

بازشناسی واژگان صوتی به کمک بیشترین میزان همبستگی خطی و نظارت عصبی

مهدی اعتمادی^۱، فرزاد حسین زاده^۲، علیرضا صداقت^۳

۱- دانشگاه آزاد اسلامی واحد بندر لنگه، گروه برق

۲- دانشگاه آزاد اسلامی واحد بندر لنگه، گروه کامپیوتر

۳- دانشگاه آزاد اسلامی لاهیجان، گروه مکانیک

چکیده

با توجه به گرایش و توسعه روز افزون استفاده از روش‌های ارتباط صوتی با ماشین، در این مقاله روشی برای بازشناسی کلمات یا گفتار منقطع با دقت بالا ارائه شده است. در این سیستم واژه‌های اداء شده دریافت و بخش حاوی اطلاعات جدا و با فرض پایایی میانگین نویز محیطی، مولفه‌های فرکانسی نویز از سیگنال نویز آلود، کاسته شده که نتیجه آن کاهش سطح نویز محیطی بوده است، سپس سیگنال به فریم‌های کوچک تقسیم و هر فریم به منظور ایجاد پیوستگی در پنجره همینگ ضرب می‌گردد و با اعمال فیلتر Mel که رفتاری شبیه گوش درونی انسان را شبیه‌سازی می‌نماید و اعمال تبدیل گسسته کسینوسی، ضرایب کپسترال استخراج می‌گردد. به ازای هر کلمه، ضرایب کپسترال به شبکه عصبی پس انتشار خطای سه لایه‌ای آموزش داده می‌شود و سپس واژه مورد آزمایش بعد از استخراج ضرایب، به تابع تعیین میزان وابستگی خطی رفته تا بیشترین شباهت با یکی از واژگان بانک آموزش داده شده تعیین گردد جهت تایید یا رد صحت شناخت، خروجی عصبی، به ازای جایگاه کلمه بازشناسی شده توسط تابع همبستگی مورد بررسی قرار می‌گیرد که پس از آزمایشات فراوان، با عمل آستانه‌گیری نتیجه شناسایی بدون خطا در گستره واژه‌های غیر مشابه، دقت بالا در تشخیص کلمات آموزش داده شده مشابه و اعلام عدم آشنایی با کلمات ناشناس غیر مشابه با بانک آموزشی، و دقت نسبی در تشخیص عدم آشنایی با واژگان ناشناس مشابه با بانک آموزشی حاصل شده است.

کلمات کلیدی: شناسایی گفتار، همبستگی خطی، شبکه عصبی، ضرایب کپسترال، پنجره همینگ، فیلتر Mel.

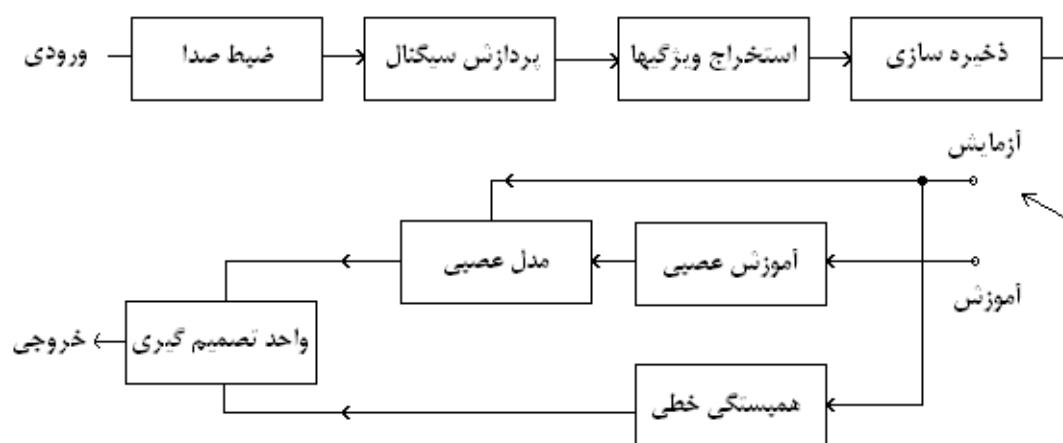
امقدمه

امروزه استفاده از سیستم‌های بازشناسی گفتار به طور قابل ملاحظه‌ای در حال رشد و توسعه است و تحقیقات فراوانی بر روی کل پروسه بازشناسی گفتار در سراسر دنیا که شامل زمینه‌های تشخیص گفتار پیوسته [۱]، تشخیص واژگان یا دستورات صوت [۲]، و زیر شاخه‌های تخصصی‌تر مانند استخراج ویژگی‌های صوتی از آوای واژه که خود به روش‌های مختلفی قابل بررسی است [۳]، انواع روش‌های حذف نویز در اطلاعات صوتی از قبیل استفاده از فیلترهای تطبیقی [۴]، تقسیم‌بندی زمانی و تخمین نویز [۵]، و تبدیل ویولت [۵]، انواع روش‌های شناسایی الگو مانند مدل مخفی مارکوف [۳]، شبکه‌های عصبی [۶]، در حال انجام است. در این مقاله با هدف

بومی سازی مباحث مطروحه و حصول دقت بیشتر در تشخیص واژگان به خصوص در کلمات با آوای مشابه، طرحی کاربردی پیشنهاد شده است. در این سیستم بعد از دریافت واژگان، عملیات حذف نویز و استخراج ویژگی ها با ساختار پیشنهاد شده انجام می گیرد و سپس جهت یادگیری و تشخیص، روشی ترکیبی مشتمل بر استفاده از حداکثر میزان وابستگی خطی و نظارت شبکه عصبی ارائه شده که منجر به دقت بالا در تشخیص واژگان بطور مستقل از گوینده می باشد. در این تحقیق ابزارهای مورد استفاده جهت نمونه گیری و پردازش یک عدد میکروفون معمولی خازنی، کارت صدای معمولی مدل AC97 با فرکانس نمونه برداری $F_s = 44100 \text{ Hz}$ ، کامپیوتر شخصی با پردازنده پنتیوم 4 و محیط نرم افزار MatLab نسخه 7.0.1 که در دپارتمان سازمان مرکزی دانشگاه آزاد اسلامی رجیستر شده است، بوده که نتایج حاصل شده مستقل از ابزار آلات مورد استفاده می باشد.

۲ ساختار کلی

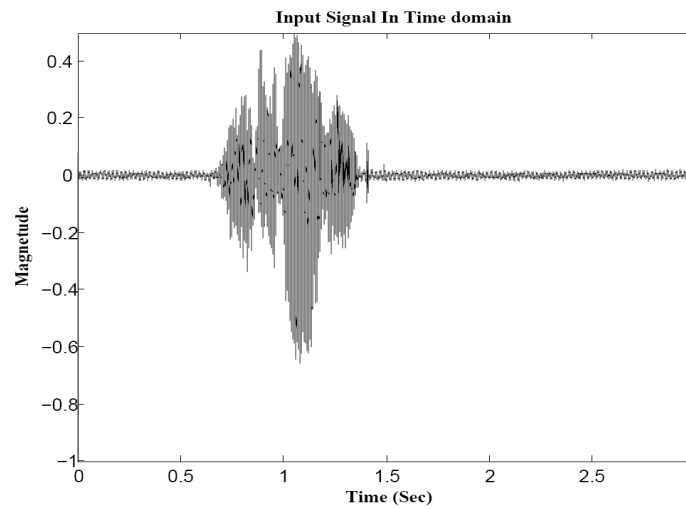
مدل و ساختار کلی سیستم از بخش های مختلف که در دیاگرام شکل (۱) نشان داده شده، تشکیل یافته است.



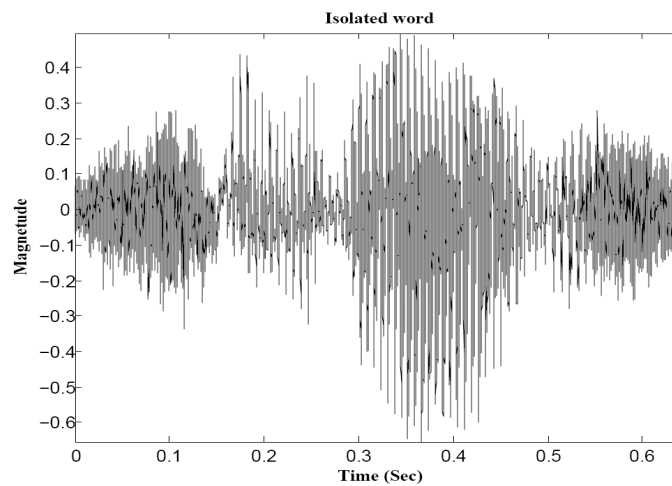
شکل ۱.

۲-۱ پیش پردازش

در ابتدا، قسمت ضبط سیگنال، اطلاعات صوتی را با فرکانس نمونه برداری $F_s = 44100 \text{ Hz}$ ذخیره می نماید و بعد از نرمال سازی به کمک عمل آستانه گیری خودکار بر اساس میزان انرژی موجود در سیگنال، قسمت حاوی اطلاعات جداسازی می گردد. شکل ۲ نمونه ای از عملکرد این بخش می باشد.



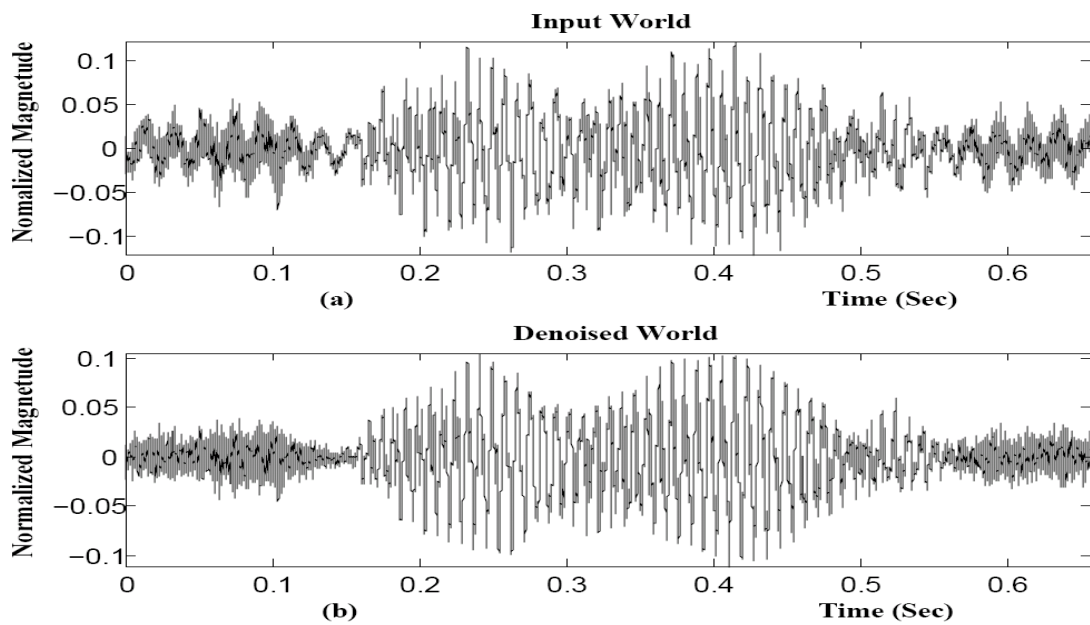
شکل ۲-الف. بخش حاوی اطلاعات



شکل ۲-ب. کل اطلاعات صوتی دریافت شده برای کلمه "شیراز"

۲-۲ حذف نویز

در سیستم‌های تشخیص گفتار یا شنوایی ماشینی یکی از مهمترین مشکلات غلبه بر نویز و اغتشاش صوتی محیط است [۷]، در این طرح ما با فرض پایایی میانگین نویز محیطی [۸]، قبل و بعد از دریافت کلمه از نویز محیطی نمونه‌برداری نموده و دامنه فرکانس‌های نویز را از دامنه فرکانس‌های سیگنال نویز آلود کم کردیم که نتیجه بهبود نسبی سیگنال استخراج شده بوده است. شکل (۳) نمونه‌ای از سیگنال قبل و بعد از حذف نویز کلمه "شیراز" در سیگنال به نویز ۸/۵ دسی بل می‌باشد.



شکل ۳.

البته روش فوق با وجود مزیت سادگی، بر فرض پایایی میانگین نویز محیطی استوار است که ممکن است در شرایط مختلف صادق نباشد، اما اثر نامطلوب این مساله بر دقت باز شناسی واژگان، تا حد زیادی توسط روش ترکیبی ارائه شده در الگوریتم بازشناسی واژگان بهبود یافته است.

۲-۳ استخراج مشخصات صوتی

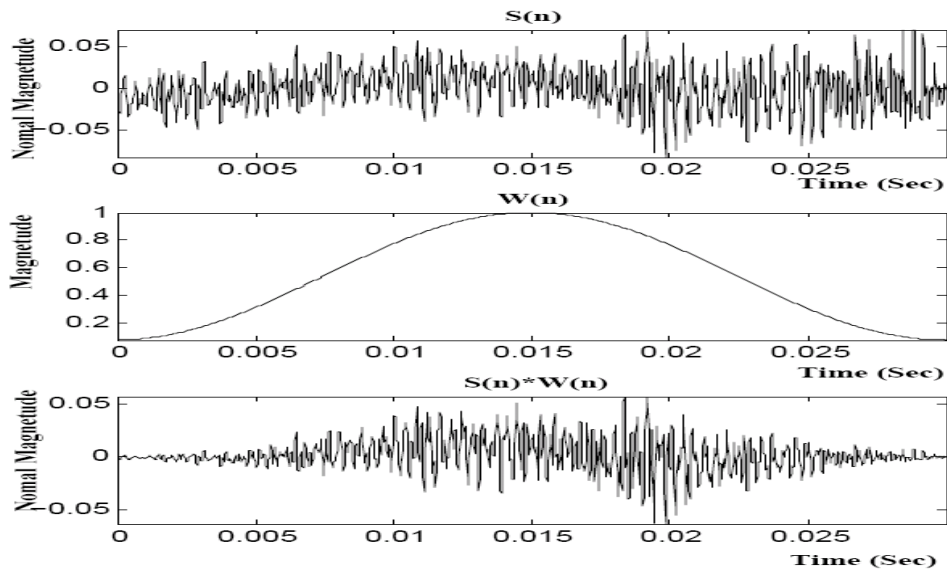
برای استخراج مشخصات اکوستیکی از سیگنال حذف نویز شده، می توان آن را به فریم های ۲۰ تا ۳۰ میلی-ثانیه ای با برهم نهی و تداخل تقریبی بین صفر تا یک سوم تقسیم بندی نمود [۹]، و سپس به منظور ایجاد پیوستگی بین فریم های مختلف، هر فریم در پنجره همینگ با سایز فریم ضرب می گردد. اگر سیگنال موجود در هر فریم $S(n)$ ، $n=0,1,\dots,N-1$ بنامیم سیگنال خروجی برابر است با:

$$Y(n) = S(n) \times W(n) \quad (1)$$

که در آن $W(n)$ پنجره همینگ می باشد که با معادله زیر بیان می گردد [۱۰]،

$$W(n, \alpha) = (1 - \alpha) - \alpha \cos\left(\frac{2\pi n}{N - 1}\right), \quad 0 \leq n \leq N - 1 \quad (2)$$

که در آن n سایز پنجره بوده و α ضریبی است که میزان انحنا پنجره را تغییر می دهد و در نرم افزار MatLab به صورت پیش فرض ۰/۴۶ می باشد و در محاسبات ما، نیز همین مقدار در نظر گرفته شده است. شکل (۴) سیگنال موجود در یک فریم انتخابی، پنجره همینگ و خروجی $Y(n)$ را نمایش می دهد.

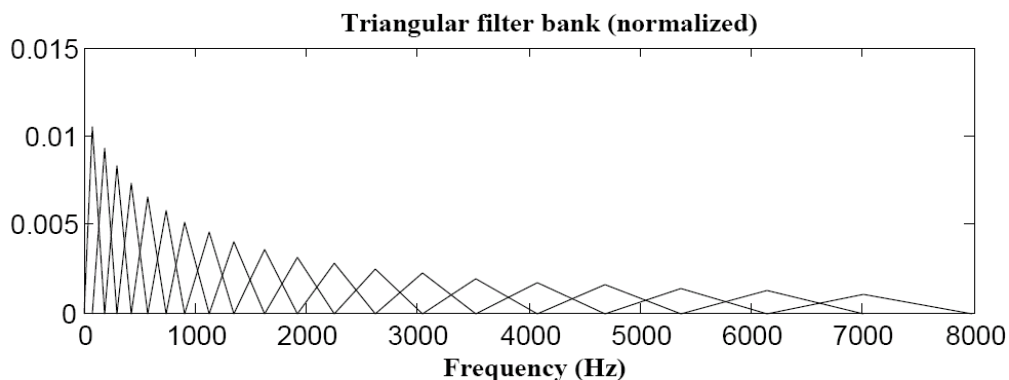


شکل ۴.

در این مرحله با استفاده از آنالیز فوریه، سیگنال‌های موجود در هر فریم $Y(n)$ را به حوزه فرکانس برده تا فرکانس‌های موجود در هر فریم مشخص گردد، آن‌گاه به جهت کاهش اثر هارمونیک‌های موجود در فریم و کاهش سایز ویژگی‌های استخراج شده که مستقیماً در زمان بازشناسی و آموزش سیستم موثر است، فرکانس‌ها را در بیست فیلتر میان‌گذر سه گوش که فرکانس مرکزی هر یک در امتداد مدل Mel است ضرب نموده و از آن لگاریتم می‌گیریم [۳ و ۲]. رابطه (۳) بیانگر ارتباط فرکانس‌های خطی با مدل Mel می‌باشد [۱۱].

$$Mel(f) = 1125 \times \ln\left(1 + \frac{f}{700}\right) \quad (3)$$

در شکل (۵) مدل نرمال شده بیست فیلتر میانگذر سه گوش در امتداد فرکانس‌های Mel نشان داده شده است.

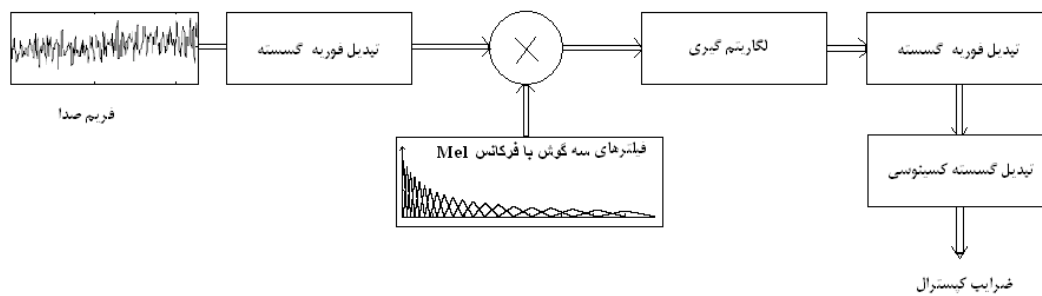


شکل ۵.

در واقع با انجام عمل فوق بخشی از گوش درونی انسان شبیه‌سازی شده است در این مرحله با اعمال تبدیل گسسته کسینوسی (DCT) بر روی لگاریتم انرژی سیگنال عبور داده شده از فیلترهای میانگذر سه گوش Mel ضرایب کپسترال طبق رابطه (۴) به دست می‌آیند [۱۱].

$$C_n = \sum_{k=1}^N x_k \cos\left[\frac{n\pi(k - 0.5)}{N}\right], \quad n = 1, 2, \dots, p \quad (4)$$

که در آن n اندیس ضرایب کپسترال، p تعداد کل ضرایب کپسترال، N تعداد فیلترهای سه گوش، k اندیس ضرایب کسینوسی و x_k امین لگاریتم انرژی خروجی فیلترهای Mel می‌باشد. شکل (۶) بلوک دیاگرام سیستم استخراج ضرایب کپسترال یا MFCC را نشان می‌دهد.



شکل ۶.

معمولاً تعداد فیلترهای سه گوش را بیست و ضرایب کپسترال را دوازده عدد در نظر می‌گیرند [۲]. در اینجا ما علاوه بر استفاده از ضرایب کپسترال، از سرعت و شتاب تغییرات ضرایب کپسترال و همچنین مشخصه اکوستیکی نرخ عبور از صفر سیگنال در هر فریم، نیز جهت ایجاد ساختار بردار مشخصه هر فریم بهره بردیم. سرعت و شتاب تغییرات ضرایب کپسترال از رابطه (۵) قابل محاسبه است [۱۲].

$$\Delta C_n(t) = \left[\sum_{\tau=-m}^m C_n(t+\tau)\tau \right] / \left[\sum_{\tau=-m}^m \tau^2 \right] \quad (5)$$

که معمولاً $m=2$ در نظر گرفته می‌شود [۱۲].

۳ آموزش و بازشناسی الگوی کلمات

۳-۱ الگوریتم ترکیبی بازشناسی

برای بازشناسی واژه‌ها و کلمات روشی ترکیبی پیشنهاد شده است که در عمل پیاده‌سازی شده و نتایج آن در جدول ۳ و ۴ مندرج است. در این قسمت ابتدا بردار ویژگی‌های به دست آمده از همه فریم‌های سیگنال ورودی

از نظر میزان همبستگی خطی یا Correlation با داده‌های موجود در بانک از پیش آموزش داده شده، مقایسه می‌گردد و کلمه‌ای که بیشترین همبستگی خطی با ورودی را دارا باشد در غالب موارد طبق نتایج مندرج در جدول (۲) تشخیص صحیح واژه ورودی است. اما نکته قابل توجه این که با وجود دقت بسیار عالی روش فوق در صورت عدم وجود واژه ورودی در بانک آموزشی یا میزان سیگنال به نویز پایین و اغتشاش صوتی بالای محیط، در حالت اول تشخیص کاملاً اشتباه و در حالت دوم احتمال خطا وجود خواهد داشت. برای رفع این مشکل از یک شبکه ناظر عصبی با ساختار پس انتشار خطا استفاده شده است که بعد از آموزش و فراگیری الگوهای ورودی، نقش آن ایجاد بانک معیار صحت شناخت برای هر کلمه است، شبکه عصبی که در شکل (۸) مشاهده می‌گردد دارای تعداد خروجی برابر با تعداد واژه‌های آموزشی است و به تناسب شباهت ورودی به یکی از الگوهای آموزشی در یک شبکه رقابتی خروجی آن الگو به یک و سایرین به صفر همگرا خواهد شد با استفاده از این نکته دنباله‌ای از ضرایب تشخیص صحیح در تست‌های متعدد از یک کلمه در صورت تایید صحت شناخت شبکه توسط اپراتور، فراهم می‌گردد که با انواع روش‌ها از جمله استفاده از شبکه عصبی یا میانگین و آستانه‌گیری که در این جا استفاده شده است می‌توان به معیاری جهت صحت شناخت از سوابق تشخیص قبلی برای هر کلمه رسید.

بنابراین تابع تصمیم‌گیری در مواجهه با یک واژه ناشناس، ابتدا از طریق میزان همبستگی خطی به سمت یکی از الگوهای موجود در بانک همگرا خواهد شد و سپس با بررسی پاسخ شبکه عصبی در جایگاه کلمه تشخیص داده شده توسط تابع همبستگی و مقایسه آن با آستانه معیار صحت شناخت کلمه مذکور پی به عدم وجود کلمه در بانک آموزشی خواهد برد.

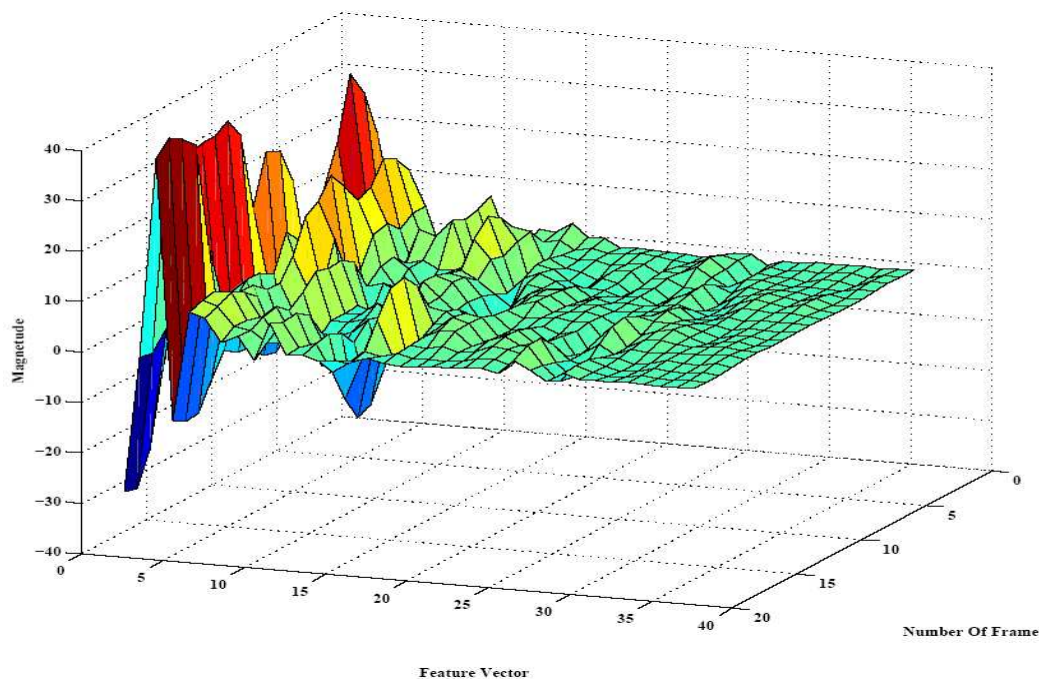
۳-۲ تابع ترکیبی بازشناسی

با توجه به موارد یاد شده در بخش (۳-۱) ابزار مورد استفاده، در زیر مورد بررسی قرار می‌گیرد. تابع همبستگی بین دو بردار هم اندازه A و B از رابطه (۶) قابل محاسبه می‌باشد [۱۳].

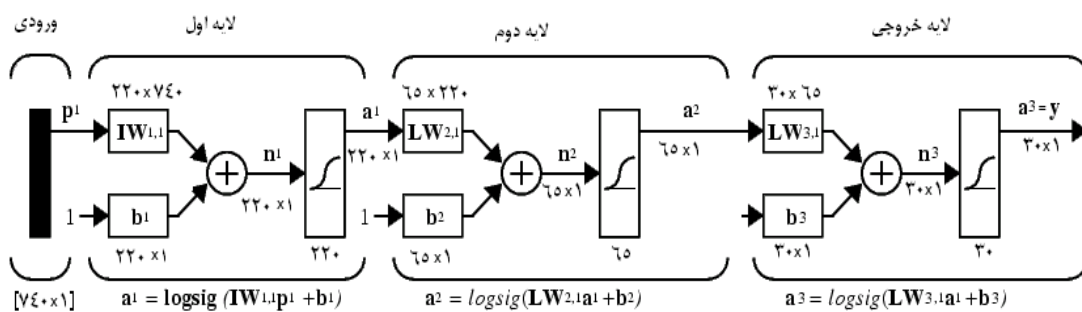
$$R = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\sum_m \sum_n (A_{mn} - \bar{A})^2} \sqrt{\sum_m \sum_n (B_{mn} - \bar{B})^2}} \quad (6)$$

که در آن R ضریب همبستگی خطی و \bar{A}, \bar{B} میانگین بردارهای A و B می‌باشد. شبکه عصبی مورد استفاده در این مقاله از نوع پس انتشار خطاست، با آزمایشات فراوان و با توجه به این که در مکانیسم استخراج ویژگی با استفاده توام از ضرایب کپسترال دوازده گانه، سرعت و شتاب تغییرات ضرایب و نرخ عبور از صفر، بردار هر فریم، ۳۷ عضو خواهد داشت و تعداد فریم‌های کلمات با توجه به گویش‌های مختلف بین ۱۰ تا ۵۰ فریم با فرکانس نمونه برداری $F_s = 44100 \text{ Hz}$ خواهد بود که نشان‌دهنده حجم نسبتاً بالای اطلاعات

مربوط به یک کلمه فرضی می‌باشد. در این جا به دلیل هم سایز نبودن کلمات آن‌ها را به ۲۰ فریم، به کمک الگوریتم‌های تغییر سایز تصاویر، هم سایز می‌نماییم که این امر تاحدودی باعث کاهش دقت در کلمات مشابه خواهد شد، سپس ماتریس ۲۰*۳۷ ویژگی‌ها به بردار ۱*۷۴۰ تغییر شکل داده می‌شود تا به شبکه اعمال گردد. در شکل (۷) بردارهای ویژگی کلمه "شیراز" که شامل ۲۰ بردار ۳۷ عضوی متناظر با ۲۰ فریم می‌باشد به صورت ۳ بعدی مشخص شده است. در شبکه عصبی مطرحه با توجه به این که با افزایش تعداد کلمات در صورت ثابت بودن سایز شبکه، از دقت آن کاسته خواهد شد لذا شبکه با یک بانک ثابت ۳۰ کلمه ای متشکل از کلمات مشابه و غیر مشابه با ساختاری ۳ لایه و ابعاد نشان داده شده در شکل (۸)، جهت جلوگیری از دقت پایین در تفکیک الگوها و همچنین اجتناب از مساله بیش برازش (Over Fitting) بر روی داده‌ها، طراحی گردید.

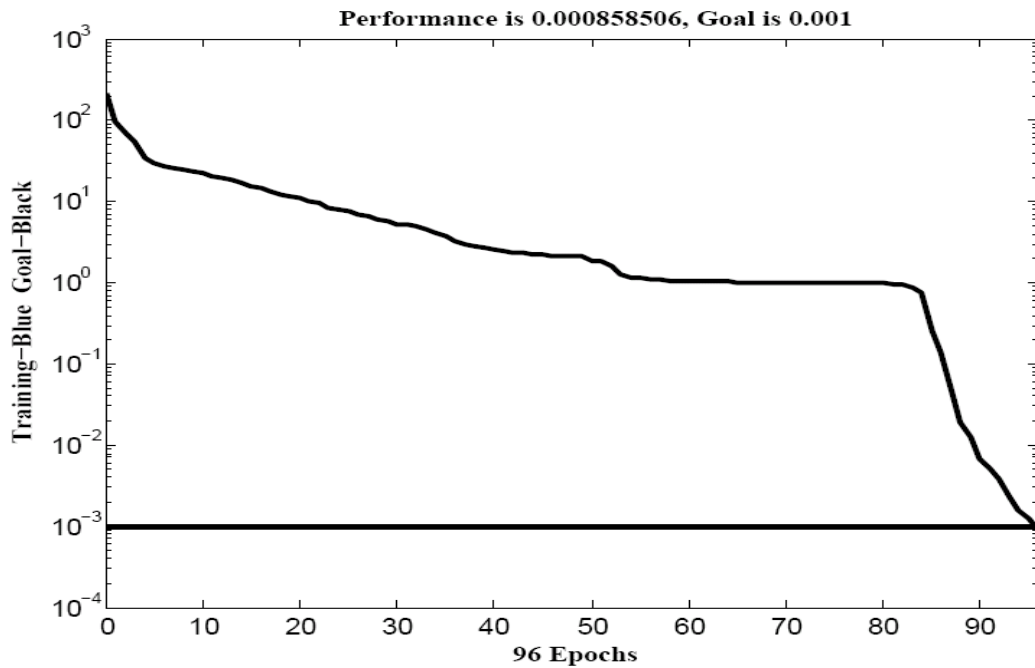


شکل ۷. نمایش سه بعدی مجموعه بردار فریم‌های کلمه "شیراز"



شکل ۸. شبکه عصبی سه لایه پس انتشار خطا

در شبکه عصبی ذکر شده از الگوریتم آموزشی Powell-Beale Restarts که موجب افزایش سرعت یادگیری می‌گردد [۱۴]، و مجموع مربعات خطا جهت اصلاح ضرایب وزن‌ها استفاده شده و شبکه برای رسیدن به حداقل خطای مجاز تعیین شده (۰/۰۰۱) با توجه به تعیین تصادفی مقادیر اولیه وزن‌ها و بایاس‌ها تعداد تکرار نشان داده شده در شکل (۹) را طی نموده است.



شکل ۹.

۳-۳ تست و آزمایش سیستم باز شناسی واژگان

در این قسمت جداول حاصل از آزمایشات مختلف شبکه آورده شده، در تمامی حالات بانک آموزشی متشکل از ۳۰ کلمه ذکر شده در ذیل است که حاوی کلمات مشابه و غیر مشابه می‌باشد و شرایط سیگنال به نویز تقریبی یکسان ذکر شده در هر آزمایش، مورد توجه قرار گرفته است.

جدول ۱. مجموعه کلمات آموزش داده شده

گل ، گم ، قم	سیستان ، چیستان ، بوستان
آمل ، زابل ، بابل	بندر انزلی ، بندر عباس
کرمانشاه	احساس ، احسان
لرستان	ریشه ، تیشه ، بیشه ، گیشه
اصفهان	گیلان ، میلان
اهواز ، شیراز ، زاهدان	مشهد ، اشهد
گرگان ، کرمان ، تهران	

جدول ۲ بیانگر نتایج حاصل از میزان دقت شناسایی تابع همبستگی خطی، در کلمات مشابه و غیر مشابه است و تست تشخیص، توسط آموزش دهنده مرد در شرایط سیگنال به نویز میانگین ۱۰ دسی بل می باشد.

جدول ۲.

کلمات دارای مشابه	کلمات غیر مشابه					
	احسان	سیستان	مشهد	گیشه	زاهدان	شیراز
میانگین سیگنال به نویز (db)	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰
تعداد تکرار	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰
تعداد اشتباه	۱	۰	۲	۸	۰	۰
تعداد صحیح	۱۹	۲۰	۱۸	۱۲	۲۰	۲۰
درصد تشخیص	%۹۵	%۱۰۰	%۹۰	%۶۰	%۱۰۰	%۱۰۰

جدول ۳ نتایج استفاده از روش ترکیبی در باز شناسی کلمات آموزش داده شده و تشخیص کلمات آموزش داده نشده در سیگنال به نویز میانگین حدود ۱۰ دسی بل و تست تشخیص، توسط آموزش دهنده مرد با کلمات مشابه جدول ۲.

جدول ۳. گوینده مرد و تعلیم دهنده شبکه عصبی

کلمات ناشناس غیر مشابه و مشابه	کلمات دارای مشابه						کلمات غیر مشابه			
	تهران	کابل	اردبیل	بوشهر	احسان	سیستان	مشهد	گیشه	زاهدان	شیراز
میانگین سیگنال به نویز (db)	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰
تعداد تکرار	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰
تعداد اشتباه	۰	۳	۰	۰	۰	۰	۰	۶	۰	۰
تعداد صحیح	۰	۰	۰	۰	۱۸	۲۰	۱۷	۱۱	۱۹	۲۰
عدم تشخیص	۲۰	۱۷	۲۰	۲۰	۲	۰	۳	۳	۱	۰
درصد تشخیص	%۱۰۰	%۸۵	%۱۰۰	%۱۰۰	%۹۵	%۱۰۰	%۸۵	%۵۵	%۹۵	%۱۰۰

جدول ۴ نتایج استفاده از روش ترکیبی در باز شناسی کلمات آموزش داده شده و تشخیص کلمات آموزش داده نشده در سیگنال به نویز میانگین حدود ۱۰ دسی بل و تست تشخیص، توسط زن، در شبکه آموزش داده شده با صدای مرد با کلمات مشابه جدول ۲.

جدول ۴. گوینده زن و تعلیم دهنده شبکه عصبی مرد

	کلمات غیر مشابه		کلمات دارای مشابه				کلمات ناشناس غیر مشابه و مشابه			
	شیراز	زاهدان	گیشه	مشهد	سیستان	احسان	بوشهر	اردبیل	کابل	تهران
میانگین سیگنال به نویز (db)	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰
تعداد تکرار	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰
تعداد اشتباه	۰	۰	۸	۰	۰	۰	۰	۰	۳	۲
تعداد صحیح	۲۰	۲۰	۱۰	۱۵	۲۰	۱۵	۰	۰	۱	۰
عدم تشخیص	۰	۰	۲	۵	۰	۵	۲۰	۲۰	۱۶	۱۸
درصد تشخیص	%۱۰۰	%۱۰۰	%۵۰	%۷۵	%۱۰۰	%۷۵	%۱۰۰	%۱۰۰	%۸۰	%۹۰

در آزمایشات فوق، گوینده وازگان و آموزش دهنده وازگان یکسان بوده است و سیگنال به نویز هر کدام از تست‌ها در حدود ۱۰ دسی بل حفظ شد و میانگین آن هم ۱۰ در نظر گرفته شد که ۰/۱۰ اعشار تقریب دارد. در شرایطی که گوینده وازگان تغییر نماید دقت تشخیص کاهش خواهد یافت.

۴ نتیجه گیری

این مقاله، از سیستم شنوایی طرح پژوهشی ربات انسانمای "دارا" با حمایت دانشگاه آزاد اسلامی واحد بندرلنگه، استخراج شده است که بدین وسیله مراتب قدردانی مولف بیان می‌گردد در این مقاله طراحی و پیاده سازی یک سیستم تشخیص کلمات و وازگان به روشی سیگنالی که مستقل از زبان بوده و در هر زبان از جمله فارسی قابل استفاده است، مطرح گردید که شامل بخش‌های مختلفی بوده و هر یک به صورت زیر شاخه تخصصی قابل بررسی است، اما از نتایج این مقاله می‌توان به ترکیب پیشنهادی بردار استخراج ویژگی در مقایسه با ترکیبات استفاده شده قبلی [۳ و ۲] نامبرد که به نظر نگارنده باعث افزایش دقت تشخیص بوده است که گواه آن ضریب تشخیص ۱۰۰٪ برای کلمات غیر مشابه آموزش داده شده مطابق جدول ۳ و ۲ می‌باشد، همچنین ارایه روشی ترکیبی به کمک تابع همبستگی و شبکه عصبی جهت تشخیص و نظارت بر صحت آن، با استخراج معیار صحت، که موجب دقت در پی بردن به عدم وجود کلمات ناشناس در بانک آموزشی شده است. استقلال سیستم تشخیص در مواجهه با گویندگان مختلف مطابق با نتایج جدول ۴ موجب بهروری سیستم در کاربردهایی نظیر منشی کامپیوتری و سیستم‌های خودپرداز می‌گردد، در پایان باید به دقت نه چندان عالی سیستم در مواجهه با واژه‌های ناشناس مشابه بانک آموزشی اشاره نمود که این مساله علاوه بر ضرورت مطالعات بیشتر، بیان گر لزوم استفاده از مدل‌های گرامری و الگوهای آوایی زبان نیز می‌باشد تا تشخیص قطعی وازگان در ساختار گرامری و آوایی زبان ممکن گردد.

منابع

1. Nelson, M., Herve A., (1995). Neural Networks for Statistical Recognition of Continuous Speech. Proc. IEEE, 83, 5.

2. Prasad, D. P., Zhou, R., Yang, J., Adnani, F., Hobson, R. S., (2001). Isolated Speech Recognition Using Artificial Neural Network. Proc. Of 23rd Annual EMBS international Conference October, 25-28.
3. Chibelushi, C. C., Deravi, F., (2002). A Review of Speech-Based Bimodal Recognition. IEEE Transactions on multimedia, 4(1).
4. Juang, C. F., Lin, C. T., (2001). TNoisy Speech Processing by Recurrently Adaptive Fuzzy Filter. IEEE Transactions on Fuzzy Systems, 9(1).
5. Hazas, M., Hall, H., (1999). TProcessing of Non-Stationary Audio Signals. Department of Engineering, University of Cambridge, p. 4.
6. Li, F. F., Cox, T. I., (2003). a neural network for blind identification of speech transmission index. Proc. IEEE.
7. Thirsch, H. G., Ehrlicher, C., (1995). Noise estimation techniques for robust speech recognition. Acoustics, Speech, and Signal Processing, ICASSP-95.
8. Kliever, Dr. J., (2007). Speech Denoising. University of Notre Dame, Department of Electrical Engineering, EE 40471: Digital Signal Processing.
9. Garrido, C. M., Murthi, M. N., Miami, U. O., Andersen, S. V., (2005). Towards iLBC SpeechCoding at Lower Rates Through a New Formulation of the Start State search. Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP '05). IEEE International Conference.
10. Newandee, D. A., Reisman, S. S., (2003). TWavelet representation comparison for heart rate variability analysis. Proc. IEEE, 2003
11. Molau, S., Pitz, M., Schluter, R., Ney, H., (2001). Acoustics, Speech, and Signal Processing. Proceedings. (ICASSP apos;01).
12. Kwon, O. W., Kwokleung, H. J., Lee, T. W., (2003). Emotion recognition by speech signals. In EUROSPEECH-2003, 125-128.
13. Vijaya Kumar, B. V. K., Mahalanobis, A., (2005). Correlation Pattern Recognition.
14. Powell, M. J. D., (1997). Restart procedures for the conjugate gradient method. Mathematical Programming, 12, 241-254.