

مدل سازی غیرخطی قطع پوانکاره سیگنال گفتار در ترکیب با تحلیل حوزه فرکانس به منظور افزایش صحت عملکرد سیستمهای بازشناسی گفتار

ایوب جعفری^{۱*}، فرشاد الماس گنج^۲، مریم نبی بیدهدنی^۳

۱- استادیار دانشکده مهندسی پزشکی دانشگاه آزاد اسلامی قزوین

۲- دانشیار دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر

۳- کارشناس ارشد دانشکده کامپیوتر دانشگاه صنعتی امیرکبیر

ajafari20@qiau.ac.ir

(دریافت مقاله: دی ۱۳۸۸، پذیرش مقاله: آذر ۱۳۸۹)

چکیده- در این مقاله روشی جدید برای افزایش صحت سیستمهای بازشناسی گفتار، با استفاده از ترکیب بردارهای ویژگی به دست آمده از مدل سازی غیرخطی فضای فاز بازسازی شده سیگنال گفتار با ویژگیهای معمول به دست آمده از تحلیل حوزه فرکانس ارائه می شود. بر اساس نظریه پذیرفته شده کنونی، در صورت انتخاب تعداد بُعد کافی برای بازسازی فضای فاز سیگنال، این فضا به صورت کامل دنبامیک سیستم تولید کننده آن را نشان می دهد و بنابراین می تواند شامل اطلاعات مفیدی باشد که در تحلیل فرکانسی معمول - مانند استخراج ضرایب کیسترال مل^۱ MFCC - در دسترس نیست. همچنین از آنجا که سیستمهای پیچیده ای مانند سیستم تولید گفتار، رفتارهایی نوسانی و تناوبی نشان می دهند، قطع پوانکاره می تواند به عنوان ابزاری مؤثر در تحلیل این رفتارها به کار رود. در این مقاله نوعی مدل سازی غیرخطی با استفاده از مدل مخلوط گوسی (GMM^۲) بر روی نقاط قطع پوانکاره سیگنال گفتار انجام می شود. بردار ویژگی نهایی از یک مرحله انتخاب ویژگی بر روی پارامترهای مدل مخلوط گوسی و بردار ویژگی های فرکانسی معمول MFCC حاصل می شود. از یک سیستم بازشناسی مبتنی بر مدل مخفی مارکوف (HMM^۳) و پایگاه داده تیمیت^۴ به منظور ارزیابی سیستم ارائه شده استفاده شده است. نتایج آزمایشها بهبودی در حدود ۵/۷ درصد در صحت بازشناسی گفتار واج گسسته را نسبت به حالت استفاده از ویژگیهای معمول MFCC نشان می دهد. این روش جدید می تواند به عنوان روشی مؤثر و جایگزین در روشهای استخراج ویژگی به ویژه برای سیستمهای با ماهیت غیرخطی شدید استفاده شود.

کلیدواژگان: فضای فاز بازسازی شده، قطع پوانکاره، بازشناسی گفتار، استخراج ویژگی، مدل سازی غیرخطی.

1. Mel Frequency Cepstral Coefficients
2. Gaussian Mixture Model
3. Hidden Markov Model
4. TIMIT

۱- مقدمه

برای سالیان متوالی بردارهای ویژگی استخراج شده در حوزه فرکانس مانند ضرایب کپسترال فرکانسی مل - که از تفسیر ادراکی عملکرد گوش انسان بر روی طیف فرکانسی سیگنال گفتار به دست می آیند - در سیستمهای بازشناسی گفتار با موفقیت محدودی استفاده شده اند.

برخی سیگنالها به ویژه سیگنالهایی که از معادلات دیفرانسیل غیرخطی به دست می آیند، مشخصات طیفی عریض و نامشخصی دارند و در این موارد، تحلیل فرکانسی ابزار چندان مناسبی به نظر نمی رسد، زیرا امکان تفکیک اطلاعات سیگنال در حوزه فرکانس وجود ندارد [۱]. به عنوان مثال شکل (۱) سیگنال زمانی، طیف فرکانسی و فضای فاز بازسازی شده را برای نگاشت لاجستیک با معادله $x_{n+1} = k \times x_n \times (1 - x_n)$ به ازای k برابر ۴ نشان می دهد [۲].

دلایل نظری و عملی مختلفی در برخی تحقیقات اخیر ([۲] [۳] [۴] [۵]) مطرح شده که حضور پدیده غیرخطی قویی را در سیستم تولید گفتار نشان می دهند که در مدل های خطی در نظر گرفته نمی شوند. این موضوع یکی از دلایل محدودیت در صحت عملکرد سیستمهای بازشناسی گفتار امروزی است. یکی از پدیده های غیرخطی موجود در سیگنال گفتار پدیده توربولانس است [۵].

از سوی دیگر موفقیتی که نظریه سیستمهای آشوبگونه در برخورد با سیستمهای دارای توربولانس داشته است، در نظر گرفتن پدیده توربولانس موجود در سیگنال گفتار به ویژه در واج های سایشی را مطرح کرده و ما را به استفاده از این نظریه در سیستمهای بازشناسی گفتار رهنمون می شود. در کاربردهای معمول نظریه آشوب در پردازش سیگنال به استخراج برخی ویژگیهای معمول در

این حوزه مانند ضرایب نمایی لیاپانوف و بُعد فراکتال پرداخته می شود. در استخراج این ویژگیها به جای استفاده از سیگنال یک بُعدی گفتار، از فضای چند بُعدی فاز بازسازی شده از سیگنال گفتار با استفاده از روشی مبتنی بر تأخیرهای زمانی - به نام فرایند پوشش - استفاده می شود [۶] [۷] [۸]. فرایند پوشش از بازسازی مجموعه ای از جفت های ورودی - خروجی بر اساس تأخیر نمونه های سیگنال زمانی یک بُعدی انجام می شود. به دلیل تعداد کم نمونه های موجود در قاب های گفتار، قابلیت تعمیم پذیری فضای بازسازی شده، یکی از مهمترین چالش ها در به کارگیری روش پوشش در سیستمهای بازشناسی گفتار است.

بسیاری از تحقیقات در زمینه استفاده از فضای فاز بازسازی شده، بر روی تخمین پارامترهای اندازه (ضرایب لیاپانوف و بُعد فراکتال)، چگالی [۹] و توپولوژی تمرکز کرده اند. مدل مخلوط گوسی روشی مناسب برای تخمین چگالی نقاط در فضای فاز است؛ اما تحقیقات بسیار محدودی در زمینه استفاده از مدل سازی فضای فاز در کاربردهای بازشناسی گفتار انجام شده است.

استفاده از ویژگیهای آشوبگونه معمول مانند ضرایب لیاپانوف در بازشناسی گفتار به خوبی در تحقیقات قبلی مطالعه شده است [۸] [۹] [۱۰] [۱۱]. این تحقیقات به روشنی بیان می کنند که استفاده از این دسته ویژگیها کمک چندان به بهبود صحت سیستمهای بازشناسی گفتار نمی کنند.

همان گونه که شکل (۱) نشان می دهد، بازسازی در فضای فاز، نمایش بسیار بهتری را نسبت به طیف فرکانسی به دست می دهد. لذا با توجه به مسائل مطرح شده و با توجه به ماهیت غیرخطی دینامیک سیستم تولید گفتار، به استفاده از فضای فاز در ترکیب با تحلیل حوزه

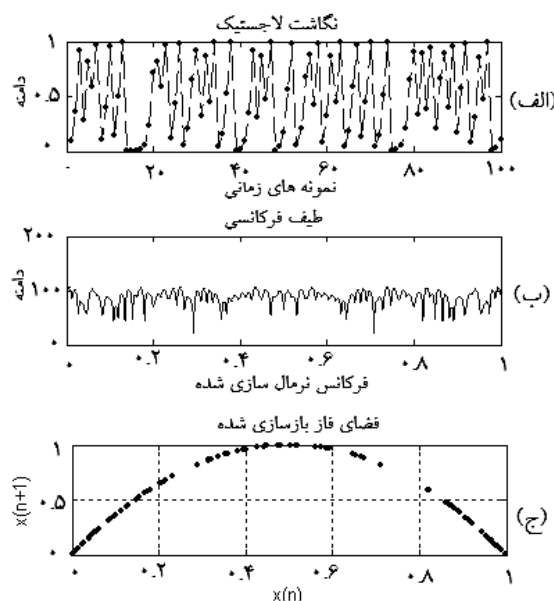
اطلاعات در حوزه فاز است. دوم آن که به جای استفاده از ویژگیهای استخراج شده از حوزه فاز به تنهایی، بردار ویژگی ترکیب این اطلاعات با اطلاعات حوزه فرکانس به کار گرفته شد.

سازمان این مقاله چنین است. در بخش ۲ کلیات سیستم پیشنهادی ارائه می شود. بخش ۳ فرایند پوشش به منظور بازسازی فضای فاز سیگنال گفتار را به صورت خلاصه بیان می کند. قطع پوانکاره و الگوریتم مورد استفاده در این تحقیق در این بخش ارائه می شود. در بخش ۴ مدل سازی غیرخطی مبتنی بر روش مدل مخلوط گوسی بیان می شود. در بخش ۵ به معرفی روش پیشنهادی به منظور ترکیب ویژگیهای فضای فاز و فرکانس پرداخته می شود. بخش ۶ آزمایشها و دادگان مورد استفاده را توضیح می دهد. بخش ۷ شامل مباحثی درباره روش پیشنهادی و مزایا و معایب آن بوده و با ارائه جمع بندی کلی در بخش ۸ مقاله به پایان می رسد.

۲- ساختار روش پیشنهادی

همانگونه که در مقدمه گفته شد، مبنای روش پیشنهادی، مدل سازی غیرخطی قطع پوانکاره به دست آمده از فضای فاز بازسازی شده سیگنال گفتار و ترکیب اطلاعات استخراج شده از این بخش با ویژگیهای فرکانسی معمول کپسترال حوزه مل است. شکل (۲) کلیات روش پیشنهادی را نشان می دهد. همانگونه که در شکل (۲) نشان داده شده، نخست فضای فاز سیگنال گفتار با استفاده از فرایند پوشش مبتنی بر تأخیر زمانی نمونه های سیگنال زمانی یک بُعدی - که در بخش سوم معرفی می شود - ساخته می شود. در مرحله بُعدی مدل سازی مبتنی بر مدل ترکیبهای گوسی بر روی نمونه های به دست آمده از محاسبه قطع پوانکاره از فضای فاز بازسازی شده سیگنال

فرکانس در مسأله بازشناسی گفتار رهنمون می شویم. پاورینی و همکاران [۱] نوعی مدل سازی آماری را بر روی فضای فاز بازسازی شده سیگنال مطرح کردند و نشان دادند که ویژگیهای استخراج شده عملکرد مناسبی را در کاربردهای طبقه بندی سیگنال نشان می دهد. با وجود این، کیفیت عملکرد روش آنها، در صورت استفاده از دادگان بزرگتر و به ویژه در بازشناسی گفتار پیوسته به شدت کاهش می یابد.



شکل ۱ الف) سیگنال زمانی، ب) طیف فرکانسی و ج) فضای فاز بازسازی شده برای سیستم لاجستیک به ازای پارامتر k برابر ۴

برای ارائه روشی مناسب برای استخراج ویژگی که قابل استفاده در سامانه های بازشناسی گفتار باشد، دو رویکرد را در این تحقیق مدنظر قرار داده ایم: اول آن که مدل سازی غیرخطی به جای اعمال به نقاط فضای فاز بازسازی شده، بر روی قطع پوانکاره سیگنال گفتار انجام شد. قطع پوانکاره، به ویژه برای سیگنال هایی با ماهیت شبه تناوب مانند سیگنال گفتار، ابزار مناسبی برای استخراج

جدایی پذیری فیشر است که در بخش ۵ معرفی می شود. استفاده از این روش بردار ویژگی ۵۲ بُعدی را به دست می دهد که بردار مناسبی برای استفاده در طبقه بندی کننده های مبتنی بر مدل مخفی ماکوف معمول در کاربردهای بازشناسی گفتار است.

۳- بازسازی فضای فاز از سیگنال زمانی و قطع پوانکاره

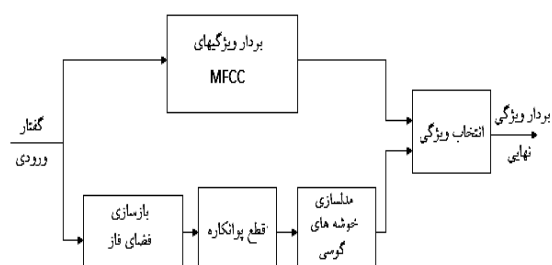
در این بخش به معرفی روش مورد استفاده در بازسازی فضای فاز سیگنال از سیگنال گفتار یک بُعدی و نیز الگوریتم مورد استفاده در استخراج قطع پوانکاره سیگنال می پردازیم.

۳-۱- بازسازی فضای فاز

بازسازی فضای فاز با استفاده از فرایندی به نام پوشش - که روشی مبتنی بر استفاده از نمونه های تأخیر یافته سیگنال زمانی است - انجام می شود. در این بخش نظریه بازسازی جاذب فاز سیگنال را مرور می کنیم. فرایند پوشش بر دستیابی به نگاهی مبتنی است که امکان دسترسی به دینامیک سیستم تولید کننده سیگنال را فراهم می سازد.

ویتنی [۱۲] نشان داد که اگر فضایی با بُعد d داشته باشیم، نگاشت $\mathbf{R}^d \rightarrow \mathbf{R}^{d+1}$ می تواند به عنوان نگاشتی پوشش از سیستم اولیه مطرح باشد. تیکنز [۱۳] نشان داد که f نگاشتی است که دینامیک سیستم اولیه را توصیف می کند. پیشرفت اصلی در مشخص شدن نگاشت پوشش از کار سائر و یورک [۱۴] به دست آمد که نشان دادند هر نگاشت تأخیر زمانی با بُعد بازسازی کافی می تواند نوعی نگاشت پوشش باشد و همخوانی با دینامیک سیستم اولیه را تضمین می کند. همچنین آنان

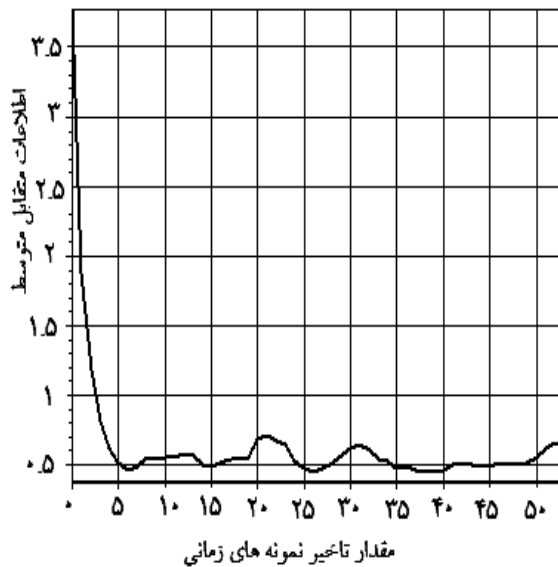
گفتار اعمال و بردار ویژگی نهایی به ورودی طبقه بندی کننده مبتنی بر مدل مخفی مارکوف اعمال می شود. بردار ویژگی به دست آمده، شامل پارامترهای مدل ترکیب گوسی مانند ضریب بهره، بردار میانگین و مقادیر واریانس خوشه های گوسی است. در این تحقیق از مدل ترکیب گوسی با چهار خوشه استفاده شده است. در بخش (۳) نشان داده می شود که بُعد بهینه برای بازسازی فضای فاز در سیگنال گفتار برابر هشت است و در نتیجه بُعد قطع پوانکاره به دست آمده برابر خواهد بود.



شکل ۲ ساختار کلی سیستم پیشنهادی

لذا برای هر خوشه گوسی موجود در مدل مخلوط گوسی، یک ضریب بهره، هفت ضریب بردار میانگین و هفت ضریب ماتریس واریانس در دست داریم. لذا به ازای هر خوشه، تعداد ویژگیهای مورد نیاز برابر ۱۵ و در مدل چهار خوشه ای مورد استفاده در این تحقیق بُعد بردار نهایی به دست آمده برابر ۶۰ خواهد بود. با توجه به بُعد ۳۹ برای بردار ویژگیهای کپسترال حوزه مل، بُعد نهایی بردار ویژگی برابر ۹۹ می شود که برای کاربردهای بازشناسی گفتار چندان مناسب نیست. لذا در ترکیب ویژگیهای فضای فاز و فضای فرکانس در مرحله نهایی، یک مرحله انتخاب ویژگی داریم. مبنای این انتخاب، معیار بیشترین تمایز پذیری با استفاده از تحلیل

معیار، مقدار بهینه پارامتر تأخیر زمانی است. شکل (۳) مقادیر محاسبه شده را به ازای مقادیر مختلف تأخیر زمانی نشان می دهد. همان گونه که در شکل (۳) نشان داده شده، مقدار بهینه پارامتر تأخیر زمانی برابر ۶ است. پس از انتخاب پارامتر تأخیر زمانی نوبت به تعیین بُعد بهینه برای فضای فاز بازسازی شده از فرایند پوشش می رسد. در این تحقیق از معیار کمینه سازی همسایگی های غلط استفاده شده است. همسایگی های غلط، نقاطی بر روی فضای فاز بازسازی شده است که از یکدیگر دور هستند. بُعد فضای پوشش به گونه ای انتخاب می شود که درصد این همسایگی ها کمینه شود. با استفاده از این ابزار، بُعد بهینه برای بازسازی فضای فاز برابر هشت انتخاب شد. شکل (۴) فضای فاز دو بُعدی بازسازی شده را برای واج $a/$ از پایگاه داده تیمیت نشان می دهد.



شکل ۳ مقادیر معیار اطلاعاتی دوسویه به ازای مقادیر مختلف تأخیر زمانی (اولین کمینه در نقطه ۶)

دریافتند که اگر d بُعد جاذب فاز به دست آمده باشد، انتخاب $d > 2d$ تضمین می کند که این نگاهت دینامیک سیستم اولیه را توصیف کند. لذا بر مبنای این نظریه، می توان از نگاهت های مبتنی بر تأخیر زمانی برای نگاهت سیگنال یک بُعدی به فضای فاز چند بُعدی استفاده کرد.

اگر $x = \{x_1, \dots, x_n\}$ نقاط سری زمانی سیگنال گفتار یک بُعدی باشد، ماتریس X با بُعد d و تأخیر زمانی τ_d به شکل زیر تعریف می شود.

$$X = \begin{bmatrix} x_{1+(d-1)\tau_d} \\ x_{2+(d-1)\tau_d} \\ \vdots \\ x_N \end{bmatrix} \quad (1)$$

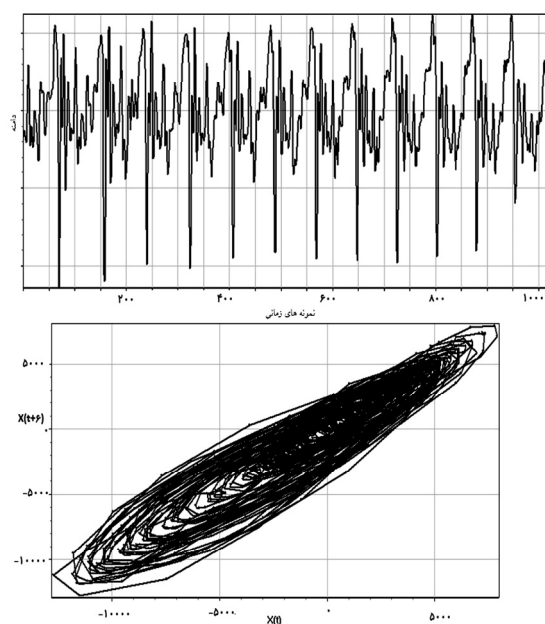
$$= \begin{bmatrix} x_{1+(d-1)\tau_d} & \dots & x_{1+\tau_d} & x_1 \\ x_{2+(d-1)\tau_d} & \dots & x_{2+\tau_d} & x_2 \\ \vdots & & \vdots & \vdots \\ x_N & \dots & x_{N-(d-2)\tau_d} & x_{N-(d-1)\tau_d} \end{bmatrix}$$

مقدار بهینه تأخیر زمانی τ_d را می توان از همبستگی میان نمونه های سیگنال گفتار به دست آورد. نارایانان [۱۵] معیار متوسط اطلاعات متقابل را برای تخمین ضریب تأخیر زمانی به صورت زیر پیشنهاد کرد:

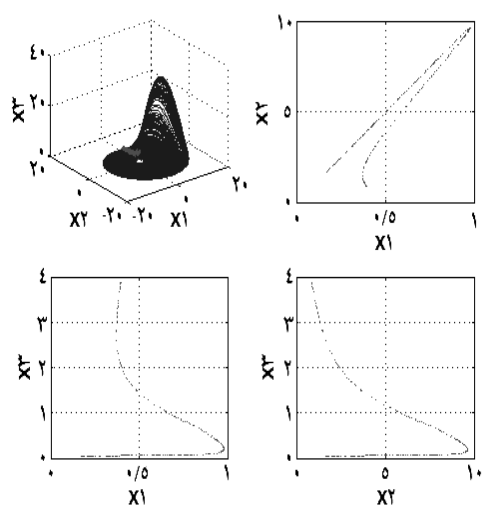
$$I(T) = \sum_{n=1}^{N-T} P(x(n), x(n+T)) \cdot \log_2 \left[\frac{P(x(n), x(n+T))}{P(x(n))P(x(n+T))} \right] \quad (2)$$

که در آن $x(n)$ نمونه سیگنال گفتار، P تابع احتمال و I معیار اطلاعاتی دوسویه است. اولین کمینه این

پوانکاره برای فضاهای فاز با بُعد بیشتر از سه، منجر به استخراج نقاطی در فضایی با بُعد $n-1$ برای فضای فازی با بُعد n می‌شود [۲۱]. با در نظر گرفتن این مسأله در واقع می‌توان گفت که قطع پوانکاره در واقع بُعد سیستم اولیه را یکی کاهش می‌دهد. با در نظر گرفتن رفتار تناوبی مشخص سیگنال گفتار به‌ویژه در بخشهای واکدار می‌توان چنین نتیجه گرفت که خط حرکت بازسازی شده در فضای فاز، بُعد از موقعیت هر پیچ گفتار تکرار می‌شود. شکل (۶) فضای فاز بازسازی شده سه‌بُعدی و قطع پوانکاره محاسبه شده را برای واج /a/ نشان می‌دهد. محاسبه قطع پوانکاره برای هر قاب از سیگنال گفتاری شامل مراحل زیر است:



شکل ۴ سیگنال زمانی و فضای فاز دوبُعدی بازسازی شده برای واج /a/ از پایگاه داده تیمیت



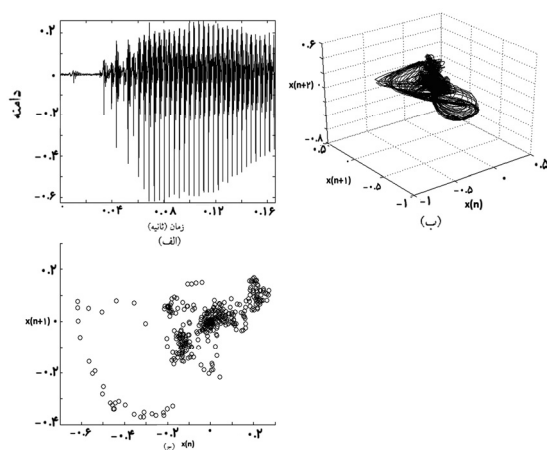
شکل ۵ قطع پوانکاره سامانه ریسر

- ۱- فضای فاز سیگنال برای قاب مورد نظر با استفاده از فرایند پوشش بازسازی می‌شود.
- ۲- برای نقطه اولیه $x(n_0)$ همسایگی $N(n_0)$ برای k نزدیکترین نقطه بر اساس فاصله اقلیدسی انتخاب می‌شود.

۳-۲- قطع پوانکاره

بررسی و نمایش خط حرکت در فضای فاز بازسازی شده با بُعد بالا - مانند فضای فاز سیگنال گفتار با بُعد هشت - معمولاً مشکل است. یکی از روشهای بر طرف ساختن این محدودیت، استفاده از تصویرسازی این خط بر روی برخی صفحات است. در میان روشهای موجود برای تصویرسازی، یکی از بهترین روشهای ارائه شده، متعلق به هنری پوانکاره است. قطع پوانکاره در واقع تصاویری از خط حرکت است که در بازه‌های خاصی برداشته می‌شود. برای سیستم دینامیکی با سه بُعد (X_1, X_2, X_3) ، انتخاب صفحه و انتخاب نقاطی که خط حرکت این صفحه را قطع می‌کند، قطع پوانکاره، یک خط را در فضای فاز ایجاد می‌کند. برای سیستمی با تعداد بُعد بالاتر از سه، سطوحی از صفحات به کار می‌رود. شکل (۵) نقاط پوانکاره سامانه ریسر را نشان می‌دهد. تعمیم قطع

سیگنال پیشنهاد کردند. پاوینلی و همکاران [۱] برخی روشهای مدل سازی آماری مانند روشهای مبتنی بر هیستوگرام را پیشنهاد کردند و این مدل ها را در آزمایشهایی برای بازشناسی واج مطالعه کردند.



شکل ۶ سیگنال زمانی، فضای فاز سه بُعدی بازسازی شده و قطع پوانکاره برای واج /a/ از پایگاه داده تیمیت

بر مبنای این تحقیقات، برای سیستم بازشناسی گفتار پیوسته مورد نظر در این تحقیق، روش مدل سازی غیرخطی قطع پوانکاره فضای فاز بازسازی شده با استفاده از مدل ترکیبات گوسی برای تحلیل قاب های گفتار انتخاب شد. از مدل مخلوط گوسی به صورت گسترده ای در پردازشهای گفتار استفاده می شود. این روش نقشی اساسی را در مدل مخفی مارکوف - که معمول ترین ابزار در بازشناسی گفتار است - بازی می کند. در این تحقیق، از مدل مخلوط گوسی به عنوان طبقه بندی کننده استفاده نشده بلکه برای مدل سازی توزیع دادگان در فضای نقاط قطع پوانکاره فاز بازسازی شده به کار رفته است. مدل ترکیبهای گوسی، دو مزیت اصلی در کاربرد مورد نظر برای مدل سازی توزیع دادگان دارد. اولین مزیت آن است

۳- جهت جریان متوسط $f(n_t)$ برای خط حرکت در همسایگی $N(n_t)$ از رابطه $f(n_t) = \text{mean}(x[n+1] - x[n]) \forall n \in N(n_t)$ محاسبه می شود.

۴- سطح پوانکاره عمود بر $f(n_t)$ در نقطه $X(n_t)$ تعیین می شود.

۵- برای به دست آوردن نقاط قطع پوانکاره، نقاط حاصل از قطع خط حرکت فضای فاز با سطح پوانکاره تعیین می شود.

۴- مدل سازی غیرخطی فضای فاز بازسازی شده

در پردازش سیگنال گفتار فرض ایستایی فقط برای سری های زمانی با طول ۲۰ تا ۲۶ میلی ثانیه اعتبار دارد. برین اساس، طول قاب های به کار رفته برای کاربردهای پردازش گفتار برابر این مقدار انتخاب می شود. با در نظر گرفتن فرکانس نمونه برداری ۱۶ کیلوهرتز برای دادگان گفتاری تیمیت، تعداد نمونه های هر قاب در حدود ۴۰۰ است. با در نظر گرفتن بُعد هشت برای فضای فاز بازسازی شده برای تشکیل هر همسایگی، به $8 * 8 + 8 = 72$ نمونه نیاز است که تقریباً ۱۷ درصد از کل نقاط را شامل می شود. بنابراین با مشکل کوتاه بودن طول سری زمانی روبه رو هستیم و این کمبود داده، یکی از مشکلات اصلی است که باعث می شود ویژگی های معمول موجود در تحلیل آشوب - مانند ضرایب لیاپانوف - عملکرد مناسبی را در بازشناسی گفتار نشان ندهند [۵]. برای غلبه بر این مشکل یکی از روشهای جالب، استفاده از مدل سازی غیرخطی توزیع دادگان در فضای فاز بازسازی شده است. ککینوس و همکاران [۵] برخی از روشهای مدل سازی غیرخطی مانند توابع چند جمله ای محلی، تخمینگرهای فازی و شبکه عصبی شعاعی را برای مدل سازی فضای فاز

۵- انتخاب ویژگی

هدف اصلی این تحقیق نشان دادن مزایای ترکیب اطلاعات حوزه فاز بازسازی شده و حوزه فرکانس در بازشناسی گفتار است. در این تحقیق برخی آزمایشهای بازشناسی واج را با استفاده از ویژگی‌های کپسترال حوزه مل و ویژگیهای استخراج شده از فضای فاز انجام شد تا ارزیابی توانایی‌های هر دسته از ویژگیها در مسأله بازشناسی گفتار امکان پذیر باشد. سپس این ویژگیها در یک بردار ویژگی واحد ترکیب شد. در ترکیب ویژگیها به یک بردار ۳۹ بُعدی ضرایب کپسترال حوزه مل و یک بردار ۶۰ بُعدی به دست آمده از پارامترهای مدل مخلوط گوسی می‌رسیم. بر این مبنا بُعد بردار نهایی به دست آمده برابر ۹۹ است. این بُعد برای بردار ویژگی، به دلایل زیر برای استفاده در کاربردهای بازشناسی گفتار پیوسته قابل قبول نیست:

- پیچیدگی محاسبات مورد نیاز در بازشناسی با افزایش بُعد بردار ورودی، به شدت افزایش می‌یابد.
- حجم حافظه مورد نیاز برای نگهداری پارامترهای مورد نیاز در محاسبات افزایش می‌یابد.
- صحت سیستم بازشناسی ممکن است بر اثر افزایش بُعد دادگان ورودی کاهش یابد.
- برای غلبه بر این محدودیت‌ها و برای افزایش درجه صحت عملکرد سیستم بازشناسی گفتار، در این تحقیق از روش تحلیل جدایی‌پذیری فیشر [۱۷] برای کاهش بُعد بردار ویژگی بردار ترکیب اطلاعات استفاده شده است.

۵-۱- تحلیل جدایی‌پذیری فیشر

همان‌گونه که گفته شد، یک بردار ۹۹ بُعدی در ورودی در دست داریم و به دلایل گفته شده، بهینه‌سازی این بردار ویژگی می‌تواند به بهبود عملکرد سیستمهای بازشناسی گفتار منجر شود، زیرا بسیاری از ویژگیها می‌توانند دارای

که در این مدل‌ها نواحی تصمیم‌دارای مرزهایی نرم هستند. مزیت دوم این مدل در این است که تعداد خوشه‌های مورد نیاز، با افزایش بُعد فضای پوشش، به صورت نمایی افزایش نمی‌یابد. این بدان معنا است که می‌توان با برازش مناسب مدل به توزیع دادگان با تعداد محدودی خوشه، مدل مناسبی را بازسازی کرد.

۴-۱- مدل مخلوط گوسی

در مدل مخلوط گوسی، توزیع احتمال نقاط از رابطه زیر تخمین زده می‌شود:

$$p(x_n) = \sum_{m=1}^M w_m p_m(x_n) = \sum_{m=1}^M w_m N(x_n; \mu_m, \sigma_m^2) \quad (3)$$

که در آن P تابع توزیع احتمال را نشان می‌دهد. پارامتر M تعداد خوشه‌ها را مشخص می‌کند و N تابع توزیع نرمال با میانگین μ_m و ماتریس کواریانس σ_m^2 است. w_m ضریب وزن خوشه m است. در مرحله آموزش، پارامترهای مدل را می‌توان با استفاده از الگوریتم بیشینه‌سازی انتظار به دست آورد [۱۶]. از آنجا که بُعد بهینه تعیین شده برای فضای فاز بازسازی شده برابر هشت انتخاب شد، برای بُعد هفت قطع پوانکاره هر خوشه گوسی دارای پارامترهای زیر است:

- یک ضریب بهره
 - هفت عدد شامل بردار میانگین خوشه
 - هفت عدد شامل عناصر قطری ماتریس کواریانس
- بنابراین برای هر خوشه گوسی تعداد ویژگیهای به دست آمده برابر ۱۵ است. با توجه به تعداد چهار خوشه گوسی استفاده شده در مدل‌سازی‌ها در نهایت، یک بردار ویژگی ۶۰ تایی، نتیجه مدل‌سازی غیرخطی فضای فاز قاب گفتار خواهد بود.

ماتریس S_B را می توان از رابطه زیر به دست آورد.

$$S_B = S_T - S_W \quad (۸)$$

پس از محاسبه ماتریس های مورد نیاز، از رابطه (۴) می توان برای مسأله ای با K کلاس خروجی (مانند مسأله بازشناسی گفتار) معیار جدایی پذیری فیشر را محاسبه کرد. بردار ویژگی دارای ضریب فیشر بالاتر، توانایی بیشتری در تمایز کلاسهای خروجی دارد. بنابراین در الگوریتم انتخاب ویژگی در این تحقیق، با حذف ویژگیهایی که ضریب فیشر پایین تری داشتند، با انجام آزمون بازشناسی گفتار، تأثیر حذف این ویژگی را در صحت سیستم بازشناسی اندازه گیری کرده و در صورت ناچیز بودن تأثیر هر حذف بر نتایج خروجی، آن ویژگی از بردار خروجی حذف می شد. با تکرار این فرایند در نهایت برداری ۵۲ بُعدی از ویژگی به دست آمد که برای آموزش و آزمون طبقه بندی کننده خروجی به عنوان بردار ویژگی نهایی به کار رفت.

۶- نتایج آزمایشها

در این بخش نتایج آزمایشهای انجام شده در مراحل مختلف آورده می شود. آزمایشهای بازشناسی انجام شده، در دو دسته اصلی: بازشناسی واج گسسته و بازشناسی گفتار پیوسته انجام شد. ترتیب مطالب در این بخش چنین است: نخست پایگاه داده گفتاری مورد استفاده و شرایط طبقه بندی کننده مبتنی بر مدل مخفی مارکوف به کار رفته در این مقاله معرفی می شود. در ادامه، آزمایشهای انجام شده در مرحله انتخاب ویژگی ارائه می شود. در قسمت بُعدی نتایج آزمایشهای بازشناسی واج گسسته آورده می شود. قسمت پایانی این بخش نتایج آزمایشهای بازشناسی گفتار پیوسته را نشان می دهد.

همپوشانی اطلاعاتی با ویژگیهای دیگری باشند که حذف آنها تغییری در توانایی بازشناسی سیستم نهایی ایجاد نمی کند. بنابراین انتخاب بردار ویژگی نهایی با حداکثر تمایز پذیری، هدف این بخش است. برای تحلیل تمایز پذیری ویژگیها، به معیاری نیاز داریم. معیار فیشر "F-Ratio" ابزار مناسبی برای تحلیل تمایز پذیری ویژگیهای مختلف در بردار ویژگی نهایی است. ضریب فیشر در مسأله ای با دو کلاس خروجی از رابطه زیر به دست می آید:

$$F_{ratio}(W) = \frac{W^T S_B W}{W^T S_W W} \quad (۴)$$

که در آن S_B ماتریس پراکندگی بین-کلاسی است که از رابطه زیر به دست می آید:

$$S_B = (\hat{\mu}_1 - \hat{\mu}_2)(\hat{\mu}_1 - \hat{\mu}_2)^T \quad (۵)$$

$\hat{\mu}_i$ میانگین N_i نمونه از کلاس w_i است. ماتریس S_W مجموع ماتریسهای پراکندگی دو کلاس است. بردار W بزرگترین بردار ویژه $S_B^{-1} S_W$ است. در مورد مسأله ای با K کلاس، عبارتهای بالا به شکل زیر تغییر می کند. از جمع زدن مقادیر ماتریس های پراکندگی کل کلاسها S_W به دست می آید.

$$S_W = \sum_{i=1}^K S_{W_i} \quad (۶)$$

اگر ماتریس S_T را به شکل زیر تعریف کنیم.

$$S_T = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (۷)$$

۱-۶- پایگاه داده و سیستم بازشناسی گفتار

سیگنال‌های گفتاری با فرکانس نمونه‌برداری ۱۶ کیلوهرتز از پایگاه داده تیمیت [۱۸] انتخاب شده است. پایگاه داده تیمیت شامل ۶۳۰۰ جمله است که شامل ۱۰ جمله از ۶۳۰ گوینده از ۸ منطقه مختلف امریکا با لهجه‌های مختلف است. دادگان گفتاری موجود در این پایگاه دارای برچسب‌های دقیق واژگانی و مرزهای واجی هستند. این پایگاه شامل دو مجموعه آزمون هسته‌ای^۱ و مجموعه آزمون کامل است که در این تحقیق از مجموعه آزمون کامل در آزمایشهای بازشناسی گفتار پیوسته استفاده شده است. این مجموعه آزمون شامل ۱۳۴۴ جمله از ۱۶۸ گوینده است و ۲۷٪ کل دادگان را شامل می‌شود.

در آزمایشهای بازشناسی واج گسسته، از مجموعه آزمون هسته‌ای استفاده شده که شامل ۱۹۲ جمله از ۲۴ گوینده است. در این آزمایشها واجها در ۳۹ کلاس خروجی طبقه‌بندی شد. به منظور ارزیابی عملکرد سیستم پیشنهادی در شرایط نویزی از پایگاه داده نویزی NoiseX.۹۲ [۱۹] استفاده شده است. چهار دسته از سیگنال‌های نویزی این پایگاه شامل نویزهای سفید، صورتی، همهمه و نویز کارخانه است.

برای بازشناسی گفتار از جعبه‌ابزار HTK [۲۰] دانشگاه کمبریج یعنی معمولترین ابزار بازشناسی گفتار و از مدل پنهان مارکوف با شش حالت، ۱۶ خوشه چپ به راست با ماتریس کواریانس قطری استفاده شده است. برای بخش سکوت از مدل سه حالت استفاده شده است. طول قاب‌های گفتار برابر ۲۵/۶ میلی ثانیه انتخاب شده است. خروجی‌های سیستم بازشناسی، رشته واجهای بازشناسی شده است.

۲-۶- انتخاب ویژگی‌ها

همان‌گونه که در بخش ۵ گفته شد، استفاده از مدل مخلوط گوسی با چهار خوشه به برداری با ۶۰ ویژگی منجر می‌شود. به منظور استفاده همزمان از مزایای اطلاعات حوزه فاز و فرکانس در فرایند بازشناسی گفتار در این مقاله از ترکیب ویژگی‌ها استفاده شده است.

جدول ۱ برچسب‌های اختصاص داده شده به ویژگی‌های ترکیبی

برچسب	شماره ویژگی	دسته ویژگی
۱-۱۳ مل	۱ الی ۱۳	ضرایب حوزه مل
۱۳- دلنا	۱۴ الی ۲۶	ضرایب دلنا
دلنا دلنا-۱۳	۲۷ الی ۳۹	ضرایب دلنا-دلنا
گوسی ۴- گوسی ۱	۴۰ الی ۴۳	ضرایب وزن خوشه‌های مدل مخلوط گوسی
گوسی ۳۲- گوسی ۵	۴۴ الی ۶۱	ضرایب میانگین خوشه‌های مدل مخلوط گوسی
گوسی ۶۰- گوسی ۳۳	۶۲ الی ۹۹	ضرایب واریانس خوشه‌های مدل مخلوط گوسی

افزودن بردار ۳۹ بُعدی کپسترال مل به بردار ۶۰ بُعدی پارامترهای مدل گوسی، برداری ۹۹ بُعدی را نتیجه می‌دهد. جدول (۱) برچسب‌های اختصاص داده شده به این ویژگی‌ها را نشان می‌دهد. همان‌گونه که در بخش ۵ گفته شد، برای انتخاب ویژگی‌های مؤثر برای فرایند بازشناسی از معیار جدایی‌پذیری فشر استفاده شده است. با استفاده از معیار فشر ضریب فشر برای تمامی ویژگی‌ها محاسبه شد. شکل (۷) ضرایب محاسبه شده را برای ۱۵ ویژگی با بیشترین ضریب تمایزپذیری نشان می‌دهد. به منظور کاهش بُعد بردار ویژگی نهایی، ویژگی‌هایی با کمترین ضریب فشر را حذف و سپس صحت بازشناسی واج گسسته را با استفاده از ویژگی‌های باقیمانده محاسبه کردیم. در صورتی که حذف یک ویژگی تأثیر چندانی بر نتایج بازشناسی نداشته باشد، آن ویژگی حذف

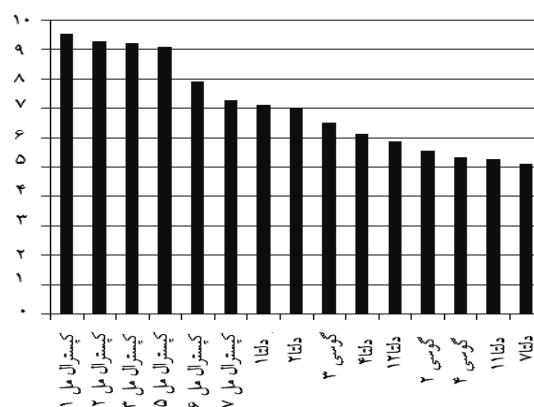
به منظور ارزیابی عملکرد سیستم پیشنهادی، آزمایشهای بازشناسی واج گسسته بر روی ۴۲۰۰ واج از مجموعه آزمون هسته‌ای پایگاه داده تیمیت انجام شد. از ضریب $COR\%$ به عنوان معیار عملکردی سیستم استفاده شد. این ضریب از درصد نسبت تعداد واجهای بازشناسی شده با برچسب صحیح به تعداد کل واجهای مورد آزمایش به دست می‌آید. جدول (۲) ضریب $COR\%$ را برای بردارهای ویژگی مختلف نشان می‌دهد. مدل پایه شامل ۳۹ بردار ویژگی معمول کپسترال مل است. در جدول (۲) نتایج درصد بازشناسی واج برای ضرایب ۶۰ بُعدی استخراج شده از مدل مخلوط گوسی، مدل پیشنهادی [۱] و بردار ۵۲ بُعدی نهایی ترکیبی را برای سیگنال تمیز نشان می‌دهد. همان‌گونه که در جدول (۲) نشان داده شده، نتایج به دست آمده از کاربرد ویژگیهای فضای فاز به تنهایی، بسیار جالب است، با وجود این ضرایب فرکانسی معمول، عملکرد بهتری را نشان داده است. بردار ترکیبی پیشنهادی شامل ۵۲ ویژگی انتخاب شده از ترکیب ویژگیهای فضای فاز و فرکانس بهترین عملکرد را نشان داده که ۵/۷٪ بهبود نسبت به مدل پایه است. به منظور ارزیابی عملکرد سیستم در شرایط نویزی، آزمایشهای بازشناسی واج را در شرایط نویز جمعی و در چند مقدار مختلف نسبت سیگنال به نویز انجام دادیم. جدول (۳) نتایج این آزمایشها را نشان می‌دهد. همان‌گونه که در جدول (۳) دیده می‌شود، بهترین عملکرد سیستم پیشنهادی برای سیگنال تمیز بوده است.

می‌شود و در غیر این صورت باقی می‌ماند. به این روش، بردار نهایی ویژگی‌های باقیمانده شامل ۵۲ ویژگی به شرح زیر بر اساس برچسب‌های اختصاص داده شده در جدول ۱ انتخاب شد:

- ۱) ضرایب کپسترال مل (۱۰، ۹، ۸، ۷، ۶، ۵، ۴، ۳، ۲، ۱)
 - ۲) ضرایب دلتای کپسترال مل (۱، ۲، ۳، ۴، ۵، ۶، ۷، ۸، ۹، ۱۰، ۱۱، ۱۲)
 - ۳) ضرایب دلتا-دلتای کپسترال مل (۱، ۲، ۳، ۴، ۵، ۶، ۷، ۸)
 - ۴) ضرایب مخلوط گوسی شامل ضرایب (۱، ۲، ۳، ۴، ۵، ۶، ۷، ۸، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹، ۲۰، ۲۱، ۲۲، ۲۳، ۲۴، ۲۵، ۲۶، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۳۹، ۴۰، ۴۱، ۴۲، ۴۳، ۴۴، ۴۵، ۴۶، ۴۷، ۴۸، ۴۹، ۵۰، ۵۱، ۵۲)
- از میان این ۵۲ ویژگی نهایی، ۲۶ ویژگی از حوزه فرکانس و ۲۶ ویژگی از حوزه فاز انتخاب شد.

۶-۳- آزمایشهای بازشناسی واج گسسته

تا اینجا روش مدل‌سازی غیرخطی توزیع نقاط فضای فاز باسازی شده سیگنال گفتار و ترکیب ویژگیهای به دست آمده از این مدل‌سازی با ویژگیهای معمول فرکانسی را معرفی کردیم.



شکل ۷ ضریب جدایی پذیری پذیرنده فشر برای ۱۵ ویژگی بالاترین تمایزپذیری

است. به دلیل حجم پایین محاسبات مورد نیاز در بازسازی بردار ترکیبی و بُعد پایین و قابل قبول آن، امکان استفاده از سیستم پیشنهادی در کاربردهای بازشناسی پیوسته گفتار وجود دارد. به منظور تکمیل آزمایشهای بازشناسی گفتار در این تحقیق، تعدادی آزمایشهای بازشناسی گفتار پیوسته انجام شد. در این بردار ویژگیهای کپسترال مل، ویژگیهای مدل مخلوط گوسی و بردار نهایی ترکیبی به صورت مجزا به دست آوردیم. نرخ بازشناسی واج با رابطه زیر تعریف می شود.

$$PRR = 1 - PER = 1 - \frac{S + I + D}{N} \quad (9)$$

جدول ۴ نتایج نرخ بازشناسی زنجیره واجی در آزمایشهای

بازشناسی گفتار پیوسته به ازای بردارهای ویژگی مختلف

بردار ویژگی	تعداد جایگزینی نادرست	تعداد درج نادرست	تعداد حذف نادرست	نرخ بازشناسی واج درصد
ضرایب کپسترال مل	۲۸۵۹	۲۴۵۰	۱۳۱۱	۶۶٫۹
ضرایب مدل مخلوط گوسی	۳۲۴۵	۲۷۰۰	۲۶۷۰	۵۶٫۹
بردار ترکیبی پیشنهادی	۲۷۸۰	۱۷۵۰	۱۳۲۰	۷۰٫۷

که در آن PER نرخ خطای بازشناسی زنجیره واجی است. S تعداد کل جایگزینی های اشتباه، I تعداد درجهای نادرست واجها و D تعداد حذف های نادرست در زنجیره واجی و N تعداد کل واجها در زنجیره واجی مورد بررسی است. جدول (۴) نرخ بازشناسی زنجیره

جدول ۲ درصد بازشناسی واج برای بازشناسی واج گسسته به ازای بردارهای ویژگی مختلف

بردار ویژگی	درصد صحت بازشناسی
بردار پایه ویژگیهای کپسترال مل ۳۹ بُعدی	۷۲٫۴
ویژگیهای مخلوط گوسی و مدل پاونلی و همکاران [۱]	۵۳٫۶
ویژگیهای مخلوط گوسی ۶۰ بُعدی	۵۷٫۱
بردار ترکیبی پیشنهادی ۵۲ بُعدی	۷۸٫۱

جدول ۳ درصد بازشناسی واج برای بازشناسی واج گسسته در

شرایط نویزی مختلف

سیگنال	ضریب سیگنال به نویز	درصد صحت بازشناسی بردار ویژگی پایه	درصد صحت بازشناسی مدل پیشنهادی
تمیز	بینهایت	۷۲٫۴	۷۸٫۱
نویز سفید	۵	۵۲٫۴	۵۷٫۵
	۱۰	۵۹٫۳	۶۳٫۸
نویز صورتی	۵	۵۱٫۸	۵۵٫۷
	۱۰	۵۸٫۶	۶۲٫۹
نویز هممه	۵	۴۸٫۷	۵۱٫۴
	۱۰	۵۷٫۱	۶۱٫۴
نویز کارخانه	۵	۴۹٫۴	۵۳٫۱
	۱۰	۵۸٫۶	۶۱٫۷

۶-۴- آزمایشهای بازشناسی گفتار پیوسته

هدف این تحقیق نشان دادن کاربرد مدلسازی غیرخطی فضای فاز بازسازی شده سیگنال گفتار در سیستمهای بازشناسی خودکار گفتار در بازشناسی گفتار پیوسته

نکته جالب دیگری که در این تحقیق مشخص شد، برتری ویژگیهای استخراج شده از قطع پوانکاره فضای فاز نسبت به ویژگیهای استخراج شده از حوزه فاز مانند [۱] است که در جدول (۴) نشان داده شده. این نکته همچنین مزیت استفاده از قطع پوانکاره را در تحلیل سیگنالهایی با ماهیت شبه تناوبی تأیید می‌کند.

در ساختار سیستم پیشنهادی، نخست فضای فاز هشت بُعدی با استفاده از فرایند پوشش برای هر قاب تشکیل و سپس قطع پوانکاره برای این فضای فاز با الگوریتم بخش ۳ محاسبه شده و توزیع نقاط قطع پوانکاره، با استفاده از مدل مخلوط گوسی مدل‌سازی می‌شود. سه دسته از پارامترها که به عنوان ویژگی استفاده می‌شوند، عبارتند از: ضرایب وزن خوشه‌ها، بردار میانگین و واریانس خوشه‌های گوسی. بر این مبنا و با در نظر گرفتن مدل‌سازی با استفاده از چهار خوشه، یک بردار ویژگی ۶۰ بُعدی از هر قاب گفتار استخراج و با بردار ۳۹ بُعدی ضرایب کپسترال مل ترکیب می‌شود. بر اساس الگوریتم انتخاب ویژگی ارائه شده بر اساس تحلیل تمایزپذیری فیشر، ویژگیهای فهرست شده در بخش ۶-۲ از بین کل ویژگیها انتخاب شده و بعنوان ورودی به طبقه‌بندی کننده وارد می‌شود. مدل مخلوط گوسی در کاربردهای مختلف پردازش گفتار استفاده شده و در این مقاله نیز به دلیل وجود مرزهای انعطاف‌پذیر تصمیم و امکان مدل‌سازی دادگان با بُعد بالا در چند خوشه محدود به عنوان روش مدل‌سازی انتخابی در این مقاله برای مدل‌سازی توزیع دادگان در فضای فاز بازسازی شده مورد استفاده قرار گرفتند.

نکته جالب دیگری که در این تحقیق ارائه شد، این است که با وجود عملکرد مناسب ویژگی‌های استخراج شده از حوزه فاز، باز هم عملکرد ویژگیهای معمول فرکانسی بهتر است و لذا مسأله ترکیب بردارهای ویژگی

واجبی را در آزمایشهای بازشناسی گفتار پیوسته برای بردارهای ویژگی مختلف به ازای زنجیره واجی ۲۰,۰۰۰ تایی نشان می‌دهد.

۷- جمع بندی

در این مقاله، به کارگیری ویژگیهای استخراج شده از مدل‌سازی غیرخطی قطع پوانکاره فضای فاز بازسازی شده سیگنال گفتار در ترکیب با ویژگیهای معمول فرکانسی در بازشناسی گفتار مطالعه شد. دیدیم که استفاده از ترکیب ویژگیهای فضای فاز با ویژگیهای معمول فرکانسی، باعث ۵/۷ درصد بهبود در صحت سیستمهای بازشناسی گفتار در آزمایشهای واج گسسته می‌شود. این تحقیق نشان می‌دهد که با آن که ویژگیهای معمول استخراج شده از فضای فاز بازسازی شده سیگنال - مانند ضرایب لیاپانوف و بُعد فراکتال - بهبود خاصی را در عملکرد سیستم بازشناسی ایجاد نمی‌کنند، اما استفاده از ترکیب مدل‌سازی قطع پوانکاره فضای فاز تشکیل شده و ویژگیهای معمول کپسترال مل، باعث بهبود قابل توجهی در عملکرد سیستم بازشناسی می‌شود. همان‌گونه که در بخش مقدمه گفته شد، طول قاب‌های گفتار و تعداد نمونه‌های آنها بسیار محدود است و با توجه به بُعد هشت به دست آمده برای بازسازی فضای فاز، تعداد نقاط همسایگی‌های بازسازی شده در فضای فاز بسیار کم خواهد بود. این مسأله و حساسیت ویژگیهای معمول استخراج شده از نظریه آشوب به نقطه شروع اولیه، مهمترین دلایل ضعف این نوع ویژگیها در کاربردهای بازشناسی گفتار است. استفاده از مدل‌سازی غیرخطی فضای فاز، امکان نمایش بهتر اطلاعات نهفته در آن را بر اساس پارامترهای مدل برآزش شده به آن فراهم می‌سازد.

به دست آمده از حوزه فاز و فرکانس برای بهبود عملکرد سیستم بازشناسی گفتار ضروری است.

به منظور ارزیابی عملکرد سیستم پیشنهادی در شرایط نویزی، تعدادی آزمایش بازشناسی واج گسسته برای دادگان نویز جمعی با چهار نوع نویز سفید، صورتی، هممه و کارخانه انجام شد. آزمایشها نشان می‌دهد که سیستم پیشنهادی حتی در شرایط نویزی باعث بالارفتن صحت بازشناسی گفتار می‌شود و بهترین عملکرد در برای نویز سفید و کمترین بهبود در نویز هممه مشاهده شده است. برای ارزیابی بهتر عملکرد سیستم در بازشناسی مقاوم گفتار می‌توان از دادگان نویزی واقعی تری مانند پایگاه داده Aurora2 به جای نویز جمعی استفاده کرد. روش پیشنهادی همچنین می‌تواند در کاربردهای دیگر پردازش سیگنال به ویژه برای سیگنالهای غیرخطی استفاده شود.

۸- نتیجه گیری

در این مقاله روشی جدید برای استخراج ویژگی بر مبنای ترکیب اطلاعات استخراج شده از مدل‌سازی غیرخطی قطع پوانکاره محاسبه شده از فضای فاز بازسازی شده سیگنال گفتار با ویژگیهای معمول فرکانسی ارائه شد. برای مدل‌سازی توزیع نقاط قطع پوانکاره از مدل مخلوط گوسی استفاده شد. انتخاب این مدل به دلیل عملکرد مناسب آن در کاربردهای پردازش گفتار، نرم بودن مرزهای تصمیم و امکان مدل‌سازی توزیع دادگان با بُعد بالا با استفاده از خوشه‌هایی با بُعد کمتر انجام شد. با انتخاب شش برای تعداد پارامترهای تأخیر زمانی و هشت برای بُعد فضای بازسازی و با استفاده از فرایند پوشش، فضای فاز سیگنال گفتار بازسازی شد. سپس با استفاده از مدل مخلوط گوسی با چهار خوشه، بردار ویژگی ۶۰

بُعدی به دست می‌آید. بر روی این بردار ویژگی، در ترکیب با بردار ۳۹ بُعدی کپسترال مل، تحلیل تمایزپذیری با استفاده از تحلیل تمایزپذیری فیشر انجام شد. از بردار ۹۹ بُعدی ورودی، بردار ۵۲ بُعدی به عنوان بردار ویژگی نهایی برای استفاده در ورودی طبقه‌بندی کننده به دست می‌آید. بردار ویژگی نهایی هیچ‌گونه محدودیتی از نظر حجم محاسبات و حجم حافظه مورد نیاز برای استفاده در کاربردهای پردازش گفتار پیوسته ندارد. برای ارزیابی توانایی عملکردی سیستم پیشنهادی، مجموعه‌ای از آزمایشهای بازشناسی واج گسسته و بازشناسی گفتار پیوسته انجام شد. نتایج آزمایشها نشان می‌دهد که سیستم پیشنهادی صحت بازشناسی رشته واجی را در بازشناسی گفتار پیوسته به اندازه ۳/۸٪ بهبود می‌بخشد. همچنین بهترین افزایش در صحت بازشناسی در آزمایشهای بازشناسی واج گسسته برای سیگنال تمیز برابر ۵,۷٪ حاصل شد. این نتایج نشان می‌دهد که از طرح پیشنهادی می‌توان در سایر کاربردهای بازشناسی و طبقه‌بندی سیگنال نیز استفاده کرد.

۹- منابع

- [1] Povinelli. R. J, Johnson. M. T, Lindgren. A. C, Roberts. F. M, Ye. J, "Statistical Models of Reconstructed Phase Spaces for Signal Classification", IEEE Trans. Signal Processing., vol. 54, pp. 2178–2186, 2006.
- [2] Theiler. J, Eubank. S, Longtin. A, Galdrikian. B, Farmer, J. D, "Testing for nonlinearity in time series: The method of surrogate data", Physica D, vol. 58, pp. 77–94, 1992.

- reconstructed phase spaces” , IEEE Trans. Knowl. Data Eng., vol. 16, pp. 779–783, 2004.
- [10] McGowan. R. S, “An aeroacoustics approach to phonation”, J. Acoust Soc. Am, vol. 83, no. 2, pp. 696–704, 1988.
- [11] Teager. H. M, Teager. S. M., “Evidence for nonlinear sound production mechanisms in the vocal tract”, in Speech Production and Speech Modeling, W. Hardcastle and A. Marchal, Eds. Bonas, France, vol. 55, 1989.
- [12] Whitney. H, “Differentiable manifolds” , Ann. Math., ser. 2nd, vol.37, pp. 645–680, 1936.
- [13] Takens. J, “Detecting strange attractors in turbulence”, in Proc Dynam. Syst. Turbulence, Warwick, pp. 366–381, 1980.
- [14] Sauer. T, Yorke. J.A, “Casdagli, Embedology” , J. Stat. Phys.,vol. 65, pp. 579–616, 1991.
- [15] Narayanan.S, Alwan. A, “A nonlinear dynamical systems Analysis of fricative consonants” , J. Acoust. Soc. Amer., vol.97, no. 4, pp.2511–2524, 1995.
- [16] Dempster. A. P, Laird. N. M., Rubin. D. B., “Maximum likelihood from incomplete data via the EM algorithm,” J.
- [3] Kaiser. J. F , “Some observations on vocal tract operation from a fluid flow point of view”, in Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control, I. R. Titze and R. C. Scherer, Eds. Denver, CO: Denver Center for Performing Arts, pp. 358–386, 1983.
- [4] Maragos. P, Dimakis. A, Kokkinos. I, “Some advances in nonlinear speech modeling using modulations, fractals, and chaos”, in Proc. Int. Conf. on DSP, Santorini, Greece, Jul. 2002.
- [5] Kokkinos. I, Maragos. P, “Nonlinear speech analysis using models for chaotic systems”, IEEE transactions on speech and audio processing, vol. 13, no. 6, 2005
- [6] Pitsikalis. V, Kokkinos. I, Maragos. P, “Nonlinear analysis of speech signals: Generalized dimensions and Lyapunov exponents”, in Proc. Of Eurospeech, Geneva, Switzerland, Sep. 2003.
- [7] Abarbanel. H, *Analysis of Observed Chaotic Data* , New York: Springer - Verlag, 1996.
- [8] Kantz. H, Schreiber. T, *Nonlinear Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1997.
- [9] Povinelli. R. J, Johnson. M. T, Lindgren. A. C, Ye. J, “Time series classification using Gaussian mixture models of

- [19] Noisex.92 Noise database, Signal Processing Information Base (SPIB) http://spib.rice.edu/spib/select_noise.html
- [20] HTK, The Hidden Markov Model Toolkit, Machine Intelligence Laboratory, Cambridge University, <http://htk.eng.cam.ac.uk/>
- [21] R. C. Hilborn, *Chaos and Nonlinear Dynamics*, Oxford University Press, Oxford, 2000
- Royal Statist. Soc., ser.B, vol. 39, pp. 1–38, 1977.
- [17] Mika, S. et al, "Fisher Discriminant Analysis with Kernels". IEEE Conference on Neural Networks for Signal Processing, 1999
- [18] Garofolo. J. S, Lame. L. Fl, Fisher. W. M, Fiscus. J. G, Pallett. D. S, Dahlgren.N, Dahlgren. N. L. "Darpa timit acoustic-phonetic continuousspeech corpus". [CD-ROM], 1993.