‌ها کاهش یابد. تبدیل فوریه زمان کوتاه، روی هر قاب، اطلاعات طیفی سیگنال گفتار را به‌صورت اعداد مختلط نمایش می‌دهد.

در دومین مرحله از روش پیشنهادی، ترکیبی از یک بردار ویژگی معمولی (یا مستخرج از اندازه اسپکتروگرام) به همراه یک ویژگی مبتنی بر زاویه، به‌عنوان ویژگی ترکیبی محاسبه می‌شود. یکی از فرضیه‌های تحقیق این است که ترکیب ویژگی مبتنی بر فاز در کنار یک ویژگی برداری معمولی، به ارتقاء شناسایی لهجه کمک می‌کند. ویژگی‌های برداری معمولی در بخش ۱-۱ معرفی شدند. همچنین به دلیل مقایسه ویژگی مبتنی بر زاویه پیشنهادی به چهار ویژگی مبتنی بر فاز در بخش ۱-۲ اشاره شد. ویژگی پیشنهادی در ادامه شرح داده می‌شود.

۳-۱- ویژگی پیشنهادیِ زاویهٔ اَبعاد کاهش یافته

برای استخراج ویژگی پیشنهادیِ زاویهٔ اَبعاد کاهش یافته، مراحل زیر انجام می‌پذیرد:

۱- تبدیل سیگنال از فضای زمان به فضای زمان-فرکانس

۲- استخراج زاویه از سیگنال

۳- نرمال‌سازی زاویه اسپکتروگرام با رابطه (۵)

٤- استخراج زاویه نرمال شده زیر ۱ کیلوهرتز

۵- اعمال تحلیل اجزاء اصلی (PCA)[4] روی زاویه مرحله چهارم

مراحل ۱ تا ۳، مشابه بخش‌های گذشته است. در مرحله چهارم، زاویه‌های زیر ۱ کیلوهرتز جدا می‌شود. در فرکانس‌های بالا (معمولا بالای ۱ کیلوهرتز) اطلاعات زاویه سیگنال به دلیل نوسانات زیاد، دقیق نیست. در نظر گرفتن اطلاعات زیر ۱ کیلوهرتز، اطلاعات مفیدتری از زاویه سیگنال گفتار در اختیار قرار می‌دهد.

سپس در مرحله ۵، اجزاء اصلی ویژگی‌ها استخراج می‌شوند. در واقع در این مرحله، اَبعاد بردار ویژگی کاهش می‌یابد و اطلاعات با ارزش این ویژگی‌ها به‌عنوان ویژگی مبتنی بر زاویهٔ اَبعاد کاهش یافته در نظر گرفته می‌شود.

تحلیل اجزاء اصلی، یک فرایند آماری است که طی آن یک مجموعه بردار (در این تحقیق بردار ویژگی ترکیبی) به یک فضای مختصاتی ناهمبسته (عمود) تبدیل می‌شوند. از این تکنیک برای فشرده‌سازی و استخراج اطلاعات موثر استفاده می‌شود.

لازم به ذکر است در این تحقیق پس از کاهش اَبعاد، از پنج مؤلفهٔ ابتدایی حاصل به‌عنوان ویژگی استفاده شده است.

۳-۲- دسته‌بندی کننده

بعد از استخراج ویژگی‌ها در مدل پیشنهادی، یک دسته‌بندی کننده به بازشناسی لهجه‌های گوینده‌ها می‌پردازد. در این مدل، از یک شبکهٔ عصبی دولایه پیشرو پرسپترون برای دسته‌بندی بردارهای ویژگی استفاده شده است. تعداد نرون‌های لایهٔ ورودی برابر تعداد ویژگی‌ها و تعداد نرون‌های لایهٔ مخفی با آزمون و خطا مشخص می‌گردد. تعداد نرون‌های لایهٔ آخر نیز برابر با تعداد لهجه‌ها می‌باشد. از تابع انتقال تانژانت سیگموید (Tansig) برای نرون‌های پنهان و خروجی شبکه استفاده شده است.

4- Principle Component Analysis (PCA)

جدول ۲: واکه‌های صدادار در دادگان تیمیت					
واکه‌های صدادار در تیمیت	iy	ih	eh	ae	aa
	ux	ey	ah	ao	oy
	er	uh	uw	ow	ax
	ix	axr	aw	ay	ax-h

تابع آموزش استفاده شده در این تحقیق نیز مبتنی بر بهینه‌سازی لونبرگ- مارکواد (TRAINLM) می‌باشد.

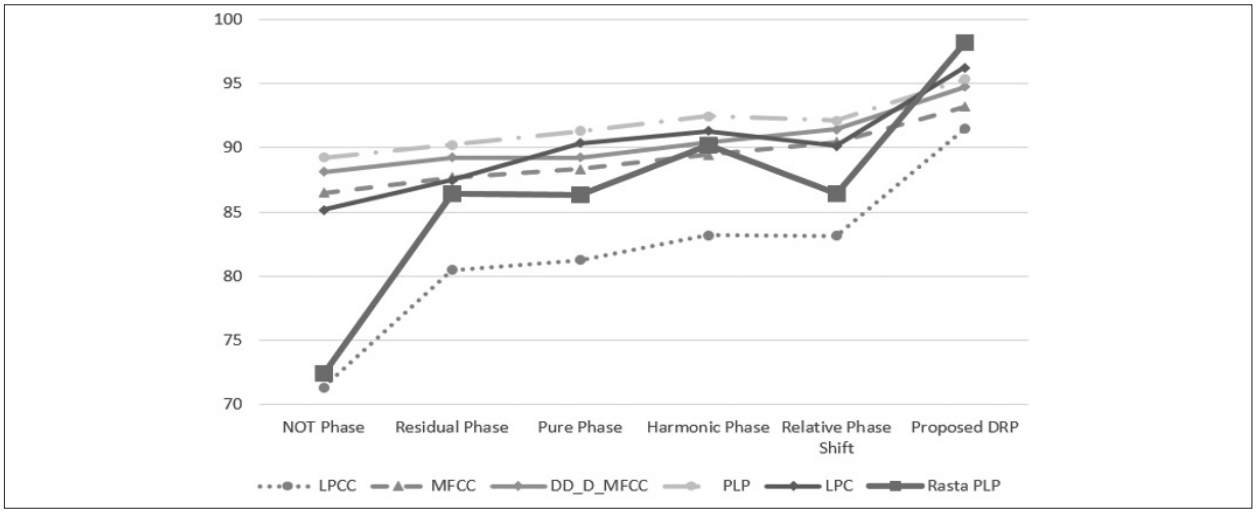
۴- آزمایش‌ها و نتایج

در کلیه آزمایش‌های این مقاله، از دادگان آماری (پایگاه داده) تیمیت [۲۲] استفاده شده است. این پایگاه داده شامل ٦۳۰ گوینده زن و مرد از ۸ منطقه یا لهجه معمول آمریکای شمالی می‌باشد و برای هر گوینده ۱۰ جمله ضبط شده است که طول هر جمله بین ۲ تا ۵ ثانیه است. جملات ضبط شده در سطح واج[5] به صورت دستی برچسب‌دهی شده است. پردازش گفتار این تحقیق، در سطح واکه[6] است و فقط از واکه‌های صدادار استفاده شده است. لیست واکه‌های صدادار استخراج شده از هر سیگنال در جدول ۲ نشان داده شده است.

در تمام آزمایش‌های این تحقیق از اعتبارسنجی[7] استفاده شده است. به این معنی که این آزمایش‌ها برای هر لهجه ۱۰ بار در شرایط بدون نوفه و با سیگنال به نوفه‌های ۱۰، ۵ و صفر تکرار شده و ارقام جداول، میانگین نتایج این ۱۰ بار آزمایش هستند. تقسیم‌بندی داده‌ها در هر بار اجرا، به صورت کاملاً تصادفی می‌باشد. از کلیه نمونه‌های استخراج شده از این واکه‌ها، ٦۰٪ مجموعه داده‌ها برای آموزش، ۲۰٪ به‌عنوان مجموعهٔ داده‌های آزمایشی و ۲۰٪ باقیمانده برای تنظیم پارامترها در نظر گرفته شده است. علاوه براین، قبل از هر فرایندی کلیه نمونه‌ها بین مقادیر ۱ تا ۱- نرمال شده‌اند.

در تمام آزمایش‌ها از ۱۳ ضریب برای هر کدام از ویژگی‌های پیشگویی خطی، کپسترال پیشگویی خطی،

5- phoneme
6- vowel
7- Cross validation



شکل۵: بازدهی دسته‌بندی با ترکیب ویژگی‌های مبتنی بر زاویه و ویژگی‌های برداری معمولی طیفی

پیشگویی خطی ادراکی، تبدیل طیفی آن، کپسترال فرکانسی مل و مشـــتقات اول و دوم آن‌ها ، در ترکیب با ویژگی‌های مبتنی بر زاویه استفاده شده است. همچنین، برای ارزیابی سیستم از ماتریس درهم‌ریختگی[8] و منحنی ROC استفاده می‌شـــود. کلیه آزمایش‌ها و نتایج برای حالت دو دسته‌ای می‌باشـــد. همچنین در ادامه، جهت سادگی، از مخفف‌های لاتین ویژگی‌ها که در جدول ۱ ذکر شد استفاده شده است.

٤-١- شرایط بدون نوفه

در آزمایش‌های این تحقیـــق، ازچهار جفت گوینده زن و مرد از دو دســـتهٔ مختلف برای استخراج واکه‌ها استفاده می‌شـــود. در این صورت، تعداد نمونه‌های دودسته ۳٦۹۳ نمونه می‌باشد. در شکل ٥ ترکیب ویژگی پیشنهادی DRP و ویژگی‌های مبتنی بر زاویه با ویژگی‌های برداری معمولی طیفی در حالت دو دســـته‌ای و بـــر روی دادگان آزمایش نشـــان داده شده اســـت. در هر آزمایش، نتایج نشان داده شـــده در شـــکل ٥، مربوط به ترکیب یک ویژگی مبتنی بر زاویه با یکی از ویژگی‌های برداری معمولی طیفی است.

همان‌طور که در شکل ٥ نشان داده شده است، Rasta PLP بیشـــترین بازدهی را در ترکیب با ویژگی پیشنهادی DRP دارد. ویژگی‌هـــای PLP و Rasta PLP از ویژگی‌هایی هستند که از سیستم شنوایی انسان الهام گرفته‌اند. ویژگی پیشـــنهادی نســـبت دیگر ویژگی‌های زاویه از جمله زاویه هارمونیک و شیفت زاویه نسبی بالاترین کارایی را داشته اســـت. بهترین بازدهی سیستم دسته‌بندی لهجه به غیر از ویژگی پیشـــنهادی، حاصل از ترکیب زاویه هارمونیک با ویژگی‌های برداری معمولی طیفی می‌باشد.

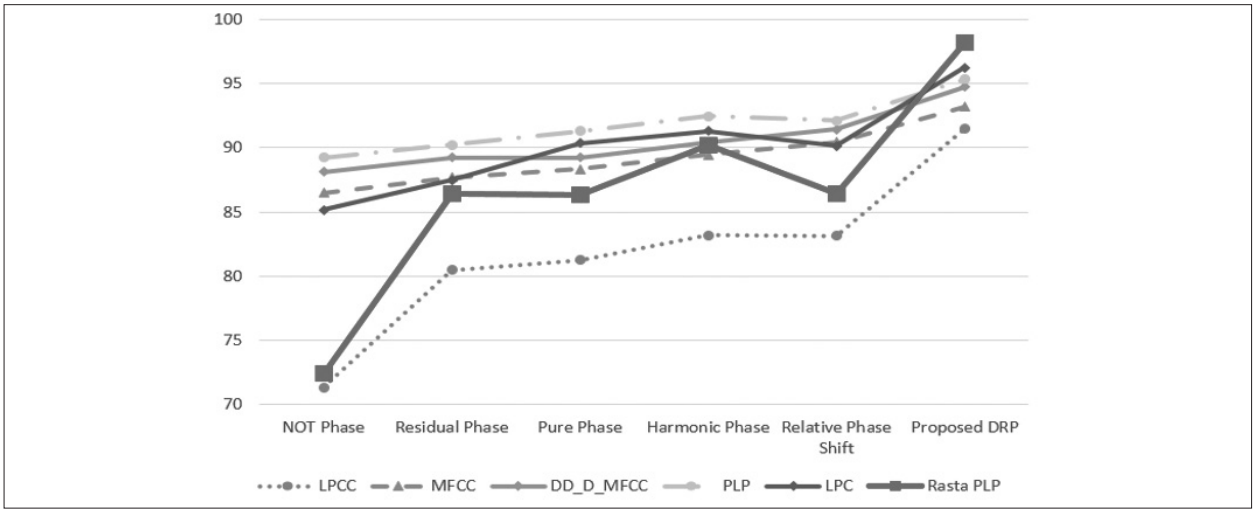
همان‌طور که شکل ٥ نشـــان می‌دهد، ویژگی مبتنی بر زاویه پیشنهادی DRP حدود ۳ تا ۸ درصد نسبت به زاویه هارمونیک بازدهی بیشـــتری داشته اســـت. همان‌طور که می‌بینید Rasta PLP در ترکیب با ویژگی پیشنهادی عملکرد بهتری نســـبت به بقیه ویژگی‌ها داشـــته و همچنین باعث افزایش کارایی سیستم تا ۸ درصد شده است.

٤-٢- اعمال نوفهٔ سفید گاوسی

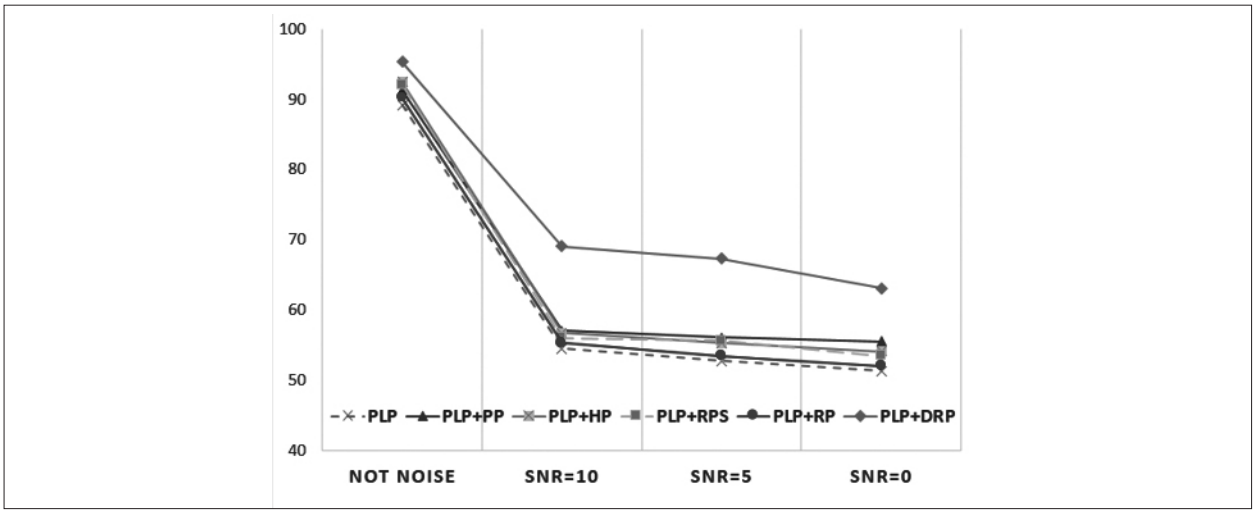
به منظور بررســـی مقاوم بـــودن ویژگی‌های مبتنی بر زاویه نســـبت به نوفه، از نوفهٔ سفید گاوسی استفاده شده اســـت. ویژگی‌های برداری معمولی طیفی در این آزمایش ویژگی‌هایی هستند که در آزمایش‌های فوق عملکرد بهتری در ترکیب با ویژگی‌های مبتنی بر زاویه داشته‌اند.

اشکال ٦، ۷، ۸ به ترتیب، نتیجه ترکیب ویژگی‌های مبتنی بر زاویه با سه ویژگی LPC، PLP و Rasta PLP و با مقادیر ســـیگنال به نوفه ۰، ٥ و ۱۰ را نشان می‌دهد. در این اشکال DRP در محیط‌های نوفه‌دار عملکرد بهتری نسبت به دیگر ویژگی‌های مبتنی بر زاویه را نشان می‌دهد. همان‌طور که در این سه شکل مشاهده می‌کنید بهترین عملکرد مربوط به

8- Confusion Matrix



شکل ۶: ترکیب ویژگی LPC با ویژگی‌های مبتنی بر زاویه در دو محیط نوفه‌دار و سالم



شکل ۷: ترکیب ویژگی PLP با ویژگی‌های مبتنی بر زاویه در دو محیط نوفه‌دار و سالم

Rasta PLP با ویژگی پیشنهادی در حالت‌های مختلف SNR نسبت به LPC و PLP می‌باشد. بازدهی سیستم پیشنهادی با Rasta PLP در محیط بدون نوفه برابر با ۷۲/٤۱ ٪ می‌باشد. اما در محیط نوفه‌دار برابر با ٥٤/۳۳٪ است. البته ترکیب با ویژگی‌های مبتنی بر زاویه و همچنین ویژگی DRP باعث افزایش عملکرد دسته‌بندی کننده شده است. همچنین، ویژگی پیشنهادی در ترکیب با PLP و LPC در محیط نوفه‌دار با مقادیر مختلف SNR نسبت به دیگر ویژگی‌ها عملکرد بهتری داشته است.

شکل ۸ ترکیب Rasta PLP با ویژگی پیشنهادی در محیط سالم و محیط نوفه‌دار با SNR مختلف نشان داده شده است. مقادیر ۹۰/۱۹ ٪، ۹۳/۹۱ ٪، ۹۵/۵۵ ٪ به ترتیب، ترکیب ویژگی مبتنی بر زاویه در حالت‌های SNR=5، SNR=0 و SNR=10 می‌باشد. تفاوت در این مقادیر اهمیت ویژگی DRP را نسبت به دیگر ویژگی‌های مبتنی بر زاویه نشان می‌دهد. نتایج به دست آمده تاثیر موثر زاویه و ویژگی پیشنهادی در محیط‌های سالم و نوفه‌دار را نشان می‌دهد.

در ادامه، به منظور ارزیابی بهتر عملکرد سیستم از دو معیار زیر نیز استفاده شده است:

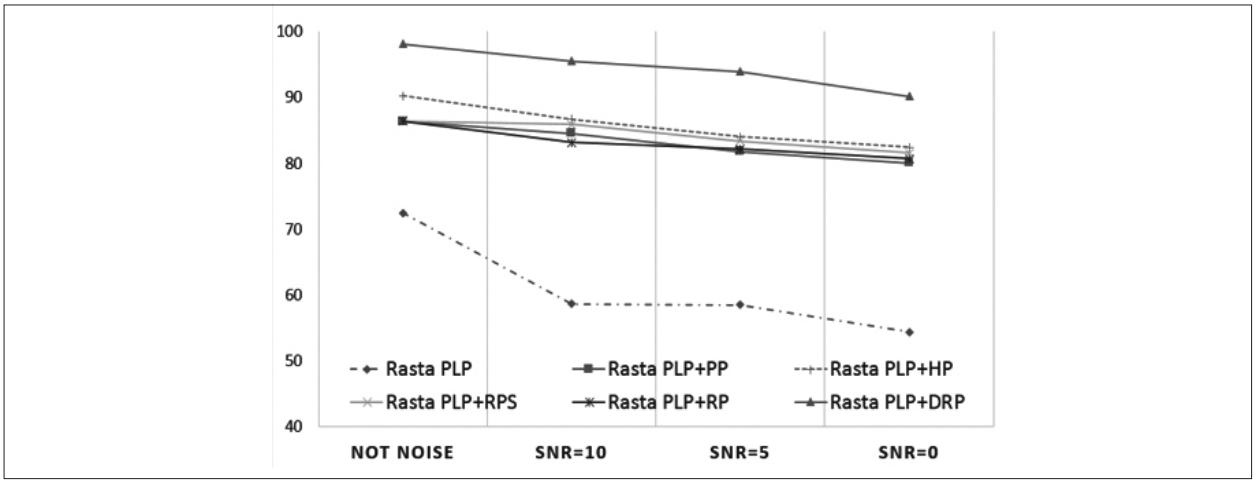
- ماتریس درهم ریختگی
- منحنی ROC

۴-۲-۱-ارزیابی عملکرد سیستم با ماتریس درهم‌ریختگی

ماتریس درهم‌ریختگی چگونگی عملکرد الگوریتم دسته‌بندی را با توجه به مجموعه داده ورودی به

جدول ۳: مقادیر ماتریس درهمریختگی با ویژگی Rasta PLP

ویژگی	دو دسته		دسته ۱		دسته ۲	
	بازدهی روی دادگان تست(٪)	تعداد دسته‌بندی غلط	بازدهی روی دادگان تست(٪)	تعداد دسته‌بندی غلط	بازدهی روی دادگان آزمایش(٪)	تعداد دسته‌بندی غلط
Rasta PLP	۷۲/۴۱	۳۱۲	۷۱/۲۲	۲۵۸	۷۶/۵۵	۱۳۵
Rasta PLP+PP	۸۶/۰۳	۲۳۴	۸۷/۲۵	۱۵۳	۸۸/۵۸	۱۰۴
Rasta PLP+HP	۹۰/۲۳	۲۰۶	۹۰/۴۱	۱۲۱	۸۹/۲۳	۱۳۸
Rasta PLP+RPS	۸۶/۴۳	۲۵۵	۸۶/۲۹	۱۹۲	۸۶/۳۵	۱۹۳
Rasta PLP+RP	۸۶/۳۸	۲۳۸	۸۶/۷۴	۱۶۱	۸۶/۷۳	۱۲۲
Rasta PLP+DRP	۹۸/۱۴	۱۶۸	۹۸/۷۰	۱۳۴	۹۷/۹۵	۵۳
Rasta PLP+DRP+SNR=10	۹۵/۵۵	۱۹۶	۹۱/۴۱	۱۶۱	۹۶/۷۱	۹۶
Rasta PLP+DRP+SNR=5	۹۳/۹۱	۲۲۶	۹۲/۳۵	۱۸۶	۹۵/۰۲	۱۳۷
Rasta PLP+DRP+SNR=0	۹۰/۱۹	۲۰۱	۹۰/۲۵	۱۶۴	۹۴/۷۱	۹۰



شکل ۸: ترکیب ویژگی Rasta PLP با ویژگی‌های مبتنی بر زاویه در دو محیط نوفه‌دار و سالم

تفکیک انواع دسته‌ها، نمایش می‌دهد [۱۵]. همان‌طور که در آزمایش‌ها نشان داده شده است. ویژگی Rasta PLP در ترکیب با ویژگی زاویه پیشنهادی بهترین عملکرد را داشته است. در جدول ۳ نمایش ماتریس درهم‌ریختگی برای حالت Rasta PLP در ترکیب با ویژگی‌های مبتنی بر زاویه و همچنین ویژگی پیشنهادی DRP در دو محیط سالم و محیط نوفه‌دار با مقادیر سیگنال به نوفه برابر ۵،۱۰ و صفر نشان داده شده است. ستون‌های این جدول، بازدهی و تعداد دسته‌بندی‌های غلط به‌صورت دو دسته‌ای و سپس در هر دسته، به تفکیک هستند.

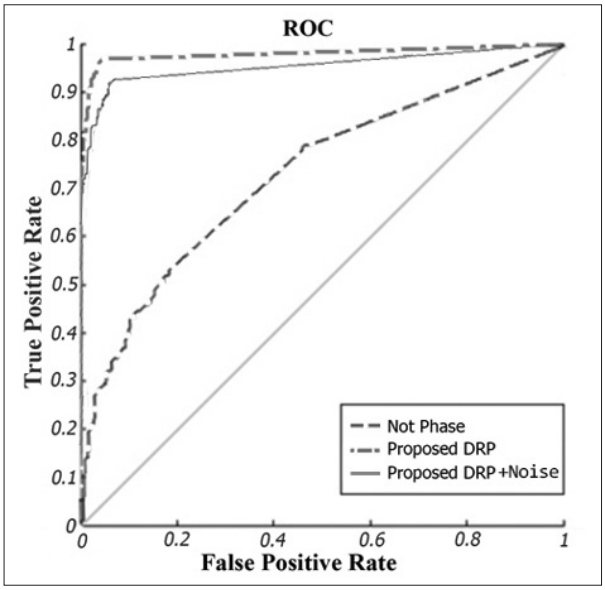
همان‌طور که در جدول ۳ نیز نشان داده شده است، ویژگی زاویه پیشنهادی نسبت به دیگر ویژگی‌ها عملکرد بهتری داشته است. ترکیب Rasta PLP با زاویه هارمونیک برابر با ۹۰/۲۳٪ است و زمانی که از زاویه پیشنهادی استفاده می‌کنیم به ۹۸/۱۴٪ می‌رسد. نرخ خطا در استفاده از زاویه پیشنهادی ۱/۸۶٪ و با سیگنال به نوفه‌های مختلف برابر با ۹/۸۱٪، ۶/۰۹٪ و ۴/۴۵٪ است. تعداد دسته‌بندی غلط درحالت عادی ۳۱۲ نمونه و در هنگام استفاده از زاویه پیشنهادی به ۱۶۸ نمونه می‌رسد، این مقدار نزدیک به ۱۹۶ نمونه در حالت SNR=10 می‌باشد.

۴-۲-۲- ارزیابی عملکرد سیستم با منحنی ROC

منحنی ROC می‌تواند مدل‌های بهینه را از غیر بهینه و

مستقل از محتوا و نوع دسته، از یکدیگر جدا کند و نشان دهنده بازدهی یک دسته‌بندی کننده باشد. شکل ۹ منحنی ROC برای زمانی که سیستم فقط از ویژگی Rasta PLP بدون زاویه استفاده می‌کند، با ویژگی زاویه پیشنهادی و با اعمال نوفهٔ سفید گاوسی با SNR=10 نشان می‌دهد.

همان‌طور که شکل ۹ نشان می‌دهد، محور افقی مربوط به هزینه و محور عمودی مربوط به مزایا و بازدهی است. بهترین دسته‌بندی کننده، مربوط به منحنی ROC ای است که هر چه بیشتر به سمت چپ و بالا نزدیک شده باشد. همان‌طور که در شکل ۹ نشان داده شده است، از نظر این معیار نیز ویژگی پیشنهادی بازدهی بازشناسی را حتی در محیط نوفه‌دار افزایش داده است.



شکل ۹: منحنی ROC برای ویژگی Rasta PLP، در ترکیب با ویژگی پیشنهادی، بدون زاویه و با اعمال نوفهٔ سفید گاوسی

۵- نتیجه‌گیری

در این تحقیق، یک ویژگی جدید مبتنی بر اطلاعات زاویه در سیگنال گفتار ارائه شده است. به طور مشخص، این مقاله به بررسی ویژگی‌های مبتنی بر زاویه (زاویه هارمونیک، شیفت زاویه نسبی، زاویه خالص شده، زاویه باقی‌مانده) و زاویه مشتق شده از تحلیل مؤلفه‌های اساسی طیف زاویه به‌عنوان ویژگی زاویه پیشنهادی پرداخته است. در ویژگی زاویه پیشنهادی، با استفاده از تحلیل اجزاء اصلی زاویه‌های موثرتر در قالب ویژگی کاهش اَبعاد یافته استخراج شده است. نتایج آزمایش‌ها نشان داد این ویژگی‌ها در ترکیب با سایر ویژگی‌های معمولی طیفی سیگنال گفتار مانند ضرایب کپسترال فرکانسی مل، ضرایب پیشگویی خطی می‌توانند کارایی سیستم دسته‌بندی و درصد دقت شناسایی لهجه‌ها را افزایش دهند.

مطالعات و نتایج به‌دست آمده از این تحقیق، نشان داد که ویژگی گفتاری Rasta PLP در ترکیب با ویژگی زاویه پیشنهادی در محیط سالم و نوفه‌دار، هرچند تعداد ورودی‌ها و پیچیدگی سیستم را افزایش می‌دهد ولی بهترین نتایج را ارائه می‌کند و بعضی از ویژگی‌ها مثل LPCC وMFCC کمترین بازدهی را در ترکیب با ویژگی‌های مبتنی بر زاویه و ویژگی زاویه پیشنهادی دارند. البته می‌توان از این ترکیبات در دیگر دسته‌بندی‌ها مانند ماشین بردار پشتیبان و kامین نزدیک‌ترین همسایه نیز استفاده کرد. در این تحقیق، اطلاعات زاویه، فقط از واج‌های صدادار استخراج شده است. یکی از پیشنهادات آینده این تحقیق، بررسی امکان استخراج اطلاعات زاویه از واج‌های بیصدای گفتار است. همچنین می‌توان در مورد اطلاعات زاویه در سیگنال‌های موسیقی نیز بررسی و مطالعه لازم را انجام داد.

مراجع

1-K. K. Paliwal and L. D. Alsteris, “Usefulness of phase spectrum in human speech perception,” in INTERSPEECH, 2003.

2-G. Shi, M. M. Shanechi, and P. Aarabi, “On the importance of phase in human speech recognition,” Audio, Speech, Lang. Process. IEEE Trans., vol. 14, no. 5, pp. 1867–1874, 2006.

3-L. Wang, K. Minami, K. Yamamoto, and S. Nakagawa, “Speaker recognition by combining MFCC and phase information in noisy conditions,” IEICE Trans. Inf. Syst., vol. E93–D, no. 9, pp. 2397–2406, 2010.

4-S. Nakagawa, K. Asakawa, L. Wang, K. Minami, and K. Yamamoto, “Speaker recognition by combin-

verification by combining information from magnitude and phase spectrum," In Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on, pp. 163-166. IEEE, 2016.
17-I. Saratxaga, J. Sanchez, Z. Wu, I. Hernaez and E. Navas, "Synthetic speech detection using phase information," Speech Communication 81 (2016): 30-41.
18-J. Deng, X. Xinzhou, Z. Zixing, F. Sascha and S. Björn, "Exploitation of Phase-Based Features for Whispered Speech Emotion Recognition," IEEE Access 4 (2016): 4299-4309.
19-A. Maly and P. Mowlaee, "On the importance of harmonic phase modification for improved speech signal reconstruction," In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, pp. 584-588. IEEE, 2016.
20-Z. Ge, "Improved accent classification combining phonetic vowels with acoustic features," In Image and Signal Processing (CISP), 2015 8th International Congress on, pp. 1204-1209. IEEE, 2015.
21-A. FalakRafar and A. Rabiee, "On the improtance of the phase of the speech signal for accent classification," in 4th Int. Con. Applied Research in Computer Engnieering and Signal Processing, Tehran, Iran, Dec. 2016.
22-J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, V. Zue, TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.
ing MFCC and phase information," spectrum, vol. 60, no. 700Hz, pp. 74–76, 2007.
5-H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. Am., vol. 87, no. 4, pp. 1738–1752, 1990.
6-S. Molau, M. Pitz, R. Schluter, and H. Ney, "Computing Mel-frequency cepstral coefficients on the power spectrum," 2001 IEEE Int. Conf. Acoust. Speech, Signal Process. Proc. (Cat. No.01CH37221), vol. 1, pp. 1764–1769, 2001.
7-P. Mowlaee and J. Kulmer, "Harmonic phase estimation in single-channel speech enhancement using phase decomposition and SNR information," IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP) 23, no. 9 (2015): 1521-1532.
8-H. Hermansky, N. Morgan, A. Bayya, and P. Kohn, "RASTA-PLP speech analysis technique," in icassp, 1992, pp. 121–124.
9-A. Rabiee and S. Setayeshi, "Robust and Optimum Features for Persian Accent Classification," in Neural Information Processing, 2012, pp. 441–449.
10-S. V Stehman, "Selecting and interpreting measures of thematic classification accuracy," Remote Sens. Environ., vol. 62, no. 1, pp. 77–89, 1997.
11-I. Saratxaga, I. Herna, D. Erro, E. Navas, J. Sa, I. Hernaez, D. Erro, E. Navas, and J. Sanchez, "Simple representation of signal phase for harmonic speech models," Electron. Lett, vol. 45, no. 7, pp. 381–383, 2009.
12-J. R. Deller, J. H. L. Hansen, and J. G. Proakis, Discrete-Time Processing of Speech Signals, IEEE Press, 2000.
13-I. Saratxaga, I. Hernáez, I. Odriozola, E. Navas, I. Luengo, D. Erro, I. Hernaez, I. Odriozola, E. Navas, I. Luengo, and D. Erro, "Using harmonic phase information to improve ASR rate.," in INTERSPEECH, 2010, no. September, pp. 1185–1188.
14-N. J. Nalini, S. Palanivel, and M. Balasubramanian, "Speech Emotion Recognition Using Residual Phase and MFCC Features," Int. J. Eng. Technol., vol. 5, no. 6, pp. 4515-4527, 2013.
15-A. M, I. Hernáez, I. Saratxaga, J. Sanchez, E. Navas, and I. Luengo, "Use of the Harmonic Phase in Speaker Recognition.," in INTERSPEECH, 2011, no. August, pp. 2757–2760.
16-K. K. Jain, K. M. G, V. R. Pai and N. K. C, "Speaker