

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

ایوب جعفری<sup>۱</sup>، فرشاد الماس گنج<sup>۲</sup>، مریم نبی بیدهندی<sup>۳</sup>

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 دینامیک سیستم تولید کننده آن را نشان می دهد و بنابراین می تواند شامل اطلاعات مفیدی باشد که در تحلیل فرکانسی معمول - مانند استخراج ضرایب کپسٹرال مل<sup>۱</sup> - 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) 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) و پیگاه داده تمیت<sup>۴</sup> به منظور ارزیابی سیستم ارائه شده استفاده شده است. نتایج آزمایشها بهبودی در حدود ۵/۷ درصد در صحت بازشناسی گفتار واج گستته 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

این حوزه مانند ضرایب نمایی لیپانوف و بُعد فراکتال پرداخته می‌شود. در استخراج این ویژگیها به جای استفاده از سیگنال یک‌بعدی گفتار، از فضای چندبعدی فاز بازسازی شده از سیگنال گفتار با استفاده از روشی مبتنی بر تأخیرهای زمانی - به نام فرایند پوشش - استفاده می‌شود [۷] [۶]. فرایند پوشش از بازسازی مجموعه‌ای از جفت‌های ورودی - خروجی بر اساس تأخیر نمونه‌های سیگنال زمانی یک‌بعدی انجام می‌شود. به دلیل تعداد کم نمونه‌های موجود در قاب‌های گفتار، قابلیت تعمیم‌پذیری فضای بازسازی شده، یکی از مهمترین چالش‌ها 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 لیپانوف و بُعد فراکتال)، چگالی [۹] و توبولوژی مرکز کرده‌اند. مدل مخلوط گوسی روشی مناسب برای تخمین چگالی نقاط در فضای فاز است؛ اما تحقیقات بسیار محدودی در زمینه استفاده از مدل‌سازی فضای فاز در کاربردهای بازشناسی گفتار انجام شده است.

استفاده از ویژگیهای آشوبگونه معمول مانند ضرایب لیپانوف در بازشناسی گفتار به خوبی در تحقیقات قبلی مطالعه شده است [۱۰] [۹] [۸]. این تحقیقات به روشنی بیان می‌کنند که استفاده از این دسته ویژگیها کمک چندانی به بهبود صحت سیستمهای بازشناسی گفتار نمی‌کنند.

همان‌گونه که شکل (۱) نشان می‌دهد، بازسازی در فضای فاز، نمایش بسیار بهتری را نسبت به طیف فرکانسی به دست می‌دهد. 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

## ۱- مقدمه

برای سالیان متوالی بردارهای ویژگی استخراج شده در حوزه فرکانس مانند ضرایب کپسٹرال فرکانسی مل - که از تفسیر ادراکی عملکرد گوش انسان بر روی طیف فرکانسی سیگنال گفتار به دست می‌آیند - در سیستمهای بازشناسی گفتار با موقیت محدودی استفاده شده‌اند.

برخی سیگنالها به ویژه سیگنالهایی که از معادلات دیفرانسیل غیرخطی به دست می‌آیند، مشخصات طیفی عریض و نامشخصی دارند و در این موارد، تحلیل فرکانسی ابزار چندان مناسبی به نظر نمی‌رسد، 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  $k = x_{n+1} \times x_n \times \dots \times x_1$  به ازای  $k$  برابر ۴ نشان می‌دهد [۲].

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

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

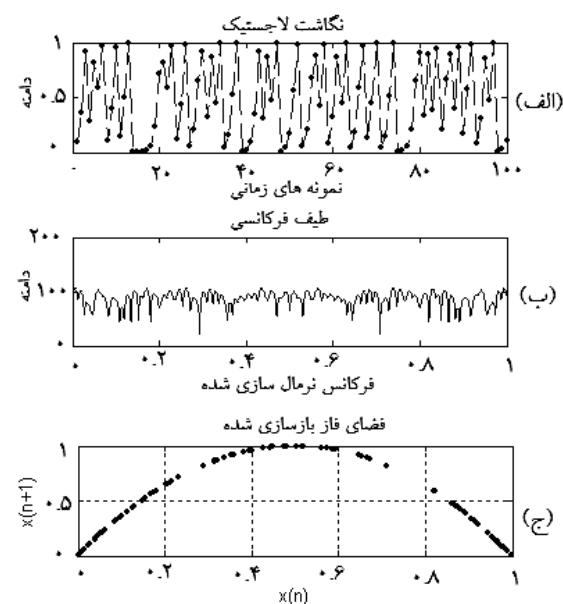
اطلاعات در حوزه فاز است. دوم آن که به جای استفاده از ویژگیهای استخراج شده از حوزه فاز به تنها، بردار ویژگی ترکیب این اطلاعات با اطلاعات حوزه فرانکنس به کار گرفته شد.

سازمان این مقاله چنین است. در بخش ۲ کلیات سیستم پیشنهادی ارائه می‌شود. بخش ۳ فرایند پوشش به منظور بازسازی فضای فاز سیگنال گفتار را به صورت خلاصه بیان می‌کند. قطع پوانکاره و الگوریتم مورد استفاده در این تحقیق در این بخش ارائه می‌شود. در بخش ۴ مدل‌سازی غیرخطی مبتنی بر روش مدل مخلوط گوسی بیان می‌شود. در بخش ۵ به معرفی روش پیشنهادی به منظور ترکیب ویژگیهای فضای فاز و فرانکنس پرداخته می‌شود. بخش ۶ آزمایشها و دادگان مورد استفاده را توضیح می‌دهد. بخش ۷ شامل مباحثی درباره روش پیشنهادی و مزایا و معایب آن بوده و با ارائه جمع‌بندی کلی در بخش ۸ مقاله به پایان می‌رسد.

## ۲- ساختار روش پیشنهادی

همانگونه که در مقدمه گفته شد، مبنای روش پیشنهادی، مدل‌سازی غیرخطی قطع پوانکاره به دست آمده از فضای فاز بازسازی شده سیگنال گفتار و ترکیب اطلاعات استخراج شده از این بخش با ویژگی‌های فرانکنس معمول کپسیتال حوزه مل است. شکل (۲) کلیات روش پیشنهادی را نشان می‌دهد. همانگونه که در شکل (۲) نشان داده شده، نخست فضای فاز سیگنال گفتار با استفاده از فرایند پوشش مبتنی بر تأخیر زمانی نمونه‌های سیگنال زمانی یک‌بعدی - که در بخش سوم معرفی می‌شود - ساخته می‌شود. در مرحله بُعدی مدل‌سازی مبتنی بر مدل ترکیبی‌گوسی بر روی نمونه‌های به دست آمده از محاسبه قطع پوانکاره از فضای فاز بازسازی شده سیگنال

فرانکنس در مسأله بازشناسی گفتار رهنمون می‌شویم. پاوینلی و همکاران [۱] نوعی مدل‌سازی آماری را بر روی فضای فاز بازسازی شده سیگنال مطرح کردند و نشان دادند که ویژگیهای استخراج شده عملکرد مناسبی را در کاربردهای طبقه‌بندی سیگنال نشان می‌دهد. با وجود این، کیفیت عملکرد روش آنها، در صورت استفاده از دادگان بزرگتر و به‌ویژه در بازشناسی گفتار پیوسته به شدت کاهش می‌یابد.



شکل ۱ (الف) سیگنال زمانی، (ب) طیف فرانکنسی و (ج) فضای فاز بازسازی شده  
با ازای پارامتر  $K$  برابر ۴

برای ارائه روشی مناسب برای استخراج ویژگی که قابل استفاده در سامانه‌های بازشناسی گفتار باشد، دو رویکرد را در این تحقیق مدنظر قرار داده‌ایم: اول آن که مدل‌سازی غیرخطی به جای اعمال به نقاط فضای فاز بازسازی شده، بر روی قطع پوانکاره سیگنال گفتار انجام شد. قطع پوانکاره، به‌ویژه برای سیگنال‌هایی با ماهیت شبکه‌تนาوب مانند سیگنال گفتار، ابزار مناسبی برای استخراج

جدایی‌پذیری فیشر است که در بخش ۵ معرفی می‌شود. استفاده از این روش بردار ویژگی ۵۲ بعدی را به دست می‌دهد که بردار مناسبی برای استفاده در طبقه‌بندی کنده‌های مبتنی بر مدل مخفی ماکوف معمول در کاربردهای بازشناسی گفتار است.

### ۳- بازسازی فضای فاز از سیگنال زمانی و قطع پوانکاره

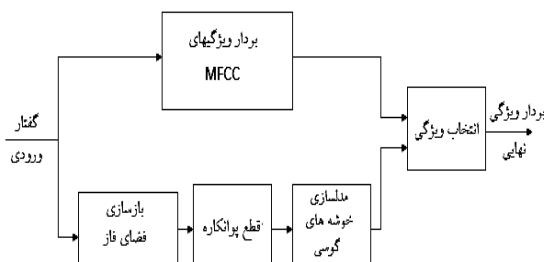
در این بخش به معرفی روش مورد استفاده در بازسازی فضای فاز سیگنال از سیگنال گفتار یک‌بعدی و نیز الگوریتم مورد استفاده در استخراج قطع پوانکاره سیگنال می‌پردازیم.

#### ۱-۳- بازسازی فضای فاز

بازسازی فضای فاز با استفاده از فرایندی به نام پوشش - که روشنی مبتنی بر استفاده از نمونه‌های تأخیر یافته سیگنال زمانی است - انجام می‌شود. در این بخش نظریه بازسازی جاذب فاز سیگنال را مرور می‌کنیم. فرایند پوشش بر دستیابی به نگاشتی مبتنی است که امکان دسترسی به دینامیک سیستم تولید کننده سیگنال را فراهم می‌سازد.

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

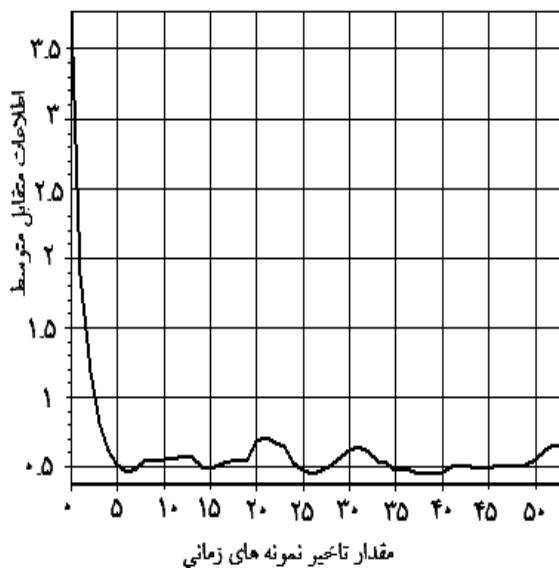
گفتار اعمال و بردار ویژگی نهایی به ورودی طبقه‌بندی کننده مبتنی بر مدل مخفی مارکوف اعمال می‌شود. بردار ویژگی به دست آمده، شامل پارامترهای مدل ترکیب گوسی مانند ضربی بهره، بردار میانگین و مقادیر واریانس خوش‌های گوسی است. در این تحقیق از مدل ترکیب گوسی با چهار خوش‌ه استفاده شده است. در بخش (۳) نشان داده می‌شود که بعد بهینه برای بازسازی فضای فاز در سیگنال گفتار برابر هشت است و در نتیجه بعد 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 ضربی بهره، هفت ضربی بردار میانگین و هفت ضربی ماتریس واریانس در دست داریم. لذا به ازای هر خوش‌ه، تعداد ویژگی‌های مورد نیاز برابر ۱۵ و در مدل چهار خوش‌های مورد استفاده در این تحقیق بعد بردار نهایی به دست آمده برابر ۶۰ خواهد بود. با توجه به بعد ۳۹ برای بردار ویژگی‌های کپسکال حوزه مل، بعد نهایی بردار ویژگی برابر ۹۹ می‌شود 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. بنای این انتخاب، معیار بیشترین تمايز 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 می‌دهد. همان‌گونه که در شکل (۳) نشان داده شده، مقدار بهینه پارامتر تأخیر زمانی برابر ۶ است. پس از انتخاب پارامتر تأخیر زمانی نوبت به تعیین بعد بهینه برای فضای فاز بازسازی شده از فرایند پوشش می‌رسد. در این تحقیق از معیار کمینه‌سازی همسایگی‌های غلط استفاده شده است. همسایگی‌های غلط، نقاطی بر روی فضای فاز بازسازی شده است که از یکدیگر دور هستند. بعد فضای پوشش به گونه‌ای انتخاب می‌شود که در صد این همسایگی‌ها کمینه شود. با استفاده از این ابزار، بعد بهینه برای بازسازی فضای فاز برابر هشت انتخاب شد. شکل (۴) فضای فاز دو بعدی بازسازی شده را برای واج  $a/a$  از پایگاه داده تیمیت نشان می‌دهد.



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

دریافتند که اگر  $d > d_{\text{threshold}}$  بُعد جاذب فاز به دست آمده باشد، انتخاب  $d > d_{\text{threshold}}$  تضمین می‌کند که این نگاشت دینامیک سیستم اولیه را توصیف کند. لذا بر مبنای این نظریه، می‌توان از نگاشتهای مبتنی بر تأخیر زمانی برای نگاشت سیگنال یک بعدی به فضای فاز چند بعدی استفاده کرد.

اگر  $\{x_1, \dots, x_n\}$  نقاط سری زمانی سیگنال گفتار یک بعدی باشد، ماتریس  $X$  با بُعد  $d$  و تأخیر زمانی  $\tau_d$  به شکل زیر تعریف می‌شود.

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

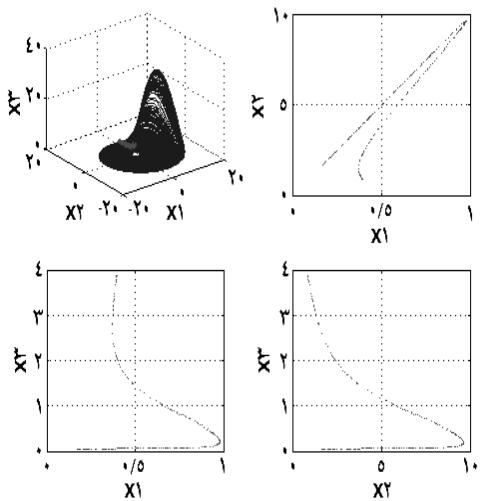
مقدار بهینه تأخیر زمانی  $\tau_d$  را می‌توان از همبستگی میان نمونه‌های سیگنال گفتار به دست آورد. نارایانان [۱۵] معیار متوسط اطلاعات متقابل را برای تخمین ضرب تأخیر زمانی به صورت زیر پیشنهاد کرد:

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

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

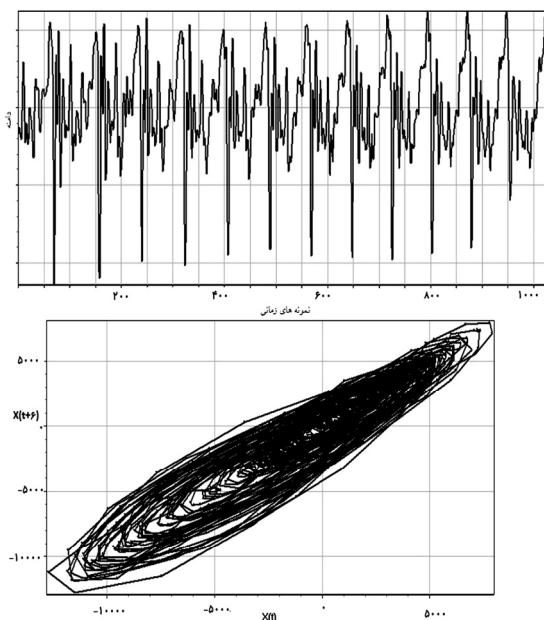
پوانکاره برای فضاهای فاز با بعد بیشتر از سه، منجر به استخراج نقاطی در فضایی با بعد  $n-1$  برای فضای فازی با بعد  $n$  می شود [۲۱]. با در نظر گرفتن این مسئله در واقع می توان گفت که قطع پوانکاره در واقع بعده سیستم اولیه را یکی کاهش می دهد. با در نظر گرفتن رفتار تنایی مشخص سیگنال گفتار به ویژه در بخش‌های واکدار می توان چنین نتیجه گرفت که خط حرکت بازسازی شده در فضای فاز، بعد از موقعیت هر پیچ گفتار تکرار می شود. شکل (۶) فضای فاز بازسازی شده سه بعدی و قطع پوانکاره محاسبه شده را برای واج /a/ نشان می دهد.

محاسبه قطع پوانکاره برای هر قاب از سیگنال گفتاری شامل مراحل زیر است:



شکل ۵ قطع پوانکاره سامانه ریسلر

- ۱- فضای فاز سیگنال برای قاب مورد نظر با استفاده از فرایند پوشش بازسازی می شود.
- ۲- برای نقطه اولیه  $(n_0, X_0)$ ، همسایگی  $(n_0, N)$  برای  $k$  نزدیکترین نقطه بر اساس فاصله اقلیدسی انتخاب می شود.

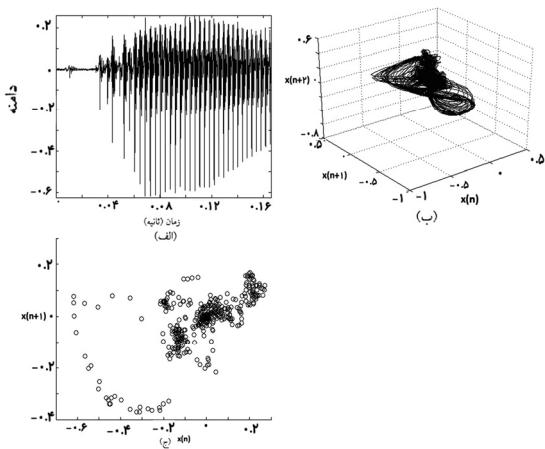


شکل ۴ سیگنال زمانی و فضای فاز دو بعدی بازسازی شده برای واج /a/ از پایگاه داده تیمیت

### ۲-۳- قطع پوانکاره

بررسی و نمایش خط حرکت در فضای فاز بازسازی شده با بعد بالا - مانند فضای فاز سیگنال گفتار با بعد 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 می شود. برای سیستم دینامیکی با سه بعد  $(X_1, X_2, X_3)$ ، انتخاب صفحه و انتخاب نقاطی که خط حرکت این صفحه را قطع می کند، قطع پوانکاره، یک خط را در فضای فاز ایجاد می کند. برای سیستمی با تعداد بعد بالاتر از سه، سطوحی از صفحات به کار می رود. شکل (۵) نقاط پوانکاره سامانه ریسلر را نشان می دهد. تعمیم قطع

سیگنال پیشنهاد کردند. پاوینلی و همکاران [۱] برخی روش‌های مدل‌سازی آماری مانند روش‌های مبتنی بر هیستوگرام را پیشنهاد کردند و این مدل‌ها را در آزمایش‌هایی برای بازناسی واج مطالعه کردند.



شکل ۶ سیگنال زمانی، فضای فاز سه‌بعدی بازسازی شده و قطع پوانکاره برای واج  $a/\omega$  از پایگاه داده تیمیت

بر مبنای این تحقیقات، برای سیستم بازناسی گفتار پیوسته مورد نظر در این تحقیق، روش مدل‌سازی غیرخطی قطع پوانکاره فضای فاز بازسازی شده با استفاده از مدل ترکیبات گوسی برای تحلیل قاب‌های گفتار انتخاب شد. از مدل مخلوط گوسی به صورت گستردگی در پردازش‌های گفتار استفاده می‌شود. این روش نقشی اساسی را در مدل مخفی مارکوف - که معمول‌ترین ابزار در بازناسی گفتار است - بازی می‌کند. در این تحقیق، از مدل مخلوط گوسی به عنوان طبقه‌بندی کننده استفاده نشده بلکه برای مدل‌سازی توزیع دادگان در فضای نقاط قطع پوانکاره فاز بازسازی شده به کار رفته است. مدل ترکیبات گوسی، دو مزیت اصلی در کاربرد مورد نظر برای مدل‌سازی توزیع دادگان دارد. اولین مزیت آن است

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

-۴- سطح پوانکاره عمود بر  $(n_i)$  در نقطه  $X(n_i)$  تعیین می‌شود.

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

#### ۴- مدل‌سازی غیرخطی فضای فاز بازسازی شده

در پردازش سیگنال گفتار فرض ایستایی فقط برای سری‌های زمانی با طول ۲۰ تا ۲۶ میلی‌ثانیه اعتبار دارد. برین اساس، طول قاب‌های به کار رفته برای کاربردهای پردازش گفتار برابر این مقدار انتخاب می‌شود. با در نظر گرفتن فرکانس نمونه برداری ۱۶ کیلوهرتز برای دادگان گفتاری تیمیت، تعداد نمونه‌های هر قاب در حدود ۴۰۰ است. با در نظر گرفتن بُعد هشت برای فضای فاز بازسازی شده برای تشکیل هر همسایگی، به  $8*8*8=512$  نمونه نیاز است که تقریباً ۱۷ درصد از کل نقاط را شامل می‌شود. بنابراین با مشکل کوتاه بودن طول سری زمانی روبرو هستیم و این کمبود داده، یکی از مشکلات اصلی است که باعث می‌شود ویژگی‌های معمول موجود در تحلیل آشوب - مانند ضرایب لیپانوف - عملکرد مناسبی را در بازناسی گفتار نشان ندهند [۵]. برای غلبه بر این مشکل یکی از روش‌های جالب، استفاده از مدل‌سازی غیرخطی توزیع دادگان در فضای فاز بازسازی شده است. کُینوس و همکاران [۵] برخی از روش‌های مدل‌سازی غیرخطی مانند توابع چندجمله‌ای محلی، تخمینگرهای فازی و شبکه عصبی شعاعی را برای مدل‌سازی فضای فاز

## ۵- انتخاب ویژگی

هدف اصلی این تحقیق نشان دادن مزایای ترکیب اطلاعات حوزه فاز بازسازی شده و حوزه فرکانس در بازناسی گفتار است. در این تحقیق برخی آزمایش‌های بازناسی واج را با استفاده از ویژگی‌های کپسیترال حوزه مل و ویژگی‌های استخراج شده از فضای فاز انجام شد تا ارزیابی توانایی‌های هر دسته از ویژگی‌ها در مسئله بازناسی گفتار امکان پذیر باشد. سپس این ویژگی‌ها در یک بردار ویژگی واحد ترکیب شد. در ترکیب ویژگی‌ها به یک بردار  $39 \times 1$  بعدی ضرایب کپسیترال حوزه مل و یک بردار  $60 \times 1$  بعدی به دست آمده از پارامترهای مدل مخلوط گوسی می‌رسیم. بر این مبنای بعد بردار نهایی به دست آمده برابر  $99 \times 1$  است. این بعد برای بردار ویژگی، به دلایل زیر برای استفاده در کاربردهای بازناسی گفتار پیوسته قابل قبول نیست:

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

## ۶- تحلیل جدایی‌پذیری فیشر

همان‌گونه که گفته شد، یک بردار  $99 \times 1$  بعدی در ورودی در دست داریم و به دلایل گفته شده، بهینه‌سازی این بردار ویژگی می‌تواند به بهبود عملکرد سیستمهای بازناسی گفتار منجر شود، زیرا بسیاری از ویژگی‌ها می‌توانند دارای

که در این مدل‌ها نواحی تصمیم دارای 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 برآش 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:

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

که در آن  $P$  تابع توزیع احتمال را نشان می‌دهد. پارامتر  $M$  تعداد خوشه‌ها را مشخص می‌کند و  $N$  تابع توزیع نرمال با میانگین  $\mu_m$  و ماتریس کواریانس  $\sigma^2_m$  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 بازناسی شده برابر هشت انتخاب شد، برای بُعد هفت قطع پانکاره 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  $60 \times 1$ ، نتیجه مدل‌سازی غیرخطی فضای فاز قاب گفتار خواهد بود.

ماتریس  $S_B$  را می‌توان از رابطه زیر به دست آورد.

$$S_B = S_T - S_W \quad (8)$$

پس از محاسبه ماتریس‌های مورد نیاز، از رابطه (۴) می‌توان برای مسئله‌ای با  $K$  کلاس خروجی (مانند مسئله بازناسنی 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 پایین‌تری داشتند، با انجام آزمون بازناسنی گفتار، تأثیر حذف این ویژگی را در صحت سیستم بازناسنی اندازه‌گیری کرده و در صورت ناچیز بودن تأثیر هر حذف بر نتایج خروجی، آن ویژگی از بردار خروجی حذف می‌شد. با تکرار این فرایند در نهایت برداری ۵۲ بعدی از ویژگی به دست آمد که برای آموزش و آزمون طبقه‌بندی کننده خروجی به عنوان بردار ویژگی نهایی به کار رفت.

## ۶- نتایج آزمایشها

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

همپوشانی اطلاعاتی با ویژگی‌های دیگری باشند که حذف آنها تغییری در توانایی بازناسنی سیستم نهایی ایجاد نمی‌کند. بنابراین انتخاب بردار ویژگی نهایی با حداقل تمایزپذیری، هدف این بحث است. برای تحلیل تمایزپذیری ویژگیها، به معیاری نیاز داریم. معیار فیشر "F-Ratio" ابزار مناسبی برای تحلیل تمایزپذیری ویژگی‌های مختلف در بردار ویژگی نهایی است. ضریب فیشر در مسئله‌ای با دو کلاس خروجی از رابطه زیر به دست می‌آید:

$$F_{ratio}(W) = \frac{W^T S_B W}{W^T S_W W} \quad (4)$$

که در آن  $S_B$  ماتریس پراکنده‌گی بین-کلاسی است که از رابطه زیر به دست می‌آید:

$$S_B = (\hat{\mu}_1 - \hat{\mu}_2)(\hat{\mu}_1 - \hat{\mu}_2)^T \quad (5)$$

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

$$S_W = \sum_{i=1}^K S_{w_i} \quad (6)$$

اگر ماتریس  $S_T$  را به شکل زیر تعریف کیم.

$$S_T = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (7)$$

## ۶-۲- انتخاب ویژگی‌ها

همان گونه که در بخش ۵ گفته شد، استفاده از مدل مخلوط گوسی با چهار خوش 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

دسته ویژگی	شماره ویژگی	برچسب
ضرایب حوزه مل	۱ الی ۱۳	مل ۱۲-مل ۱
ضرایب دلتا	۱۴ الی ۲۶	دلتا ۱۲- دلتا ۱
ضرایب دلتا- دلتا	۲۷ الی ۳۹	دلتا دلتا ۱۳- دلتا دلتا ۱
ضرایب وزن خوش‌های مدل مخلوط گوسی	۴۰ الی ۴۳	گوسی ۴- گوسی ۱
ضرایب میانگین خوش‌های مدل مخلوط گوسی	۴۴ الی ۶۱	گوسی ۳۲- گوسی ۵
ضرایب واریانس خوش‌های مدل مخلوط گوسی	۶۲ الی ۹۹	گوسی ۶۰- گوسی ۳۳

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

## ۶-۱- 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<sup>۱</sup> 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 ۰.۲۷٪ کل دادگان را شامل می‌شود.

در آزمایشهای بازشناسی واج گستته، از مجموعه آزمون هسته‌ای استفاده شده که شامل ۱۹۲ جمله از ۲۴ گوینده است. در این آزمایشهای واژها 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 خوش 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، رشتہ واجههای بازشناسی شده است.

به منظور ارزیابی عملکرد سیستم پیشنهادی، آزمایش‌های بازشناسی واج گستته 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 بعده استخراج شده از مدل مخلوط گوسی، مدل پیشنهادی [۱] و بردار ۵۲ بعده 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 ۵/۰٪ 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.

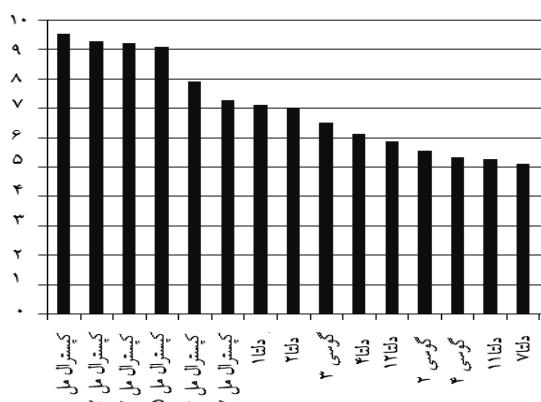
می‌شود و در غیر این صورت باقی می‌ماند. به این روش، بردار نهایی ویژگی‌های باقیمانده شامل ۵۲ ویژگی به شرح زیر بر اساس برچسب‌های اختصاص داده شده در جدول ۱ انتخاب شد:

- (۱) ضرایب کپسترال مل (۴، ۵، ۶، ۷، ۸، ۹ و ۱۰)
- (۲) ضرایب دلتای کپسترال مل (۱، ۲، ۳، ۴، ۹، ۷، ۳، ۴، ۱۱، ۹، ۷، ۲، ۱)
- (۳) ضرایب دلتای دلتای کپسترال مل (۱، ۲، ۳، ۴، ۵، ۶، ۷، ۱۱، ۱۲، ۱۵، ۱۶، ۱۷، ۲۰، ۲۱، ۲۷، ۲۹، ۳۰، ۳۳، ۳۴، ۴۰، ۴۵، ۴۷، ۴۸، ۵۳، ۵۹)
- (۴) ضرایب مخلوط گوسی شامل ضرایب (۱، ۲، ۳، ۴، ۵، ۶، ۷، ۱۱، ۱۲، ۱۵، ۱۶، ۱۷، ۱۸، ۲۰، ۲۱، ۲۷، ۲۹، ۳۰، ۳۳)

از میان این ۵۲ ویژگی نهایی، ۲۶ ویژگی از حوزه فرکانس و ۲۶ ویژگی از حوزه فاز انتخاب شد.

### ۳-۳- آزمایش‌های بازشناسی واج گستته

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



است. به دلیل حجم پایین محاسبات مورد نیاز در بازسازی بردار ترکیبی و بعد 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 آزمایشها نرخ بازشناسی زنجیره واجی (PRR) را برای بردار ویژگی‌های کپسٹرال مل، ویژگی‌های مدل مخلوط گوسی و بردارنهایی ترکیبی به صورت مجزا به دست آورديم. نرخ بازشناسی واج با رابطه زير تعريف می‌شود.

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

جدول ۴ نتایج نرخ بازشناسی زنجیره واجی در آزمایش‌های بازشناسی گفتار پیوسته به‌ازای بردارهای ویژگی مختلف

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

که در آن  $PER$  نرخ خطای بازشناسی زنجیره واجی است.  $S$  تعداد کل جایگزینی‌های اشتباه،  $I$  تعداد درجهای نادرست واجها و  $D$  تعداد حذف‌های 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 کپسٹرال مل ۳۹ بعده	۷۲,۴
ویژگی‌های مخلوط گوسی و مدل پاوینی و همکاران [۱]	۵۳,۶
ویژگی‌های مخلوط گوسی ۶۰ بعده	۵۷,۱
بردار ترکیبی پیشنهادی ۵۲ بعده	۷۸,۱

جدول ۳ درصد بازشناسی واج برای بازشناسی واج گسسته در شرایط نویزی مختلف

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

۶-۴-آزمایش‌های بازشناسی گفتار پیوسته هدف این تحقیق نشان دادن کاربرد مدل سازی 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 سیگنانهایی با ماهیت شبه تناوبی تأیید می کند.

در ساختار سیستم پیشنهادی، نخست فضای فاز هشت بُعدی با استفاده از فرایнд 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 می شود. سه دسته از پارامترها که به عنوان ویژگی استفاده می شوند، عبارتند از: ضرایب وزن خوشه ها، بردار میانگین و واریانس خوشه های گوسی. بر این مبنای 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 کپسیتال مل ترکیب می شود. بر اساس الگوریتم انتخاب ویژگی ارائه شده بر اساس تحلیل تمایز پذیری فیشر، ویژگیهای فهرست شده در بخش ۲-۶ از بین کل ویژگیها انتخاب شده و عنوان ورودی به طبقه بندی کننده وارد می شود. مدل مخلوط گوسی در کاربردهای مختلف پردازش گفتار استفاده شده و در این مقاله نیز به دلیل وجود مرزهای انعطاف پذیر تصمیم و امکان مدلسازی دادگان با بعد 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 ویژگی های استخراج شده از حوزه فاز، باز هم عملکرد ویژگیهای معمول فرکانسی بهتر است و لذا مسئله ترکیب بردارهای ویژگی

واجی را در آزمایش های بازشناسی گفتار پیوسته برای بردارهای ویژگی مختلف به ازای زنجیره واجی ۲۰,۰۰۰ تایی نشان می دهد.

## ۷- جمع‌بندی

در این مقاله، به کارگیری ویژگیهای استخراج شده از مدلسازی غیرخطی قطع پوانکاره فضای فاز بازسازی شده سیگنان 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 آزمایش های واج گسته می شود. این تحقیق نشان می دهد که با آن که ویژگیهای معمول استخراج شده از فضای فاز بازسازی شده سیگنان - مانند ضرایب لیاپانوف و بُعد فراکتال - بهبود خاصی را در عملکرد سیستم بازشناسی ایجاد نمی کنند، اما استفاده از ترکیب مدلسازی قطع پوانکاره فضای فاز تشکیل شده و ویژگیهای معمول کپسیتال مل، باعث بهبود قابل توجهی در عملکرد سیستم بازشناسی می شود. همان گونه که در بخش مقدمه گفته شد، طول قاب های گفتار و تعداد نمونه های آنها بسیار محدود است و با توجه به بُعد هشت به دست آمده برای بازسازی فضای فاز، تعداد نقاط همسایگی های بازسازی شده در فضای فاز بسیار کم خواهد بود. این مسئله 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 برآش شده به آن فراهم می سازد.

بعدی به دست می‌آید. بر روی این بردار ویژگی، در ترکیب با بردار ۳۹ بعدی کپسٹرال مل، تحلیل تمایزپذیری با استفاده از تحلیل تمایزپذیری فیشر انجام شد. از بردار ۹۹ بعدی ورودی، بردار ۵۲ بعدی 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 گستته 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 گستته برای سیگنال تمیز برابر ۰,۷٪ 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.

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 گستته برای دادگان نویز جمعی با چهار نوع نویز سفید، صورتی، همه‌مه 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 همه‌مه مشاهده شده است. برای ارزیابی بهتر عملکرد سیستم در بازشناسی مقاوم گفتار می‌توان از Aurora2 دادگان نویزی واقعی تری مانند پایگاه داده به جای نویز جمعی استفاده کرد. روش پیشنهادی همچنین می‌تواند در کاربردهای دیگر پردازش سیگنال به ویژه برای سیگنالهای غیرخطی استفاده شود.

## ۸- نتیجه‌گیری

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 نقطه 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 بعد بالا با استفاده از خوشه‌هایی با بعد کمتر انجام شد. با انتخاب شش برای تعداد پارامترهای تأخیر زمانی و هشت برای بعد 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

- 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 snalysis 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](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, Lam. 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.