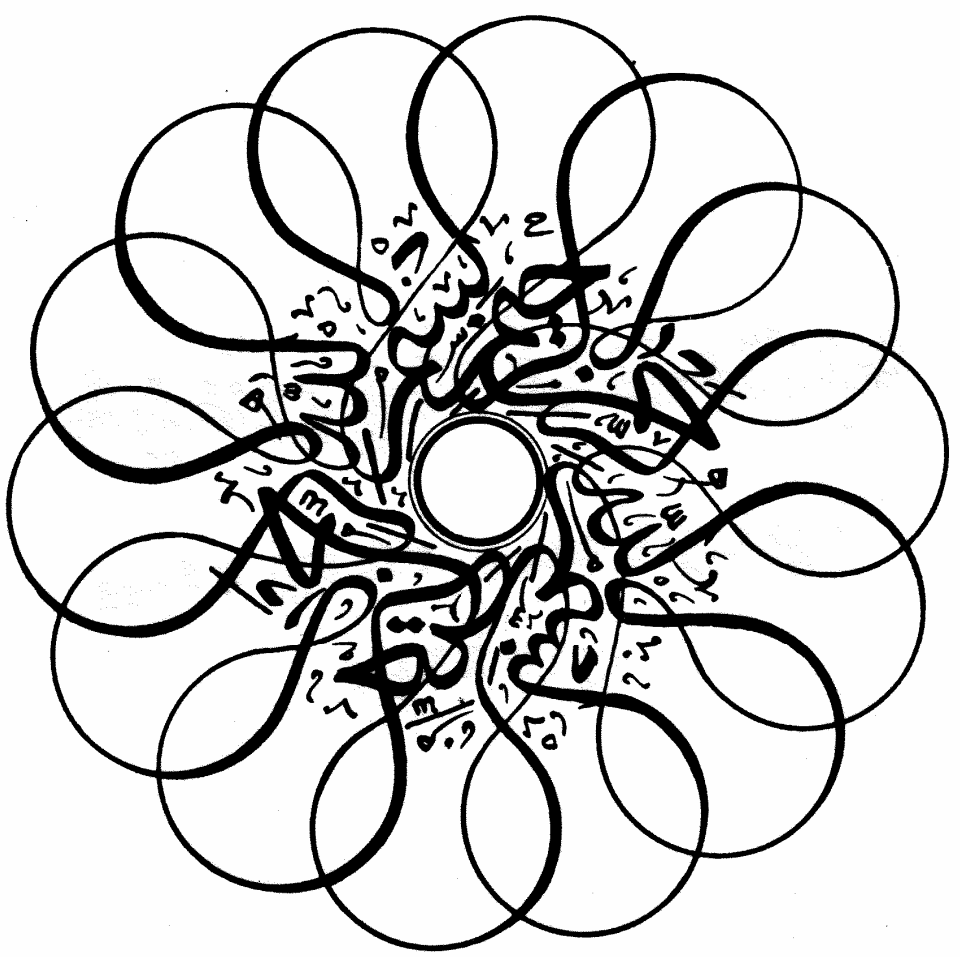

دانشگاه صنعتی امیرکبیر
(پلی‌تکنیک تهران)

دانشکده مهندسی برق
گروه مخابرات - سیستم

پایان نامه کارشناسی ارشد

بهبود بازشناسی واجهای مصوت در گفتار پیوسته به کمک روش ترکیب GMM و SVM

نگارش:
محمد نظری

استاد راهنما:
دکتر ابوالقاسم صیادیان

خرداد ماه ۱۳۸۷

سپاس و ستایش خدایی را که آفریننده قلم است و بدان سوگند یاد می کند
و هیچ کاری بدون لطف و اراده او غایت و سرانجامی نخواهد داشت.

نویسنده این اثر بر خود لازم می داند
تا از زحمـــات اساتیدی که بی شک بدون کمک و راهنمائی شان انجام این کار تحقیقی ممکن نبود،
خصوصاً استاد راهنمای محترم این پروژه جناب آقای دکتر صیادیان که با صبر و بردباری بسیار از
هیچ کمکی کوتاهی نکردند و همچنین جناب آقای دکتر احدی و جناب آقای دکتر همایون‌پور که در
ارزیابی و بهبود این اثر تاثیر فراوان داشته‌اند، تشکر نماید.

چکیده

با پیشرفت سریع در کاربردهای پردازش صوت اعم از سیستمهای بازشناسی گفتار زمان واقعی، سیستمهای بازشناسی گوینده، تطبیق گوینده و ... اهمیت وجود یک سیستم بازشناسی واجهای مصوت (واکه ها) هر چه بیشتر هویدا می‌شود. این اهمیت تا آنجاست که بخشی از تحقیقات امروز دانشمندان در زمینه پردازش صوت بر این موضوع متمرکز شده است. ما در این پایان نامه یک سیستم بازشناسی واکه‌ها را پیشنهاد کرده و افزایش قدرت بازشناسی آن را نسبت به سیستمهای موجود نشان داده‌ایم.

در این تحقیق، توجه عمده ما بر روی مدلسازی آکوستیکی و آماری واکه‌ها و در حقیقت ترکیب یک مدل مولد با یک مدل تمایزی برای کاهش خطای بازشناسی و تعیین مرزهای دقیقتر واکه‌ها معطوف شده است. در سیستم پیشنهادی جهت بازشناسی واکه‌ها، از تلفیق مدل مخلوط گوسی(GMM) با پنجره نرم و پارامترهای آکوستیکی مانند انرژی میان گذر با مدل تمایزی ماشین بردار پشتیبان احتمالی (PSVM) بهره گرفته‌ایم.

این سیستم دارای دو بخش اصلی شامل آشکارسازی و طبقه‌بندی واکه‌ها می‌باشد. مهمترین دستآورد این تحقیق در قسمت آشکارسازی حل مساله قدیمی مقدار آستانه برای مرزهای واکه می‌باشد. در سیستم پیشنهادی مرز واکه به صورت هوشمندانه و بسیار دقیق نسبت به داده‌های آموزشی اولیه انتخاب می‌شوند و سعی شده است سیستم تا حد ممکن نسبت به نویز محیطی مقاومت داشته باشد. در حقیقت در این گام از یک طبقه‌بندی کننده ترکیبی برای جداسازی واکه از غیر واکه استفاده می‌شود. در گام طبقه‌بندی نیز از روشی همانند قسمت آشکارسازی کمک می‌گیریم با این تفاوت که طبقه‌بندی‌کننده قبلی را به حالت چند کلاسی تعمیم می‌دهیم.

قابل ذکر است که برای افزایش سرعت در آشکار سازی و همچنین تخمین اولیه توان نویز محیطی از یک آشکارساز فعالیت گفتار(VAD) ساده استفاده شده است. در انتهای پایان نامه نتایج با الگوریتمهای دیگر که از تلفیق پارامترهای آکوستیکی و مدلهای آماری بر روی پایگاه داده موجود پیاده سازی شده اند، مقایسه گردیده و نتایج مقایسه نشان دهنده کارایی بالاتر نسبت به سایر روشهای موجود می‌باشد.

کلمات کلیدی:

پردازش صوت، بازشناسی واج مصوت، گفتار پیوسته، مدل مخلوط گوسی(GMM)، ماشین بردار پشتیبان احتمالی (PSVM).

فهرست مطالب

لیست واژه‌های انگلیسی

Basic-Level ..... سطح اولیه

Bayes ..... بیز

Discriminant Function ..... تابع تمایز

Empirical ..... تجربی

Entry-Level ..... سطح ورودی

Vowel Detection ..... تشخیص واکه

Generalization ..... قابلیت عمومیت دادن

Hyper-Plane ..... ابر صفحه

Kernel ..... توابع پایه

Linear Projection ..... تصویرکردن خطی

Loss Function ..... تابع تلف

Observation Vectors ..... بردارهای مشاهده

Reject ..... رد

Representation ..... ارائه

Structural Risk Minimization ..... کمینه‌سازی ساختاری خطا

Supervised ..... تحت نظارت

Support Vector Machines (SVM) ..... ماشین بردار پشتیبان

لیست واژه‌های اختصاری

Maximum A posteriori Probability.....MAP

Gaussian mixture models .....GMM

Support Vector Machines.....SVM

Probability Support Vector Machines .....PSVM

Principle Component Analysis.....PCA

Linear Discriminant Analysis.....LDA

Kullback-Leibler.....KL

Maximum Likelihood.....ML

Radial Basis Function.....RBF

Least Squares.....LS

Receiver Operator Characteristics.....ROC

پیشگفتار

زبان از اجزای بسیار زیادی تشکیل شده است که کوچکترین جزء یا واحد آن "واج" است. همه زبانهای بشری به واج تجزیه می شوند؛ بنابراین در نخستین مرحله لازم است واجها با هم ترکیب شوند تا واحدهای بزرگتری مانند تکواژ و یا واژه ساخته شوند.

در مقوله پردازش گفتار، شناسایی واجهای مصوت (واکه‌ها[1]) به دلیل خصوصیات ویژه آنها در گفتار پیوسته، نقش کلیدی بازی می کنند. برای مثال می‌توان به کاربردهای بازشناسایی واجهای مصوت برای استفاده در سیستمهای بازشناسایی خودکار گفتار (ASR[2]) و شناسایی گوینده از روی گفتار اشاره کرد. این امر باعث گردیده تا طیف وسیعی از مطالعات امروز دانشمندان این رشته به روشهای جدید برای بهبود روشهای بازشناسی واجهای مصوت در گفتار پیوسته معطوف گردد [۲ و ۳].

گفتار واکدار نتیجهٔ تحریک لولهٔ صوتی توسط جریان پریودیک هوا در دهانهٔ حنجره می‌باشد. واکه ها از مهمترین گروه اصوات می‌باشند که در همهٔ زبانها موجود و دارای اهمیت زیادی در بازشناسی گفتار می‌باشند. این دسته اصوات پایدارترین گروه از نظر فرکانسی و زمانی می‌باشند ومعمولاً نسبت به همخوانها دارای طول خوبی هستند؛ از این رو راحت تر و دقیق تر بازشناسی می‌شوند. در محیط های با نسبت سیگنال به نویز بالا، انرژی واکه ها، ۱۰dB تا ۲۰dB بالاتر ازمتوسط انرژی سیگنال گفتار است. همچنین بخشهای مصوت گفتار به دلیل دارا بودن ساختار هارمونیکی، نسبت به شرایط نویزی مقاوم تر می باشند و تخمین محل وقوع آنها نسبت به بی واک ها آسان تر می باشد. به همین دلیل بسیاری از سیستمهای بازشناسی برای دستیابی به عملکرد بهتر، بر بازشناسی دقیق واکه ها تأکید دارند. مهمترین ویژگی آکوستیکی که برای تشخیص محل واکه استفاده می‌شود، انرژی می‌باشد؛ زیرا در گفتار، تمامی واکه ها صرف نظر از گوینده، دارای انرژی بالاتری نسبت به همخوانهای[3] مجاور می‌باشند.

---

[1] *Vowels*
[2] *Automatic Speech Recognition*
[3] *Consonant*

در دنیای واقعی هنگامی که واجهای مصوت یکسان توسط افراد مختلف با جنسیتهای مختلف ، سنین مختلف و در شرایط مختلف ادا می شود و یا حتی هنگامی که یک واج مشخص توسط یک فرد خاص در زمانهای مختلف و یا در شرایط مختلف تلفظ می گردد ، با یکدیگر تفاوت چشمگیر دارد. در این صورت باید تشخیص واجهای مصوت به درستی و با دقت زیاد انجام شود . بنابراین در اینجا مسئله برای بازشناسی واجهای مصوت در گفتار پیوسته مطرح می شود. هنگامی ماشین توانایی انجام درست این امر را دارد که توسط یک مدل آموزش پذیر مناسب آموزش دیده باشد. الگوریتمهای مختلف برای بازشناسی واجهای مصوت در گفتار پیوسته وجود دارد که بخش مهمی از الگوریتم های موجود در این زمینه مبتنی بر مدلهای آماری می باشد.

از دیدگاه سیستمی، در یک سیستم بازشناسی الگو، پس از فرایندهای پیش پردازش نمونه های زمانی گفتار، ویژگی های مناسب از نمونه های زمانی استخراج می‌شود. سپس بردار ویژگی های استخراج شده به مدلی عرضه می‌شود تا تطبیق آن با مدل بررسی شود. میزان این تطبیق و مقایسه با تطابق با دیگر الگوها، به عنوان معیار بازشناسی شناخته می‌شود. همانند بسیاری از سیگنالهای طبیعی دیگر، الگوهای مورد بازشناسی که دارای معانی یکسان می‌باشند، بسیار متنوع هستند و همچنین مرز دسته الگوهای متفاوتی که سیستم بازشناسی، وظیفهٔ طبقه بندی آنها را بر عهده دارد، به خوبی روشن نیست. اگرچه استخراج ویژگی های مناسب می‌تواند پیچیدگی این مسأله را کاهش دهد، اما در حال حاضر اکثر سیستمهای بازشناسی گفتار بر پایهٔ مدلسازی آماری هر الگو بنا شده اند؛ به طوریکه پارامترهای مدل، در حین آموزش (مشاهده نمونه های مناسب از الگوی مورد نظر) تخمین زده می‌شوند.

در این پایان نامه، هدف ارایه روش جدیدی برای بهبود بازشناسی واجهای مصوت در گفتار پیوسته به کمک روشهای ترکیبی GMM و SVM می باشد. در اینجا مساله بازشناسی واجهای مصوت در گفتار پیوسته مورد بررسی قرار می دهیم و از یک مدل آماری برای حل این مساله استفاده می کنیم و سپس به بهبود روشهای موجود می پردازیم.

ابزار کلی در حل مساله استفاده از دو گام اصلی می باشد. در ابتدا، گام اول که شامل یک گام آموزشی بر روی نمونه های گرفته شده است، انجام می پذیرد. این گام آموزشی بر روی یک مدل ترکیبی مدل مخلوط

گوسی (GMM) و ماشین بردار پشتیبان(SVM) اجرا می گردد. در گام دوم که گام بازشناسی واجهای مصوت در گفتار پیوسته می‌باشد ، ورودی اعمال می شود تا واجهای مصوت مورد نظر شناسایی شود. در این پروژه ضمن پیاده سازی روشهای مختلف سعی می کنیم با استفاده ازروشی ترکیبی که مورد طراحی قرار می گیرد عمل بازشناسی واجهای مصوت در گفتار پیوسته را بهبود دهیم.

این سیستم دارای دو بخش اصلی شامل آشکارسازی و طبقه‌بندی واکه‌ها می‌باشد. مهمترین دستآورد این تحقیق در قسمت آشکارسازی حل مساله قدیمی مقدار آستانه برای مرزهای واکه می‌باشد. در سیستم پیشنهادی مرز واکه به صورت هوشمندانه و بسیار دقیق نسبت به داده‌های آموزشی اولیه انتخاب می‌شوند و سعی شده است سیستم تا حد ممکن نسبت به نویز محیطی مقاومت داشته باشد. در حقیقت در این گام از یک طبقه‌بندی کننده ترکیبی برای جداسازی واکه از غیر واکه استفاده می‌شود. در گام طبقه‌بندی نیز از روشی همانند قسمت آشکارسازی کمک می‌گیریم با این تفاوت که طبقه‌بندی‌کننده قبلی را به حالت چند کلاسی تعمیم می‌دهیم.

به این منظور، در فصل اول به بیان ویژگی های یک سیستم بازشناسی الگو در کاربردهای پردازش گفتار پرداخته و در ادامه، مدلهای معمول جهت بازشناسی، نقاط ضعف و قوت آنها را مورد بررسی قرار داده ایم. فصل دوم به بررسی مدل ماشین بردار پشتیبان (SVM) اختصاص دارد. در این فصل ایدهٔ اصلی، مدلسازی تحلیلی و الگوریتمهای آموزش و بازشناسی این مدل بررسی می‌شوند و در خاتمه با تعمیم آن، مدل ماشین بردار پشتیبان احتمالی (PSVM) را معرفی می‌کنیم. فصل سوم به بررسی انواع مدلهای ترکیب مدل مخلوط گوسی (GMM) و مدل ماشین بردار پشتیبان(SVM) اختصاص دارد. در فصل چهارم الگوریتمهای پیشنهادی و نتیجهٔ پیاده سازی آنها آورده شده است. در فصل پنجم، با نتیجه گیری نهایی و پیشنهاداتی برای ادامهٔ کار، به مطالب پایان نامه خاتمه می‌دهیم.

فصل اول

الگوهای گفتاری و روشهای بازشناسی آنها

۱-۱- مقدمه

شنوایی، به عنوان یکی از حسهای اصلی انسان، نقش بسیار مهمی در به دست آوردن اطلاعات محیطی بر عهده دارد. گوش انسان با شنیدن گفتار اطلاعات فراوانی از قبیل محتوای گفتار، جنس، سن، وضعیت احساسی و حتی هویت فرد گوینده به دست می‌آورد. این توانایی، دانشمندان را بر آن داشته است تا با تحقیقات فراوان سعی در مدل کردن این توانایی بشری با الگوهای موجود نمایند تا هر چه بیشتر بتوانند در مسیر نزدیکی انسان و ماشین قدمی بردارند. این ایده باعث شده است تا دانشمندان علوم اعصاب شناختی، علوم ریاضی و علوم کامپیوتر دست در دست هم به سمت این مدل‌سازی ها حرکت کنند.

اولین و مهمترین زمینه در شناسایی الگوهای گفتاری، بازشناسی خودکار گفتار (ASR[1]) می‌باشد. بازشناسی خودکار گفتار به عنوان یکی از آرزوهای پژوهشگران سیستمهای پردازشی، هم اکنون به یکی از شاخه های اصلی پردازش سیگنال مبدل شده است. سیستمهای مترجم خودکار، سیستمهای محاورهٔ ماشینی، منشی های ماشینی و رابط های گفتاری، کاربردهای مختلف در این زمینه می‌باشند. اگرچه بسیاری از سیستمهای بازشناسی گفتار به موفقیتهایی دست یافته اند ولی همواره فاصلهٔ انسان و بهترین روشهای بازشناسی گفتار حفظ شده است. به طوریکه در عمل، نه تنها سیستمهای کاربردی نتوانسته اند در این زمینه از توانائیهای انسان پیشی بگیرند، بلکه هیچگاه جانشین مطلوبی نیز محسوب نشده اند. از این رو گسترهٔ وسیعی از پژوهش برای بهبود کارائی سیستمهای ماشینی بازشناسی گفتار گشوده شده است.

در سیستمهای گفتاری اعم از بازشناسی و یا سنتز، گفتار به صورت رشته ای از اجزای تشکیل دهنده اش مورد بررسی قرار می گیرد. این اجزای تشکیل دهنده می توانند جملات، عبارات، کلمات، هجاها و یا واجها باشند. برای مثال در بازشناسی لازم است میزان شباهت بین سیگنال ورودی و گفتار (یا مدلهای گفتار) موجود در لغت نامه سنجیده شود. این مقایسه می تواند در سطح هرکدام از اجزای تشکیل دهنده گفتار انجام شود و یا در سنتز گفتار، باید با استفاده از این اجزای تشکیل دهنده بیان مورد نظر ساخته شود.

---

[1] *Automatic Speech Recognition*

در این فصل ابتدا به صورت مختصر به الگوهای گفتاری و مخصوصاً واحدهای زبانی می‌پردازیم سپس خواص واکه ها و مزایای بازشناسی دقیق آنها را مورد بررسی قرار می‌دهیم. آنگاه ویژگیهای یک سیستم بازشناسی الگوی گفتار را بررسی می‌کنیم. سپس در مورد مدلهای مختلف و خصوصاً آماری به کار رفته برای بازشناسی الگوی گفتار توضیحاتی داده شده و نقاط قوت و ضعف این مدلها مورد بررسی قرار خواهد گرفت.

۱-۲- الگوها و واحدهای گفتاری

یک سیستم بازشناسی الگوی گفتاری متشکل از اجزای زیادی است که به صورت متوالی به کار رفته‌اند و دقت خروجی سیستم به کارایی کلیه اجزای آن بستگی دارد. اما حساس‌ترین بخش این سیستم، استفاده از مدلی است که تنوع بردارهای ویژگی استخراج شده را تحمل می‌کند، به واحدهای زبانی متفاوت، امتیاز تطابق منتسب می‌کند و آنها را از لحاظ تطابق با ورودی رتبه بندی می‌کند.

پیچیدگی این مدل‌سازی به توانایی بخش‌های دیگر سیستم وابسته است. به عنوان مثال، در صورتی که ویژگی‌های استخراج شده از گفتار آنقدر تفکیک‌کننده باشد که هر دنباله بردارهای ویژگی مستقیماً به یک واحد زبانی اشاره کند، نیازی به این مدل‌سازی نیست. همچنین در صورتی که مجموعه هدف خروجی آنقدر کوچک باشد که اشتباه در بازشناسی اکوستیکی با مدل زبانی جبران شود، مدل‌سازی اکوستیکی نقش حساس خود را از دست می‌دهد. اما در عمل، نه ویژگی‌های موجود تفکیک‌پذیری کافی را ارائه می‌دهند و نه کاربردهایی با مجموعه‌ای چنین محدود مورد علاقه هستند. لذا پیچیدگی زیادی به بخش مدل‌سازی آکوستیکی واحدهای زبانی تحمیل می‌گردد. در واقع، مدل‌سازی آکوستیکی، باید آنچه را که توسط بخش‌های دیگر بازنمایی نشده است، به شیوه‌ای آماری مدل کند. این مسایل شامل تنوع در گوینده، محیط گفتار، وابستگی به متن گفتار، تلفظ، لهجه و احساسات بیانی گوینده و حتی تنوع حاصل از دو بیان یک واحد گفتاری توسط یک گوینده می‌شود [۱].

واحدهای زبانی معمول در بازشناسی گفتار شامل آوا، واج، هجا و اجزای آن،(ترکیب‌های نیم هجایی) و کلمه است که از لحاظ آکوستیکی به ترتیب ساده به پیچیده ارائه شده‌اند. آوا کوتاه‌ترین واحد آکوستیکی در

بازشناسی گفتار پیوسته است که شکل لوله صوتی در طول آن تقریباً ثابت می‌ماند و واج کوتاه‌ترین واحد مورد استفاده زبان‌شناسان است.

در بازشناسی گفتار پیوسته، ترکیب‌های مجاز گفتاری بسیار متنوع است و معمولاً از واحدهای زبانی کوتاه‌تر و کم‌تنوع‌تر استفاده می‌شود تا با الفبایی محدود، تمام ترکیب‌های مجاز قابل استخراج و مدل‌سازی باشد. برای استفاده از مزیت مدل‌های مورد استفاده زبان‌شناسان، معمولاً از آواهای متناظر با واج‌های پایه به عنوان واحد آگوستیکی استفاده می‌شود، اگرچه در بعضی پژوهش‌ها برای مدل‌سازی تاثیر متقابل واج‌ها به هم از ترکیب‌های دو آوایی و سه آوایی نیز بهره گرفته شده است. رویکرد اخیر مستلزم مدل‌سازی تعداد زیادی واحد آکوستیکی است که خود باعث پیچیدگی بازشناسی آکوستیکی و نیاز به بانک اطلاعاتی بسیار وسیع برای تخمین پارامترهای سیستم می‌شود. واضح است که بیان چنین حجم پیچیدگی، تنها به صورتی آماری قابل مدل‌سازی است.

در بازشناسی گفتار، انتخاب واحد گفتاری برای سیستم، از اهمیت ویژه ای برخورداراست. برای سیستمهای با تعداد لغات کم، معمولاً از واحد کلمه استفاده می شود. استفاده از کلمه به عنوان واحد گفتار، برای سیستمهای بازشناسی کلمات گسسته وبازشناسی کلمات بدنبال هم به کار می رود. از مزایای انتخاب واحد کلمه به عنوان واحد گفتاراین است که نمایش آکوستیکی کلمات به خوبی معین شده وتغییرات آوایی عمدتاً در ناحیهٔ شروع و انتهای کلمه قرار دارد. مزیت دیگر اینکه، استفاده از واحد کلمه، نیاز به فرهنگ لغت را مرتفع می سازد و بنابراین ساختار بازشناسی ذاتاً ساده تر می‌شود. اما معایب استفاده از مدلهای گفتاری کلمه برای بازشناسی گفتار پیوسته هنگامی نمایان می‌گردد که در فاز آموزش نیاز به انواع حالات و ترکیبات کلمه در جمله به وجود می‌آید، که این امر آموزش در سیستمهای عملی را غیر ممکن می‌سازد.

تعدادی از واحدهای آکوستیکی عبارتند از آوا، دوآوایی، سه آوایی، هجا و نیم هجا، که در ادامه در مورد آنها توضیحاتی داده می‌شود.

- آوا : کوچکترین واحد آکوستیکی است که به دو گروه واکه[1] و همخوان[2] تقسیم می‌شود. بر طبق[۱] زبان فارسی ۸ واکه و ۲۳ همخوان دارد. این ۸ واکه شامل ۶ واکه اصلی و ۲ واکه مرکب است.

[1] Vowel
[2] Consonant

همخوان‌ها شامل دو زیر گروه واکدار[1] و بی واک[2] می‌باشند. در میان زیر کلمه‌ها تعداد آواهای زبان فارسی از سایر واحدها کمتر می‌باشد. با توجه به اثر آواها بر روی یکدیگر، نمی‌توان آنها را مستقل از هم در نظرگرفت و در واقع باید اثرات درون هجایی و برون هجایی آواها بر یکدیگر لحاظ شود. یعنی بسته به کلمه و محل قرار گرفتن آواها، باید مدل آوای مورد نظر را تغییر داد که در این صورت، کارائی سیستم بازشناسی به میزان زیاد افزایش پیدا می‌کند [۱ و ۲].

- دو آوائی[3] : هر آوا با توجه به آوای سمت راست یا سمت چپ به طور جداگانه مدل می‌شود.
- سه آوایی[4] : هر آوا با توجه به آوای سمت راست وسمت چپ، به طور جداگانه مدل می‌شود.
- هجا[5] : واحدهای آوایی- مانند واکه و همخوان- مواد خامی هستند که بر طبق قواعد و الگوهای معین گرد هم آمده و واحدهای پیچیده تر از خود، یعنی هجاها را تشکیل می‌دهند. هجا در فارسی عبارت است از یک رشتهٔ آوایی پیوسته که از یک واکه و یک تا سه همخوان تشکیل یافته است. منظور از رشتهٔ آوایی پیوسته این است که اجزاء سازندهٔ هجا طی یک فرآیند تولیدی و بدون مکث تولید می‌گردند. واکه به منزلهٔ مرکز یا هستهٔ هجا می‌باشد. بدین معنی که موجودیت هجا بستگی به وجود واکه دارد و اگر واکه را حذف کنیم دیگر هجایی باقی نمی ماند. در صورتیکه می توان از یک هجا، یک یا دو همخوان را حذف کرد بدون اینکه به موجودیت آن لطمه ای وارد شود.
- نیم هجا[6] : هر هجا به دو نیم هجای ابتدایی و انتهایی تقسیم بندی می‌شود. نیم هجای ابتدایی شامل همخوان ابتدایی و بخشی از هسته هجا و نیم هجای انتهایی شامل باقیمانده هسته تا انتهای هجا می‌باشد. نیم هجاهای موجود در زبان فارسی به صورت CV ، VC و VCC می‌باشند که تعداد آنها حدود ۷۰۰ است.

ساختار هجاها به این نحو می‌باشد که هر هجا از سه بخش شروع، هسته و انتها تشکیل می شود. هر چند که بسیاری از هجاها دارای هر سه قسمت می باشند، اما بخشهای با اهمیت هجا معمولاًشامل یک یا دو قسمت

---

1 Voiced
2 Unvoiced
3 Biphone
4 Triphone
5 Syllable
6 Demi-Syllable

از بخشهای ذکر شده می باشد. اگر هجایی دارای یک جزء تکی باشد، الزاماً آن جزء هسته می باشد. بعضی از زبانهای دنیا[۵]، از جمله ژاپنی[۶] و چینی[۷]، از ساختار هجایی نسبتاً روشن و شفافی بر خوردارند. زبان فارسی از دسته زبانهایی می باشد که دارای ساختار هجایی ساده ای است. زبان فارسی دارای سه نوع مختلف هجا می باشد:

۱. ساختار همخوان- واکه*(CV)*
۲. ساختار همخوان- واکه- همخوان*(CVC)*
۳. ساختار همخوان- واکه- همخوان- همخوان*(CVCC)*

بررسی ها نشان می دهد که نوع اول بیشترین تعداد وقوع در گفتار فارسی را دارد و پس از آن نوع دوم و سوم به ترتیب بیشترین تعداد وقوع را دارند. نشان داده شده است که تغییرات تلفظی در سطح هجا بسیار منظم تر از سطح قطعات آوایی می باشد [۱].

درسیستمهای بازشناسی مبتنی بر نیم هجا، در صورتیکه ابتدا واکه، که هستهٔ هجا می باشد، آشکار سازی شود، بازشناسی نیم هجای قبل و بعد از واکه با دقت بالاتر و مقاوم تر نسبت به شرایط محیطی انجام می-گردد. در زبانهای مبتنی بر هجا، برای افزایش دقت بازشناسی و تقطیع گفتار به هجاها از آشکارسازی واکه استفاده می‌شود[۸-۳]، در ادامه در مورد خواص واکه ها بیشتر توضیح می‌دهیم.

۱-۳- واکه

گفتار واکدار نتیجهٔ تحریک لولهٔ صوتی توسط جریان پریودیک هوا در دهانهٔ حنجره می‌باشد. واکه ها از مهمترین گروه اصوات می‌باشند که در همهٔ زبانها موجود و دارای اهمیت زیادی در بازشناسی گفتار می-باشند. این دسته اصوات پایدارترین گروه از نظر فرکانسی و زمانی می‌باشند ومعمولاً نسبت به همخوانها دارای طول خوبی هستند؛ از این رو راحت تر و دقیق تر بازشناسی می‌شوند. در محیط های با سیگنال به نویز پایین، انرژی واکه ها، ۱۰dB تا ۲۰dB بالاتر ازمتوسط انرژی سیگنال گفتار است. همچنین بخشهای مصوت گفتار به دلیل دارا بودن ساختار هارمونیکی، نسبت به شرایط نویزی مقاوم تر می باشند و تخمین

محل وقوع آنها نسبت به بی واک ها آسان تر می باشد. به همین دلیل بسیاری از سیستمهای بازشناسی برای دستیابی به عملکرد بهتر، بر بازشناسی دقیق واکه ها تأکید دارند.

جدول ۳-۱ : واکه های زبان فارسی

آوا	علامت	مثال
اَ	@	نم، n@m
آ	a	مات، mat
اِ	e	چک، cek
ای	i	میز، miz
اُ	o	کت، kot
او	u	نور، nur
اُو	ow	لوح، lowh
اِی	ei	شیخ، $eix

در [۹] برای افزایش دقت بازشناسی هجاها در گفتار پیوسته، بر یافتن محل دقیق واکه ها تأکید شده و از ترکیب HMM، شبکه عصبی و انرژی برای این آشکارسازی استفاده شده است. مهمترین ویژگی آکوستیکی که برای تشخیص محل واکه استفاده می‌شود، انرژی می‌باشد؛ زیرا در گفتار، تمامی واکه ها صرف نظر از گوینده، دارای انرژی بالاتری نسبت به همخوانهای مجاور می‌باشند. هشت واکهٔ موجود در زبان فارسی - شش واکهٔ ساده و دو واکهٔ مرکب - در جدول(۳-۱) آورده شده اند.

۱-۴- مسایل اصلی در مدل‌سازی آکوستیکی گفتار

گفتار، پدیده پیچیده‌ای است که علی‌رغم تنوع بسیار در ابعاد گوناگون می‌تواند معنای یکسانی را منتقل کند. اگرچه استخراج ویژگی مناسب می‌تواند این تنوع را کاهش دهد، اما طبیعت گفتار پیچیده‌تر از آن است که با دنباله‌ای از بردارهای ویژگی بازنمایی گردد. بنابراین یک مدل اکوستیکی موفق باید با شناخت طبیعت گفتار آنچنان طراحی گردد که تا حد ممکن به این طبیعت نزدیک شود. در این فصل به مسایلی پرداخته می‌شود که یک مدل‌سازی آکوستیکی موفق باید به شکلی آنها را تحمل کند. مدل‌سازی مسایل زیر که همگی ابعاد گوناگون سیگنال طبیعی گفتار هستند، مستقیماً بر روی نرخ بازشناسی یک سیستم بازشناسی گفتار موثر است.

۱-۴-۱- تنوع بردارهای ویژگی

مهمترین مشخصه بردارهای ویژگی یک گفتار، عدم یکسانی آنها به ازای بیان آنها در شرایط مختلف یا بردارهای لحظات مختلف یک قطعه گفتاری است. تا زمانی که نظم این تنوع به صورت معین قابل مدل‌سازی نباشد، بازنمایی این تنوع به تخمین توزیع چگالی احتمال بردارها می‌انجامد.

۱-۴-۲- دنباله غیر ایستان بردارهای ویژگی

بردارهای ویژگی یک گفتار نه تنها متنوعند، بلکه مشخصات آماری آنها در طول زمان تغییر می‌کند و مدل آکوستیکی باید بتواند این تغییر مشخصات را بازنمایی کند. این بازنمایی معمولاً با قطعه‌بندی دنباله بردارهای ویژگی یک واحد زبانی انجام می‌پذیرد.

۱-۴-۳- اثر کشش زمانی

یک ویژگی بارز سیگنال گفتار، تنوع در طول گفتارهایی است که همگی بیانگر یک واحد زبانی هستند. این کشش زبانی که معمولاً نشانه لهجه، تاکید، یا احساسات گفتاری گوینده است، نباید موجب اشتباه در نتیجه بازشناسی واحدهای زبانی گردد.

اثر کشش زمانی از دیدگاه کمی نیز قابل بررسی است. برای تحمل صحیح کشش زمانی هر قطعه در یک واحد زبانی، باید رفتار آماری کشش زمانی هر قطعه مورد بررسی قرار گیرد. مدل‌سازی مناسب طول هر قطعه آکوستیکی می‌تواند راه‌کار مناسبی برای این منظور باشد.

۱-۴-۴- توالی قطعات

قطعات مدل‌سازی شده در هر واحد گفتاری به طور طبیعی به همان توالی مدل‌سازی شده ظاهر می‌شوند، اما در بیان یک واحد زبانی ممکن است بعضی قطعات اصلاً بیان نشوند که مدل مناسب باید بتواند این پدیده را بازنمایی کند. چنانچه مشاهده خواهد شد، عدم مدل‌سازی این پدیده تاثیر جدی در نرخ بازشناسی ندارد و علاوه بر مدل‌های آماری، از مدل‌های معین نیز برای مدل‌سازی توالی قطعات استفاده می‌شود.