مجلهٔ علمی- پژوهشی
«مهندسی برق مدرس»
دورهٔ دهم، شمارهٔ ۳، پاییز ۱۳۸۹

مدل‌سازی غیرخطی قطع پوانکاره سیگنال گفتار در ترکیب با تحلیل حوزهٔ فرکانس به‌منظور افزایش صحت عملکرد سیستمهای بازشناسی گفتار

ایوب جعفری[۱*]، فرشاد الماس گنج[۲]، مریم نبی بیدهندی[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 به‌دست آمده از مدل‌سازی غیرخطی فضای فاز بازسازی شده سیگنال گفتار با ویژگیهای معمول به‌دست آمده از تحلیل حوزهٔ فرکانس ارائه می‌شود. بر اساس نظریه پذیرفته شده کنونی، در صورت انتخاب تعداد بُعد کافی برای بازسازی فضای فاز سیگنال، این فضا به‌صورت کامل دینامیک سیستم تولید کننده آن را نشان می‌دهد و بنابراین می‌تواند شامل اطلاعات مفیدی باشد که در تحلیل فرکانسی معمول - مانند استخراج ضرایب کپسترال مل MFCC[۱] - در دسترس نیست. همچنین از آنجا که سیستمهای پیچیده‌ای مانند سیستم تولید گفتار، رفتارهایی نوسانی و تناوبی نشان می‌دهند، قطع پوانکاره می‌تواند به‌عنوان ابزاری مؤثر در تحلیل این رفتارها به کار رود. در این مقاله نوعی مدل‌سازی غیرخطی با استفاده از مدل مخلوط گوسی (GMM[۲]) بر روی نقاط قطع پوانکاره سیگنال گفتار انجام می‌شود. بردار ویژگی نهایی از یک مرحله انتخاب ویژگی بر روی پارامترهای مدل مخلوط گوسی و بردار ویژگی‌های فرکانسی معمول MFCC حاصل می‌شود. 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 می‌دهد. این روش جدید می‌تواند به‌عنوان روشی مؤثر و جایگزین در روشهای استخراج ویژگی به‌ویژه برای سیستمهای با ماهیت غیرخطی شدید استفاده شود.

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

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

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

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 به‌دست می‌آیند - در سیستمهای بازشناسی گفتار با موفقیت محدودی استفاده شده‌اند.

برخی سیگنالها به‌ویژه سیگنالهایی که از معادلات دیفرانسیل غیرخطی به‌دست می‌آیند، مشخصات طیفی عریض و نامشخصی دارند و در این موارد، تحلیل فرکانسی ابزار چندان مناسبی به‌نظر نمی‌رسد، زیرا امکان تفکیک اطلاعات سیگنال در حوزه فرکانس وجود ندارد [۱]. به‌عنوان مثال شکل (۱) سیگنال زمانی، طیف فرکانسی و فضای فاز بازسازی شده را برای نگاشت لاجستیک با معادله $x_{n+1} = k \times x_n \times (1 - x_n)$ به‌ازای k برابر ٤ نشان می‌دهد [۲].

دلایل نظری و عملی مختلفی در برخی تحقیقات اخیر ([۲] [۳] [٤] [٥]) مطرح شده که حضور پدیده غیرخطی قویی را در سیستم تولید گفتار نشان می دهند 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 درسیگنال گفتار پدیده توربولانس است [٥].

از سوی دیگر موفقیتی که نظریهٔ سیستمهای آشوبگونه در برخورد با سیستمهای دارای توربولانس داشته است، در نظر گرفتن پدیده توربولانس موجود در سیگنال گفتار به‌ویژه در واج‌های سایشی را مطرح کرده و ما را به استفاده از این نظریه در سیستمهای بازشناسی گفتار رهنمون می‌شود. در کاربردهای معمول نظریهٔ آشوب در پردازش سیگنال به استخراج برخی ویژگیهای معمول در این حوزه مانند ضرایب نمایی لیاپانوف و بُعد فراکتال پرداخته می‌شود. در استخراج این ویژگیها به‌جای استفاده از سیگنال یک‌بُعدی گفتار، از فضای چندبُعدی فاز بازسازی شده از سیگنال گفتار با استفاده از روشی مبتنی بر تأخیرهای زمانی - به نام فرایند پوشش - استفاده می‌شود [٦] [۷] [۸]. فرایند پوشش از بازسازی مجموعه‌ای از جفت‌های ورودی - خروجی بر اساس تأخیر نمونه‌های سیگنال زمانی یک‌بُعدی انجام می‌شود. به‌دلیل 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.

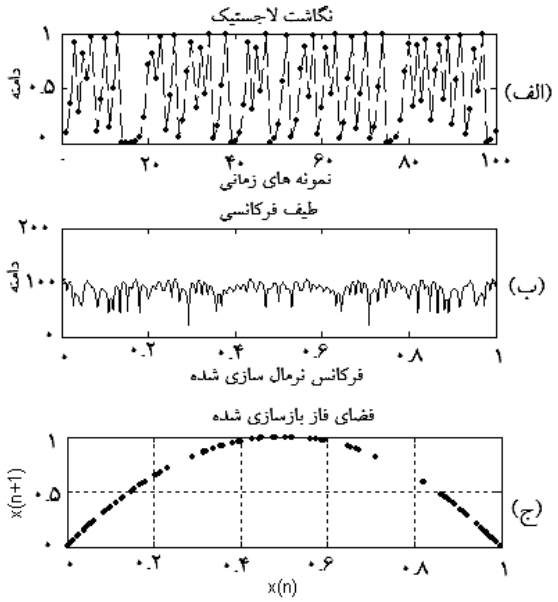
بسیاری از تحقیقات در زمینهٔ استفاده از فضای فاز بازسازی شده، بر روی تخمین پارامترهای اندازه (ضرایب لیاپانوف و بُعد فراکتال)، چگالی [۹] و توپولوژی تمرکز کرده‌اند. مدل مخلوط گوسی روشی مناسب برای تخمین چگالی نقاط در فضای فاز است؛ اما تحقیقات بسیار محدودی در زمینهٔ استفاده از مدل‌سازی فضای فاز در کاربردهای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

استفاده از ویژگیهای آشوبگونه معمول مانند ضرایب لیاپانوف در بازشناسی گفتار به‌خوبی در تحقیقات قبلی مطالعه شده است [۸] [۹] [۱۰] [۱۱]. 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

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

فرکانس 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 به‌شدت کاهش می‌یابد.



شکل ۱ الف) سیگنال زمانی، ب) طیف فرکانسی و (ج) فضای فاز بازسازی شده برای سیستم لاجستیک به ازای پارامتر k برابر ٤

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 به‌جای اعمال به نقاط فضای فاز بازسازی شده، بر روی قطع پوانکاره سیگنال گفتار انجام شد. قطع پوانکاره، به‌ویژه برای سیگنال‌هایی با ماهیت شبه‌تناوب مانند سیگنال گفتار، ابزار مناسبی برای استخراج اطلاعات در حوزهٔ فاز است. دوم آن که به‌جای استفاده از ویژگیهای استخراج شده از حوزهٔ فاز به تنهایی، بردار ویژگی ترکیب این اطلاعات با اطلاعات حوزهٔ فرکانس به کار گرفته شد.

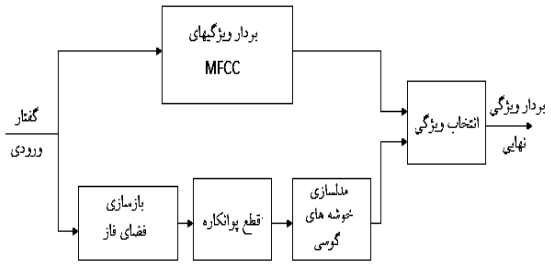
سازمان این مقاله چنین است. در بخش ۲ کلیات سیستم پیشنهادی ارائه می‌شود. بخش ۳ فرایند پوشش به‌منظور بازسازی فضای فاز سیگنال گفتار را به‌صورت خلاصه بیان می‌کند. قطع پوانکاره و الگوریتم مورد استفاده در این تحقیق در این بخش ارائه می‌شود. در بخش ٤ مدل‌سازی غیرخطی مبتنی بر روش مدل مخلوط گوسی بیان می‌شود. در بخش ۵ به معرفی روش پیشنهادی به‌منظور ترکیب ویژگیهای فضای فاز و فرکانس پرداخته می‌شود. بخش ٦ آزمایشها و دادگان مورد استفاده را توضیح می‌دهد. بخش ۷ شامل مباحثی دربارهٔ روش پیشنهادی و مزایا و معایب آن بوده و با ارائه جمع‌بندی کلی در بخش ۸ مقاله به پایان می رسد.

۲- ساختار روش پیشنهادی

همانگونه که در مقدمه گفته شد، مبنای روش پیشنهادی، مدل‌سازی غیرخطی قطع پوانکاره به‌دست 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 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 یک‌بُعدی - که در بخش سوم معرفی می‌شود - ساخته می‌شود. در مرحله بُعدی مدل‌سازی مبتنی بر مدل ترکیبهای گوسی بر روی نمونه های به‌دست آمده از محاسبه قطع پوانکاره از فضای فاز بازسازی شده سیگنال

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

گفتار اعمال و بردار ویژگی نهایی به ورودی طبقه‌بندی کننده مبتنی بر مدل مخفی مارکوف اعمال می‌شود. بردار ویژگی به‌دست آمده، شامل پارامترهای مدل ترکیب گوسی مانند ضریب بهره، بردار میانگین و مقادیر واریانس خوشه‌های گوسی است. در این تحقیق از مدل ترکیب گوسی با چهار خوشه استفاده شده است. در بخش (۳) نشان داده می‌شود که بُعد بهینه برای بازسازی فضای فاز در سیگنال گفتار برابر هشت است و در نتیجه بُعد قطع پوانکاره به‌دست 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 به‌ازای هر خوشه، تعداد ویژگیهای مورد نیاز برابر ۱۵ و در مدل چهار خوشه‌ای مورد استفاده در این تحقیق بُعد بردار نهایی به‌دست آمده برابر ٦۰ خواهد بود. با توجه به بُعد ۳۹ برای بردار ویژگیهای کپسترال حوزه مل، بُعد 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. مبنای این انتخاب، معیار بیشترین تمایز پذیری با استفاده از تحلیل جدایی‌پذیری فیشر است که در بخش ۵ معرفی می‌شود. استفاده از این روش بردار ویژگی ۵۲ بُعدیی را به دست می‌دهد که بردار مناسبی برای استفاده در طبقه‌بندی کننده‌های 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 یک‌بُعدی و نیز الگوریتم مورد استفاد در استخراج قطع پوانکاره سیگنال می‌پردازیم.

۳-۱- بازسازی فضای فاز

بازسازی فضای فاز با استفاده از فرایندی به نام پوشش – که روشی 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 بُعد d داشته باشیم، نگاشت $f : \mathbf{R}^d \to \mathbf{R}^{2d+1}$ می‌تواند به‌عنوان نگاشتی پوشش از سیستم اولیه مطرح باشد. تیکنز [۱۳] نشان داد که f نگاشتی است که دینامیک سیستم اولیه را توصیف می‌کند. پیشرفت اصلی در مشخص شدن نگاشت پوشش از کار سائُر و یورک [۱٤] به‌دست آمد که نشان دادند هر نگاشت تأخیر زمانی با بُعد بازسازی کافی می‌تواند نوعی نگاشت پوشش باشد و همخوانی با دینامیک سیستم اولیه را تضمین می‌کند. همچنین آنان

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

دریافتند که اگر $d_.$ بُعد جاذب فاز به‌دست آمده باشد، انتخاب $d > ۲d_.$ 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 یک‌بُعدی به فضای فاز چندبُعدی استفاده کرد.

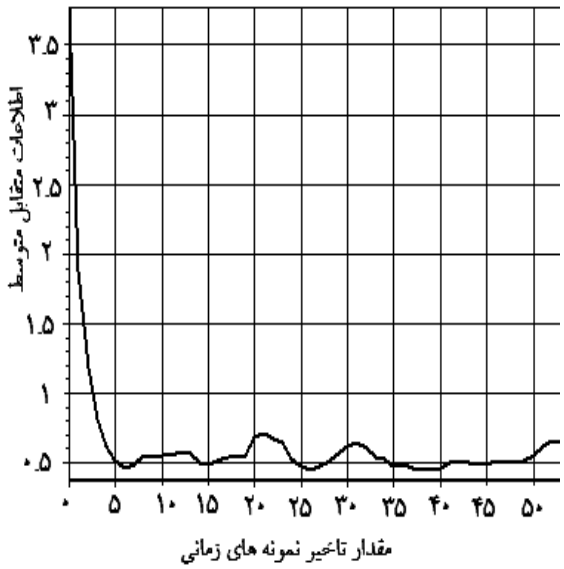
اگر $x = \{x_۱,...,x_n\}$ نقاط سری زمانی سیگنال گفتار یک‌بُعدی باشد، ماتریس X با بُعد d و تأخیر زمانی τ_d به‌شکل زیر تعریف می‌شود.

$$\mathbf{X}=\begin{bmatrix} x_{۱+(d-۱)\tau_d} \\ x_{۲+(d-۱)\tau_d} \\ . \\ . \\ x_N \end{bmatrix} = \begin{bmatrix} x_{۱+(d-۱)\tau_d} & \cdots & x_{۱+\tau_d} & x_۱ \\ x_{۲+(d-۱)\tau_d} & \cdots & x_{۲+\tau_d} & x_۲ \\ . & & & \\ . & & & \\ x_N & \cdots & x_{N-(d-۲)\tau_d} & x_{N-(d-۱)\tau_d} \end{bmatrix} \quad (۱)$$

مقدار بهینه تأخیر زمانی τ_d را می‌توان از همبستگی میان نمونه‌های سیگنال گفتار به‌دست آورد. نارایانان [۱۵] معیار متوسط اطلاعات متقابل را برای تخمین ضریب تأخیر زمانی به صورت زیر پیشنهاد کرد:

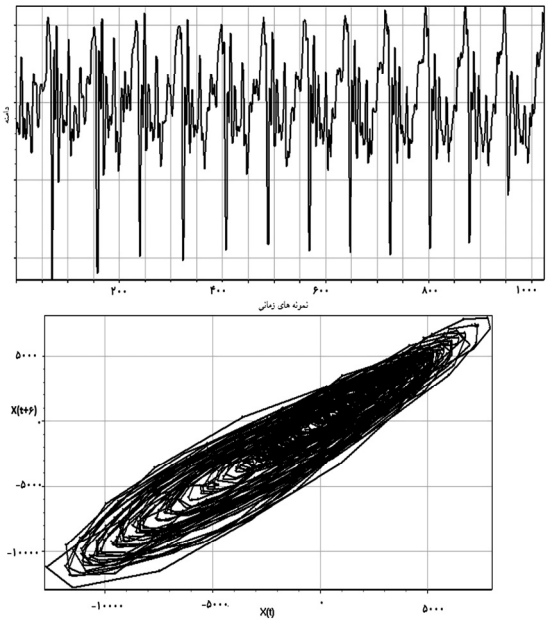
$$I(T) = \sum_{n=۱}^{N-T} P(x(n), x(n+T)).\log_۲\left[\frac{P(x(n), x(n+T))}{P(x(n))P(x(n+T))}\right] \quad (۲)$$

که در آن x(n) نمونهٔ سیگنال گفتار، P تابع احتمال و I معیار اطلاعاتی دوسویه است. اولین کمینهٔ این معیار، مقدار بهینهٔ پارامتر تأخیر زمانی است. شکل (۳) مقادیر محاسبه شده را به‌ازای مقادیر مختلف تأخیر زمانی نشان می‌دهد. همان‌گونه که در شکل (۳) نشان داده شده، مقدار بهینهٔ پارامتر تأخیر زمانی برابر ٦ است. پس از انتخاب پارامتر تاخیر زمانی نوبت به تعیین بُعد 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. بُعد فضای پوشش به‌گونه‌ای انتخاب می‌شود که درصد این همسایگی‌ها کمینه شود. با استفاده از این ابزار، بُعد بهینه برای بازسازی فضای فاز برابر هشت انتخاب شد. شکل (٤) فضای فاز دوبُعدی بازسازی شده را برای واج /a/ از پایگاه داده تیمیت نشان می‌دهد.



شکل ۳ مقادیر معیار اطلاعاتی دوسویه به‌ازای مقادیر مختلف تأخیر زمانی (اولین کمینه در نقطهٔ ٦)

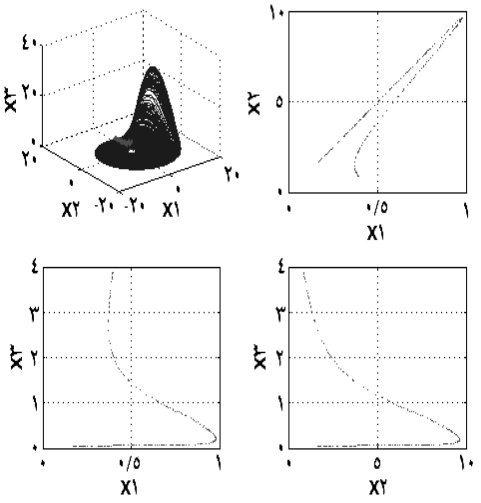
[Downloaded from mjee.modares.ac.ir on 2024-05-05]



شکل ٤ سیگنال زمانی و فضای فاز دوبُعدی بازسازی شده برای واج /a/ از پایگاه داده تیمیت

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 - معمولا 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 سامانهٔ ریسلر را نشان می‌دهد. تعمیم قطع پوانکاره برای فضاهای فاز با بُعد بیشتر از سه، منجر به استخراج نقاطی در فضایی با بُعد n-۱ 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 (٦) فضای فاز بازسازی شده سه‌بُعدی و قطع پوانکاره محاسبه شده را برای واج / a/ نشان می‌دهد. محاسبهٔ قطع پوانکاره برای هر قاب از سیگنال گفتاری شامل مراحل زیر است:



شکل ۵ قطع پوانکاره سامانهٔ ریسلر

۱- فضای فاز سیگنال برای قاب مورد نظر با استفاده از فرایند پوشش بازسازی می‌شود.

۲- برای نقطهٔ اولیه $X(n_0)$، همسایگی $N(n_0)$ برای k نزدیکترین نقطه بر اساس فاصله اقلیدسی انتخاب می‌شود.

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

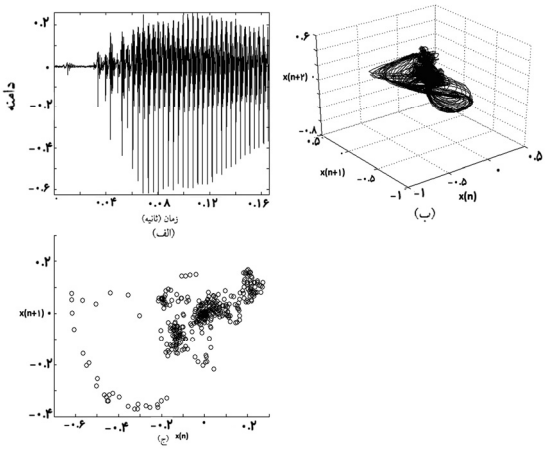
۳- جهت جریان متوسط $f(n_i)$ برای خط حرکت در همسایگی $N(n_i)$ از رابطه $f(n_i)=mean(x[n+1]-x[n])\ \forall n\in N(n_i)$ محاسبه می‌شود.

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

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

٤- مدل‌سازی غیرخطی فضای فاز بازسازی شده

در پردازش سیگنال گفتار فرض ایستایی فقط برای سری‌های زمانی با طول ۲۰ تا ۲٦ 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 ۱٦ کیلوهرتز برای دادگان گفتاری تیمیت، تعداد نمونه‌های هر قاب در حدود ٤۰۰ است. با در نظر گرفتن بُعد هشت برای فضای فاز بازسازی شده برای تشکیل هر همسایگی، به ۷۲=۸+۸+۸*۸ نمونه نیاز است که تقریبا ۱۷ درصد از کل نقاط را شامل می‌شود. بنابراین با مشکل کوتاه بودن طول سری زمانی روبه‌رو هستیم و این کمبود داده، یکی از مشکلات اصلی است که باعث می‌شود ویژگیهای 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 مدل‌سازی غیرخطی توزیع دادگان در فضای فاز بازسازی شده است. کُکینوس و همکاران[۵] برخی از روشهای مدل‌سازی غیرخطی مانند توابع چندجمله‌ای محلی، تخمینگرهای فازی و شبکه عصبی شعاعی را برای مدل‌سازی فضای فاز 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 بازشناسی واج مطالعه کردند.



شکل ٦ سیگنال زمانی، فضای فاز سه‌بُعدی 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 مدل‌سازی غیرخطی قطع پوانکاره فضای فاز بازسازی شده با استفاده از مدل ترکیبات گوسی برای تحلیل قاب‌های گفتار انتخاب شد. از مدل مخلوط گوسی به‌صورت گسترده‌ای در پردازشهای گفتار استفاده می‌شود. این روش نقشی اساسی را در مدل مخفی مارکوف - که معمول‌ترین ابزار در بازشناسی گفتار است - بازی می‌کند. در این تحقیق، از مدل مخلوط گوسی به‌عنوان طبقه‌بندی کننده استفاده نشده بلکه برای مدل‌سازی 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 مدل‌سازی توزیع دادگان دارد. اولین مزیت آن است

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

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، به‌صورت 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.

٤-١- مدل مخلوط گوسی

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

$$p(x_n) = \sum_{m=1}^{M} w_m p_m(x_n) = \sum_{m=1}^{M} w_m \mathrm{N}(x_n; \mu_m, \sigma^2{}_m) \quad (۳)$$

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 ام است. در مرحله آموزش، پارامترهای مدل را می‌توان با استفاده از الگوریتم بیشینه‌سازی انتظار به‌دست آورد[١٦]. 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 به‌دست آمده برابر ١٥ است. با توجه به تعداد چهار خوشه گوسی استفاده شده در مدل‌سازی‌ها در نهایت، یک بردار ویژگی ٦٠ تایی، نتیجهٔ مدل‌سازی غیرخطی فضای فاز قاب گفتار خواهد بود.

٥- 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 ٣٩ بُعدی ضرایب کپسترال حوزه مل و یک بردار ٦٠ بُعدی به‌دست آمده از پارامترهای مدل مخلوط گوسی می‌رسیم. بر این مبنا بُعد بردار نهایی به‌دست آمده برابر ٩٩ است. این بُعد برای بردار ویژگی، به‌دلایل زیر برای استفاده در کاربردهای بازشناسی گفتار پیوسته قابل قبول نیست:

- پیچیدگی محاسبات مورد نیاز در بازشناسی با افزایش بُعد بردار ورودی، به‌شدت 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 درجهٔ صحت عملکرد سیستم بازشناسی گفتار، در این تحقیق از روش تحلیل جدایی‌پذیری فیشر [١٧] برای کاهش بُعد بردار ویژگی بردار ترکیب اطلاعات استفاده شده است.

٥-١- تحلیل جدایی‌پذیری فیشر

همان‌گونه که گفته شد، یک بردار ٩٩ 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

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

همپوشانی اطلاعاتی با ویژگیهای دیگری باشند که حذف آنها تغییری در توانایی بازشناسی سیستم نهایی ایجاد نمی‌کند. بنابراین انتخاب بردار ویژگی نهایی با حداکثر تمایزپذیری، هدف این بخش 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 مسأله‌ای با دو کلاس خروجی از رابطه زیر به‌دست می‌آید:

$$F_{ratio}(W) = \frac{W^T S_B W}{W^T S_W W} \qquad (٤)$$

که در آن S_B ماتریس پراکندگی بین-کلاسی است که از رابطه زیر به‌دست می‌آید:

$$S_B = (\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2)(\underline{\hat{\mu}}_1 - \underline{\hat{\mu}}_2)^T \qquad (٥)$$

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

$$S_W = \sum_{i=1}^{K} S_{W\,i} \qquad (٦)$$

اگر ماتریس S_T را به شکل زیر تعریف کنیم.

$$S_T = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T \qquad (٧)$$

ماتریس S_B را می‌توان از رابطه زیر به‌دست آورد.

$$S_B = S_T - S_W \qquad (٨)$$

پس از محاسبه ماتریس‌های مورد نیاز، از رابطه (٤) می‌توان برای مسأله‌ای 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 پایین‌تری داشتند، با انجام آزمون بازشناسی گفتار، تأثیر حذف این ویژگی را در صحت سیستم بازشناسی 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 بُعدی از ویژگی به‌دست آمد که برای آموزش و آزمون طبقه‌بندی کننده خروجی به‌عنوان بردار ویژگی نهایی به کار رفت.

٦- نتایج آزمایشها

در این بخش نتایج آزمایشهای انجام شده در مراحل مختلف آورده می‌شود. 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 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 بازشناسی واج گسسته آورده می‌شود. قسمت پایانی این بخش نتایج آزمایشهای بازشناسی گفتار پیوسته را نشان می‌دهد.

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

٦-١- پایگاه داده و سیستم بازشناسی گفتار

سیگنال‌های گفتاری با فرکانس نمونه‌برداری ١٦ کیلوهرتز از پایگاه داده تیمیت [١٨] انتخاب شده است. پایگاه داده تیمیت شامل ٦٣٠٠ جمله است که شامل ١٠ جمله از ٦٣٠ گوینده از ٨ منطقهٔ 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[1] 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 ١٣٤٤ جمله از ١٦٨ گوینده است و ٢٧٪ کل دادگان را شامل می‌شود.

در آزمایشهای بازشناسی واج گسسته، از مجموعه آزمون هسته‌ای استفاده شده که شامل ١٩٢ جمله از ٢٤ گوینده است. در این آزمایشها واجها در ٣٩ کلاس خروجی طبقه‌بندی شد. به‌منظور ارزیابی عملکرد سیستم پیشنهادی در شرایط نویزی از پایگاه داده نویزی Noisex.٩٢ [١٩] 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 [٢٠] دانشگاه کمبریج یعنی معمولترین ابزار بازشناسی گفتار و از مدل پنهان مارکوف با شش حالت، ١٦ خوشه چپ به راست با ماتریس کواریانس قطری استفاده شده است. برای بخش سکوت از مدل سه حالته استفاده شده است. طول قاب‌های گفتار برابر ٢٥/٦ میلی ثانیه انتخاب شده است. خروجی‌های سیستم بازشناسی، رشته واجهای بازشناسی شده است.

٦-٢- انتخاب ویژگی‌ها

همان‌گونه که در بخش ٥ گفته شد، استفاده از مدل مخلوط گوسی با چهار خوشه به برداری با ٦٠ ویژگی منجر می‌شود. به‌منظور استفادهٔ همزمان از مزایای اطلاعات حوزه فاز و فرکانس در فرایند بازشناسی گفتار در این مقاله از ترکیب ویژگیها استفاده شده است.

جدول ١ برچسب‌های اختصاص داده شده به ویژگیهای ترکیبی

بر چسب	شماره ویژگی	دسته ویژگی
مل ١٣– مل ١	١ الی ١٣	ضرایب حوزه مل
دلتا ١٣ – دلتا١	١٤ الی ٢٦	ضرایب دلتا
دلتا دلتا١٣ – دلتا دلتا١	٢٧ الی ٣٩	ضرایب دلتا-دلتا
گوسی٤ – گوسی١	٤٠ الی ٤٣	ضرایب وزن خوشه‌های مدل مخلوط گوسی
گوسی ٣٢ – گوسی ٥	٤٤ الی ٦١	ضرایب میانگین خوشه‌های مدل مخلوط گوسی
گوسی ٦٠ – گوسی ٣٣	٦٢ الی ٩٩	ضرایب واریانس خوشه های مدل مخلوط گوسی

افزودن بردار ٣٩ بُعدی کپسترال مل به بردار ٦٠ بُعدی پارامترهای مدل گوسی، برداری ٩٩ بُعدی را نتیجه می‌دهد. جدول (١) برچسب‌های اختصاص داده شده به این ویژگیها را نشان می‌دهد. همان‌گونه که در بخش ٥ گفته شد، برای انتخاب ویژگیهای مؤثر برای فرایند بازشناسی از معیار جدایی‌پذیری فیشر استفاده شده است. با استفاده از معیار فیشر، ضریب فیشر برای تمامی ویژگیها محاسبه شد. شکل (٧) ضرایب محاسبه شده را برای ١٥ ویژگی با بیشترین ضریب تمایزپذیری نشان می‌دهد. به‌منظور کاهش بُعد بردار ویژگی نهایی، ویژگیهایی با کمترین ضریب فیشر 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

1. Core Test

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

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

۱) ضرایب کپسترال مل (۱،۲،۳،٤،٥، ٦، ۷، ۸، ۹ و ۱۰)

۲) ضرایب دلتای کپسترال مل (۱، ۲، ۳،٤، ۷، ۹، ۱۱، ۱۲، ۱۳)

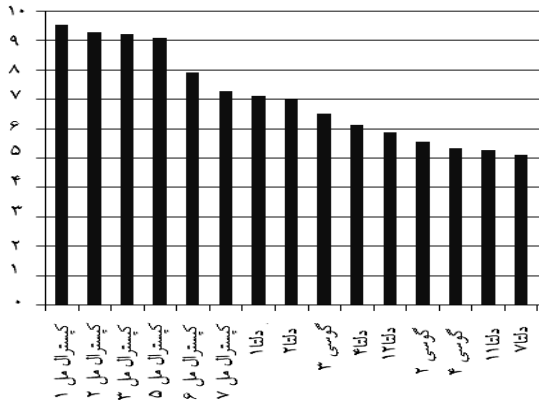
۳) ضرایب دلتا-دلتای کپسترال مل (۱، ۲، ۳، ٤، ٥، ٦، ۸)

٤) ضرایب مخلوط گوسی شامل ضرایب (۱، ۲، ۳، ٤، ٥، ٦،۷، ۱۱، ۱۲، ۱٥، ۱٦، ۱۷، ۱۸، ۲۰، ۲۱، ۲۷، ۲۹، ۳۰، ۳۳، ۳٤، ٤۰، ٤٥،٤۷، ٤۸،٥۳، ٥٥، ٥۹)

از میان این ۵۲ ویژگی نهایی، ۲٦ ویژگی از حوزهٔ فرکانس و ۲٦ ویژگی از حوزه فاز انتخاب شد.

٦-۳- آزمایشهای بازشناسی واج گسسته

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



شکل ۷ ضریب جدایی پذیری فیشر برای ۱۵ ویژگی با بالاترین تمایزپذیری

به‌منظور ارزیابی عملکرد سیستم پیشنهادی، آزمایشهای بازشناسی واج گسسته بر روی ٤۲۰۰ واج از مجموعه آزمون هسته‌ای پایگاه داده تیمیت انجام شد. از ضریب %COR به‌عنوان معیار عملکردی سیستم استفاده شد. این ضریب از درصد نسبت تعداد واجهای بازشناسی شده با برچسب صحیح به تعداد کل واجهای مورد آزمایش به‌دست 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 ٦۰ 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 به‌دست آمده از کاربرد ویژگیهای فضای فاز به‌تنهایی، 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. به‌منظور 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.

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

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	۷۲,٤
ویژگیهای مخلوط گوسی و مدل پاوینلی و همکاران [۱]	۵۳,٦
ویژگیهای مخلوط گوسی ٦۰ بُعدی	۵۷,۱
بردار ترکیبی پیشنهادی ۵۲ بُعدی	۷۸,۱

جدول۳ 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	۷۲,٤	۷۸,۱
نویز سفید	۵	۵۲,٤	۵۷,۵
	۱۰	۵۹,۳	٦۳,۸
نویز صورتی	۵	۵۱,۸	۵۵,۷
	۱۰	۵۸,٦	٦۲,۹
نویز همهمه	۵	٤۸,۷	۵۱,٤
	۱۰	۵۷,۱	٦۱,٤
نویز کارخانه	۵	٤۹,٤	۵۳,۱
	۱۰	۵۸,٦	٦۱,۷

٦-٤- 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 است. به‌دلیل 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. به‌منظور 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 کپسترال مل، ویژگیهای مدل مخلوط گوسی و بردارنهایی ترکیبی به‌صورت مجزا به‌دست آوردیم. نرخ بازشناسی واج با رابطه زیر تعریف می‌شود.

$$PRR = ۱ - PER = ۱ - \frac{S + I + D}{N} \qquad (۹)$$

جدول ٤ نتایج نرخ بازشناسی زنجیره واجی در آزمایشهای بازشناسی گفتار پیوسته به‌ازای بردارهای ویژگی مختلف

بردار ویژگی	تعداد جایگزینی نادرست	تعداد درج نادرست	تعداد حذف نادرست	درصد نرخ بازشناسی واج
ضرایب کپسترال مل	۲۸۵۹	۲٤۵۰	۱۳۱۱	٦٦,۹
ضرایب مدل مخلوط گوسی	۳۲٤۵	۲۷۰۰	۲٦۷۰	۵٦,۹
بردار ترکیبی پیشنهادی	۲۷۸۰	۱۷۵۰	۱۳۲۰	۷۰,۷

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 (٤) نرخ بازشناسی زنجیره

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

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

۷- جمع‌بندی

در این مقاله، به‌کارگیری 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 به‌دست 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 نظریهٔ 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 برازش شده به آن فراهم می‌سازد.

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 (٤) نشان داده شده. این نکته همچنین مزیت استفاده از قطع پوانکاره را در تحلیل سیگنالهایی با ماهیت شبه‌تناوبی 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 به‌عنوان 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 ٦۰ 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 ٦-۲ 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 به‌دلیل وجود مرزهای انعطاف‌پذیر تصمیم و امکان مدل‌سازی دادگان با بُعد بالا در چند خوشه محدود به‌عنوان 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 حوزهٔ فاز، باز هم عملکرد ویژگیهای معمول فرکانسی بهتر است و لذا مسأله ترکیب بردارهای ویژگی

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

به‌دست آمده از حوزهٔ فاز و فرکانس برای بهبود عملکرد سیستم بازشناسی گفتار ضروری است.

به‌منظور ارزیابی عملکرد سیستم پیشنهادی در شرایط نویزی، تعدادی آزمایش بازشناسی واج گسسته 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 دادگان نویزی واقعی‌تری مانند پایگاه داده Aurora2 به‌جای نویز جمعی استفاده کرد. روش پیشنهادی همچنین می‌تواند در کاربردهای دیگر پردازش سیگنال به‌ویژه 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 نقاط قطع پوانکاره از مدل مخلوط گوسی استفاده شد. انتخاب این مدل به‌دلیل 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 به‌دست 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 بُعدیی به‌عنوان بردار ویژگی نهایی برای استفاده در ورودی طبقه‌بندی کننده به‌دست 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 اندازهٔ 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.

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

[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 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.

[Downloaded from mjee.modares.ac.ir on 2024-05-05]

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.

[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

[Downloaded from mjee.modares.ac.ir on 2024-05-05]