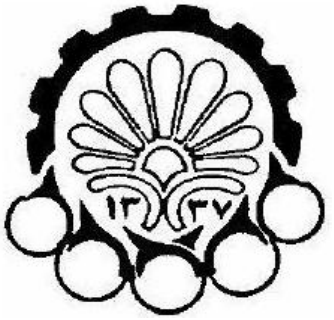

دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

دانشکده مهندسی پزشکی

پایان نامه کارشناسی ارشد

مهندسی پزشکی-گرایش بیوالکتریک

عنوان:

استخراج مؤلفه‌های مستقل غیرخطی سیگنال گفتار

استاد راهنما:

دکتر سیدعلی سیدصالحی

تهیه کننده :

مهدیه قاسمی

بهمن‌ماه ۱۳۸۴

بسمه تعالی

شماره:

تاریخ:

پیشنهاد پروژه تحصیلات تکمیلی

(رساله کارشناسی ارشد و دکترا)×

فرم پروژه تحصیلات تکمیلی ۱

دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

نام و نام خانوادگی: مهدیه قاسمی — شماره دانشجوئی:۸۲۱۳۳۲۳۱

رشته تحصیلی: بیوالکتریک — دانشکده: مهندسی پزشکی

آدرس: تهران- خ سمیه — خ پورموسی — نبش کوچهٔ شیرین - خوابگاه شهید صنعتی — تلفن : ۸۸۹۵۰۲۱ — مقطع: کارشناسی ارشد

۲-مشخصات اساتید راهنما :

نام و نام خانوادگی : دکتر سیدعلی سیدصالحی

۱- سمت، مرتبه علمی و محل خدمت:
استادیار دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر
تلفن: ۶۴۵۴۲۳۸۸

۳-مشخصات استاد مشاور

نام و نام خانوادگی: — سمت، مرتبه علمی: — تلفن :

۴-عنوان پایان نامه یا رساله

فارسی: استخراج مؤلفه های مستقل غیرخطی سیگنال گفتار

انگیسی:

Nonlinear independent component analysis of speech signal

نوع پروژه: کاربردی ☒ بنیادی ☒ توسعه ای ☐ **تعداد واحد : ۶**

۵- خلاصه پایان نامه: (مسئله فرضیات، هدف از اجراء، توجیه ضرورت انجام طرح)

تحقیقات در زمینهٔ عملکرد مغز انسان در ادراک گفتار از جنبه های مختلف مورد بررسی قرار گرفته است.یکی از موارد قابل توجه در این زمینه واحدهای پایهٔ صوتی است که بعنوان ورودی های گفتار مورد پردازش و ادراک قرار می گیرند.مطالعات روی نحوهٔ ادراک گفتار انسان نشان می دهد که مغز انسان به وقایع خاصی در سیگنال گفتار حساسیت بیشتری نشان می دهد. ماهیت این وقایع از دیدگاه مهندسی و زبان شناسی بعنوان واحدهای پایه حاوی بیشترین اطلاعات در سیگنال گفتار در دست تحقیق و بررسی است.

بسیاری از مدلهایی که برای بازشناسی واجها بکار می روند ،متکی به تقطیع سیگنال هستند و نواحی با مشخصات طیفی سریعاً گذرای سیگنال گفتار در تقطیع و یا برچسب دهی در مرز بین واجها قرار می گیرند.وجود خطاهای متعدد، ایجاد تنوعات و چندگونگی های زائدی می نماید که موجب اتلاف و ریزش اطلاعات صوتی مفید موجود در سیگنال گفتار در هنگام بازشناسی می گردد.برای رفع این مشکل لازم است که روشهای رایج بر برچسب دهی خروجی مدل های باز شناسی واحدهای صوتی را تغییر داده و در خود واحدهای صوتی پایه نیز اصلاحاتی را بعمل آوریم. در آنالیز مؤلفه های مستقل مطلوب آن است که اطلاعات اضافی ، تکراری و یا کم اهمیت را از سیگنال حذف کنیم تا سیگنال فقط حاوی اطلاعاتی باشد که از دید هدف نهایی ما دارای اهمیت و ارزش است .هدف این پروژه آن است که سعی شود با استفاده از انواع روش های استخراج مولفه های مستقل (و اساسی) غیرخطی سیگنالها توسط شبکه های عصبی ،مواضع حاوی بیشترین اطلاعات گفتاری را در روی سیگنال مشخص نماییم و آنها را بعنوان واحدهای پایهٔ بازشناسی گفتار مورد بررسی قرار دهیم .میزان انطباق این مؤلفه های مستقل غیرخطی برخواسته از داده ها با واحدهای نشأت گرفته از دانسته های زبان شناختی (و سایر روشها) نظیر آواها، دوآواها، وقایع و حالات بررسی خواهند شد، به‌عبارتی کارآیی این مؤلفه ها از نظر در برداشتن اطلاعات کاملتری از سیگنال گفتار در مقایسه با آواها و موارد مشابه مورد ارزیابی قرار خواهند گرفت، تا از این طریق واحدهای پایهٔ زبانی مورد استفاده در بازشناسی گفتار بهبود داده شوند. بعنوان مبنای ارزیابی، از ملاکهای میزان تحمل در مقابل اغتشاشات ناخواسته و تولید پاسخهای متنوع قابل قبول استفاده خواهد شد.

×**توضیحات:** لطفا تایپ شود.

و قل ربّ زدنی علما

تقدیر و تشکر

با سپاس و ستایش پروردگار بی همتایی که تنها یاور انسان در میدان رویارویی با جلوه های حیات است ، لازم می دانم تا از کلیه عزیزانی که در دوران تحصیل تا به این مرحله همراه و یاریگر من بوده اند، سپاسگزاری و تشکر نمایم.

از استاد ارجمندم ، جناب آقای دکتر سیدصالحی که راهنمای من در انجام این پروژه بوده اند و همواره نصایح علمی و معنوی ایشان راهگشای من بوده است ، کمال تشکر و امتنان را دارم.

همچنین مراتب سپاس خود را از اساتید و دوستانی که از نظرات و راهنمایی هایشان بهره بردام و خانم ساره زنده روح که زحمت مطالعه پایان نامه را متقبل شدند ، اعلام می دارم.

از پدر و مادر مهربانم به خاطر همه زحمت ها ، تشویق ها و از خودگذشتگی های بی دریغ و خالصانه شان تشکر و قدردانی می نمایم. همچنین از همسر عزیزم که نهایت همراهی و صبر و شکیبایی را در مراحل انجام این پروژه ابراز داشته است ، متشکرم.

در انتها خاطر نشان می کنم که این پروژه از طرح حمایت مرکز تحقیقات مخابرات ایران از پروژه های کارشناسی ارشد طبق قرارداد شماره ۵۰۰/۸۲۲۵/ت مورخ ۸۴/۷/۱۶ دانشگاه صنعتی امیرکبیر با آن مرکز بهره مند گردیده است.

تقدیم به

پدر و مادر عزیز و بزرگوارم که وجودشان ترجمان محبت و ایثار است و طی این راه جز با یاری دستان گرم آنها میسر نبود

و همسر مهربانم که با همراهی صمیمانه اش راه را بر من هموار نمود

چکیده

مطالعات بر روی نحوه ادراک گفتار انسان نشان می‌دهد که مغز انسان به وقایع خاصی در سیگنال گفتار حساسیت بیشتری نشان می‌دهد و این نواحی حاوی اطلاعات متمایزکننده مفیدی برای واحدهای صوتی پایه است. ماهیت این وقایع بعنوان واحدهای پایهٔ واقعی حاوی اطلاعات غنی و مهم گفتار، از نظر زبان‌شناسی و مهندسی در دست بررسی است.

از سوی دیگر در بررسی جنبه‌های زیستی ادراک گفتار توسط مغز انسان دیدگاههایی وجود دارد که نشان‌دهندهٔ آن است که اطلاعات دریافتی مغز، خلاصه‌شده و چکیده‌هایی از آن بصورت مؤلفه‌های اساسی ادراک به نواحی بالاتر مغز ارسال می شوند و برخی از قوانین محلی آنالیز مؤلفه‌های مستقل در تحلیل آنها بکار گرفته می‌شود. این مطالب مؤید ایدهٔ روش استفاده از مؤلفه‌های اساسی و مستقل است که بتواند اطلاعات نواحی پراهمیت سیگنال گفتار را استخراج کند.

در این تحقیق بر خلاف سایر روش‌هایی که بصورت باسرپرستی و بر مبنای دانش افراد خبره مثلاً نواحی گذرا و ایستان و یا وقایع و حالات گفتار را بعنوان نشانه‌های گفتار در نظر می‌گیرند، با استفاده از شبکه‌های عصبی بصورت بدون سرپرستی به شبکه این امکان را می‌دهیم تا نواحی پراطلاعات سیگنال گفتار را بطور خودکار استخراج کند. پس از استخراج مؤلفه‌ها با بهره‌گیری از راهکارهای مختلف به بررسی ماهیت این مؤلفه‌ها می‌پردازیم تا روشن شود که این مؤلفه‌ها بیان‌کننده چه اطلاعاتی از سیگنال گفتار هستند.

ابتدا با کمک روشهای PCA آماری ، مؤلفه‌های متعامد گفتار و سپس با استفاده از شبکه عصبی جلوسوی سه لایه پنهان با لایه گلوگاه، مؤلفه‌های اساسی خطی و غیرخطی در حالت استاتیک (تک فریم در ورودی) استخراج شده‌اند. همچنین جهت بررسی مؤلفه‌های حالت دینامیک گفتار، استخراج مؤلفه‌ها با ورودی ۱۵ فریم نیز انجام شده است.در هر دو حالت استاتیک و دینامیک تعداد واحدهای لایه پنهان متغیر انتخاب شده‌اند.

سپس جهت بررسی ماهیت این مؤلفه‌ها و میزان انطباق آنها با واحدهای نشأت‌گرفته از دانش زبان‌شناختی نظیر آواها آنالیزهای مختلفی انجام شده و میزان کارآیی مؤلفه‌ها بررسی شده است. راهکارهای بکار گرفته در این قسمت عبارتند از : بازسازی پارامترها در خروجی و شنیدن سیگنال گفتار بازسازی‌شده از آن، محاسبه درصد صحت بازشناسی آوا به ازای تعداد مؤلفه‌های مختلف استخراج شده ، رسم فضای پارامترهای بازنمایی ورودی و مانیفولد تصویرشدهٔ آن در فضای خروجی، بررسی مقادیر آنالوگ کدهای لایه پنهان و یافتن ارتباط بین مقادیر استخراج‌شده و مفاهیم آشنای گفتار فارسی.

بررسی‌ها نشان می‌دهد که در روش *PCA* غیرخطی مؤلفه اول استخراج شده نشان‌دهنده ویژگی واکدار یا بی‌واک بودن گفتار است و مؤلفه دوم سعی دارد که بین واکه‌ها تمایز قرار دهد. همچنین بعلت شکل‌گیری مانیفولد غیرخطی، میزان اطلاعات عبوری به ازای یک نورون در لایه گلوگاه قابل توجه است. در تعلیم شبکه *PCA* غیرخطی مؤلفه‌های اساسی غیرخطی به ترتیب بزرگی شکل می‌گیرند و سبب افت پله‌ای منحنی خطا می‌شوند.

اگر سیگنال گفتار را ترکیبی از منابع و فاکتورهای مجزا در نظر بگیریم، با استفاده از روشهای استخراج مؤلفه‌های مستقل می‌توانیم به منابع اولیه تولید گفتار دست یابیم و آنها را بعنوان مؤلفه‌های پایه گفتار برگزینیم. بدین منظور مؤلفه‌های مستقل خطی و غیرخطی گفتار با استفاده از روش‌های شبکه عصبی استخراج شده‌اند و در هر مرحله مزایا و مشکلات روشها مورد بررسی قرار گرفته است.

کلمات کلیدی: بازشناسی گفتار، شبکه عصبی ، تحلیل مؤلفه‌های مستقل ، تجزیه و تحلیل غیرخطی سیگنال گفتار

فهرست مطالب

عنوان **صفحه**

فصل اول- مقدمه‌ای بر واحدهای پایه در بازشناسی گفتار

۱-۱- مقدمه ۱

۱-۲- دیدگاه زبان‌شناسی در مورد واحدهای پایه گفتار ۲

۱-۳- بررسی نواحی گذرای مرز واجها ۳

۱-۴- نقش وقایع گسسته در بازشناسی گفتار ۴

۱-۵- شرح مسئله ۵

فصل دوم- مبانی تئوری آنالیز مؤلفه‌های اساسی

۲-۱- مقدمه ۸

۲-۲- *PCA* آماری ۹

۲-۲-۱- مروری بر جبر خطی *PCA* آماری ۹

۲-۲-۲- محاسبه بردارهای ویژه ۱۱

۲-۳- تئوری *PCA* از دیدگاه شبکه خودانجمنی ۱۳

۲-۳-۱- نگاشت خطی بر مبنای شبکه *PCA* ۱۳

۲-۳-۲- نگاشت غیرخطی در ساختار شبکه گلوگاه ۱۴

۲-۴- شبکه سلسله مراتبی جهت استخراج مؤلفه‌های اساسی بر حسب مرتبه ۱۸

۲-۴-۱- مشکلات روشهای *PCA* غیرخطی ۱۸

۲-۴-۲- معرفی شبکه مؤلفه‌های اساسی سلسله مراتبی ۱۹

۲-۴-۳-تحلیل عملکرد شبکه سلسله مراتبی ۲۱

۲-۵- جمع‌بندی ۲۲

فصل سوم- آنالیز مؤلفه‌های مستقل با استفاده از شبکه‌های عصبی

۳-۱- مقدمه ۲۳

۳-۲- مبانی آنالیز مؤلفه‌های مستقل ۲۴

۳-۲-۱- تفاوت استقلال و همبستگی ۲۴

۳-۲-۲- آنالیز مؤلفه‌های مستقل خطی و تفکیک منابع کور ۲۵
۳-۳- قوانین یادگیری *PCA* غیرخطی در آنالیز مؤلفه‌های مستقل ۲۶
۳-۳-۱- قوانین یادگیری *PCA* خطی و غیرخطی ۲۶
۳-۳-۲- پیش‌پردازش در استفاده از الگوریتم یادگیری *PCA* ۲۸
۳-۳-۳- فرضهای مورد نیاز در آنالیز مؤلفه‌های مستقل ۲۸
۳-۳-۴- تحلیل قانون یادگیری *PCA* غیرخطی در توسعه به *ICA* ۲۹
۳-۳-۵- مثالی جهت بررسی مسئله چرخش مؤلفه‌ها در *ICA* ۳۰
۳-۴- مدل دادگان در *ICA* خطی ۳۱
۳-۵- مدل شبکه عصبی برای *ICA* خطی ۳۲
۳-۶- سفیدکنندگی با استفاده از روش آماری و شبکه عصبی در مدل *ICA* خطی ۳۳
۳-۷- مروری بر الگوریتم‌های شبکه عصبی جهت *ICA* خطی ۳۵
۳-۸- تخمین بردارهای پایه در *ICA* خطی ۳۸
۳-۹- اطلاعات متقابل به عنوان معیار *ICA* ۳۸
۳-۱۰- معیار اندازه‌گیری آمارگان‌های مرتبه‌های بالاتر ۴۰
۳-۱۱- *ICA* خطی و غیرخطی بر مبنای اطلاعات متقابل (روش *MISEP*) ۴۱
۳-۱۱-۱- روش *INFOMAX* ۴۲
۳-۱۱-۲- اصول تئوری روش *MISEP* ۴۳
۳-۱۱-۲-۱- تعلیم خواص غیرخطی در خروجی ۴۳
۳-۱۱-۲-۲- تعمیم روش به *ICA* غیرخطی ۴۴
۳-۱۲- جمع‌بندی ۴۷

فصل چهارم- روشها و شبکه‌های عصبی جهت استخراج مؤلفه‌های اساسی گفتار در حالت استاتیک

۴-۱- مقدمه ۴۸
۴-۲- دادگان مورد استفاده ۴۹
۴-۲-۱- پارامترهای بازنمایی ۵۰
۴-۲-۲- آماده‌سازی، نرمالیزه کردن و برچسب‌دهی دادگان ۵۰
۴-۳- معرفی روشها و شبکه‌های عصبی استفاده شده جهت استخراج ویژگی تک فریم گفتار ۵۲
۴-۳-۱- روش *PCA* آماری ۵۲
۴-۳-۲- شبکه *PCA* خطی ۵۴
۴-۳-۳- شبکه *PCA* غیرخطی ۵۹

۴-۴- استخراج ویژگی از روی پارامترهای طیف سیگنال گفتار ۶۵
۴-۵- جمع‌بندی..... ۶۸

فصل پنجم- تحلیل و ارزیابی مؤلفه‌های استخراج شده در شبکه *PCA* غیرخطی در حالت استاتیک

۵-۱- مقدمه ۶۹
۵-۲- شبکه بازشناسی آوا ۷۰
۵-۲-۱- معرفی شبکه بازشناسی آوا ۷۰
۵-۲-۲- نتایج بازشناسی حاصل از مؤلفه‌های اساسی خطی ۷۱
۵-۲-۳- نتایج بازشناسی حاصل از مؤلفه‌های اساسی غیرخطی ۷۳
۵-۲-۴- شبکه بازشناس مرجع ۷۵
۵-۲-۵- تحلیل نتایج بازشناسی ۷۶
۵-۳- تبدیل پارامترهای بازنمایی فیلترشده به سیگنال زمانی و ارزیابی آنها ۷۹
۵-۴- تحلیل ماهیت مقادیر لایه گلوگاه به ازای یک مؤلفهٔ اساسی در آن ۸۲
۵-۴-۱- ارتباط بین مقادیر لایه پنهان و برچسب واک‌دار/ بی‌واک در فارس‌دات ۸۲
۵-۴-۲- مقادیر تک نورون لایه گلوگاه به ازای هر آوا ۸۴
۵-۴-۳- از میانگین مقدار تک نورون هر آوا به *LHCB* خروجی و بازسازی آن ۸۸
۵-۵- بردارهای بازنمایی ورودی و خروجی هر آواها ۹۱
۵-۶- نحوه عملکرد نورون‌های لایه پنهان در فریمهای متوالی ۹۵
۵-۶-۱- نقش تک نورون لایه پنهان در فریم‌های متوالی ۹۵
۵-۶-۲- نقش۶ نورون لایه پنهان در فریم‌های متوالی ۱۰۰
۵-۷- نقش هر یک از نورون‌ها در لایه پنهان میانی شبکه *PCA* غیرخطی ۱۰۵
۵-۷-۱- نگاشت حاصل از ۲ نورون در فضای *LHCB* ۱۰۶
۵-۷-۲- نگاشت حاصل از ۶ نورون در فضای *LHCB* ۱۰۸
۵-۸- کددهی به مؤلفه‌های اساسی و ایجاد ماتریس اختلاط ۱۱۰
۵-۹- جمع‌بندی ۱۱۲

فصل ششم- روشها و شبکه‌های عصبی جهت استخراج و آنالیز مؤلفه‌های اساسی گفتار در حالت دینامیک

۶-۱- مقدمه ۱۱۴
۶-۲- استخراج مؤلفه‌های اساسی گفتار در حالت دینامیک ۱۱۵
۶-۲-۱- روش *PCA* آماری ۱۱۶
۶-۲-۲- روش شبکه *PCA* غیرخطی ۱۱۷
۶-۳- شبکه بازشناس آوا ۱۲۱
۶-۳-۱- معرفی شبکه بازشناس آوا در حالت دینامیک ۱۲۱
۶-۳-۲- شبکه مرجع دینامیک ۱۲۳
۶-۳-۳- نتایج بازشناسی ۱۲۵
۶-۴- بازسازی خروجی شبکه استخراج مؤلفه‌های اساسی حالت دینامیک در حوزه زمان ۱۲۶
۶-۵- تحلیل اطلاعات شبکه در پرش‌های منحنی خطا ۱۲۷
۶-۶- نحوه تغییرات تک نورون لایه پنهان با حرکت شبکه روی فریم‌های سیگنال ورودی ۱۳۰
۶-۷- اثر نویز ۱۳۳
۶-۸- اثر تغییر گوینده ۱۳۵
۶-۹- نگاشت از فضای *LHCB* به آوا و بالعکس ۱۳۷
۶-۱۰- شبکه سلسله مراتبی جهت استخراج مؤلفه‌های اساسی برحسب مرتبه ۱۳۹
۶-۱۱- شبکه *PCA* غیرخطی با اتصال بازگشتی ۱۴۲
۶-۱۲- جمع‌بندی ۱۴۴

فصل هفتم- استخراج مؤلفه‌های مستقل گفتار و وارزیابی آنها

۷-۱- مقدمه ۱۴۶
۷-۲- استخراج مؤلفه‌های مستقل خطی و ارزیابی آنها ۱۴۷
۷-۲-۱- پیش‌پردازش در *ICA* خطی ۱۴۸
۷-۲-۲- استفاده از شبکه *PCA* غیرخطی ۱۴۹
۷-۲-۳- استفاده از الگوریتم *Bell* و *Sejnowski* در شبکه *ICA* خطی ۱۵۱
۷-۲-۴- استفاده از الگوریتم*EASI* (یا *PES*) در شبکه *ICA* خطی ۱۵۲
۷-۳- استخراج مؤلفه‌های مستقل غیرخطی و ارزیابی آنها ۱۵۷
۷-۴- جمع‌بندی ۱۶۱

فصل هشتم- جمع‌بندی، نتایج و پیشنهادات

۸-۱- جمع‌بندی و نتایج ۱۶۲
۸-۲- پیشنهادات برای ادامه کار ۱۷۱

مراجع ۱۷۳

فهرست شکل‌ها

عنوان **صفحه**

شکل ۲-۱- بردارهای ویژه متعامد فضای گفتار ۱۲
شکل۲-۲- واحد فعالیت شبکه، x بردار ورودی n بعدی و شبکه دارای n بردار وزن است ۱۴
شکل۲-۳- ساختار دوبخشی شبکه خودانجمنی، فشرده‌سازی دادگان در سمت چپ و بازسازی دوباره دادگان در سمت راست ۱۵
شکل ۲-۴- ساختار کلی شبکه گلوگاه سه لایه پنهان با تعداد معینی نورون در لایه پنهان ۱۶
شکل۲-۵- ساختار شبکه سلسله مراتبی استخراج مؤلفه‌های اساسی ۲۰
شکل ۳-۱-ترکیب و تفکیک سیگنال‌ها با *ICA* ۲۵
شکل۳-۲- شماتیک یک لایه از *PCA* غیرخطی ۲۷
شکل۳-۳- دادگان با چگالی یکنواخت، بردارهای مختصات اولیه *x(1)* و *x(2)* بردارهای مختصات مستقل چرخیده شده *y(1)* و *y(2)*، بردارهای *w(1)* و *w(2)* تابع هزینه *J(W)* را می‌نیمم می‌کنند ۳۰
شکل۳-۴- ساختاراصلی شبکه *ICA* ۳۳
شکل۳-۵- ساختار شبکه دولایه جهت سفیدکنندگی دادگان ورودی ۳۴
شکل ۳-۶- بلوک دیاگرام روش *INFOMAX* ۴۲
شکل ۳-۷- بلوک دیاگرام شبکه محاسبه ژاکوبین ۴۵
شکل ۴-۱- مقادیر ویژه مربوط به دادگان تعلیم ۵۲
شکل ۴-۲- مقادیر ویژه مربوط به دادگان تست ۵۲
شکل ۴-۳- بردارهای ویژه در فضای دو بعدی ۵۳
شکل۴-۴- وضعیت پخش شدن دادگان در راستای هر یک از مؤلفه‌ها ۵۳
شکل ۴-۵- کل فضای دادگان و تصویر فضا روی هر یک از دو مؤلفه ۵۴
شکل ۴-۶- ساختار شبکه *PCA* خطی ۵۵
شکل۴-۷- منحنی خطای تعلیم شبکه *PCA* خطی به ازای ۱ نورون در لایه پنهان میانی ۵۵
شکل۴-۸- مسیر سه مؤلفه(ستونهای ۳ و۴ و۵) از پارامتر *LHCB* ورودی (نقاط پخش شده در فضای محورها مختصات) و خروجی تصویرشده روی یک خط به ازای ۱ نورون در لایه پنهان و بکارگیری تابع خطی در شبکه خودانجمنی ۵۶
شکل ۴-۹- ساختار شبکه مشرف به هدف جهت همگرا کردن شبکه *PCA* خطی با دو نورون در لایه گلوگاه .۵۷
شکل۴-۱۰- منحنی خطای شبکه مشرف به هدف ۵۷
شکل۴-۱۱- منحنی خطای شبکه *PCA* خطی به ازای دو نورون در لایه پنهان با وزنهای اولیه مشرف به هدف ۵۸

شکل ۴-۱۲- مسیر سه مؤلفه(ستونهای ۱۳ و۱۴ و۱۵) از پارامتر *LHCB* ورودی و خروجی به ازای ۲ نورون در لایه پنهان و بکارگیری تابع خطی در شبکه خودانجمنی ۵۸
شکل ۴-۱۳- ساختار شبکه *PCA* غیرخطی سه لایه پنهان ۵۹
شکل ۴-۱۴- منحنی خطا به ازای ۱ نورون در لایه پنهان میانی، خطا بعد از ۵۰۰ بار تکرار به ۰,۳۶ رسیده ۶۰
شکل ۴-۱۵- مسیر سه مؤلفه (ستونهای ۵ و۴ و۳) از پارامتر *LHCB* ورودی و خروجی در شبکه *PCA* غیرخطی با ۱ نورون در لایه پنهان ۶۱
شکل ۴-۱۶- در سمت چپ سیستم مختصات دادگان خطی و در سمت راست غیرخطی است ۶۲
شکل ۴-۱۷- منحنی خطا به ازای ۲ نورون در لایه پنهان میانی ۶۲
شکل ۴-۱۸- مسیر سه مولفه (ستونهای ۱۳ و۱۴و ۱۵) از پارامتر *LHCB* ورودی و خروجی به ازای ۲ نورون در لایه پنهان میانی ۶۳
شکل ۴-۱۹-منحنی خطا و مسیر سه مؤلفه (ستونهای ۳ و۴و ۵) از پارامتر *LHCB* ورودی و خروجی به ازای ۶ نورون در لایه پنهان میانی.در منحنی خطا ، خطا بعد از ۴۰۰ بار تکرار به ۰,۱۲ رسیده است ۶۳
شکل ۴-۲۰- منحنی خطا و مسیر سه مؤلفه (ستونهای ۳ و۴و ۵) از پارامتر *LHCB* ورودی و خروجی به ازای ۲۴ نورون در لایه پنهان میانی شبکه *PCA* غیرخطی. خطا بعد از ۸۹۴ بار تکرار به ۰,۰۰۹۹ رسیده است ۶۴
شکل ۴-۲۱- فرآیند استخراج مؤلفه‌های اساسی گفتار ۶۵
شکل ۴-۲۲- ساختار شبکه خودانجمنی جهت استخراج مؤلفه‌ها از روی مقادیر طیف سیگنال ۶۶
شکل ۴-۲۳-ساختار شبکه مشرف به هدف جهت همگرا کردن شبکه استخراج مؤلفه‌ها از مقادیر طیف ۶۶
شکل ۴- ۲۴- منحنی خطای شبکه مشرف به هدف ۶۷
شکل ۴- ۲۵- منحنی خطای شبکه خودانجمنی با ورودی پارامترهای طیف به ازای ۲۴ نورون در لایه گلوگاه .۶۷
شکل ۵-۱- بلوک دیاگرام نمایش جایگاه شبکه بازشناس ۷۰
شکل ۵-۲- ساختار شبکه بازشناس آوا ۷۱
شکل۵-۳- منحنی خطای شبکه بازشناس آوا به ازای ۱ مؤلفهٔ اساسی خطی در ورودی آن ۷۲
شکل۵-۴- منحنی خطای شبکه بازشناس آوا به ازای ۲ مؤلفهٔ اساسی خطی در ورودی آن ۷۲
شکل۵-۵- منحنی خطای شبکه بازشناس آوا به ازای ۱ مؤلفهٔ اساسی غیرخطی در ورودی آن ۷۳
شکل۵-۶- منحنی خطای شبکه بازشناس آوا به ازای ۲ مؤلفهٔ اساسی غیرخطی در ورودی آن ۷۳
شکل۵-۷- منحنی خطای شبکه بازشناس آوا به ازای ۶ مؤلفهٔ اساسی غیرخطی در ورودی آن ۷۴
شکل۵-۸- منحنی خطای شبکه بازشناس آوا به ازای ۲۴ مؤلفهٔ اساسی غیرخطی در ورودی آن ۷۴
شکل ۵-۹- ساختار شبکه مرجع بازشناس آوا ۷۵
شکل ۵-۱۰- منحنی خطای شبکه بازشناس مرجع ۷۵
شکل۵-۱۱- مانیفولد غیرخطی گفتار مربوط به یک شخص ۷۸
شکل۵-۱۲- موقعیت دادگان وشکل‌گیری مانیفلد غیرخطی مطلوب ۷۸
شکل ۵-۱۳- بلوک دیاگرام بازسازی سیگنال گفتار از بردارهای بازنمایی فیلتر شده توسط شبکهٔ خودانجمنی ۷۹

شکل ۵-۱۴- سیگنال حوزه زمان اولیه(بالا)- محتوای فرکانسی آن قبل از تبدیل به پارامتر بازنمایی(پایین) ۸۰
شکل ۵-۱۵- سیگنال حوزه زمان به ازای ۶ مؤلفه غیرخطی(بالا)- محتوای فرکانسی آن بعد از تبدیل به پارامترهای بازنمایی ، عبور از شبکه و بازسازی سیگنال (پایین) ۸۱
شکل ۵-۱۶- سیگنال حوزه زمان به ازای ۲ مؤلفه خطی(بالا)- محتوای فرکانسی آن بعد از تبدیل به پارامترهای بازنمایی ، عبور از شبکه و بازسازی سیگنال (پایین) ۸۲
شکل ۵-۱۷- ارتباط بین مقادیر تک نورون لایه پنهان و ویژگی واکدار و بی‌واک بودن برای فریم‌های ۸۰۰-۹۹۵ دادگان تست ۸۳
شکل۵-۱۸- مقادیر تک نورون لایه پنهان به ازای هریک از آواها ۸۶
شکل ۵-۱۹- بردارهای *LHCB* ورودی و بردار *LHCB* ساخته شده از روی میانگین مقادیر تک نورون پنهان .... ۸۹ به ازای هر آوا ۸۹
شکل ۵-۲۰- بردارهای *LHCB* ورودی و خروجی شبکه در یک فضای سه بعدی و به ازای یک مؤلفه اساسی .. ۹۱
شکل ۵-۲۱-بردارهای *LHCB* ورودی و خروجی شبکه در فضای سه بعدی به ازای ۶ مؤلفه اساسی غیرخطی .. ۹۳
شکل۵-۲۲- میانگین مربعات خطای بین پارامترهای *LHCB* ورودی و خروجی و مقدار تک نورون لایه گلوگاه در فریمهای متوالی عبارت " بذل و بخشش " ۹۶
شکل۵-۲۳- میانگین مربعات خطای بین پارامترهای *LHCB* ورودی و خروجی و مقدار تک نورون لایه گلوگاه در فریمهای متوالی عبارت " این چه سبک " ۹۸
شکل۵-۲۴- میانگین مربعات خطای بین پارامترهای *LHCB* ورودی و خروجی و مقادیر ۶ نورون لایه گلوگاه در فریمهای متوالی عبارت " بذل و بخشش " ۱۰۱
شکل ۵-۲۵- فضای *LHCB* خروجی حاصل از نگاشت مؤلفه اول *PCA* (سمت چپ) ۱۰۶
و نگاشت مؤلفه دوم *PCA*(سمت راست) به ازای دو مؤلفه در لایه پنهان میانی ۱۰۶
شکل ۵-۲۶- نگاشت کلی حاصل از هر دو نورون در لایه پنهان میانی شبکه *PCA* غیرخطی ۱۰۶
شکل ۵-۲۷- سیگنال بازسازی‌شده در حوزه زمان جملات تست به ازای نگاشت مؤلفه اساسی غیرخطی اول ..۱۰۷
شکل ۵-۲۸- سیگنال بازسازی‌شده در حوزه زمان جملات تست به ازای نگاشت مؤلفه اساسی غیرخطی دوم ..۱۰۷
شکل ۵-۲۹- سیگنال بازسازی‌شده در حوزه زمان جملات تست به ازای هر دو مؤلفه اساسی غیرخطی ۱۰۷
شکل ۵-۲۹- فضای *LHCB* خروجی حاصل از نگاشت هر یک از مؤلفه‌ها به ازای ۶ مؤلفه در لایه پنهان ۱۰۸
شکل ۵-۳۰- سیگنال بازسازی‌شده در حوزه زمان جملات تست به ازای هر یک از ۶ مؤلفه اساسی غیرخطی بدون اثر مؤلفه‌های دیگر ۱۰۹

شکل ۶-۱- مقادیر ویژه مربوط به دادگان تعلیم ۱۱۶
شکل ۶-۲- بزرگنمایی مربوط به مقادیر ویژه بزرگتر ۱۱۶
شکل ۶-۳- ساختار استخراج مؤلفه‌های اساسی در حالت دینامیک ۱۱۸
شکل ۶-۴- منحنی خطا به ازای ۱ نورون در لایه پنهان میانی ۱۱۸

شکل ۶-۵- منحنی خطا به ازای ۲ نورون در لایه پنهان میانی ۱۱۹
شکل ۶-۶- منحنی خطا به ازای ۶ نورون در لایه پنهان میانی ۱۱۹
شکل ۶-۷- منحنی خطا به ازای ۱۲ نورون در لایه پنهان میانی ۱۱۹
شکل ۶-۸- منحنی خطا به ازای ۲۴ نورون در لایه پنهان میانی ۱۲۰
شکل ۶-۸- منحنی خطا به ازای ۶۴ نورون در لایه پنهان میانی ۱۲۰
شکل ۶-۹- ساختار شبکه بازشناس آوا در حالت دینامیک ۱۲۲
شکل ۶-۱۰- (سمت چپ)منحنی خطای شبکه بازشناس آوا به ازای ۱ مؤلفهٔ اساسی غیرخطی (سمت راست) منحنی خطای شبکه بازشناس آوا به ازای ۲ مؤلفهٔ اساسی غیرخطی ۱۲۲
شکل ۶-۱۱- (سمت چپ)منحنی خطای شبکه بازشناس آوا به ازای ۶ مؤلفهٔ اساسی غیرخطی (سمت راست) منحنی خطای شبکه بازشناس آوا به ازای ۱۲ مؤلفهٔ اساسی غیرخطی ۱۲۳
شکل ۶-۱۲- (سمت چپ) منحنی خطای شبکه بازشناس آوا به ازای ۲۴ مؤلفهٔ اساسی غیرخطی (سمت راست) منحنی خطای شبکه بازشناس آوا به ازای ۶۴ مؤلفهٔ اساسی غیرخطی ۱۲۳
شکل ۶-۱۳- ساختار شبکه مرجع در حالت دینامبک ۱۲۴
شکل ۶-۱۴- منحنی خطای شبکه مرجع دینامیک بازشناس آوا ۱۲۴
شکل ۶-۱۵- سیگنال حوزه زمان به ازای ۱ مؤلفه غیرخطی و ورودی دینامیک ۱۲۷
شکل ۶-۱۶- سیگنال حوزه زمان به ازای ۲۴ مؤلفه غیرخطی و ورودی دینامیک ۱۲۷
شکل ۶-۱۷- سطوح و پرش‌ها در منحنی خطای شبکه *PCA* غیرخطی به ازای۲مؤلفه اساسی حالت دینامیک ۱۲۸
شکل ۶-۱۸- خروجی بازسازی‌شده سطح ۱ منحنی خطای دو مؤلفه اساسی ۱۲۸
شکل ۶-۱۹- خروجی بازسازی‌شده سطح ۲ منحنی خطای دو مؤلفه اساسی ۱۲۹
شکل ۶-۲۰- خروجی بازسازی‌شده سطح ۳ منحنی خطای دو مؤلفه اساسی ۱۲۹
شکل۶-۲۱- میانگین مربعات خطای بین پارامترهای *LHCB* ورودی و خروجی و مقدار تک نورون لایه گلوگاه در فریمهای متوالی عبارت " این چه سبک " در حالت دینامیک ۱۳۱
شکل ۶-۲۲- سیگنال گفتار دادگان تست تمیز ۱۳۳
شکل ۶-۲۳- سیگنال تست جمع‌شده با نویز ایستان با نسبت سیگنال به نویز *0dB* ۱۳۳
شکل ۶-۲۴- سیگنال عبور داده شده از شبکه ۱۳۴
شکل ۶-۲۵- (بالا) سیگنال تست جمع‌شده با نویز غیرایستان با نسبت سیگنال به نویز *0dB*- (پایین) سیگنال بازسازی شدهٔ خروجی شبکه ۱۳۴
شکل ۶-۲۷- ساختار شبکه نگاشت از فضای *LHCB* به آوا و بالعکس ۱۳۸
شکل ۶-۲۸- نمودار کاهش خطای بازشناس آوا ۱۳۸
شکل ۶-۲۹- نمودار کاهش خطای خروجی تخمین سیگنال ۱۳۸
شکل ۶-۳۰- ساختار شبکه سلسله مراتبی جهت استخراج مؤلفه‌های اساسی دینامیک گفتار ۱۴۰
شکل۶-۳۱- منحنی‌های خطای مربوط به هر یک از زیرشبکه‌های شبکه سلسله مراتبی با افزایش .....

نورون‌های لایه پنهان میانی ۱۴۱
شکل ۶-۳۲-ساختار شبکه *NLPCA* سه لایه پنهان با اتصال بازگشتی در لایه پنهان میانی ۱۴۲
شکل ۶-۳۳- مقادیر تک نورون در لایه پنهان بدون دور زدن و بعد از دور زدن در لایه پنهان .....
برای فریم‌های متوالی آوای /*a*/ در عبارت " بخشش " ۱۴۳
شکل ۶-۳۴ - مقادیر تک نورون در لایه پنهان بدون دور زدن و بعد از دور زدن در لایه پنهان .....
برای فریم‌های متوالی کل عبارت " بخشش " ۱۴۳

شکل ۷-۱- ساختار شبکه *PCA* غیرخطی جهت استخراج مؤلفه‌های مستقل ۱۴۹
شکل ۷-۲- منحنی خطای شبکه به ازای ۱۸ نورون در لایه وسط، خطا بعد از ۶۰ بار تکرار به ۰/۰۷۵ رسیده است ۱۵۰
شکل ۷-۳- منحنی خطای شبکه به ازای ۱۸ پارامترهای مستقل در ورودی آن، خطا بعد از ۶۰۰۰ بار تکرار به ۰/۰۵۹ رسیده است ۱۵۰
شکل ۷-۴- منحنی خطای شبکه به ازای ۶ پارامتر مستقل الگوریتم *Bell* و *Sejnowski* در ورودی آن، خطا بعد از ۵۵۰۰ بار تکرار به ۰/۰۵۹ رسیده است ۱۵۲
شکل ۷-۵- منحنی خطای شبکه به ازای ۶ پارامتر مستقل الگوریتم*EASI* در ورودی آن، خطا بعد از ۶۰۰۰ بار تکرار به ۰/۰۶۵ رسیده است ۱۵۳
شکل ۷-۶- ساختار شبکه *ICA* خطی ۱۵۴
شکل ۷-۸- بلوک دیاگرام روش*MISEP* ۱۵۸
شکل ۷-۹- ساختار لایه‌های روش *MISEP* ۱۵۹
شکل ۷-۱۰- وضعیت متغیر اول و متغیر دوم ۱۶۰
شکل ۷-۱۱- وضعیت متغیر سوم و متغیر چهارم ۱۶۱

شکل ۸-۱ - بلوک دیاگرام کلی پیاده‌سازی‌ها جهت استخراج مؤلفه‌های پایه گفتار ۱۶۳

فهرست جدول‌ها

عنوان **صفحه**

جدول ۵-۱- نتایج بازشناسی آوا به ازای مؤلفه‌های *PCA* خطی در ورودی آن ۷۳
جدول ۵-۲- نتایج بازشناسی آوا به ازای مؤلفه‌های *PCA* غیرخطی در ورودی آن ۷۴
جدول ۵-۳- نتایج بازشناسی آوای شبکه مرجع را به ازای دادگان تست ۷۶
جدول۵-۴- کلیه نتایج بازشناسی مربوط به مؤلفه‌های اساسی مختلف ۷۶
جدول ۵-۵- نماد و شماره معادل آواها ۸۵
جدول۵-۶- میانگین مربعات خطای بین پارامترهای *LHCB* ورودی و خروجی و مقدار تک نورون لایه گلوگاه در فریمهای متوالی عبارت " بذل و بخشش " ۹۶
جدول۵-۸- میانگین مربعات خطای بین پارامترهای *LHCB* ورودی و خروجی و مقادیر ۶ نورون لایه گلوگاه در فریمهای متوالی عبارت " بذل و بخشش " ۱۰۱
جدول ۵-۹- ماتریس اختلاط به ازای یک نورون در لایه پنهان میانی ۱۱۱
جدول ۵-۱۰- ماتریس اختلاط به ازای دو نورون در لایه پنهان میانی ۱۱۲
جدول ۶-۱- نتایج بازشناسی شبکه مرجع و شبکه بازشناسی به ازای تعداد مؤلفه‌های اساسی مختلف ۱۲۵
جدول۶-۲- میانگین مربعات خطای بین پارامترهای *LHCB* ورودی و خروجی و مقدار تک نورون لایه گلوگاه در فریمهای متوالی عبارت " این چه سبک " در حالت دینامیک ۱۳۱
جدول ۶-۳-نتایج بازشناسی آوا برای دادگان تست گوینده شماره ۱۷۳(گفتار جلسه اول) ۱۳۵
جدول ۶-۴- نتایج بازشناسی آوا برای دادگان تست گوینده شماره ۱۷۳(گفتار جلسه اول) با استفاده از شبکه‌های تعلیم‌یافته دادگان تعلیم گوینده ۲۹۴ (گفتار جلسه دوم) ۱۳۶
جدول۷-۱- مقایسه نتایج بازشناسی مربوط به شبکه مرجع و ۱۸ پارامتر مستقل خطی ۱۵۱
جدول۷-۲- مقایسه نتایج بازشناسی مربوط به شبکه مرجع و مؤلفه‌های مستقل خطی حاصل از ۱۵۴
الگوریتم‌های *Bell*، *Sejnowski* و *EASI* ۱۵۴
جدول۷-۳- اختلاف پارامترهای *LHCB* ورودی و خروجی و مقدار تک نورون شبکه *ICA* خطی در فریمهای متوالی عبارت " بذل و بخشش " ۱۵۵
جدول۸-۱- کلیه نتایج بازشناسی مربوط به مؤلفه‌های اساسی مختلف ۱۶۵
جدول ۸-۲- نتایج بازشناسی شبکه مرجع و شبکه بازشناسی به ازای تعداد مؤلفه‌های اساسی مختلف ۱۶۸
جدول ۸-۳- نتایج بازشناسی آوا برای دادگان تست گوینده شماره ۱۷۳(گفتار جلسه اول) با استفاده از شبکه‌های تعلیم‌یافته دادگان تعلیم گوینده ۲۹۴ (گفتار جلسه دوم) ۱۶۹
جدول۸-۴- مقایسه نتایج بازشناسی مربوط به شبکه مرجع و مؤلفه‌های مستقل خطی حاصل از ۱۷۰
الگوریتم‌های Bell، Sejnowski و EASI..... ۱۷۰

فصل اول

مقدمه‌ای بر واحدهای پایه در بازشناسی گفتار

۱-۱- مقدمه

در دنیای امروز که ارتباط با کامپیوتر در حال رشد روزافزون است ، ارتقاء سطح ارتباط کلامی انسان با کامپیوتر از اهمیت خاصی برخوردار است. در این میان بهبود سیستم‌های بازشناسی گفتار در سراسر دنیا در حال تحقیق و بررسی است و هر روز سیستم‌های نوینی همراه با ایده‌های جدید به این عرصه وارد می‌شود. اصلی‌ترین راهنمای محققان در امر تکامل سیستم‌های بازشناسی گفتار، مدل کردن دریافت و ادراک توسط مغز است. ازاین لحاظ که مغز ، سیگنال گفتار را علی‌رغم آنکه تحت تأثیر عوامل مختلف تغییر یافته باشد، به خوبی تشخیص می‌دهد.

تحقیقات در زمینهٔ عملکرد مغز انسان در ادراک گفتار از جنبه‌های مختلف مورد بررسی قرار گرفته است. یکی از موارد قابل توجه در این زمینه واحدهای پایهٔ صوتی است که بعنوان ورودی‌های گفتار مورد پردازش و ادراک قرار می گیرند.با توجه به آنکه مغز انسان توانایی بسیار مطلوبی در بازشناخت گفتار[1] دارد، به نظر می رسد داشتن نگرش زیستی به سیگنال گفتار و شناسایی نحوهٔ ادراک گفتار در انسان و بهره‌جویی از آن در مدلسازی سیستم‌های مصنوعی بازشناخت گفتار، بتواند در بالا بردن میزان صحت بازشناسی در سیستم‌های مصنوعی کمک شایانی بکند[۱].

مدل آماری یا مدل مخفی مارکوف و همچنین بسیاری از مدلهای شبکه عصبی که برای بازشناسی واج بکار می‌روند ، غالباً متکی به تقطیع سیگنال می‌باشند. دقت در نتایج عملکرد این مدلها نشان می‌دهد که در این روشها به دلیل نحوهٔ برچسب دهی در نواحی مرزی آواها ، اطلاعات صوتی موجود و نهفته در این نواحی عمدتاً مورد کم‌توجهی قرار می‌گیرد.

[1] Speech recognition

در روشهای فوق برای بازشناسی واجها، نواحی با مشخصات طیفی سریعاً گذرای سیگنال گفتار در تقطیع و یا برچسب‌دهی در مرز بین واجها قرار می‌گیرند و سبب ایجاد خطاهای متعددی نظیر خطا در تشخیص محل تقطیع و یا برچسب‌دهی توسط انسان و نیز خطای در انطباق زمانی مدل بازشناسی با سیگنال صوتی در هنگام تعلیم و بازشناسی می‌شوند.این خطاها ایجاد تنوعات و چندگونگی‌های زائدی می‌نماید که موجب اتلاف و ریزش اطلاعات صوتی مفید موجود در سیگنال گفتار در هنگام بازشناسی می‌گردد. برای رفع این مشکل لازم است که روشهای رایج بر برچسب‌دهی خروجی مدل‌های بازشناسی واحدهای صوتی را تغییر داده و در خود واحدهای صوتی پایه نیز اصلاحاتی بعمل آوریم.[۱]

نتایج مجموعهٔ تحقیقاتی که تاکنون در این زمینه انجام شده است، نشان می‌دهد که ذهن انسان نیز به وقایع خاصی در سیگنال دریافتی حساسیت بیشتری نشان می دهد و اینگونه نیست که توجه به اطلاعات ورودی در همهٔ لحظه ها یکسان باشد[۲] . بر این مبنا نواحی مهم سیگنال گفتار که بعنوان واحدهای پایهٔ صوتی، ورودی سیستم پردازشی مغز محسوب می‌شوند، از دیدگاههای مختلف مورد بررسی قرار گرفته‌اند. در بخش‌های بعد چندین مورد از این دیدگاهها به همراه تحقیقاتی که در آن زمینه انجام شده است مورد بحث و بررسی قرار می‌گیرد.

۱-۲- دیدگاه زبان‌شناسی در مورد واحدهای پایه گفتار

جهت انتخاب واحدهای پایه بهینه می‌توان از راهکارهایی بر مبنای اطلاعات زبان‌شناختی و زیستی بهره گرفت. در این زمینه برخی از زبان‌شناسان به این مطلب توجه داشته و بیان کرده‌اند که سیگنال گفتار حاوی اطلاعات مهم ، نشانه ها[1] و نقاط برجسته‌ای[2] است که از دید سیستم پردازشی مغز بعنوان ورودی دریافت می‌شوند. استیونس، محققی است که از دید زبان‌شناسی ادعا کرده است که محل تولید واج‌ها به نحو منحصر بفردی می‌تواند بعنوان یک مشخصه در نظر گرفته شود. این زبان‌شناس صاحب‌نام در این زمینه برای سیگنال گفتار مدلی ارائه کرده است که در آن به جای بیان یک کلمه بر اساس واج یا آوا ، هر کلمه بصورت رشته‌ای از مشخصه‌ها که بیانگر محل و نحوه تولید واج هستند نشان داده می‌شود.

فرض می‌شود که کلمات در حافظه بصورت دنباله قطعه‌های آوایی ذخیره شده‌اند که هر کدام از این قطعه‌ها بوسیلهٔ یک مشخصه متمایز بیان می‌شوند. این مشخصه‌ها به دو نوع تقسیم می‌شوند: مشخصهٔ نواحی آزاد[3] که با یک قطعه مانند واکه[4] یا همخوان[5] مشخص می‌شوند و مشخصهٔ نواحی محدود[6] که

[1] cues
[2] landmark
[3] Articulator-free features
[4] vowels
[5] consonant
[6] Articulator-bound features