DOI: 10.22041/ijbme.2012.13096

Iranian Journal of Biomedical Engineering 6 (2012) 17- 33

Feature Extraction based on Linear Modeling of Embedded Speech Trajectory in the Reconstructed Phase Space for Speech Recognition System

Y. Shekofteh[1]*, F. Almasganj [2]

[1] Ph.D Candidate, Bioelectric Department, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran, y_shekofteh@aut.ac.ir

[2] Associate Professor, Bioelectric Department, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran, almas@aut.ac.ir

Abstract

Recent researches show that nonlinear and chaotic behavior of the speech signal can be studied in the reconstructed phase space (RPS). Delay embedding theorem is a useful tool to study embedded speech trajectories in the RPS. Characteristics of the speech trajectories have rarely used in the practical speech recognition systems. Therefore, in this paper, a new feature extraction (FE) method is proposed based on parameters of vector AR (VAR) analysis over the speech trajectories. In this method, using filter and reflection matrices obtained from applying VAR analysis on static and dynamic information of the speech trajectory in the RPS, a high-dimensional feature vector can be achieved. Then, different transformation methods are utilized to attain final feature vectors with appropriate dimension. Results of discrete and continuous phoneme recognition over FARSDAT speech corpus show that the efficiency of the proposed FE method is better than other time-domain-based FE methods such as LPC and LPREF.

Key words: Speech recognition, Feature extraction, Reconstructed phase space, Signal Embedding, Linear Prediction, Vector AR.

***Corresponding author**

Address: *Faculty of Biomedical Engineering, Amirkabir University of Technology (Tehran Polytechnic), P.O.Box: 15875-3413, I.R. Iran., Postal Code: 15914, Tehran, I.R. Iran*
Tel: *+982164542372*
Fax: *+982166495655*
E-mail: *y_shekofteh@aut.ac.ir*

DOI: 10.22041/ijbme.2012.13096

استخراج ویژگی‌های مبتنی بر مدل‌سازی خطی تراژکتوری گفتار جاسازی شده در فضای بازسازی شده فاز برای سیستم بازشناسی گفتار

یاسر شکفته [۱*]، فرشاد الماس‌گنج [۲]

[۱] دانشجوی دکتری مهندسی پزشکی، گروه بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر (پلی‌تکنیک ایران)، تهران

[۲] دانشیار، گروه بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر (پلی‌تکنیک ایران)، تهران almas@aut.ac.ir

چکیده

تحقیقات اخیر نشان می‌دهد که تظاهرات غیرخطی و آشوبی سیگنال گفتار می‌تواند در حوزهٔ فضای بازسازی شده فاز (RPS) مطالعه شود. تئوری جاسازی برمبنای محورهای تأخیری، ابزار مناسبی برای بررسی تراژکتورهای گفتاری در RPS است. تاکنون از مشخصه‌های تراژکتورهای گفتاری به ندرت در سیستم‌های کاربردی بازشناسی گفتار استفاده شده است. از اینرو در این مقاله روش استخراج ویژگی جدیدی براساس پارامترهای مدلسازی خطی مبتنی بر روش AR برداری (VAR) پیشنهاد شده است. در این روش بوسیله ماتریس ضرایب فیلتر و یا ضرایب انعکاسی به دست آمده از اعمال روش VAR بر مشخصه‌های استاتیک و دینامیک تراژکتوری های گفتاری شکل یافته در RPS، یک بردار ویژگی با بُعد زیاد حاصل می‌شود که می‌توان از روش‌های نگاشت خطی برای کاهش بُعد مناسب آن استفاده کرد. نتایج آزمایش‌های بازشناسی واج مجزا و پیوسته بر مجموعه دادگان گفتاری فارس‌دات نشان می‌دهد که کارایی این روش در مقایسه با دیگر روش‌های متداول استخراج ویژگی مبتنی بر حوزهٔ زمان مانند روش LPC و LPREF بیشتر است.

کلیدواژگان: بازشناسی گفتار، استخراج ویژگی، فضای بازسازی شده فاز، جاسازی سیگنال، پیش‌بینی خطی، AR برداری.

[*] عهده‌دار مکاتبات

نشانی: تهران، خیابان حافظ، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی **صندوق پستی:** ۱۵۸۷۵-۴۴۱۳

تلفن: ۰۲۱۶۴۵۴۲۳۷۲، **دورنگار:** ۰۲۱۶۶۴۹۵۶۵۵ ، **پیام نگار:** y_shekofteh@aut.ac.ir

۱- مقدمه

سیستم تولید گفتار انسان شامل فرایندی غیرخطی، پیچیده و چند متغیره است که خروجی آن به طور معمول به شکل یک مشاهدهٔ تک متغیره[1] یا تک بُعدی[2] (اسکالر) تحت عنوان سیگنال زمان- گسسته گفتار بوسیله میکروفون ثبت می‌شود. در فرایند تولید سیگنال گفتار عواملی مانند ارتعاش غیرخطی تارهای صوتی و یا حرکات لایه‌ای هوا در مجرای صوتی، منجر به بروز تظاهرات آشوبگونه[3] در شکل سیگنال گفتار می‌شود [۱-۵]. از این‌رو فرض‌های مبتنی بر مدل‌سازی خطی فرایند تولید گفتار که در روش‌های متداول منبع-فیلتر (مانند روش‌های استخراج ویژگی LPC، LFBE، PLP و MFCC) استفاده می‌شود، منطبق با واقعیت نخواهد بود. از طرف دیگر، با توجه به ویژگی‌های سیگنال‌های غیرخطی و آشوبگونه، می‌توان سیگنال گفتار را به صورت یک سری زمانی با توصیف چندمتغیره، در فضای بازسازی شده فاز[4] (RPS) بیان کرد [٦-٨]. این انتقال سیگنال به RPS می‌تواند به صورتی انجام شود که رفتار تراژکتوری جاسازی شده در آن از لحاظ هندسی معادل با تراژکتوری واقعی سیستم تولیدکنندهٔ گفتار باشد [۹].

در همین زمینه، تیکنز[5] روش‌های جاسازی فضایی[6] در RPS را یکی از تکنیک‌های مطالعه دینامیک‌های غیرخطی و آشوبی معرفی کرد[۱۰]. بر مبنای این روش‌ها، ثبت یک متغیر زمانی از رفتار سیستم دینامیکی می‌تواند برای کسب اغلب خواص دینامیکی آن سیستم "کافی" باشد[۹]. در نتیجه با جاسازی سیگنال گفتار در حوزهٔ RPS، می‌توان امید داشت اطلاعات بیشتر و متمایزکننده‌تری از سیگنال، خصوصاً از لحاظ رفتار دینامیک و پیچیده آن در مقایسه با دیگر روش‌های متداول کسب کرد.

در دو دههٔ اخیر تحقیقات متعددی در حوزهٔ پردازش سیگنال گفتار بر مبنای استفاده از روش جاسازی سیگنال گفتار در RPS انجام شده است [۱۱-۱۳]. اساس این تحقیقات، نمایش و استخراج اطلاعات مربوط به حوزهٔ RPS سیگنال گفتار است که می‌تواند حاوی اطلاعات متمایزکننده‌تری در مقایسه با روش‌های متداول تحلیل سیگنال گفتار باشد. بخشی از این اطلاعات متمایز متأثر از مشخصه‌هایی است که بواسطه حذف اطلاعات فاز در روش‌های استخراج ویژگی مبتنی بر اندازهٔ طیف سیگنال از بین رفته‌اند. در برخی تحقیقات الهام گرفته شده از سیستم شنوایی انسان ادعا می‌شود این سیستم به اطلاعات فاز طیف سیگنال حساس نیست؛ اما بخشی از تحقیقات منتشر یافتهٔ اخیر نشان‌دهندهٔ مؤثر بودن اطلاعات فاز در بهبود کارایی سیستم بازشناسی گفتار است[١٤-١٦]. بر این اساس اطلاعات فاز طیف هنگامی حائز اهمیت است که اثر اندازه و شکل پنجره اعمالی بر قطعات گفتاری به دقت بازبینی شود.

در حوزهٔ پردازش سیگنال گفتار، علیرغم انجام تحقیقات متعدد مبتنی بر روش جاسازی سیگنال در RPS، متأسفانه تاکنون روش‌های محدودی برای استفاده از قابلیت آن در سیستم‌های خودکار بازشناسی گفتار (ASR) ارائه شده است. عمدهٔ فعالیت‌های انجام شده در این بخش را می‌توان به پنج دسته کلی- که در ذیل آمده است- تقسیم کرد:

الف: استخراج ویژگی‌های مبتنی بر مشخصه‌های نامتغییر آشوبگونه متداول در RPS.

به طور مثال در منابع [۱۷-۲۰] از مشخصه‌هایی مانند بُعد همبستگی[7]، بُعد فرکتال[8]، بُعد فرکتال تعمیم یافته[9] و نماهای لیاپانوف[10] به عنوان ویژگی‌های جدید در طبقه‌بندی واجی و یا بازشناسی گفتار استفاده شده است.

[1] Univariate
[2] One-dimensional
[3] Chaotic
[4] Reconstructed phase space
[5] Takens
[6] Embedding space
[7] Correlation
[8] Fractal dimension
[9] Generalized fractal
[10] Lyapunov exponents

ب: استخراج ویژگی‌های تجربی و جدید مبتنی بر خصوصیات رفتار دینامیکی تراژکتوری گفتاری در RPS.
به طور مثال درمنبع [۲۱] دو مشخصه جدید از سیگنال جاسازی شده در RPS معرفی شده است. این ویژگی‌ها مقادیر اسکالری هستند که اندازه جابجایی و یا دایروی بودن مسیر تراژکتوری سیگنال را در RPS بازنمایی می‌کنند. از طرف دیگر در منابع [۲۳،۲۲] نیز طبقه‌بندی واحدهای CV گفتاری با توسعه روشی مبتنی بر هیستوگرام (SSPD) و طبقه‌بندهای شبکهٔ عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) پیشنهاد شده است.
ج: استخراج ویژگی‌های مبتنی بر توزیع مشخصه‌های استاتیک و دینامیک تراژکتوری‌های گفتاری در حوزهٔ RPS.
به طور مثال در منبع [۲٤] از توزیع آماری تراژکتوری‌های گفتاری جاسازی شده در RPS بوسیله روش هیستوگرام در کلاس‌بندی واج‌های مجزا[11] در مجموعه دادگان گفتاری TIMIT استفاده شده است. در منبع [۲۵] نیز استفاده از تحلیل مؤلفه‌های اساسی (PCA) در متعامد کردن و کاهش بُعد سیگنال جاسازی شده در RPS بررسی شده است. مشابه روش آماری هیستوگرام، در منابع [۲۷،۲٦] از مدل پارامتری مخلوط گوسی[12] (GMM) برای مدل‌سازی جداگانه توزیع مشخصه‌های استاتیک و دینامیک هر واج در RPS برای کاربرد بازشناسی واج مجزا استفاده شده است.
د: استخراج ویژگی‌های مبتنی بر پارامترهای مدل‌سازی تراژکتوری‌های گفتاری در حوزهٔ RPS.
به طور مثال در منابع [۲۹،۲۸] ویژگی‌های مبتنی بر RPS از پارامترهای مدل GMM یادگرفته شده از توزیع سیگنال جاسازی شده در RPS و یا قطع پوانکارهٔ[13] آن به دست آمده است. در این روش‌ها با انتخاب برخی پارامترهای مدل GMM تعلیم یافته، بردار ویژگی گفتاری نهایی تولید شده است. متأسفانه هزینهٔ محاسباتی زیاد در اجرای این روش‌ها یکی از ضعف‌های اساسی آن‌ها در راه‌اندازی سیستم‌های زمان حقیقی ASR بشمار می‌رود.
ه: استخراج ویژگی‌های مبتنی بر امتیازهای صوتی به دست آمده از مدل‌سازی تراژکتوری‌های گفتاری در حوزهٔ RPS.
به طور مثال در منبع [۳۰] روشی پیشنهاد شده است؛ مبنی بر اینکه ویژگی‌های گفتاری به گونه‌ای از مدل‌سازی تراژکتوری‌های گفتاری استخراج شوند تا علاوه بر هزینهٔ محاسباتی مقبول و کارایی مناسب، قابلیت استفاده در سیستم‌های بازشناسی گفتار پیوسته[14] (CSR) را داشته باشند. در این روش ابتدا مجموعه‌ای از مدل‌های از پیش تعلیم یافته گفتاری مرجع آماده می‌شوند. این مدل‌ها- که می‌توان آن‌ها را مانیفولد[15] یا جاذب گفتاری تلقی کرد- وظیفهٔ مدل‌سازی توزیع مشخصه‌های هر یک از واحدهای گفتاری واجی را در RPS برعهده دارند. سپس برای هر قاب گفتاری، ویژگی‌های PPRPS، بر مبنای محاسبهٔ مقدار شباهت (درست‌نمایی شرطی[16]) نمونه‌های قاب گفتاری جاسازی شده در RPS با هر یک از مدل‌های مرجع محاسبه می‌شوند. در منبع [۳۰] نشان داده شده است که استفاده از این ایده منجر به بهبود عملکرد سیستم بازشناسی واج مجزا در مقایسه با روش معرفی شده در منبع [۲۷] خواهد شد. همچنین کاربرد این روش برای استفاده در سیستم CSR با مدل مخفی مارکوف[17] (HMM) بررسی شده است.

در این مقاله روش جدید برای استخراج ویژگی گفتاری پیشنهاد می‌شود که در دستهٔ روش‌های "استخراج ویژگی مبتنی بر پارامترهای مدل‌سازی تراژکتوری‌های گفتاری در حوزهٔ RPS" قرار می‌گیرد. در این روش، از پارامترهای به دست آمده از مدل‌سازی خطی مبتنی بر روش پیش‌بینی

[11] Isolated Phoneme
[12] Gaussian mixture model
[13] Poincare section
[14] Continuous speech recognition
[15] Manifold
[16] Conditional likelihood
[17] Hidden Markov model

خطی AR چندبُعدی (MVAR) یا AR برداری (VAR) برای مدل‌سازی تراژکتوری‌های گفتاری در RPS استفاده می‌شود. برای این منظور از ماتریس‌های ضرایب فیلتر و یا ضرایب انعکاسی به دست آمده از اعمال روش VAR بر مشخصه‌های استاتیک و یا دینامیک تراژکتوری‌های گفتاری شکل یافته در RPS، یک بردار ویژگی با بُعد زیاد حاصل می‌شود که می‌توان با اعمال روش‌های پس‌پردازش، بردار ویژگی گفتاری مناسبی تولید کرد.

ادامه مقاله به این صورت است: در بخش دوم تئوری جاسازی سیگنال در RPS معرفی خواهد شد. بخش سوم روش پیش‌بینی خطی و انواع آن را شرح می‌دهد. در بخش چهارم الگوریتم روش استخراج ویژگی پیشنهادی آورده شده است. بخش‌های پنجم و ششم به ترتیب مجموعه دادگان مورد استفاده و آزمایش‌های اجرا شده را ارائه می‌کنند. در بخش هفتم بحث و بررسی نتایج به دست آمده آورده شده است و در انتهای مقاله نیز نتیجه‌گیری آن ارائه خواهد شد.

۲- جاسازی سیگنال در فضای بازسازی شدهٔ فاز (RPS)

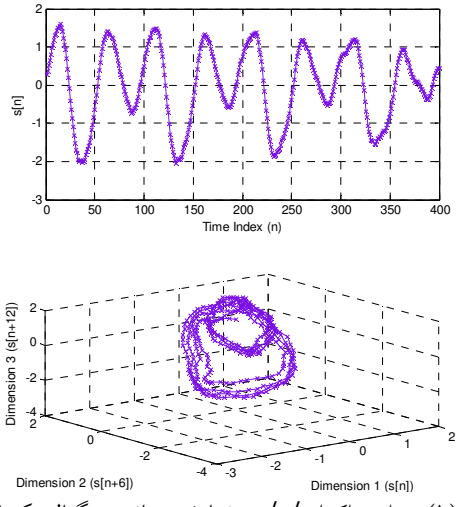
جاسازی سیگنال تک متغیره زمانی با روش محورهای تأخیری[18] بر اساس تئوری های مطرح شده در [۳۱،۱۰] است. در این روش، پس از تعیین پارامترهای مناسب تأخیر زمانی (t) و بُعد جاسازی (d)، نمونه‌های سیگنال زمانی اولیه را می‌توان به گونه‌ای به فضای چند بُعدی RPS انتقال داد که در آن حوزهٔ، دینامیک واقعی سیستم تولیدکنندهٔ آن سیگنال به صورت مناسب نمایش داده شود [۹]. در ادامه، روش جاسازی سیگنال در حوزهٔ RPS بیان می‌شود.

فرض کنیم که سری زمانی s ($s=\{s[1],..,s[N]\}$) (نمونه‌های سیگنال زمانی یک قاب گفتاری) با تعداد N نمونه، سیگنال زمانی تک متغیرهٔ مورد بحث باشد. متناظر با این سری زمانی، مجموعه‌ای متوالی از نقاط جاسازی شده در فضای چندبُعدی RPS قابل تعریف است. به طور مثال می‌توان $\mathbf{S_i}$ را نمونه iام جاسازی شده در فضای RPS با رابطهٔ (۱) و به صورت یک بردار d بُعدی نشان داد:

$$\mathbf{S}_i = [s[i], s[i+t], \dots, s[i+(d-1)t]], \qquad (۱)$$

پارامتر d، بُعد جاسازی و پارامتر t معادل با زمان تأخیر است. واضح است که برای قابی با طول N، تعداد نقاط جاسازی شده در حوزهٔ RPS برابر با $L = N - (d-1)t$ خواهد بود.

روش‌های مختلفی برای انتخاب مقدار بهینهٔ بُعد جاسازی و تأخیر زمانی وجود دارند [۳۴-۲۸،۹،۶،۳۲]. انتخاب مناسب برای سیگنال گفتار میکروفونی با آهنگ نمونه‌برداری ۱۶ کیلوهرتز، مقادیر $d=8$ و $t=6$ است [۲۸-۳۰]. شکل (۱) نمونه‌ای از یک قاب سیگنال گفتار (واج /u/) و همچنین تراژکتوری متناظر با آن را - که در فضای سه بُعدی RPS جاسازی شده است- نشان می‌دهد.



شکل (۱)- واج واکدار /u/ در نمایش زمانی سیگنال تک بُعدی (شکل بالا) و نمایش دینامیک تراژکتوری سه بعدی آن (شکل پایین) بوسیله روش جاسازی در RPS.

[18] delay-coordinat

رفتار تراژکتوری سیگنال گفتار در حوزهٔ RPS برای اکثر واج‌های واکدار مشابه فرایند قبض[19] در سیگنال‌های آشوبی است. به عنوان مثال در شکل (۱) نمونه‌ای از این رفتار برای تراژکتوری واج واکدار /u/ نشان داده شده است. از طرف دیگر برای واج‌های گفتاری انفجاری (مانند واج‌های /b/ و /t/) رفتار بسط[20] سیگنال آشوبی در حوزهٔ RPS قابل مشاهده است [۳۵،۳۶].

۳- روش پیش‌بینی خطی

پیش‌بینی خطی[21] (LP) یک روش مدل‌سازی سیگنال براساس مدل تمام قطب[22] است [۳۷]. ضرایب بدست آمده از این روش، منجر به تخمین نمونه‌های سیگنال، بر اساس ترکیبی خطی از چند نمونه قبلی آن‌ها می‌شود. در این بخش اصول عملکرد روش پیش‌بینی خطی در حالت یک بُعدی (مورد استفاده در روش‌های متداول استخراج ویژگی گفتاری) و چند بُعدی (مورد استفاده در روش پیشنهادی) معرفی و بررسی خواهد شد.

۳-۱- روش پیش‌بینی خطی تک بُعدی

روش پیش‌بینی خطی (LP) از قدیمی‌ترین روش‌های پردازش سیگنال گفتار است که غالباً به عنوان روش‌های LPC[23] یا AR[24] در پردازش سری‌های زمانی شناخته می‌شود [۳۸-۴۰]. کاربردهای آن شامل کدینگ گفتار پیش‌بین [۴۱]، بهسازی گفتار [۴۲]، بازشناسی گفتار و شناسایی گوینده [۴۱،۴۳] است. در روش استخراج ویژگی مبتنی بر تحلیل پیش‌بینی خطی، سیگنال پنجره گذاری شده گفتاری با فیلتری تمام قطب H(z) مدل سازی خواهد شد. تابع انتقال این مدل در رابطهٔ (۲) نشان داده شده است:

$$H(Z)=\frac{1}{1-\sum_{i=1}^{p}a_i z^{-i}}, \qquad (۲)$$

در این معادله p تعداد قطب‌های مدل یا همان درجه مدل[25] است. ضرایب فیلتر (a_i) این مدل به گونه‌ای انتخاب می‌شوند که مجموع میانگین مجذورخطای پیش‌بینی در پنجره مورد تحلیل، کمینه شود. تعداد قطب‌ها (p)، در واقع همان تعداد نمونه‌های قبلی از سیگنال است که برای پیشگویی نمونه جدید استفاده می‌شود:

$$x[n]=\sum_{j=}^{p}a_j x[n-j]+e[n], \qquad (۳)$$

که در آن p درجه مدل، $x[n]$ نمونه nام از سیگنال گفتار پنجره‌گذاری شده و $e[n]$ مقدار خطای پیش‌بینی است.

محاسبه ضرایب فیلتر a_j (ضرایب LP)، به طور متداول با روش بازگشتی لوینسن[26] براساس محاسبات دنباله خودهمبستگی تعیین می‌شود [۳۸،۴۰] که در ادامه توضیح داده می‌شود.

فرض کنیم که یک قاب گفتاری شامل N نمونه به صورت $\{x[j], j = 1,2,..,N\}$ داده شده است. تعداد p+1 عبارت اول دنباله خودهمبستگی به صورت زیر محاسبه می‌شود:

$$r_i=\sum_{j=1}^{N-i}x[j]x[j+i],\text{i=0,1,2,…,p.} \qquad (۴)$$

ضرایب فیلتر[27] به طور بازگشتی و با محاسبهٔ تعدادی ضرایب کمکی[28]- که اغلب ضرایب انعکاسی[29] نامیده می‌شوند- بدست می‌آیند [۴۰]. فرض کنیم $a_j^{(i-1)}$ و $k_j^{(i-1)}$ به ترتیب ضرایب jام فیلتر و انعکاسی برای یک مدل با درجه i-1 باشد. در این حالت ضرایب مدل با درجه i به طور بازگشتی از مقدار i=1 تا i=p از روابط زیر محاسبه می‌شود.

[19] Folding/Squeezing
[20] Stretching
[21] Linear prediction
[22] All pole
[23] Linear predictive coding
[24] auto-regressive modeling
[25] Model order
[26] Levinson
[27] Filter coefficients
[28] Auxiliary coefficients
[29] Reflection coefficients

$$
\begin{aligned}
&k_j^{(i)} = k_j^{(i-1)}, for\ j = 1,2,\dots,i-1,\\
&k_i^{(i)} = \frac{\left(r_i + \sum_{j=1}^{i-1} a_j^{(i-1)} r_{i-j}\right)}{E^{(i-1)}},\\
&E^{(i)} = \left(1 - k_i^{(i)} k_i^{(i)}\right) E^{(i-1)}, E^{(0)} = r_0,\\
&a_j^{(i)} = a_j^{(i-1)} - k_i^{(i)} a_{i-j}^{(i-1)}, for\ j = 1,\dots,i-1,\\
&a_i^{(i)} = -k_i^{(i)}.
\end{aligned}
\qquad (۵)
$$

از ضرایب فیلتر و یا ضرایب انعکاسی محاسبه شده برای هر قاب گفتاری می‌توان به عنوان ویژگی‌های گفتاری آن قاب در بازشناسی گفتار استفاده کرد. در جعبه ابزار پیاده‌سازی سیستم بازشناسی گفتار HTK، این ویژگی‌ها به ترتیب LPC و LPREFC نامیده می‌شوند[٤٤].

۳-۲- روش پیش‌بینی خطی برداری (چندبُعدی)

روش مدل‌سازی LP را- که مختص سیگنال‌های تک بُعدی (اسکالر) است- می‌توان به سیگنال‌های چندبُعدی توسعه داد، که در این صورت روش AR برداری یا VAR[۳۰] حاصل می‌شود. روش VAR برای تعیین وابستگی‌های بین عناصر و نمونه‌های یک سیگنال چند متغیره بسیار مناسب است [٤٥،٤٦]. در اجرای VAR با درجهٔ P، فرض می‌شود سیگنال چند متغیره $\mathbf{X}[n]$ با بُعد K، به صورت مجموع وزن‌داری از P نمونه قبلی سیگنال به علاوه یک بردار خطا $\mathbf{E}[n]$، قابل بیان است:

$$\mathbf{X}[n] = \sum_{j=1}^{P} \mathbf{A}[j]\mathbf{X}[n-j] + \mathbf{E}[j], \qquad (٦)$$

که در آن $\mathbf{X}[i]$ و $\mathbf{E}[i]$ بردارهایی با اندازه K و ماتریس‌های ضرایب مدل $\mathbf{A}[i]$ در ابعاد K×K هستند. برای محاسبه ضرایب مدل VAR روشی با استفاده از ماتریس خودهمبستگی سیگنال چندبُعدی مشابه روش AR معمولی پیشنهاد شده است [٤٨،٤٧]. در جعبه ابزار TSA، برای تخمین بهینه ماتریس ضرایب فیلتر و انعکاسی، استفاده از الگوریتم تخمین کواریانس جزئی[۳۱] بر مبنای روش کواریانس بدون بایاس *Vieira-Morf* پیشنهاد شده است[٤٩].

٤- الگوریتم روش پیشنهادی

در این بخش الگوریتم روش استخراج ویژگی پیشنهادی بر مبنای مدلسازی خطی تراژکتوری‌های گفتاری حاصل شده از انتقال نمونه‌های یک قاب گفتاری به RPS پیشنهاد و معرفی خواهد شد. این ایده برگرفته از فرضی است که بر طبق آن سیستم تولید گفتار حاوی دینامیک غیرخطی با بُعد-کم است و می‌تواند به صورت یک تابع غیرخطی H درجه یک، در فضای زمان گسسته n بررسی شود:

$$\mathbf{Z}[n] = H(\mathbf{Z}[n-1]) \qquad (۷)$$

به طوری که بردار چند بُعدی Z[n] در فضای حالت واقعی سیستم قرار دارد [٦]. در اینجا استفاده از تقریب خطی با بکارگیری روش پیش‌بینی خطی برداری (VAR) برای تابع غیرخطی H پیشنهاد می‌شود که در واقع اساس ایده الگوریتم پیشنهادی است. استفاده از این ایده منجر به مدل‌سازی دقیق‌تر دینامیک گفتار در حوزهٔ RPS در مقایسه با شکل عادی آن در حوزهٔ زمانی می‌شود.

در روش پیشنهادی، ابتدا نمونه‌های سیگنال گفتار متناظر با هر قاب گفتاری بهنجار و سپس به حوزهٔ RPS منتقل می‌شود. استفاده از فرایند هنجارسازی (نرمالیزاسیون) منجر به مقاوم‌سازی روش پیشنهادی نسبت به تنوعات مربوط به اندازه شدت سیگنال گفتار خواهد شد. سپس تراژکتوری به دست آمده از سیگنال جاسازی شده در RPS به وسیله روش VAR مدل‌سازی خواهد شد. ضرایب به دست آمده از این روش بیانگر خصوصیات خطی دینامیک مانیفولدهای گفتاری در حوزهٔ RPS هستند. تعداد این ضرایب نسبتاً زیاد است و قابلیت استفادهٔ مستقیم در

[30] Vector AutoRegressive

[31] Partial correlation estimation

سیستم‌های بازشناسی گفتار متداول را ندارد. لذا از روش‌های کاهش بُعد مانند نگاشت‌های خطی[۳۲] (LT) برای تولید بردار ویژگی نهایی استفاده می‌شود. در ادامه این بخش به توضیح مراحل و جزئیات الگوریتم روش استخراج ویژگی پیشنهادی خواهیم پرداخت. این الگوریتم شامل پنج مرحله زیر است:

الف) بهنجارکردن نمونه‌های هر قاب گفتاری به مقادیر میانگین و واریانس نمونه‌های آن قاب (تولید سری زمانی با مقدار میانگین صفر و مقدار واریانس واحد).

ب) جاسازی سری زمانی بهنجار شده در RPS، با مقادیر بُعد جاسازی ۸ (*d=8*) و پارامتر تأخیر زمانی ٦ (*t=6*)، که منجر به تشکیل یک تراژکتوری d بُعدی گفتاری در RPS برای هر قاب گفتاری می‌شود.

ج) محاسبهٔ تعداد P ماتریس ضرایب فیلتر یا ضرایب انعکاسی با استفاده از روش VAR برای تراژکتوری تشکیل شده از هر قاب گفتاری در RPS. لازم است ذکر شود که این مدل‌سازی ضمن اعمال بر ماتریس حاصل از مشخصه استاتیک ۸ بُعدی تراژکتوری (S)، می‌تواند بر ماتریس‌های حاصل از مشخصه ۸ بُعدی گذر یا دینامیک تراژکتوری (D) و یا حتی ماتریس حاصل از الحاق توأم مشخصه‌های استاتیک و دینامیک (SD) تراژکتوری گفتاری (مشخصه ١٦ بُعدی) در RPS، اعمال شود.

ماتریس‌های حاوی مشخصه‌های استاتیک (S) و دینامیک (D) تراژکتوری در RPS شامل مجموعه‌ای برداری از تعداد L نقطهٔ جاسازی شده در RPS (تعریف شده در رابطهٔ ۱) هستند و بصورت زیر محاسبه می‌شوند:

$$\mathbf{S}_{L\times d}=\begin{bmatrix}\mathbf{S}_1\\ \vdots\\ \mathbf{S}_L\end{bmatrix},\ \&\ \mathbf{D}_{L\times d}=\begin{bmatrix}\mathbf{S}_1-\mathbf{S}_0\\ \vdots\\ \mathbf{S}_L-\mathbf{S}_{L-1}\end{bmatrix}, \qquad (۸)$$

که در آن S_0 می‌تواند به عنوان آخرین نقطه جاسازی شده از قاب گفتاری قبلی تعریف شود. در منبع [۵۰] نشان داده شده که ماتریس $L\times 2d$ بُعدی حاوی اطلاعات توأم استاتیک و دینامیک (**SD=[S,D]**) نیز می‌تواند دربرگیرنده اطلاعات مفیدی از RPS باشد.

د) انتقال عناصر درون ماتریس‌های ضرایب (فیلتر یا انعکاسی) به یک بردار ویژگی اولیهٔ Q. در این حالت اگر از ماتریس مشخصهٔ استاتیک تراژکتوری (S) و یا ماتریس مشخصه دینامیک تراژکتوری (D) استفاده شده باشد، اندازه بردار اولیه Q برابر با $Dq=P\times d\times d$ و اگر از ماتریس توأم مشخصه‌های استاتیک و دینامیک (SD) استفاده شود، اندازه بردار Q برابر با $Dq=P\times 2d\times 2d$ خواهد بود.

ه) محاسبه ماتریس نگاشت خطی W^{LT} با استفاده از روش‌های متداول نگاشت خطی مانند LDA[۳۳]، HLDA[۳۴]، LPP[۳۵] و OLPP[۳۶] به منظور غیرهمبسته کردن و کاهش بُعد بردار ویژگی اولیه Q (*Dq*) به بردارهای ویژگی نهایی Y (با اندازه بُعد دلخواه *Dy*).

تبدیل بردارها با استفاده از ضرب ماتریسی $Y_{[Dy\times 1]}=W^{LT}_{[Dy\times Dq]}Q_{[Dq\times 1]}$ انجام می‌شود. در تحقیق حاضر، از دو مقدار مختلف *Dy=13* (در صورت استفاده از ضرایب دینامیک مشتقات اول و دوم بردار ویژگی Y) و یا *Dy=39* (در صورت عدم استفاده از ضرایب دینامیک مشتقات اول و دوم بردار ویژگی Y) استفاده شده است. شایان ذکر است که در منابع [۵۱،۵۲] روش‌های نگاشت خطی مذکور در کاربردهای بازشناسی گفتار و پردازش تصویر به طور کامل معرفی شده و مقایسه‌ای میان این روشها انجام شده است.

همانطور که اشاره شد، با انتخاب مناسب پارامتر بُعد جاسازی *d=8* و درجه مدل پیش بینی خطی چند متغیره P، پس از اِعمال روش مدلسازی VAR، تعداد P ماتریس با المان‌های زیاد تولید می‌شود. با توجه به اینکه ویژگی‌های

[32] Linear transform
[33] Linear Discriminant Analysis
[34] Heteroscedastic LDA
[35] Locality Preserving Projection
[36] Orthogonal LPP

استخراج شده از سیگنال گفتار، در سیستم‌های CSR، معمولاً در ابعادی مانند ۳۹، ٤٥ و یا ٥۲ بکار گرفته شده‌اند؛ از این‌رو استفاده از یک روش مناسب کاهش بُعد، برای تولید بردار ویژگی نهایی ضروری به نظر می‌رسد. در اینجا از روش‌های متداول نگاشت خطی (LT) استفاده شده است تا علاوه بر کاهش تعداد مؤلفه‌های بردار ویژگی نهایی، عناصر درون هر بردار را نسبت به هم غیرهمبسته نماید. توجه شود که غیرهمبسته بودن عناصر بردارهای ویژگی، فرضی است که برای تعلیم مدل‌های صوتی مبتنی بر ساختار GMM/HMM، زمانی‌که از بردار واریانس بجای ماتریس کوواریانس در مدل‌سازی GMM برای توزیع هر حالت[۳۷] استفاده می‌شود (فرض قطری بودن ماتریس کواریانس)، بسیار مطلوب است.

۵- مجموعه دادگان

دادگان گفتاری مورد استفاده در این مقاله، مجموعه "دادگان فارس‌دات کوچک میکروفونی" با حجم زمانی حدود ٦ ساعت است که آهنگ نمونه‌برداری آن به kHz ۱٦ تقلیل یافته است [٥۳]. این دادگان شامل جملات پیوسته بیان شده از ۳۰٤ گویندهٔ زن و مرد با ده لهجه غالب فارسی است به طوری که هر فرد ۲۰ جمله را به زبان فارسی بیان کرده است. جملات دادگان شامل برچسب‌دهی آوایی با ٤۳ برچسب مختلف (بدون سکوت) است.

در پژوهش حاضر با تجمیع تعدادی از برچسب‌ها از ۳۰ برچسب واجی (۲۹ واج متداول فارسی به همراه یک برچسب سکوت) استفاده شده است [٥٤]. بنابراین واحدهای به‌کار رفته در مدل‌های صوتی بازشناسی شامل ۲۹ واج فارسی، همراه یک مدل برای سکوت است. در اینجا از داده‌های مربوط به ۲٥۰ گوینده اول (حدود ۲۸۲ دقیقه) در بخش تعلیم مدل بازشناس و از داده‌های ٥٤ گوینده آخر (حدود ۷٥ دقیقه) در بخش آزمون استفاده شده است. ازاین‌رو بازشناسی به دست آمده از این تحقیق، به صورت بازشناسی گفتار مستقل از گوینده است. مجموعه آزمون این داده‌ها شامل حدود ٦ هزار و پانصد کلمه از یک مجموعه ۱۰۷٦ کلمه‌ای (دایره لغات متوسط) و تعداد واج‌های آن حدود ۳۳ هزار واج است. همچنین برای انجام دادن آزمایش‌های بازشناسی واج مجزا، در بخش تعلیم مدل‌های صوتی از داده‌های ۲٥۰ گویندهٔ اول فارس‌دات (شامل ۱۲۹٥۲۷ واج) و در بخش آزمون از داده‌های٥٤ گویندهٔ باقیمانده (شامل ۲۸٦٥۸ واج) استفاده شده است. در مجموعه آزمون بیشترین فراوانی واج به ترتیب متعلق به واکه /a/ با ۳۷٥۸ نمونه، واکه/e/ با ۲۷۸۲ نمونه و واکه /A/ با ۱۷۸٥ نمونه است. کمترین فراوانی نیز به ترتیب مربوط به واج‌های/zh/ (۲۰۰ نمونه)، /j/ (۲۰۳ نمونه) و /ch/ (۳۰۱ نمونه) است.

٦- آزمایش‌ها

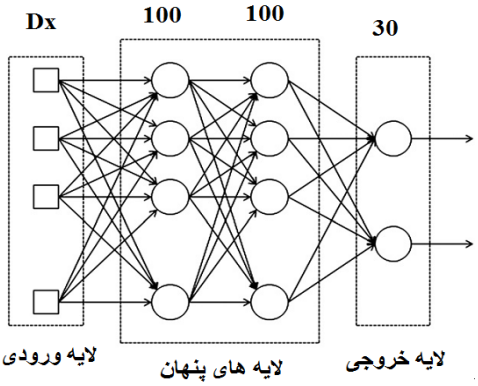
در این بخش بوسیله طراحی آزمایش‌های بازشناسی واج مجزا و واج پیوسته به ارزیابی عملکرد روش استخراج ویژگی پیشنهادی و مقایسه کارایی آن با دیگر روش‌های استخراج ویژگی متداول خواهیم پرداخت.

٦-۱- آزمایش‌های بازشناسی واج مجزا

در این قسمت آزمایش‌های مربوط به اجرای روش پیشنهادی در بازشناسی واج مجزا آورده شده است. در آزمایش بازشناسی واج مجزا در این بخش از یک طبقه‌بند شبکهٔ عصبی دولایه پنهان MLP با معیار آموزش کمینه کردن مجذور خطا (MSE) استفاده شده است. با این

[37] State

ساختار شبکه عصبی- که در شکل (۲) نشان داده شده است- خروجی شبکه می‌تواند تخمینی از مقدار احتمال پسین طبقه‌های واجی را به ازای بردار ویژگی ورودی به شبکه عصبی تولید کند.

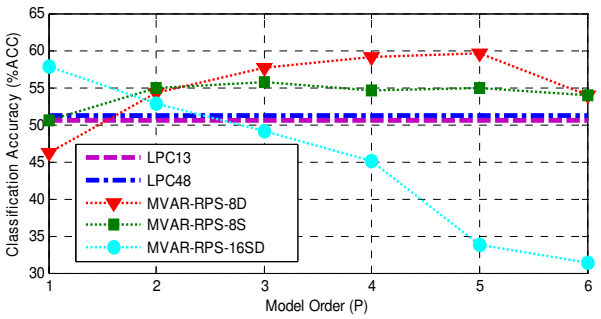


شکل (۲)- ساختار شبکه عصبی MLP با دولایه پنهان در آزمایش بازشناسی واج مجزا.

در تعلیم این شبکه از معیار توقف زود هنگام برای جلوگیری از برازش بیش از حد[38] آموزش شبکه برای بخش ثابتی از داده‌های تعلیم (به نام داده‌های توسعه) استفاده می‌شود. همچنین برای هر روش استخراج ویژگی، تعداد ۱۰ مرتبه تعلیم شبکه با مقادیر وزن‌دهی اولیه تصادفی و مختلف انجام شده است تا در نهایت عملکرد هر روش براساس میانگین‌گیری نتایج حاصل شده از ۱۰ آزمایش گزارش شود. برای تعلیم شبکه عصبی، در ورودی آن، بردار ویژگی مربوط به هر واج و در خروجی آن برچسب‌دهی سخت (One Hot) متناظر با آن قرار دارد.

در آزمایش‌های این بخش (آزمون بازشناسی واج مجزا) به علت استفاده از طبقه‌بند شبکه عصبی با قابلیت اندازه لایه ورودی متغیر، عملیات تبدیل خطی (LT) بر بردار ویژگی نهایی (مرحله ۵ از الگوریتم روش پیشنهادی) انجام نشده است. نتایج این بخش مشخص می‌کند که چه نوع اطلاعات گفتاری در حوزهٔ RPS (بردار چندبُعدی X در رابطهٔ (٦)) مانند اطلاعات به دست آمده از مشخصه‌های استاتیک (S) و یا دینامیک (D) تراژکتوری گفتاری و یا ترکیب توأم آن‌ها (SD) از سیگنال جاسازی شده در RPS می‌تواند در بازشناسی مؤثر باشد.

در اجرای آزمون واج مجزای روش پیشنهادی (VAR-RPS)، متناظر با درجه مدل P از روش VAR، برای کل نمونه‌های جاسازی شده یک واج گفتاری در RPS، تعداد P ماتریس ضرایب فیلتر تخمین زده می‌شود. پس از محاسبه عناصر ماتریس‌های ضرایب فیلتر، آن‌ها را در قالب یک بردار، به عنوان بردار ویژگی اعمالی متناظر با آن واج به شبکهٔ عصبی MLP استفاده می‌کنیم. به این ترتیب از مشخصه‌های استاتیک، بردار ویژگی VAR-RPS-S؛ مشخصه‌های دینامیک، بردار ویژگی VAR-RPS-D؛ و از مشخصه‌های توأم استاتیک و دینامیک، نیز بردار ویژگی VAR-RPS-SD تولید خواهند شد. در شکل (۳) کردارهای مربوط به نتایج درصد دقت طبقه‌بندی واج مجزا برای روش‌های مختلف معرفی شده برحسب درجه مدل روش VAR (P) نشان داده شده است.



شکل (۳)- نتایج درصد دقت بازشناسی (%ACC) واج مجزا برحسب درجه مدل (P) در روش VAR.

[38] Overfitting

همچنین در شکل (۳) نتایج بازشناسی مربوط به روش‌های زمانی متداول LPC (براساس ضرایب فیلتر) با درجه‌های مدل p=13 و p=48 به ترتیب به عنوان بردار ویژگی‌های LPC13 و LPC48 نشان داده شده است. انتخاب درجه مدل ۱۳ انتخابی متداول برای روش LPC در کاربرد بازشناسی گفتار است [٤٠]. همچنین درجه مدل ٤٨ بر مبنای محتوای دربرگرفته نمونه‌های گفتاری در حوزهٔ RPS با پارامترهای جاسازی $d=8$ و $t=6$ انجام شده است.

شکل (۳) نشان می‌دهد روش VAR-RPS-D که حاوی اطلاعات ضرایب فیلتر در مدل‌سازی مشخصه‌های دینامیک تراژکتوری گفتاری جاسازی شده در RPS است، در درجه مدل P=5 توانسته بالاترین نتیجه بازشناسی را کسب کند. همچنین برای تمامی مقادیر P>1، همواره نتایج ACC% روش‌های پیشنهادی حاوی مشخصه‌های استاتیک (VAR-RPS-S) و دینامیک (VAR-RPS-D) از نتایج بازشناسی روش‌های متداول LPC13 و LPC48 بیشتر است. برای روش VAR-RPS-SD نیز بیشترین درصد بازشناسی برای مقادیر کم درجه مدل (P=1,2) به دست آمده است. همانطور که مشاهده می‌شود کردار دقت بازشناسی برای این روش با افزایش مقدار P، نزولی است که توجیه آن می‌تواند در افزایش سریع تعداد پارامترهای مورد نیاز برای تخمین مناسب مدل آن (به علت بزرگ بودن اندازه بردار Q (Dq) در مقایسه با حالت فقط مشخصه استاتیک یا دینامیک) و نیاز به تعداد نمونه زیاد برای تخمین مناسب آن‌ها در مقدارهای بزرگ P است. برای بررسی دقیقتر نتایج این بخش، در جدول (۱) نتایج دقت بازشناسی (ACC) به دست آمده از بهترین درجه مدل برای روش‌های معرفی شده، آورده شده است.

جدول (۱)- نتایج آماری به دست آمده از بازشناسی واج مجزا برای روش‌های مختلف استخراج ویژگی.

بردار ویژگی	حوزهٔ پردازش سیگنال	بُعد بردار ویژگی	درجه مدل	ACC%
LPC48	**زمانی**	٤٨	٤٨	٥١/١٤
LPC13	**زمانی**	١٣	١٣	٥٠/٥٣
VAR-RPS-D (P=5)	**RPS – D دینامیک**	٨×٨×٥	**٥**	**٥٩/٦٧**
VAR-RPS-S (P=3)	**RPS - S استاتیک**	٨×٨×٣	٣	٥٥/٦٩
VAR-RPS-SD (P=1)	**RPS - SD استاتیک و دینامیک**	١٦×١٦×١	١	٥٧/٩٠

٦-٢- آزمایش‌های بازشناسی واج پیوسته

در این بخش به ارائه نتایج حاصل از سیستم بازشناسی گفتار واج پیوسته مبتنی بر HMM و روش استخراج ویژگی پیشنهادی می‌پردازیم. برای اجرای HMM از جعبه ابزار HTK استفاده شده است [٥٥]. این نرم‌افزار حاوی پیاده‌سازی کاملی از سیستم بازشناسی گفتار CSR است. در اکثر فعالیت‌های آزمایشگاهی و تحقیقاتی مربوط به سیستم‌های بازشناسی گفتار از این نرم‌افزار استفاده می‌شود. ابزارهای اصلی HTK برای آموزش مدل‌ها عبارتند از: HInit، HRest و HERest. در اینجا از مدل‌های سه واجی[39] و همچنین قالب گره‌زده آن با روش وابسته به درخت‌واره برای گره زدن حالت‌های مدل سه واجی استفاده شده است. از ابزار Hvite به عنوان بخش بازشناس HTK بهره‌برداری و از یک مدل زبانی بایگرم واجی نیز در بازشناسی دنبالهٔ واج‌ها استفاده شده است. نتایج بازشناسی

[39] Triphone

بیان شده در این بخش شامل درصد دقت (%Acc) بازشناسی واج است:

$$\%Acc = \frac{N - D - I - S}{N} \times 100, \qquad (۹)$$

که در آن N تعداد واج در برچسب‌های مرجع، D تعداد واج حذف شده، I تعداد واج درج شده و S تعداد واج جانشین شده هستند [٤٤].

در آزمایش‌های این بخش، کارایی روش پیشنهادی با دیگر روش‌های شناخته شده استخراج ویژگی زمانی مانند روش LPC و LPREF مقایسه شده است. این ویژگی‌ها که از اِعمال روش LP به سیگنال تک بُعدی گفتار حاصل می‌شوند، به ترتیب روش ضرایب پیشگویی خطی (LPC) و ضرایب انعکاسی (LPREF) هستند. در اینجا مشابه آزمون بازشناسی واج مجزا، درجهٔ مدل در رابطهٔ (۳) مقدار $p=13$ انتخاب شده که در کاربردهای پردازش گفتار میکروفونی مناسب است. علاوه بر این با انتخاب این درجه مدل و افزودن ضرایب مشتقات مرتبه اول و دوم دینامیک این ویژگی‌ها به بردار ویژگی نهایی، تعداد کل ویژگی‌های ورودی سیستم بازشناسی گفتار برابر ۳۹ می‌شود، که یک اندازه بردار ویژگی متداول در سیستم‌های بازشناسی گفتار است. رابطهٔ (۱۰) نحوه محاسبه ضرایب دینامیک مرتبه اول (دلتا) را نشان می‌دهد:

$$\Delta F[k] = \sum_{i=-W}^{W} iF[k+i] \Big/ \sum_{i=-W}^{W} i^2, \qquad (۱۰)$$

در رابطهٔ فوق $F[k]$، بردار ویژگی متناظر با قاب گفتاری kام است. پارامتر W نیز معمولاً برابر ۲ انتخاب می‌شود[٤٤]. محاسبهٔ مشتق زمانی دوم بردار ویژگی‌ها (بردار ویژگی دلتادلتا)، از اعمال مجدد رابطهٔ (۱۰) بر ضرایب دلتای بردار ویژگی‌ها به دست می‌آید.

٦-۲-۱- انتخاب نوع ضرایب

در این بخش به بررسی و مقایسه عملکرد نوع ضرایب ویژگی‌های مبتنی بر روش LP می‌پردازیم. همانطور که در بخش (۳-۱) اشاره شد، در حین محاسبات روش LP، دو نوع ضرایب فیلتر یا انعکاسی تولید می‌شود. در جدول (۲) نتایج دقت بازشناسی واج ویژگی‌های به دست آمده از روش‌های LPC (ضرایب فیلتر) و LPREF (ضرایب انعکاسی) آورده شده است. در این جدول نماد DA نشان‌دهندهٔ الحاق ویژگی‌های دینامیک مشتقات اول و دوم (دلتا و دلتادلتا) به بردار ویژگی اصلی است.

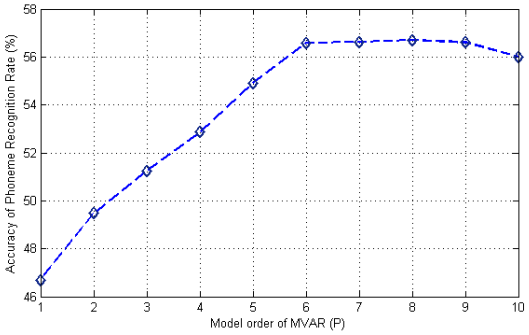
جدول (۲)- نتایج دقت بازشناسی واج (%Acc) برای روش‌های متداول استخراج ویژگی LPC و LPREF.

روش استخراج ویژگی	درجه مدل پیشگویی خطی (p)	بُعد بردار ویژگی	%Acc
LPC_DA	۱۳	۳۹	۵۱/٤۱
LPREF_DA	۱۳	۳۹	٦۲/۸۰

با توجه به نتایج نشان داده شده در جدول (۲)، دقت بازشناسی (%Acc) واج روش LPC حدود ۱۱ درصد کمتر از نتایج دقت بازشناسی واج به دست آمده از بردار ویژگی LPREF است. این تفاوت می‌تواند ناشی از حساسیت اندازه ضرایب LPC به اندازه زمانی سیگنال و از طرف دیگر بهنجار بودن اندازه ضرایب انعکاسی باشد. بنابراین با توجه به کارایی برتر ضرایب انعکاسی در مقایسه با ضرایب فیلتر پیش‌بینی خطی، در اجرای روش پیشنهادی در آزمون بازشناسی واج پیوسته تنها از ماتریس ضرایب انعکاسی حاصل از روش VAR (VLPREF) استفاده کرده‌ایم.

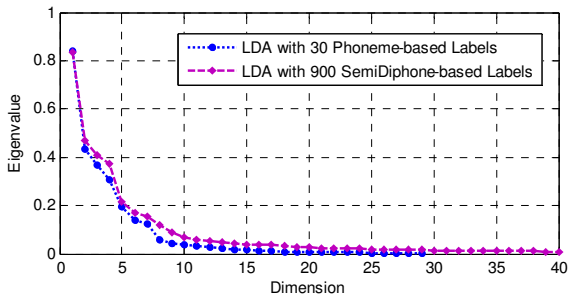
٦-۲-۲- انتخاب درجه مدل

یکی از عوامل مهم برای اجرای روش پیشنهادی تعیین بهینه درجه مدل (P) در رابطهٔ (٦) است که برای انتخاب بهینه آن مجموعه‌ای آزمایش مبتنی بر نتیجه دقت بازشناسی (Acc) طراحی شده است. در این آزمایش‌ها از مشخصه‌های استاتیک (S) نقاط جاسازی شده در RPS در مدل‌سازی تراژکتوری گفتاری با روش VAR، مقدار Dy=39 به عنوان اندازه بُعد کاهش یافته ویژگی‌ها در الگوریتم روش پیشنهادی و نگاشت خطی LDA به عنوان روش کاهش بُعد LT استفاده شده است. در شکل (٤) کردار نتایج دقت بازشناسی واج با استفاده از ویژگی‌های به دست آمده از روش پیشنهادی در مقدارهای درجه مدل (P) مختلف نشان داده شده است.



شکل(٤)- نتایج دقت بازشناسی واج پیوسته برحسب درجه‌های مختلف مدل (P) و بُعد کاهش‌یافته Dy=39.

براساس شکل (٤) انتخاب درجه مدل P=6 در آزمایش بازشناسی گفتار پیوسته (CSR) مناسب است. در شکل (٥) کردار اندازه مقادیر ویژه مرتب شده برحسب ابعاد بردار ویژگی (در تخمین وزن روش کاهش بُعد خطی LDA از بردار ویژگی اولیه Q با مقدار بُعد Dq=384 با P=6) با استفاده از دو شیوهٔ برچسب‌گذاری ۳۰ طبقهٔ واجی و ۹۰۰ طبقهٔ شبه دایفونی (معرفی شده در منبع[٥١]) نشان داده شده است.



شکل(٥)- کردار اندازه مقادیر ویژه مرتب شده برحسب بُعد بردار ویژگی در تخمین روش کاهش بُعد خطی LDA با درجه مدل P=6 و استفاده از دو شیوهٔ برچسب‌گذاری ۳۰ طبقهٔ واجی و ۹۰۰ طبقهٔ شبه‌دایفونی.

براساس کردارهای شکل (٥)، بردار ویژگی پیشنهادی، می‌تواند اطلاعات مفید خود را در تعداد محدودی از بُعدهای ویژگی (اندازه بُعد کمتر از ۱٥) نمایش دهد.

٦-۲-۳- انتخاب نوع اطلاعات به دست آمده از RPS

پس از انتخاب مقدار مناسب درجه مدل (P)، به مقایسه عملکرد روش استخراج ویژگی پیشنهادی (VLPREF) با دیگر روش‌های زمانی متداول استخراج ویژگی گفتاری می‌پردازیم. برای مقایسهٔ مستقیم روش‌ها، در این بخش و در اجرای روش پیشنهادی، از مقدار بُعد کاهش یافته ۱۳ (Dy=13) استفاده شده است که با افزودن مشتقات اول و دوم آن، طول بردار ویژگی نهایی آن نیز برابر ۳۹ می‌شود. مشابه آزمون بازشناسی واج مجزا، در اینجا نیز مدل‌سازی خطی در حوزهٔ RPS به صورت مجزا بر مشخصه‌های استاتیک (S) و یا دینامیک (D) نقاط جاسازی شده در RPS انجام شده است که به ترتیب بردارهای ویژگی

VLPREF_8S_DA و VLPREF_8D_DA نامگـــــــذاری شده‌اند. نتایج در جدول (۳) آورده شده است.

جدول (۳)- مقایسه نتایج دقت بازشناسی واج (Acc%) روش‌های استخراج ویژگی مختلف بر روی داده‌های فارس‌دات کوچک.

روش استخراج ویژگی	حوزه پردازش سیگنال	درجه مدل	بُعد بردار ویژگی اولیه	بُعد بردار ویژگی نهایی	درصد دقت بازشناسی واج
LPREF_DA	زمانی	۱۳	۱۳	۳۹	٦۲/۸۰
LPREF_LDA_DA	زمانی	٤۸	٤۸ ٤۸LDA۱۳	۳۹	٦۷/۲٤
VLPREF_S_DA	RPS – S استاتیک	٦	٦۰۸۰۸ ۳۸٤LDA۱۳	۳۹	٦۹/٤۹
VLPREF_D_DA	RPS - D دینامیک	٦	٦۰۸۰۸ ۳۸٤LDA۱۳	۳۹	۷۱/۲۱

با توجه به نتایج جدول فوق، دقـت بازشناسـی واج بـه دست آمـده از روش پیشـنهادی مبتنـی بـر مشخصـه‌هـای اســـــــــتاتیک (VLPREF_S_DA) و دینامیـــــــــک (VLPREF_D_DA) به ترتیـب ٦۹/٤۹٪ و ۷۱/۲۱٪ اسـت که حدود ٦/۷٪ تا ۸/۳٪ بیشتر از عملکـرد روش مبتنـی بـر ضرایب انعکاسی (LPREF_DA) اسـت. عـلاوه بـر نتـایج روش‌های مطرح شده، در جدول (۳) نتایج روش ضـرایب انعکاسی(LPREF) با درجه مدل ٤۸- که ضرایب اولیـه آن با روش تبدیل خطی LDA به اندازه بُعد ۱۳ کـاهش یافتـه شده و در نهایت با افزودن بردار ویژگی‌های دینامیک اندازه آن به ۳۹ رسیده اسـت- بـا عنـوان (LPREF_LDA_DA) آورده شده است. با این تعریف، کارایی ایـن روش حـدود ٤/٤ درصد بهبود یافته است که ناشی از مدل‌سازی دقیق‌تـر و غیرهمبسته کردن ویژگی‌هـا بـا بکـارگیری تبـدیل LDA است. با ایـن حـال عملکـرد ایـن روش همچنـان از روش پیشنهادی کمتر است. از طرف دیگر، بـا توجـه بـه کـارایی برتر روش پیشنهادی بر مشخصه دینامیـک (D) در مقایسـه با مشخصه استاتیک (S) تراژکتوری گفتاری جاسازی شـده در RPS (حدود ۱/۷٪) نتیجه به دست آمده از ایـن بخـش نیز مؤید نتایج به دست آمده از آزمون بازشناسی واج مجزا است.

٦-۲-٤- انتخاب نوع روش نگاشت و کاهش بُعد

در یک بررسی دیگـر عملکـرد بخـش کـاهش بُعـد روش پیشنهادی بوسیله اعمال روش‌هـای متـداول تبـدیل خطـی ارزیابی شد. روش‌های استفاده شـده شـامل نگاشـت‌هـای LDA، HLDA، LPP و OLPP اسـت [٥٦،٥۲،٥۱]. نتـایج این شبیه‌سازی‌ها در جدول (٤) آورده شده است.

جدول (٤)- مقایسه نتایج دقت بازشناسی واج (Acc%) از اعمال روش‌های مختلف تبدیل خطی بر روی بردار ویژگی اولیه ۳۸٤ بُعدی به دست آمده از بکارگیری روش پیشنهادی با P=6.

روش استخراج ویژگی	درجه مدل	روش تبدیل خطی	بُعد بردار ویژگی نهایی	درصد دقت بازشناسی واج
VLPREF_D_DA	٦	LDA	۳۹	۷۱/۲۱
VLPREF_D_HLDA_DA	٦	HLDA	۳۹	۷۳/٦۸
VLPREF_D_LPP_DA	٦	LPP	۳۹	٦۸/٤۲
VLPREF_D_OLPP_DA	٦	OLPP	۳۹	۷۰/۷۱

با توجه به نتایج به دست آمده از این جدول می‌توان نتیجه گرفـت کـارایی تبـدیل خطـی HLDA بـا درصـد دقـت بازشناسی واج ۷۳/٦۸٪ ، از دیگر روش‌های تبـدیل خطـی بهتر بوده است.

۷- بحث و بررسی

نتایج بازشناسی به دست آمده از آزمایش‌های اجرا شـده در بخش ٦ نشان می‌دهد که درجه مدل روش VAR در آزمون بازشناسی گفتار پیوسته مقدار ٦ تعیین شد و بـرای آزمـون بازشناســـی واج مجـــزا مقـــدارهای ۱ و ۳ و ٥ در حالـــت

مختلف بکارگیری اطلاعات کسب شده از مشخصه‌های استاتیک (S)، دینامیک (D) و توأم استاتیک و دینامیک (SD) از تراژکتوری گفتاری جاسازی شده در RPS به دست آمد. کم بودن مقدار درجه مدل‌های به دست آمده از این روش (VAR) در مقایسه با درجه مدل‌های پرکاربرد در روش‌های متداول حوزهٔ زمانی و مبتنی بر LPC گفتاری- که غالباً بالای مقدار ۱۰ هستند- نشان می‌دهد نقاط تراژکتوری جاسازی شده در RPS حاوی اطلاعات مفیدی از محدوده وسیعی از سیگنال گفتاری است که منجر به کاهش درجه مدل پیش‌بینی خطی (LP) شده است. از طرف دیگر در اجرای روش بازشناسی واج مجزا بر ویژگی‌های به دست آمده از اعمال روش VAR بر ماتریس SD، از آنجایی که این روش از اطلاعات توأم استاتیک و دینامیک تراژکتوری گفتاری استفاده می‌کند تعیین درجه مدل P=1 برای بهترین مقدار دقت بازشناسی واج برای آن قابل توجیه است. این نتیجه که اولین بار در تأیید فرض مدل درجه یک برای تابع تولید گفتار H در رابطهٔ (۷) گزارش شده است؛ در حالی حاصل شده که هنوز تابع چندمتغیره مورد بررسی، خطی در نظر گرفته شده است.

نتایج دقت بازشناسی واج پیوسته حاصل از اجرای روش‌های متنوع استخراج ویژگی مبتنی بر الگوریتم VAR، نشان داد که ویژگی‌های به دست آمده از ماتریس ضرایب انعکاسی از مدل‌سازی خطی مشخصه‌های استاتیک تراژکتوری سیگنال گفتار (روش VLPREF_S_DA)، توانسته است افزایش دقت بازشناسی واج را بطور مطلق، در حدود ٦/٦٩ درصد نسبت به روش استخراج ویژگی پایه (حوزهٔ زمانی) مبتنی بر ضرایب انعکاسی(LPREF) افزایش دهد. این نتیجه نشان‌دهندهٔ مؤثر بودن ویژگی‌های حاصل از الگوریتم روش پیشنهادی است. بنابراین اجرای روش پیشنهادی نشان می‌دهد که استفاده همزمان از تئوری جاسازی سیگنال (انتقال به RPS) و پردازش چند بُعدی آن (روش مدل‌سازی VAR)، می‌تواند منجر به استخراج اطلاعات مفید و متمایزکننده‌تری در مقایسه با حالت استخراج ویژگی از سیگنال زمانی تک بُعدی گفتار شود. از خصوصیات دیگر روش پیشنهادی، قابلیت استفادهٔ عملی آن در سیستم‌های بازشناسی گفتار پیوسته است؛ چرا که اغلب روش‌هایی که تاکنون برای استخراج ویژگی از سیگنال گفتار جاسازی شده در RPS ارائه شده‌اند، این قابلیت را به علت حجم و هزینه زیاد محاسبات ندارند. لذا این روش این نوید را می‌دهد که با تکمیل رویکرد دنبال شده، امکان رسیدن به کارایی برتر وجود دارد و بعلاوه این روش می‌تواند جایگزین روش‌های زمانی استخراج ویژگی از سیگنال گفتار تک بُعدی شود

از مقایسهٔ نتیجه دقت بازشناسی واج به دست آمده از روش پیشنهادی (روش VLPREF_S_DA) با مقدار بُعد کاهش یافته *Dy=13*- که با اضافه کردن مشتقات زمانی اول و دوم آن، بُعد برداروِیژگی نهایی‌اش به ۳۹ رسیده است- با روش اجرا شده با مقدار بُعد کاهش یافته *Dy=39* (بدون استفاده از مشتقات زمانی اول و دوم بردار ویژگی) این نتیجه حاصل می‌شود که استفاده از ٢٦ ویژگی دینامیک از نوع مشتقات زمانی اول و دوم، در مقابل استفاده از ٢٦ ویژگی استاتیک اضافی، منجر به افزایش نتایج دقت بازشناسی واج از مقدار حدود ٥٦/٧ به ٦٩/٤٩ شده است. این مقدار اختلاف زیاد در نتایج دقت بازشناسی واج، نشان‌دهندهٔ مهم بودن استفاده از اطلاعات دینامیک بردارهای ویژگی گفتاری، نسبت به اطلاعات استاتیک بردارهای ویژگی گفتاری، در فرایند استخراج ویژگی گفتاری و استفاده از آن در فرایند بازشناسی گفتار پیوسته

است.

از طرف دیگر با اجرای روش پیشنهادی بر مشخصه‌های دینامیک تراژکتوری گفتاری، دقت بازشناسی گفتار پیوسته واجی افزایشی حدود ۱/۷۲٪ در مقایسه با اجرای این روش بر مشخصه‌های استاتیک تراژکتوری گفتاری، نشان داده است. این نتیجه بیانگر مفیدتر بودن اطلاعات دینامیک در اجرای این روش است. همچنین استفاده از نگاشت HLDA بجای روش کاهش بُعد خطی LDA توانست بالاترین مقدار کارایی را در حالت مدل‌سازی مشخصه‌های دینامیک تراژکتوری گفتاری در RPS با مقدار دقت بازشناسی ۷۳/۶۸ تولید کند به طوری که این مقدار حدود ۲ درصد از پیاده‌سازی این روش با نگاشت LDA بیشتر است.

با توجه به اینکه روش پیشنهاد شده در این مقاله از فضای اطلاعاتی متمایز در مقایسه با روش‌های متداول در استخراج ویژگی استفاده می‌کند، بکارگیری آن می‌تواند در سیستم‌های ترکیب خروجی سیستم‌های بازشناسی گفتار چندمرحله‌ای مؤثر باشد. همانطور که می‌دانیم یکی از روش‌های بهبود کارایی سیستم بازشناس گفتاری، استفاده از روش ترکیب خروجی‌های به دست آمده از سیستم‌های ASR متنوع است. در این حالت سیستم‌های متنوع مورد استفاده بایستی مقدار اندازه خطای مناسب و خروجی متفاوت داشته باشند تا بتوانند در نتیجه کلی سیستم نهایی مؤثر باشند. همچنین برای مقاوم‌سازی عملکرد سیستم‌های ASR، می‌توان به طور توأم از عملیات پیش‌پردازشی در بهسازی تراژکتوری گفتاری در حوزهٔ RPS با روش‌هایی مانند نگاشت محلی استفاده کرد [۸] و سپس از روش پیشنهادی این مقاله در حوزهٔ RPS برای تولید ویژگی‌های مقاوم به نویز بهره گرفت.

۸- نتیجه‌گیری

در این مقاله روشی مبتنی بر استفاده از اطلاعات فضای بازسازی شده فاز سیگنال گفتار برای بهبود کارایی سیستم‌های بازشناسی گفتار واج مجزا و پیوسته مطرح و بررسی شد. در روش پیشنهادی این مقاله از روش پردازش چندبُعدی VAR، به عنوان توسعه‌ای بر روش پیش‌بینی خطی LPC گفتاری و تقریب خطی فرض مدل‌سازی غیرخطی فرایند تولید گفتار در RPS استفاده شد. در این روش فرض شد که نمونه‌های جاسازی شده سیگنال در فضای RPS (نقاط تراژکتوری گفتاری) به صورت ترکیب خطی از چند نمونه قبلی آن به دست خواهند آمد. در الگوریتم روش پیشنهادی، تراژکتوری‌های گفتاری به دست آمده از سیگنال گفتار جاسازی شده در RPS بوسیله مشخصه‌های چند بُعدی آنها مانند مشخصه‌های استاتیک، دینامیک و یا توأم استاتیک و دینامیک با روش VAR، مدل‌سازی و سپس از روی ماتریس‌های ضرایب فیلتر یا انعکاسی به دست آمده از آن‌ها یک بردار ویژگی اولیه گفتاری تهیه شد. این رویکرد منجر به تولید بردار ویژگی با اندازه بسیار زیاد شد که این بردار ویژگی اولیه توانایی تولید مانیفولد گفتاری در فضای ویژگی با بُعد بسیار زیاد را دارد. در ادامه بوسیله روش‌های کاهش بُعد، عملکرد آن در سیستم‌های بازشناس گفتار واج مجزا و پیوسته بررسی شد. سپس با اجرای آزمایش‌های متنوعی نشان داده شد که استفاده از روش پیشنهادی می‌تواند منجر به افزایش کارایی سیستم‌های بازشناسی گفتار شود.

۹- مراجع

[1] Awrejcewicz J., Bifurcation portrait of the human vocal cord oscillation; Journal of Sound Vibrations, 1990; 136: 151–156.

[2] Berry, D.A., Herzel, H., Titze, I.R., Krischer K., Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions; The Journal of the Acoustical Society of America, 1994; 95: 3595–3604.
[3] Herzel, H., Berry, D., Titze, I., Steinecke, I., Nonlinear dynamics of the voice: signal analysis and biomechanical modeling; Chaos, 1995; 5: 30–34.
[4] Jiang, J.J., Zhang, Y., Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds; The Journal of the Acoustical Society of America,2002; 112: 2127–2133.
[5] Jiang, J.J., Zhang, Y., McGilligan, C., Chaos in voice, from modeling to measurement; Journal of Voice, 2006; 20(1): 2006; 2-17.
[6] Kokkinos, I., Maragos, P., Nonlinear speech analysis using models for chaotic systems; IEEE Trans. Speech Audio Processing, 2005; 13: 1098–1109.
[7] Hagmuller, M., Kubin, G., Poincare pitch marks. Speech Communication; 2006; 48: 1650–1665.
[8] Sun, J., Zheng, N., Wang, X., Enhancement of Chinese speech based on nonlinear dynamics; Signal Processing, 2007; 87: 2431–2445.
[9] Kantz, H., Schreiber, T., Nonlinear Time Series Analysis Cambridge University Press, Cambridge, England. 1997.
[10] Takens, F., Detecting strange attractors in turbulence; In Proc. Dynamical System Turbulence, 1980; pp. 366–381.
[11] Narayanan, S.S., Alwan, A.A., A nonlinear dynamical systems analysis of fricative consonants; Acoustical Society of America Journal, 1995; 97: 2511-2524.
[12] Shekofteh, Y., Almasganj, F., Using phase space based processing to extract proper features for ASR systems; In Proc. 5th International Symposium on Telecommunications (IST), 2010; pp. 596-599.
[13] Vaziri, G., Almasganj, F., Behroozmand, R., Pathological assessment of patients' speech signals using nonlinear dynamical analysis; Computers in Biology and Medicine, 2010; 40(1): 54-63.
[14] Paliwal, K., Alsteris, L., On the usefulness of STFT phase spectrum in human listening tests; Speech Communication, 2005; 45: 153–170.
[15] Hegde, R. M., Murthy, H.A., Gadde, V.R.R., Significance of the modified group delay feature in speech recognition; IEEE Trans. Audio, Speech and Language Processing, 2007; 15(1): 190–202.
[16] Alsteris, L.D., Paliwal, K.K., Short-time phase spectrum in speech processing: A review and some experimental results; Digital Signal Processing, 2007; 17: 578–616.
[17] Pitsikalis, V., Maragos, P., Speech analysis and feature extraction using chaotic models. In Proc. ICASSP, Orlando, Florida, 2002; pp. 533-536.
[18] Pitsikalis, V., Maragos, P., Filtered dynamics and fractal dimensions for noisy speech recognition; Signal Processing Letters, 2006; 13(11): 711-714.
[19] Pitsikalis, V., Maragos, P., Analysis and classification of speech signals by generalized fractal dimension features; Speech Communication, 2009; 51(12): 1206-1223.
[20] Ezeiza, A., Ipina, K.L., Hernández, C., Barroso, N., Enhancing the feature extraction process for automatic speech recognition with fractal dimensions; Cognitive Computation, 2012; pp. 1-6.
[21] Yu, S., Zheng, D., Feng, X., A new time domain feature parameter for phoneme classification. In Proc. WESPAC IX 2006, Seoul, Korea. 2006.
[22] Narayanan, N.K., Thasleema, T.M., Prajith, P., Reconstructed state space model for recognition of consonant - vowel utterances using support vector machines; International Journal of Artificial Intelligence and Applications, 2012; 3(2): 101-119.
[23] Thasleema, T.M., Prajith, P., Narayanan, N.K., Time–domain non-linear feature parameter for consonant classification; International Journal of Speech Technology, 2012; 15(2): 227-239.
[24] Ye, J., Povinelli, R.J., Johnson, M.T., Phoneme classification using naive Bayes classifier in reconstructed phase space; In Proc. IEEE Digital Signal Processing Workshop, Atlanta, Georgia. 2002.
[25] Ye, J., Johnson, M.T. M.T., Povinelli, R.J., Phoneme classification over reconstructed phase space using principal component analysis; In Proc. NOLISP, Le Croisic, France, 2003; pp. 11–16.
[26] Povinelli, R.J., Johnson, M.T., Lindgren, A.C., Ye, J., Time series classification using Gaussian mixture models of reconstructed phase spaces; IEEE Trans. Knowledge and Data Engineering, 2004; 16:779–783.

with backward-adaptive algorithms for postfiltering and noise feedback; IEEE Journal on Selected Areas in Communications, 1988; 6(2): 364-382.

[43] Lee, K.F., Hon, H.W., Reddy, R., An overview of the SPHINX speech recognition system; IEEE Trans. Acoustics, Speech and Signal Processing, 1990; 38(1): 35-45.

[44] Young, S. J., Evermann, G., Gales, M.J.F., Kershaw, D., Moore, G., Odell, J.J., Woodland, P.C., The HTK book (version 3.4). 2006.

[45] Kamiński, M., Determination of transmission patterns in multichannel data; Philosophical Transactions of the Royal Society B: Biological Sciences, 2005; 360(1457): 947-952.

[46] Stock, J.H., Watson, M.W., Vector autoregressions. The Journal of Economic Perspectives, 2001; 15(4): 101-115.

[47] Schlogl, A., A comparison of multivariate autoregressive estimators; Signal Processing, 2006; 86(9): 2426-2429.

[48] Hytti, H., Takalo, R., Ihalainen, H., Tutorial on multivariate autoregressive modeling; Journal of clinical monitoring and computing, 2006; 20(2): 101-108.

[49] Marple, S.L., Digital spectral analysis with applications; Englewood Cliffs, NJ, Prentice-Hall. 1987.

[50] Lindgren, A.C., Johnson, M.T., Povinelli, R.J., Joint frequency domain and reconstructed phase space features for speech recognition; In Proc. ICASSP, Montreal, Canada, 2004; pp. I-533–I-536.

[51] Shekofteh, Y., Almasganj, F., Goodarzi, M.M., Comparison of linear based feature transformations to improve speech recognition performance; In Proc. 19th Iranian Conference on Electrical Engineering (ICEE), pp. 2011; 1-4.

[52] Cai, D., He, X., Han, J., Zhang, H.J., Orthogonal laplacianfaces for face recognition; IEEE Trans. Image Processing, 2006; 15(11): 3608-3614.

[53] FARSDAT, Persian speech database: <http://catalog.elra.info/product_info.php?products_id=18>.

[54] Bijankhan, M., Sheykhzadegan, J., Roohani, M.R., Zarrintare, R., Ghasemi, S.Z., Ghasedi, M.E., TFarsDat - The telephone farsi speech database; In Proc. EuroSpeech, Geneva, Switzerland, 2003; pp. 1525-1528.

[55] HTK, Hidden Markov Model Toolkit: <http://htk.eng.cam.ac.uk/>

[56] Shekofteh, Y., Almasganj, F., Using linear models of speech trajectory in the reconstructed phase space to extract useful features for speech recognition system; In Proc. Iranian Conf. Biomedical Engineering (ICBME), Tehran, Iran, 2012; pp.233–236

[27] 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, 2006; 54: 2178–2186.

[28] Jafari, A., Almasganj, F., NabiBidhendi, M., Statistical modeling of speech Poincaré sections in combination of frequency analysis to improve speech recognition performance; Chaos, 2010; 20(033106):1-11.

[29] Jafari, A., Almasganj, F., Using nonlinear modeling of reconstructed phase space and frequency domain analysis to improve automatic speech recognition performance; International Journal of Bifurcation and Chaos, 2012; 22(3).

[30] Shekofteh, Y., Almasganj, F., Feature extraction based on speech attractors in the reconstructed phase space for automatic speech recognition systems; ETRI Journal, 2013; 35(1): 100-108.

[31] Sauer, T., Yorke, J.A., Casdagli, M., Embedology; Journal of Statistical Physics, 1991; 65: 579–616.

[32] Kennel, M.B., Brown, R., Abarbanel, H.D.I., Determining embedding dimension for phase-space reconstruction using a geometrical construction; Physical review A, 1992; 45(6): 3403–3411.

[33] Abarbanel, H.D.I., Analysis of observed chaotic data; Springer, New York. 1996.

[34] Johnson, M.T., Povinelli, R.J., Lindgren, A.C., Ye, J., Liu, X., Indrebo, K.M., Time-domain isolated phoneme classification using reconstructed phase spaces; IEEE Trans. Speech Audio Processing, 2005; 13(4): 458–466.

[35] Banbrook, M., McLaughlin, S., Dynamical modelling of vowel sounds as a synthesis tool; In Proc. ICSLP, 1996; pp. 1981-1984.

[36] Indrebo, K.M., Povinelli, R.J., Johnson, M.T., Sub-banded reconstructed phase spaces for speech recognition; Speech Communication, 2006; 48: 760-774.

[37] Rabiner, L.R., Schafer, R.W., Digital processing of speech signals (vol. 19). New York: Prentice-hall. 1979.

[38] Markel, J.E., Gray, A.H., Linear prediction of speech. Springer-Verlag New York. 1982.

[39] Ramachandran, R.P., Zilovic, M.S., Mammone, R.J., A comparative study of robust linear predictive analysis methods with applications to speaker identification; IEEE Trans. Speech and Audio Processing, 1995; 3(2): 117-125.

[40] Huang, X., Acero, A., Hon, H.W., Reddy, R., Spoken Language Processing: A Guide to Theory, Algorithm & System Development. 2001.

[41] Atal, B.S., Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification; The Journal of the Acoustical Society of America, 1974; 55, 1304.

[42] Ramamoorthy, V., Jayant, N.S., Cox, R.V., Sondhi, M.M., Enhancement of ADPCM speech coding