3rd. International Conference on September 2017
Information Technology, Computer and Telecommunications
سومین کنفرانس بین المللی شهریورماه ۱۳۹۶
فناوری اطلاعات، مهندسی کامپیوتر و مخابرات

مروری بر سیستم های تشخیص صوت، مفاهیم و روش ها

[۱]مژده مطالبی ، [۲] اعظم باستان فرد

[۱] دانشکده مهندسی کامپیوتر و فناوری اطلاعات، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، Mzhd_Mti@yahoo.com
[۲] دانشگاه صدا و سیمای جمهوری اسلامی ایران، bastanfard03@gmail.com

چکیده

سیستم های تشخیص صوت سیستم های الکترونیکی هستند که به دستگاه های الکترونیکی اجازه می دهند تا در برابر صدا از خود واکنش نشان دهند. این سیستم ها بسیار کاربردی هستند و اهداف زیادی را دنبال می کنند ولی تا به امروز به همه اهدافشان نرسیده اند و از تمامی پتانسیل هایشان استفاده نشده است. سیستم های تشخیص صوت کلمات گفتار و صداهای موجود را به سیگنال هایی ترجمه می کنند که قابلیت پردازش دارند و می توان آن ها را به وظیفه ای مشخص تبدیل نمود. مباحث مطرح در پردازش گفتار و تشخیص صوت عبارتند از بازشناسی گفتار، تبدیل متن به گفتار، بازشناسی گوینده، رمزگذاری گفتار، بهسازی گفتار، بازشناسی زبان، نمایه گذاری اسناد صوتی

واژه های کلیدی : تشخیص صوت، بازشناسی گفتار

۱- مقدمه

امروزه تعداد زیادی سیستم تشخیص صوت در بازارها وجود دارند. قوی ترین آن ها امکان پردازش و شناسایی هزاران کلمه را دارد. تشخیص صوت در کامپیوتر به معنای توانایی یک سیستم کامپیوتری، برنامه نرم افزاری یا یک سخت افزار در رمزگشایی سیگنال های صوتی به صداهای دیجیتالیست که بتوان آن را توسط کامپیوتر یا سخت افزار تعبیر کرد و مورد پردازش قرار داد. تشخیص صوت معمولی برای انجام یک عملیات در یک دستگاه، انجام دستورات، نوشتن بدون نیاز به کیبورد و موس و انجام فعالیت هایی نظیر آن ها مورد استفاده قرار می گیرد. [4] به طور کلی بیشترین حوزه تشخیص صوت در تشخیص صدای انسان و گفتار است. یک سیستم تشخیص صوت پایه نیاز به ۲۲ مگاهرتز پردازنده، حداقل۴۶مگابایت رم، یک میکروفن پایه و یک کارت صدای حداقل ۶۴ بیتی نیاز دارد. افزایش سایز رم، پردازنده و کارت صدا و افزایش توان میکروفن می تواند در افزایش دقت و کارایی سیستم های تشخیص صوت کمک شایانی نماید . علاوه بر این حداقل نیازمندی های سخت افزاری ،سیستم های تشخیص صوت نیاز به نرم افزار دارند تا بتواند داده ها را جمع آوری، تحلیل و تفسیر نمایند. نرم افزارهای متفاوت از روش های متفاوتی به این اهداف نائل میآیند. مدل های وابسته به صوت و زبان دارای مدل پردازش پایه هستند که در آن ها صدا از میکروفن گرفته می شود و توسط کامپیوتر پردازش می شوند . [6] در مدل های صوتی صدا آنالیز و تحلیل می شوند و بعد از آن که کاربر در میکروفن صحبت می کند، صدای مورد نظر توسط میکروفن گرفته می شود و نویزها و صداهای اضافی موجود در پس زمینه صدا که روی حجم صدا و کیفیت آن تاثیر گذارند، حذف می شوند . از توابع ریاضی برای دریافت صدا و تبدیل آن به رنج و فرکانس مورد نیاز استفاده می شود. سپس داده های بدست آمده تحلیل می شوند و تبدیل به نمایش های دیجیتالی در می آیند . مدل های زبانی محتوای صدا را مورد بررسی قرار می دهند. از این مدل بیشتر با اهداف تشخیص گفتار استفاده می شود. این مدل به مقایسه بین صدای دریافتی و لغت های موجود در دایره المعارف می پردازد که بزرگترین و رایج ترین پایگاه داده ی لغات موجود در زبان انگلیسی اند . اگر چه مدل های صدایی و زبانی مدل های پایه سیستم های تشخیص صوتی اند اما ممکن است هر سیستم مدل خاص خودش را داشته باشد و نحوه طراحی نرم افزار وابسته به دیدگاه برنامه نویس آن دارد که از چه تکنیک هایی در طراحی سیستم استفاده نماید. [7]

۲- تاریخچه ی توسعه سیستم های تشخیص صوت

پیشرفت تکنولوژی تشخیص صوت در ۵ دهه ی گذشته را می توان به شرح زیر سازمان دهی کرد:

نسل اول: اولین تلاش در سال های ۱۹۵۰ و ۱۹۶۰

نسل دوم : فن آوری مبتنی بر قالب در اواخر ۱۹۶۰ و ۱۹۷۰

نسل سوم: فن آوری های مبتنی بر مدل های آماری در ۱۹۸۰

آخرین نسل سوم: پیشرفت در ۱۹۹۰ و ۲۰۰۰ [32]

قدیمی ترین تلاش: اولین تلاش برای ابداع سیستم تشخیص صوت در سال های ۱۹۵۰ و ۱۹۶۰ صورت گرفت. زمانی که محققان مختلف به تلاش برای بهره برداری از ایده های اساسی مرتبط با آواشناسی آکوستیک پرداخته اند . از آنجا که پردازش سیگنال و تکنولوژی های کامپیوتر در آن زمان بسیار بدوی بودند، بسیاری از سیستم های تشخیص گفتار مورد بررسی قرار داده شده از تشدید طیفی استفاده می کردند که از سیگنال های خروجی بانکهای فیلتر و مدار های منطقی استخراج می شد.[8] ,[9] ,[30] ,[32] .

سیستم های اولیه: در سال ۱۹۵۲ و در آزمایشگاه بل در آمریکا محققان برای ایزوله سازی تشخیص رقمی را برای یک سخنران، با استفاده از اندازه گیری فرکانس های فومنت و برای هر یک سیستم ساختند. در یک تلاش که در سال ۱۹۶۱ و در یک آزمایشگاه دیگر در آمریکا صورت گرفت، محققان به تلاش برای به رسمیت شناختن ده سیلاب یا هجای مجزا در یک رشته گفتار پرداختند . با وارد کردن اطلاعات آماری مربوط به توالی واج مجاز در زبان انگلیسی، آنها دقت تشخیص واج را برای کلمات متشکل از دو یا چند واج افزایش داده اند . [8] ,[9] ,[30] ,[32]

این فعالیت اولین استفاده از ترکیب نحوی آماری در تشخیص گفتار بوده است. در سال ۱۹۶۱ از آنجا که هنوز کامپیوتر ها به اندازه کافی سریع نبودند تعدادی سخت افزار خاص با هدف و منظور خاص ساخته شد .این سخت افزار ها وظیفه تشخیص واج را به عهده داشتند. که دارای قابلیت جداسازی سخنرانی و تجزیه و تحلیل صفر را در مناطق مختلف دارا بودند. روند ساخت این سخت افزار ها تا سال ۱۹۶۱ پا برجا بود. نرمال سازی و بهینه سازی زمان از مشکلات مطرح و مهم بود که به علت غیر یکنواختی مقیاس های زمانی در رخداد های کلام وجود داشتند. در سال ۱۹۶۱ در آزمایشگاه RCA محققان موفق به ساخت متدهایی ابتدایی نرمال سازی زمان شدند که مبتنی بر توانایی قابل اعتماد تشخیص ابتدا و انتهای کلام هستند.[8] ,[9] ,[30] ,[32]

تشخیص گفتار پیوسته: در اواخر ۱۹۶۱ تحقیقاتی در حوزه ی تشخیص گفتار پیوسته با استفاده از ردیابی پویای واج انجام پذیرفت در آزمایشگاه IBM محققان تحقیقاتی را در زمینه واژه های وسیع مورد استفاده در تشخیص گفتار را بر روی سه عملیات مجزا بر روی یک پایگاه داده انجام دادند. [8] ,[9] ,[30] ,[32]

تشخیص کلمات متصل: مشکل ایجاد یک سیستم قوی بود تا بتواند رشته ای از کلمات متصل به هم را تشخیص دهدکه تمرکز تحقیقات در سال های ۱۹۸۰ بر روی این موضوعات بود . طیف گسترده ای از الگوریتم های مبتنی بر تطبیق الگو کلمات مجزا هستند که شامل دو مرحله برنامه نویسی پویا هستند[8] ,[9] ,[30] ,[32]

مدل سازی آماری: تحقیقات انجام شده در ۱۹۸۰ با یک تغییر در روش از روش مبتنی بر الگو به سمت مدلسازی آماری شخصیت جدیدی یافت. بسیاری از سیستم های تشخیص صوت امروزی مبتنی بر مدل هاو فریم ورک های آماری هستند که در سال ۱۹۸۰ بنا نهاده و توسعه داده شد. یکی از کلیدی ترین تکنولوژی هایی که در سال ۱۹۸۰ ایجاد شد مدل مخفی مارکوف بود. این مدل یک فرایند مضاعف تصادفی است که در آن یک فرایند تصادفی اساسی است که قابل مشاهده نیست اما می تواند از طریق یکی دیگر از فرایند های تصادفی که دنباله ای از مشاهدات را تولید می کنند، آن را مشاهده نمود. [8] ,[9] ,[30] ,[32]

۳- الگوریتم های تشخیص صوت

علاوه بر این حداقل نیازمندی‌های سخت افزاری، سیستمهای تشخیص صوت نیاز به نرم افزار دارند تا بتواند داده‌ها را جمع آوری، تحلیل، تفسیر نمایند. نرم‌افزارهای متفاوت از روش‌های متفاوتی به این اهداف نائل آیند.

مدل‌های وابسته به صوت و زبان دارای مدل پردازش پایه هستند که در آن‌ها صدا از میکروفن گرفته می‌شود و توسط کامپیوتر پردازش می‌شوند . در مدل‌های صوتی صدا آنالیز و تحلیل می‌شوند و بعد از آن که کاربر در میکروفن صحبت می‌کند، صدای مورد نظر توسط میکروفن گرفته می‌شود، نویزها و صداهای اضافی موجود در پس زمینه صدا که روی حجم صدا و کیفیت آن تاثیر گذارند، حذف می‌شوند. از توابع ریاضی برای دریافت صدا و تبدیل آن به بازه ی فرکانسی مورد نیاز استفاده می‌شود. سپس داده‌های بدست آمده تحلیل می‌شوند و تبدیل به نمایش‌های دیجیتالی می‌شوند. مدل‌های زبانی محتوای صدا را مورد بررسی قرار می‌دهند. از این مدل بیشتر با اهداف تشخیص گفتار استفاده می‌شود. این مدل به مقایسه بین صدای دریافتی و لغت‌های موجود در دایره المعارف می‌پردازد که بزرگترین و رایج ترین پایگاه داده‌ی لغات موجود در زبان انگلیسی اند. اگر چه مدل‌های صدایی و زبانی مدل‌های پایه سیستم‌های تشخیص صوتی‌اند اما ممکن است هر سیستم مدل خاص خودش را داشته باشد و نحوه طراحی نرم افزار وابسته به دیدگاه برنامه نویس آن دارد که از چه تکنیک‌هایی در طراحی سیستم استفاده نماید.[10]

تجزیه و تحلیل صدا زمانی انجام می پذیرد که صدای ورودی از طریق میکروفن و یا سایر دستگاه های ورودی دریافت شود. طراحی سیستم ها شامل ایجاد تغییر و نفوذ در سیگنال ورودی می شود. در سطوح متفاوت عملیات های متفاوتی بر روی سیگنال های ورودی انجام می پذیرد .عملیات هایی چون پیش تاکید، قاببندی، پنجره بندی، آنالیز و تحلیل . الگوریتم های صدا شامل دو مرحله اند :فاز اول آموزش و فاز دوم مرحله تست است. مراحل کلی تشخیص صوت به شرح زیر است:

آماده سازی سیگنال: اولین قدم در آماده سازی سیگنال گفتار و پردازش آن، تبدیل آن از فرم آنالوگ به فرم دیجیتال است.گفتاری که به وسیله انسان بیان می شود، به صورت یک موج در هوا انتشار می یابد .این موج توسط میکروفون دریافت و به سیگنال الکتریکی که یک سیگنال آنالوگ و دارای تغییرات پیوسته در زمان است تبدیل می شود. رقمی سازی نتیجه گسسته سازی سیگنال در حوزه زمان و دامنه است. رقمی کردن گفتار باعث کاهش حجم ذخیره سازی، پردازش و انتقال آسانتر و کاهش هزینه می شود . وقتی یک سیگنال آنالوگ به رقمی تبدیل می شود، هم در زمان و هم در اندازه گسسته می شود [11] .

نمونه برداری: نمونه برداری به معنای عملی است که توسط آن سیگنال آنالوگ گفتار به یک سری نمونه ها تبدیل می شود . عمل نمونه برداری معمولا توسط یک نمونه بردار بنام Hold&Sample انجام می شود.[12]

چندی کردن: چندی کردن بر روی نمونه های سیگنال صورت می گیرد. چندی کردن، سیگنال را در حوزه دامنه گسسته می نماید .هر چه تعداد سطوح چندی کردن بیشتر باشد، سیگنال به سیگنال آنالوگ اصلی نزدیکتر می شود .به تعداد بیت مورد استفاده برای چندی کردن رزولوشن یا دقت چندی سازی گفته می شود . پایین بودن دقت چندی سازی موجب می شود که نتوان پارامتر شیمر را به درستی محاسبه نمود .

استخراج ویژگی ها: استخراج بهترین نماینده پارامتری از میان سیگنال های صوتی، یکی از وظایف مهم برای داشتن بهترین عملکرد است. انجام نتیجه گیری درست این مرحله برای عملیاتی شدن فاز های دیگر تاثیر خواهد داشت [12] .

انتخاب استراتژی شناسایی و تشخیص صوت: در این مرحله یکی از روش های تشخیص صدا نظیر شبکه عصبی؛ ماشین SVM یا .... انتخاب و بر اساس روشها و متدولوژی های مرتبط سیگنال های ورودی مورد پردازش قرار می گیرند.

۴- معرفی برخی روش ها و رویکردها

تطابق الگو

تکنیکی است که از ورودی های کاربر استفاده می نماید و وابسته به آن است که کاربر از چه متدی استفاده نماید. دارای بیشترین دقت است .دقتی در حدود ۹۹ ٪ ولی محدودیت های بسیاری را داراست .این الگوریتم به این شکل آغاز به کار می نماید که از کاربر درخواست پخش صدا در میکروفن را می کند. صدا ها مکررا تکرار می شوند و متوسط معیار ها به عنوان نمونه ذخیره می شود. زمانی که صدایی به عنوان ورودی از میکروفن و به شکل سیگنالی دیجیتال دریافت می شود، تبدیل شده و در حافظه ذخیره می گردد سیگنال های ذخیره شده با نمونه ها مقایسه می شوند و منطبق ترین به عنوان خروجی بازگردانده می شوند.[16]

مبتنی بر مدل مخفی مارکوف

تشخیص صدا یا شناسایی گوینده یکی از مسایل علوم رایانه و هوش مصنوعی است که هدف آن شناسایی یک فرد تنها از روی صدای شخص است. یکی از اصلیترین ابزارهای ریاضی برای حل این مسئله مدل های پنهان مارکوف هستند. برای حل این مسئله با استفاده از مدل پنهان مارکوف این مدل های آماری ابتدا باید مورد آموزش قرار بگیرند. برای این مرحله ابتدا مقدار قابل توجهی از صدای ضبط شده افراد پردازش میشود.

دادههای پردازش شده که در حقیقیت مجموعه عظیمی از اعداد میباشند متناوباً مورد استفاده قرار می گیرند تا مدل پنهان مارکوف. برای هر گوینده به دست آید. در حقیقت مدل پنهان مارکوف مانند یک ماشین عمل میکنند که ورودی آنها یک سری داده است و خروجیشان یک عدد برای هر مجموعهای از دادهها، به این صورت که آن عدد نشان دهنده اختلاف داده های ورودی با مدل پنهان مارکوف هر ماشین است. برای آموزش مدل پنهان مارکوف در هر تناوب داده ها به مدل پنهان مارکوف داده میشود و پارامترهای مدل پنهان مارکوف ذرهای تغییر داده میشود تا عدد خروجی که نشان دهنده اختلاف داده ها با مدل پنهان مارکوف است کوچکتر شود. برای اطمینان از اینکه تغییر پارامترهای مدل پنهان مارکوف در جهت درست انجام میگیرد و نهایتا به حداقل شدن عدد خروجی میانجامد از یک روش ریاضی استفاده میشود. در نهایت بعد از آموزش این مدلها که با استفاده از صدای مرجع انجام شده، میتوان برای آزمایش سامانه صدای یکی از افرادی که قبال از صدای وی برای آموزش مدل پنهان مارکوف استفاده شده را به هر یک از مدل های پنهان مارکوف داد مدل پنهان مارکوف ای که کوچکترین عدد را تولید میکند به عنوان فرد شناسایی شده در نظر گرفته میشود.[16]

مدل های ترکیبی

از مدل اصلی کانال منبع و یا بخشی و یا نوعی از مدل های آماری مولد معمولا استفاده می شود تا مشکلات تشخیص گفتار را فرموله کنند. ذهن گوینده تصمیم می گیرد و دنباله کلمات را انتخاب می نماید و آن را از طریق ژنراتور های متن تحویل می دهد . منبع از کانال های ارتباطی پر سر و صدا که شامل بلند گوی سخنگو است که برای تولید موج های گفتار یا کلام و همچنین سیگنال های پردازشی الزم برای بازشناس گفتار است عبور می کند و در نهایت وظیفه رمز گشا است که سیگنال های صوتی X را به W رمز گشایی نماید که W ایده آل ترین حالت و نزدیک به

سیگنال اصلی صوت می باشد . برنامه های کاربردی با رمز گشا مرتبط می شوند تا نتایج بازشناسی را بدست بیاورند که ممکن است آن ها را به سایز بخش های سیستم وقف دهند یا سازگار نمایند [17] .

مدل انطباق زمانی پویا: مدلی ساده و قدیمی که در گوشیهای تلفن همراه برای شماره گیری صوتی با بیان نام فرد به کار میرود.

شبکه عصبی مصنوعی: مدلی ساده و کارا با سرعت تشخیص بالا و عملکرد بال درنگ که در برابر نویزهای محیطی مقاوم است و فرایند آموزش آن زمان بر است .

مدل آکوستیک: مسئله دقت در تشخیص گفتار بعد از سال ها تحقیق و توسعه به عنوان یکی از اساسی ترین چالش ها باقی مانده است. عوامل شناخته شده و معروفی هستند که میزان سیستم های تشخیص صوت را تعیین می کنند. از قابل توجه ترین آن ها می توان به نغییرات زمینه، تغییرات سخن گو و تغییرات محیط را برشمرد .. مدل سازی آکوستیک گفتار به طور معمول برمی گردد به روند ایجاد بازنمود ها یا نمایش های آماری برای دنباله ای از بردار های ویژگی محاسبه شده از شکل موج گفتار مدل سازی آکوستیک هم چنین شامل مدل سازی تلفظی است که در آن توضیح میدهند که چگونه یک توالی از واحد های اساسی گفتار استفاده می شود تا واحد های بزرگتری از گفتار ایجاد شوند نظیر کلمات یا عبارات که این ها هدف تشخیص گفتار هستند . مدل سازی آکوستیک هم چنین شامل استفاده از اطلاعات باز خورد از تشخیص دهنده باشد تا بردارهای ویژگی های کلام در محیط های پر سر و صدا را تغییر شکل دهد.[15]

مدل صوتی یا آکوستیک شامل بخش هایی از علم در مورد صدا ها و آکوستیک ،فونوتیک یا آوا شناسی ،تنوع زیست محیطی و تفاوت های مهم در نوع گویش سخنران است . مدل زبان مربوط به دانش یک سیتم در مورد کلمات ممکن است و اینکه چه کلماتی به احتمال زیاد در چه رشته ای از کلمات استفاده می شود . توابع و قواعد معنا شناسی بسته به عملیات کاربر ممکن است در بخش مدل زبان قرار گیرند . بسیاری از ابهامات در این زمینه ، در ارتباط با ویژگی های سخنران ،سبک گفتار و سرعت آن ،به رسمیت شناختن بخش های اساسی بیان ، کلمات ممکن یا احتمالی .کلمات متشابه ،کلمات ناشناخته ،تنوع گرامری ،نویز و سر و صدا ، لهجه های غیر بومی و ... روی نتیجه تاثیر خواهد گذاشت . یک سیستم تشخیص صوت موفق باید با تمامی ابهامات ستیزه و مقابله کند . ابهامات آکوستیک ناشی از لهجه های متفاوت و سبک های صحبت کردن هر فرد ،توسط پیچیدگی ها واختلافات لغوی و گرامری ناشی از تغییرات زبان محاوره ای که در مدل زبانی وجود دارند ،تشدید می شود [17] .

،سیگنال های گفتار در ماژول های پردازش سیگنال پردازش می شوند که در این مرحله بردار ویژگی ها ی برجسته برای رمز گشا ،استخراج می شود . رمزگشا از هر دو مدل آکوستیک و زبانی برای ساختن دنباله ای از کلمات که دارای اولویت بیشتری با توجه به ویژگی های استخراج شده ورودی دارند استفاده می کند.

مدل زمانی: نقش و وظیفه مدل زبانی یافتن مقدار wPدر معادله اصلی تشخیص گفتار است. این مدل یکی از انواع مدلهای زبانی گرامر با دستور زبان است که یک ساختار رسمی و مجاز برای زبان بحساب می آید . تکنیک های تجزیه یکی از انواع متد هایی است که با آن میتوان جملات را آنالیز کرد و تحلیل نمود و تطابق ساختار جمله را با دستور زبان بررسی کرد . با ورود قالب های متنی که هر کدام ارای ساختار مختص به خود هستند ،امروزه امکان تعمیم دستور زبان اصلی وجود دارد . علاوه بر این روابط احتمالی میان دنباله ها را می توان به صورت مستقیم بدست آورد . Corporaیا مدل های زبانی تصادفی مانند gram-N از نیاز به ایجاد پوشش گسترده ی دستور زبان رسمی اجتناب می کنند . از انواع شایع مدل های دیگر زبان، مدل زبان های تصادفی هستند که نقش مهمی را در سیستم های زبانی گفتاری ایفا می کند. [21]

۵- مزایای تشخیص صدا در جنبه های متفاوت

آموزشی

از مزایای سیستم های تشخیص صوت در زمینه آموزش است. چرا که می توان از آن ها در کمک به پیشرفت کاربران در زمینه گفتار و تلفظ استفاده نمود. از آن جا که سیستم های تشخیص صوت از الگوریتم ها و نمونه استفاده می نمایند با تعریف پیام های هشدار در صورت تلفظ غلط واژه می توان به کاربران کمک کرد تا تلفظ صحیح لغات را آموزش ببینند. از این سیستم ها نتها می توان به کمک در آموزش و یادگیری زبان اول فرد پرداخت بلکه می توان از آن ها در آموزش زبان های دیگر به کاربران کمک گرفت.

مصارف عادی و روزمره

افراد و کاربران عادی میتوانند از نرم افزار های تشخیص صدا به منظور انجام کارهای عادیشان نظیر دیکته ایمیل بهره ببرند. همچنین به منظور ایجاد حرکت در برنامه های کاربردی که بر روی دستگاه های الکترونیکی قرار دارند، ساخت اسناد، جست و جو در اینترنت و... بهره ببرند .

این سیستم ها هم چنین می توانند به افراد معلول کمک نمایند.کسانی که در اختلالات گفتاری مبتلا هستند و یا افراد ناشنوا یا کم شنوا می توانند از سیستم های تشخیص صوت بهره ببرند . این سیستم ها با تسهیل تعاملات سریعتر میان اشخاص نا شنوا یا کم شنوا با دیگران ،به خصوص کسانی که زبان اشاره نمی دانند ،به این افراد کمک می کنند .

سیستم های تشخیص صوت هم چنین می توانند به افرادی که مبتلا به مشکلات حرکتی و معلولیت های جسمی اند، در یادگیری کمک نمایند چرا که به آنها اجازه می دهند تا از دستگاه های الکترونیکی بدون استفاده از دست استفاده نمایند. این کمک می کند تا با دستگاه های صوتی سریعتر تعامل کنند و دیگر برای استفاده از دستگاه های الکترونیکی نیاز به داشتن گزینه هایی نظیر الزام به وجود مانیتور بزرگ و یا سایر خصوصیات نخواهد بود . استفاده از سیستم ها سبب می شود تا در زمان کاربران صرفه جویی شود. برای مثال دیگر نیازی به تایپ نخواهد بود که این سبب کاهش زمان می شود. ابزار های ارتباطی حاصل از تکنولوژی های امروزی نظیر وسیله های کمک رسان با خروجی صوتی معمولا وابسته به یک سوئیچ یا صفحه کلید برای دریافت ورودی هستند چرا که آن ها توانایی خواندن و داشتن ارتباط طبیعی را ندارند چرا که بسیار ضعیف و کند هستند . بر اساس پژوهش های صورت گرفته توسط محققان، می توان دریافت که کاربران نیاز به وسیله ای دارند که بوسیله آن بتوان به سادگی در شرایط و محیط های متفاوت و متنوع استفاده کرد . به منظور وارد کردن اطالعات ،از کلام و صحبت استفاده می نمایند [30] .

تجاری

بسیاری از حرفه ها می توانند از سیستم های تشخیص صوت به منظور افزایش بهره وری استفاده نمایند . سیستم های تشخیص صوت می توانند در گزارش گیری و گزارش دهی ،سیستم های مراقبتی و بهداشتی و در مراکز تماس مورد استفاده قرار گیرند . سیستم بهداشت و درمان اغلب از نرم افزار های تشخیص صوت برای دریافت سریع علایم سلامت مورد استفاده می کنند . استفاده از این سیستم ها در مراکز تماس سبب می شود که کاربران بدون نیاز به انتظار برای صحبت با اپراتور ،کار های مد نظرشان را انجام دهند .

در مراکز گزارش گیری نیاز به تایپ سریع و استفاده از سیستم های صوتی می تواند به سرعت بخشیدن به عملیات ها کمک بخشد .مزیت اصلی استفاده از سیستم های تشخیص صوت در اهداف شرکتی و عمده افزایش بهره وری است [30]

امنیتی

از سیستم های تشخیص صوت و نرم افزار های تشخیص صدا می توان برای کمک به دولت استفاده کرد. از تشخیص صدا در ارگان ها و نهاد های دولتی با اهداف اجرای قانون، ادارات حقوقی استفاده می شود . کارکنان می توانند به سرعت و به راحتی اسناد را جست و جو نمایند و با سرعت قابل قبول از اینترنت و رایانه استفاده نمایند از این سیستم ها در ارتش هم استفاده می شود.شایع ترین استفاده از نرم افزار های تشخیص صدا در ارتش در بخش کنترل و فرماندهی است که اجازه می دهد تا به کنترل دستگاه و یا سیستم ها با استفاده از دستورات گفته شده پرداخت.هم چنین در ارتش از این سیستم ها در ارتش برای تشخیص هویت استفاده می شود. [30]

صنایع رباتیک

در دنیای امروز ،تعداد افراد مسن در جامعه به میزان فوق العاده ای افزایش یافته است.برای مثال تخمین زده شده که تا سال ۲۰۳۵ نرخ این افراد در توکیو به ۲۵۰۲ درصد از جمعیت برسد .در هرحال مراقبت جسمی و روانی از این گروه از افراد از اهمیت ویژه ای برخوردار است تا از زوال عقل درافراد سالخورده که به تنهایی در خانه زندگی می کنند جلوگیری شود چرا که افراد مسن از شانس کمتری برای صحبت با دیگر افراد برخوردارند و کمتر به انجام فعالیت های روزانه می پردازند.از این رو مراقبان این افراد نقش مهمی را در سلامت و مراقبت روحی و جسمی آن ها بر عهده دارند . اما نکته قابل توجه آن جاست که تعداد مراقبان و درمانگران در شرایط فعلی جامعه به بزرگی . مقدار تعداد افراد مسن نخواهد بود .

به همین دلیل استفاده از ربات هایی که قابلیت ارتباط آسان با انسان ها را دارند برای حمایت و مراقبت های جسمی و روانی برای افراد مسن و همچنین کمک به مراقبت کنندگان سالمندان ضروری است. برای مثال استفاده از یک ربات برای صحبت کردن می تواند به جلوگیری از زوال عقل در سالمندان تاثیر گذار باشد. گفت و گو ی رباتیک می تواند به افراد سالمند کمک کند تا حافظه و توانایی های خود را فعال و تمرکز خود را بهبود بخشند. علاوه بر این یک شریک ربات می تواند همچنان که با سالمند صحبت می کند رفتار او را در انجام کار های خانه نظارت کند. با این حال صحبت کردن با یک ربات دشوار به نظر می رسد حتی اگر محتوای بسیاری از صحبت ها از پیش طراحی شده باشد که این به دلیل عملکرد تشخیص صداست. علاوه بر ارتباط کلامی، ربات باید ارتباطات غیر شفاهی، بعنوان مثال حالت صورتاحساسات و عواطف و حرکات و حرکات اشاره را درک کند. با توجه به نظریه ارتباط هر فرد دارای محیط شناختی مربوط به خود است و ارتباطات بین افراد بوسیله محیط های شناختی مخصوص به هر فرد محدود شده است. [19]

۶- نتیجه گیری

امروزه سیستم های تشخیص صوت نقش مهم و عمده ای در زندگی بشری بر عهده دارند. از این رو اهمیت این سیستم ها در بهبود کیفیت زندگی انسان بر کسی پوشیده نیست. و مطالعات انجام شده بر این سیستم ها را می توان در چند جهت دسته بندی نمود. ا مروزه انتظارات از این سیستم ها بیشتر در جهت بهبود کیفیت و همچنین ساده سازی و سرعت دهی به این سیستم هاست. پس بسیاری از محققان در تلاش اند تا تکنیک های موجود را در جهت افزایش سرعت و بهبود کیفیت و یا ساده تر کردن نحوه عملکرد بهبود دهند.گروه دیگری از تحقیقات امروزی در سمت و سوی موضوعات پزشکی صورت می پذیرند. چرا که استفاده از این سیستم ها می تواند کیفیت زندگی بسیاری از انسان ها را تحت پوشش قرار دهد. با پیشرفت روز افزون علم و استفاده هرچه بیشتر از تکنولوژی،کیفیت زندگی روز به روز رو به بهبود است و استفاده از سیستم های تشخیص صوت در وسایل و تکنولوژی هایی که بشر به طور روزمره از آن استفاده می کند و روز به روز وابستگی اش به آن بیشتر می شود، رو به افزایش است. از آنجا که استفاده از حداکثر توان سیستم های تشخیص صوت در این ابزارها را می توان سبب ساز ایجاد جهشی بزرگ در این تکنولوژی دانست،ب سیاری از تحقیقات امروزی به بررسی عملکرد و بهبود و سازگار نمودن سیستم ها با چالش ها و شرایط روزمره زندگی بشر می پردازند .با توجه به چالش های مطرح در هریک از حوزه ها و توجه به این نکته که علم پردازش و تشخص صدا علمی نوپاست و سوالات و مشکلات بسیاری در بخش های متفاوت آن موجود است ،تحقیق در هر یک از حوزه های بیان شده نقش به سزایی در زندگی بشر خواهد داشت.

۷- مراجع

[1] WRONISZEWSKA, MARTA,"VOICE COMMAND RECOGNITION USING HYBRID GENETIC ALGORITHM", TASK QUARTERLY 14 No 4, 377–396, 2010
[2] http://www.wikipedia.org
[3] Furui S.,"Digital Speech Processing, Synthesis, and Recognition, Marcel Dekker", E. Keller, 2001
[4] Prabhakar, Om Prakash," A Survey On: Voice Command Recognition Technique", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013
[5] J.Kreiman, D.Sidtis, "Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception", Blackwell Publishing Ltd, 2011
[6] A, Perrakis, W, Hohenberger,"Integrated operation systems and voice recognition in minimally Invasive surgery: comparison of two systems", springer, 2132
[7] Kubota," Multimodal Communication for Human-Friendly Robot Partners in Informationally Structured Space", IEEE Transactions on, 2012
[9] Sadaoki Furui," 51 Years of Progress in Speech and Speaker Recognition Research", ECTI TRANSACTIONS ON COMPUTER AND INFORMATION TECHNOLOGY VOL.1, NO.2 NOVEMBER, 2005
[10] Sadaoki Furui," History and Development of Speech Recognition", 2010
[11] G. Bailly, A. Monaghan, J. Tekren, M. Huckvale," Improvements in speech Synthesis" John Wiley & Sons, Inc., 393P, 2002
[12] D. G. Childers, Speech Processing and Synthesis Toolboxes, John Wiley & Sons, Inc., 482P 2111
[13] X. Huang, A. Acero, H. W. Hon, "Spoken Language Processing, A Guide to Theory, Algorithm, And System Development", Chapters 14, 15, and 16, Prentice Hall, 935P, 2000
[14] Karaali et al0, "A High Quality Text-To-Speech System Composed of Multiple Neural Networks", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing 2:1237-1240, 1998
[15] L. R. Rabiner. "A tutorial on hidden Markov models and selected applications in Speech recognition". Proceeding of the IEEE, 1989
[16] Muda, Lindasalwa," Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, 2010
[17] Yan Zhang" Speech Recognition Using Deep Learning Algorithms" 62
[18] Mark S. Hawley, Stuart P," A Voice-Input Voice-Output Communication Aid for People with Severe Speech Impairment, IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 21, NO. 1, 2013
[19] Kubota," Multimodal Communication for Human-Friendly Robot Partners in Informationally Structured Space", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETIC, 2012
[20] Xuedong Huang," An Overview of Modern Speech Recognition", 2009
[21] Hermansky,H., Sharma, S0," Temporal patterns (TRAPS) in ASR of noisy speech" In Proceedings of the International Conference on Acoustics Speech and Signal Processing,Phoenix, Arizona, 1999
[22] Allen, B0" How do humans process and recognize speech", IEEE fiansactions on Speech and Audio Processing, 2(4):567-577, October 1994
[23] Chen, B., Sivadas, S.," Learning discriminative temporal Patterns in speech: Development of novel TRAPSlike Classifiers" In Proceedings of Eurospeech, Geneva, Switzerland, September 2003.
[24] Sohn J., Kim N. S. and Sung W., "A statistical model-based voice activity detection," IEEE Signal Process Letters, vol. 6, pp. 1-3, 1999
[25] Mohammdi M., Nasersharif B., Rahmani M. and Akbari A., "The New Sub-band Level Voice Activity .Detector Based on Wavelet Transform," ICEE, 2007
[26] E.Grivel, M.Gabrea and M.Najim, "speech enhancement az a realization issue", Signal Processing, Vol82, pp.1963-1978, Dec 2002.
[27] K.K.Paliwal and A.Basu, "A Speech Enhancement Method Based on Kalman Filtering", in Prec.ICASSP, 87, PP.177-

180, 1981
[28] Marcel Gabrea,"A SINGLE MICROPHONE NOISE CANCELLER BASED ON AN ADAPTIVE KALMAN FILTER ", 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2012
[30] Sadaoki Furui," 40 Years of Progress in Automatic Speaker Recognition", Springer, 2009
[31] Sadaoki Furui, Li Deng," Fundamental Technologies in Modern Speech Recognition" Signal Processing Magazine, IEEE, 2012
[32] Terutaka Marukami, Shoko Tani," A Basic Study on Application of Voice Recognition Input to an Electronic Nursing Record System -Evaluation of the Function as an Input Interface", Springer, 2010
[33] Xiaodong He and Li Deng "Speech Recognition, Machine Translation, and Speech Translation—A Unified Discriminative Learning Paradigm", IEEE SIGNAL PROCESSING MAGAZINE, 2011
[34] Buttermore, Ronald," DEVELOPMENTS IN VOICE REGONITION TECHNOLOGY", 2013
[35] Janet M. Baker, Li Deng," Historical Development and Future Directions in Speech Recognition and Understanding", 2007
[36] Vishweshwara Rao, Preeti Rao" Vocal Melody Extraction in the Presence of Pitched Accompaniment in Polyphonic Music", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, 2010
[37] Hyeopwoo Lee, Sukmoon Chang, "A Voice Trigger System using Keyword and Speaker Recognition for Mobile Devices", Contributed Paper IEEE, 2009
[38] Javier Ramírez, José C, "Improved Voice Activity Detection Using Contextual Multiple Hypothesis Testing for Robust Speech Recognition", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, 2007
61
[39] Takeshi Yamada, "Performance Estimation of Speech Recognition System under Noise Conditions Using Objective Quality Measures and Artificial Voice", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, 2006

A Review of Voice Recognition Systems, Concepts and Methods

Mozhdeh Matalebi, Azam Bastanfard

Faculty of computer and information technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran.,
E-mail Mzhd_Mti@yahoo.com
Iran Broadcasting University ,E-mail: bastanfard03@gmail.com.

Abstract. Voice recognition systems are electronic systems that allow electronic devices to respond to voice. These systems are very functional and pursue many goals, but to date they have not achieved their full potential and they have not used all their potential. Voice recognition systems translate speech and speech sounds into signals that are capable of processing and can be converted to a specific task. The topics discussed in speech processing and voice recognition include speech recognition, text-to-speech, speech recognition, speech encoding, speech enhancement, language recognition, audio document indexing.

Keywords: voice recognition, speech recognition