

بسمه تعالی

تشخیص گفتار چیست؟

تشخیص گفتار یا **تبدیل صوت به متن** (speech recognition) به معنای استفاده از رایانه و هوش مصنوعی برای تشخیص کلمات و عبارت موجود در صوت انسان و تبدیل آن ها به متن به عنوان خروجی است.

تشخیص گفتار (Speech Recognition) به این معناست که کاری کنیم کامپیوتر بتواند گفتار را تشخیص دهد، تفسیر کند و آن را به متن تبدیل کند. تشخیص گفتار زیرشاخه‌ای از زبان‌شناسی محاسباتی (Computational Linguistic) است. این زیرشاخه با تکنولوژی‌هایی کار می‌کند که داده‌های صوتی (گفتار) را به عنوان ورودی دریافت و تجزیه و تحلیل می‌کنند. تشخیص گفتار به سرعت در حال تبدیل شدن به یک اصل در تعامل انسان و کامپیوتر است. امروزه ابزارهای تشخیص احساسات (Scenes Recognition) در کارهای مختلفی از جمله نوشتن پیام متنی، پخش موسیقی، دستیارهای مجازی ((Virtual Assistants)) و غیره استفاده می‌شوند.

یک نمونه از سیستم تشخیص گفتار را در تصویر زیر مشاهده می‌کنید. در این سیستم با دریافت صوت به عنوان ورودی، صوت تحلیل می‌شود و متن حاصل از آن به عنوان خروجی به کاربر نشان داده می‌شود.



تاکنون تکنولوژی‌های تشخیص گفتار متعددی توسعه پیدا کرده اند که از جمله آن ها می‌توان به Siri، Google Assistant، Cortana، Alexa اشاره کرد که نحوه ی تعامل افراد با ابزارها، خانه ها، ماشین و کارشان را تغییر داده اند. از طریق فناوری تشخیص گفتار می‌توان با رایانه ها و ابزارها صحبت کرد، در مقابل، این ابزارها مفهوم صحبت شما را متوجه شده و پاسخی به آن می‌دهند.

معرفی دستیارهایی که با صدا کنترل می شوند یا دستیاران دیجیتال به بازارهای تشخیص گفتار منجر به تغییر چشم انداز این تکنولوژی در قرن ۲۱ شد.

در چند سال گذشته پیشرفت در زمینه‌ی یادگیری ماشین و زبان‌شناسی محاسباتی به پیشرفت‌های چشمگیری در نحوه‌ی تعامل ما با جهان پیرامون انجامیده است. یکی از شاخص‌ترین این پیشرفت‌ها حوزه‌ی تشخیص گفتار است.

البته تشخیص گفتار موضوع کاملاً جدیدی نیست؛ درواقع از زمانی که کامپیوترها به وجود آمدند تشخیص گفتار هم وجود داشت. مفهوم تشخیص گفتار اولین بار به شکل نظریه‌ای علمی-تخیلی در فیلم‌های مانند ادیسه فضایی (Space Odyssey) در سال ۲۰۰۱ به عموم معرفی شد..

تشخیص گفتار به سرعت در حال تبدیل شدن به یک فناوری پرکاربرد و موثر است. شرکت‌های بزرگ برای قدرتمند کردن سیستم‌های خود از مجموعه داده‌های بزرگ برای بلندگوها و منابع نويز مختلف استفاده می کنند. معماری کلاسیک برای محاسبه ویژگی‌های الهام گرفته از سیستم شنوایی انسان برای تجزیه و تحلیل طیفی ورودی صوتی ، از پردازش سیگنال استفاده می کند و آنها را به یک سیستم بزرگ از Hidden markov model که دارای مدل‌های ترکیبی Gaussian هستند ، منتقل می کند. مدل‌های پیچیده پردازش گفتار برای کمک به جدا کردن کلمات موجود در صدا استفاده می شود. درپردازش زبان و گفتار روشهای یادگیری عمیق تأثیر بسزایی داشته و بسیاری از گروههای بزرگ صنعتی روشهای یادگیری عمیق مانند شبکه های عصبی را با روش های کلاسیک تشخیص گفتار را با را جایگزین کردند.

سیستم تشخیص گفتار را دستیار دیجیتال یا دستیار هوشمند می نامند. دستیارهای دیجیتال برای کمک به مردم در انجام کارهای اساسی و پاسخ به سوالات طراحی شده اند. این ربات ها با توانایی دسترسی به اطلاعات پایگاه های گسترده و منابع دیجیتالی مختلف ، به حل مشکلات در زمان کمتر و افزایش تجربه کاربر و بهره وری انسان کمک می کنند.

کاربردهای تشخیص گفتار:

تکنولوژی تشخیص گفتار و دستیاران دیجیتال راه خود را از گوشی های هوشمند به خانه های ما پیدا کرده اند و کاربرد آن ها در زمینه هایی مانند کسب و کارها، بانکداری، بازاریابی، مراکز بهداشتی در حال افزایش است.

۱- محیط کار:

تکنولوژی تشخیص گفتار در محیط کار به ترکیب وظایف ساده برای افزایش بهره وری و حتی انجام وظایفی که به صورت سنتی توسط انسان انجام می شد، می پردازد. مثال هایی از کارهای اداری که دستیاران دیجیتال در حال حاضر یا در آینده می توانند انجام دهند. در واقع همه ی این وظایف می توانند با تبدیل صوت به متن و کنترل و هدایت کارها با استفاده از صوت انجام شوند.

- جستجو برای گزارش یا سند در رایانه با استفاده از جستجوی صوتی
- ایجاد نمودار یا جدول با استفاده از داده های صوتی و تبدیل آن ها به متن
- نوشتن اطلاعاتی که باید با سند ترکیب شود
- پرینت اسناد درخواستی
- شروع کنفرانس های ویدئویی
- برنامه ریزی ملاقات ها
- فراهم کردن مقدمات سفر

۲- بانکداری:

صنعت بانکداری از تشخیص گفتار برای کاهش ناسازگاری های مشتریان استفاده می کند. این نوع بانکداری نیاز به خدمت رسانی توسط نیروی انسانی به مشتری را کاهش داده در نتیجه هزینه ی نیروی انسان را کم می کند. یک دستیار بانکی شخصی می تواند وفاداری و رضایت مشتری را افزایش دهد. تشخیص گفتار از طریق موارد زیر به بانکداری کمک می کند:

- درخواست اطلاعاتی در مورد باقی مانده ی حساب و تراکنش ها بدون نیاز به باز کردن تلفن همراه
- انجام پرداخت ها
- دریافت اطلاعاتی در مورد تاریخچه ی تراکنش ها

۳-بازاریابی:

جستجوی صوتی روش جدیدی برای دسترسی بازاریابان به مشتری ها ایجاد می کند. با تغییر در نحوه ی تعامل مردم با وسایل خود، بازاریاب ها باید به دنبال گرایش های در حال توسعه در داده و رفتار کاربران باشند. تا سال ۲۰۲۰ بسیاری از برندها، سایت های خود را به جستجوی صوتی مجهز می کنند تا سود تجارت دیجیتال خود را تا ۳۰ درصد افزایش دهند.

جستجو بر اساس صوت که سرعت بیشتری دارد می تواند کاربران را بی صبر کند و همچنین آن ها را به اینترنت به عنوان منبعی از اطلاعات بیشتر وابسته کند. به همین دلیل مقدار زمان صرف شده برای نگاه کردن به صفحه ممکن است کاهش یابد. بازاریابان باید به این مسئله توجه کنند که این امر ممکن است بر روی استفاده از وسایل بصری تاثیر گذارد و افراد بیشتر به مطالب شنیداری روی بیاورند.

۴-مراکز بهداشتی:

در مراکز بهداشت و درمان حتی چند ثانیه هم مهم و حیاتی است و شرایط عملیاتی استریل در اولویت قرار دارد. در نتیجه دسترسی به اطلاعات بدون نیاز به استفاده از دستان و به صورت سریع می تواند تواند تاثیر مثبتی بر ایمنی و بازده عملیات پزشکی داشته باشد.

مزایای تشخیص گفتار برای مراکز درمانی:

- یافتن اطلاعات از بین گزارشات پزشکی به صورت سریع
- یادآوری به پرستاران در مورد فرایندها و یا دادن دستورالعمل های خاص به آن ها
- پرستاران می توانند اطلاعات مدیریتی، مانند تعداد بیماران روی زمین و تعداد واحدهای موجود را پرس و جو کنند.
- در خانه، والدین می توانند علائم شیوع بیماری، زمان مراجعه به پزشک و چگونگی مراقبت از یک کودک بیمار را سوال کنند.
- کاهش کاغذبازی
- کاهش زمان ورود اطلاعات

- بهبود جریان کاری

مهم ترین نگرانی در مورد استفاده از تشخیص گفتار در مراکز مراقبت های بهداشتی، محتوایی است که دستیار دیجیتال به آن دسترسی دارد. این محتوا باید توسط موسسات پزشکی تایید و تولید شود تا قابل اعتماد باشد.

۵-اینترنت اشیاء:

اینترنت اشیا به معنای اتصال اینترنت به ابزارهای فیزیکی و همه ی اشیا است. در واقع از این طریق ما شبکه ای از اشیا خواهیم داشت که همه به هم متصل هستند. با به کارگیری تراشه هایی به منظور اتصال به اینترنت، همه ی اشیا و ابزار می توانند از طریق اینترنت با هم ارتباط و تعامل داشته باشند.

یکی از کاربردهای مهم تشخیص گفتار در اینترنت اشیا در خودروها دیده می شود. چنین پیشرفتی منجر به تغییر روش رانندگی و تعامل افراد با خودروهایشان می شود که هدف نهایی آن محدود کردن مداخلات راننده است. کاربردهای دستیار دیجیتال در خودروها:

- گوش دادن به پیام ها بدون نیاز به استفاده از دست
- کنترل کردن رادیو
- راهنمایی و جهت یابی
- پاسخ به درخواست های صوتی

۶-یادگیری زبان:

تشخیص گفتار می تواند ابزار سودمندی برای یادگیری زبان دوم باشد به این صورت که تلفظ صحیح را به کاربر آموزش می دهد و در کنار آن، او را در یادگیری مهارت صحبت کردن یاری می رساند. یکی از کاربردهای تشخیص گفتار از دید انسان توانایی آن در از میان برداشتن موانع زبانی و فرهنگی در زندگی اجتماعی و محیط کاری است. دنیایی بدون موانع زبانی، فرصت های زیادی را برای همکاری میان کشورها و فرهنگ های مختلف ایجاد می کند که شاید در نتیجه ی تنوع بیشتر نرخ رشد نیز سریع تر شود. دانش آموزان نابینا یا کم بینا و یا معلول می توانند از این تکنولوژی برای یادگیری

استفاده نمایند. آن ها می توانند بدون نگرانی در مورد تایپ کردن یا انجام تکالیف خود این کار را انجام داده و حتی به راحتی به جستجو در اینترنت بپردازند

۷-موتور های جست و جو

وقتی شخصی از موتور جستجو برای یافتن پاسخ خود استفاده می کند، اینکه درخواست خود را به صورت متنی یا صوتی بنویسد تفاوت هایی با هم دارد. فرد ممکن است در نوشتن درخواست دچار مشکل شود ولی وقتی به صورت صوتی این کار را انجام می دهد، نتایج بهتری می باید.

۸-ارائه ی خدمات به مشتری

مشتری ها ممکن است به ارتباط با سیستم تشخیص گفتار تمایل بیشتری نسبت به ارتباط با یک انسان داشته باشند. این روش موجب بهبود فرایندها و زمان پاسخگویی می شود. برای مثال یکی کاربرد تشخیص گفتار در فرودگاه ها به منظور تایید برنامه های مسافرتی است.

کاربردهای آینده:

فناوری تشخیص گفتار هنوز تا حدود زیادی در ابتدای راه خود است اما با تمایل بیشتر مردم به استفاده از تکنولوژی ها، به احتمال زیاد به سرعت رشد خواهد کرد. در این مرحله از چرخه حیات فن آوری تشخیص گفتار، داشتن یک ایده ی روشن از پتانسیل آن بسیار مهم است. کسب و کارها باید در رویکرد خود برای معرفی یا ترکیب تکنولوژی تشخیص گفتار در استراتژی های بازاریابی دیجیتال خود دقت لازم را داشته باشند. افراد عادی هم می توانند مزایای تشخیص گفتار در فعالیت های روزمره خود را بررسی کنند. با افزایش میزان دقت و فروش این فناوری، صنایع باید سازگاری بیشتر با آن پیدا کنند تا بیشتر مبتنی بر صوت باشند.

فیلدهای مشابه تشخیص گفتار:

یکی از فیلدهای مشابه تشخیص گفتار ، تشخیص صدا می باشد که دو زمینه کاملاً متفاوت و جدا از هم هستند.

نکته‌ی مهم در اینجا این است که عبارت تشخیص گفتار (Speech Recognition) و تشخیص صدا (Voice Recognition) اغلب به‌جای هم استفاده می‌شوند. این در حالی است که این دو کاملاً با هم فرق دارند. تشخیص گفتار به معنای شناسایی کلمه‌های استفاده‌شده در یک گفتار است، ولی تشخیص صدا (Voice Recognition) به معنای شناسایی صدای شخص صحبت‌کننده است

تشخیص صدا چیست؟

تشخیص صدا یا بلندگو توانایی یک برنامه برای شناسایی افراد بر اساس صدای منحصر به فرد آنهاست. با اسکن گفتار و ایجاد تطابق با اثر انگشت صدای مورد نظر کار می‌کند. توسعه هوش مصنوعی فرصت‌های گسترده‌ای را برای این زیرشاخه از علوم کامپیوتر باز کرد. ما را قادر می‌سازد تا بدون دست زدن به ماشین‌ها با آنها ارتباط برقرار کنیم. به سرعت در حال رشد است و توسعه دهندگان راه‌های بیشتری برای اعمال آن در زمینه‌های مختلف پیدا می‌کنند.

تفاوت بین تشخیص گفتار و تشخیص صدا چیست؟

درک تفاوت بین این دو رشته ضروری است. هدف از تشخیص صدا شناسایی صاحب صدا است. هدف تشخیص گفتار شناسایی کلمات گوینده است. در حالت اول، برنامه برای مقایسه نیاز به یک صدای منحصر به فرد از بلندگو دارد. در حالت دوم، برنامه به یک دیکشنری عظیم برای شناسایی کلمات گوینده نیاز دارد

ابزارهای روش تشخیص گفتار:

اگر نمی‌خواهید سیستم تشخیص گفتار خود را بسازید، ابزارهای متن‌باز مختلفی وجود دارد. از جمله آنها عبارتند از:

- [CMU Sphinx](#) یک سیستم تشخیص گفتار مستمر و مستقل از بلندگو که در دانشگاه کارنگی ملون توسعه یافته است. CMU Sphinx شامل گروهی از محصولات است که برای اهداف مختلف طراحی شده‌اند. برای دانلود از [صفحه وب](#) GitHub در دسترس است. علاوه بر این، در آنجا می‌توانید اسنادی را برای کاربران پیدا کنید. این برنامه از بسیاری از زبان‌های برنامه‌نویسی محبوب مانند C/C++، C#، Java و Python پشتیبانی می‌کند.

- **HTK Toolkit** یک جعبه ابزار برای کار با مدل های پنهان مارکوف . در دانشگاه کمبریج توسط آزمایشگاه هوش ماشینی توسعه یافته است و عمدتاً برای تحقیقات تشخیص گفتار استفاده می شود . این به طور کامل منبع باز نیست . کاربران می توانند اطلاعات مربوط به استفاده از محصول را در [وب سایت](#) رسمی HTK بیابند . زبان های برنامه نویسی پشتیبانی شده C و Python هستند.

معروف ترین کارها در حوضه تشخیص گفتار:

۱- خانه و اسپیکرهای هوشمند

آمازون اکو و الکسا

تا همین اواخر، **کاربرد تشخیص گفتار** الکسا آمازون را تنها می شد در محصولات تجاری ساخته شده توسط این شرکت مشاهده نمود. با این حال ، خدمات وب آمازون، دستیار صوتی را در اختیار سایر شرکت ها قرار داده است. با همکاری آمازون با اینتل، امکان توسعه یک کیت نرم افزاری فراهم شد که خدمات صوتی ویژه ای ایجاد می کرد. این خدمات به شرکت های شخص ثالث اجازه می دهد قابلیت های الکسا را در دستگاه های خود تعبیه نمایند. این همکاری نتیجه استراتژی خاص آمازون "الکسا در همه جا" است.

در CES2018 در لاس وگاس، سونی ، TiVo و Hisense از مهارت های یکپارچه سازی شده هوشمند الکسا رونمایی کردند، این مهارت جدید این امکان را فراهم می آورد تا تلویزیون را از طریق صدا کنترل کرد. سازندگان لوازم خانگی مانند گرداب، دلتا، ال جی و هایر (Hair) نیز مهارت های تشخیص صدا الکسا را برای کمک به افراد در کنترل همه جوانب خانه خود ، از تلویزیون و مایکروویو گرفته تا واحد تهویه مطبوع و شیر آب اضافه کرده اند. طبق سایت آمازون الکسا، بیش از ۱۳۰۰۰ دستگاه خانه هوشمند از بیش از ۲۵۰۰ مارک تجاری با الکسا قابل کنترل هستند. بدین ترتیب، **کاربرد تشخیص گفتار** سود بی سابقه ای را روانه این شرکت ساخته است.

به عنوان یک دستیار مجازی ، آمازون ادعا می کند که Alexa می تواند به افراد کمک کند تا برنامه خود را مدیریت و وظایف خود را پیگیری نمایند. وقتی در دستگاه هایی مانند کنسول های جلسات ادغام شوند، می توانند تنظیمات اتاق کنفرانس را با صدای بلندگو کنترل کنند. دستگاه های دارای امکان الکسا همچنین می توانند به عنوان دستگاه کنفرانس صوتی در اتاق های کنفرانس کوچکتر، یا تجهیزات کنترل در اتاق های بزرگتر عمل کنند.

۲- کاربرد تشخیص گفتار در Google home and assistant

دستیار گوگل یک دستار مجازی صوتی است که توانایی‌های آن شامل ارسال و درخواست پرداخت از طریق google pay، یا عیب‌یابی تلفن‌های x12 پیکسلی است.

دستیار صوتی گوگل در دستگاه‌های اندروید و ios، ساعت‌های هوشمند، لپ‌تاپ‌های pixelbook، تلویزیون‌های و صفحه نمایش‌های هوشمند اندروید، اتوموبیل‌های هوشمند در دسترس است.

برای کودکان و خانواده‌ها دستیار صوتی گوگل قریب به ۵۰ بازی برپایه صدا و گفتار را دارد.

گوگل به تازگی برنامه دستیار سرمایه‌گذاری (Assistance Investment) را رونمایی کرده که در استارت‌آپ‌هایی در زمینه فناوری‌های پیشرفته صوتی چه سخت‌افزاری و چه نرم‌افزاری فعالیت دارند، سرمایه‌گذاری می‌کند.

از دیگر محصولات **کاربرد تشخیص گفتار گوگل**، ابزار مبتنی بر هوش مصنوعی تبدیل **ابری گفتار به نوشتار** است که توسعه‌دهندگان امکان می‌دهد از طریق الگوریتم‌های **شبکه عصبی** یادگیری صوتی را به متن تبدیل کنند. با کار بر روی ۱۲۰ زبان مختلف، این ابزار قادر توانایی همچون کنترل و فرمان صوتی، انتقال فایل‌های صوتی از مراکز تماس، پردازش جریان در زمان واقعی یا فایل‌های صوتی از پیش ضبط شده را دارد

۳- کاربرد تشخیص گفتار در کورتانا مایکروسافت (Cortana by microsoft):

مایکروسافت در اکتبر سال ۲۰۱۷ دستیار مجازی صوتی خود با نام **کورتانا (Cortana)** منتشر ساخت. بلندگوی خانگی و برنامه تلفن همراه کورتانا به کاربر یادآوری می‌کند تا یادداشت‌ها و لیست‌های خود را نگاه دارد. طبق گفته مایکروسافت این برنامه می‌تواند به مدیریت یک تقویم کمک کند. کورتانا قابل دریافت و بارگیری از فروشگاه اپل و Google Play است و می‌تواند بر روی رایانه‌های شخصی، بلندگوهای هوشمند و تلفن‌های همراه اجرا شود.

کورتانا می‌تواند در بلندگو خانگی مایکروسافت بنام Invoke، صدای موسیقی را کنترل نماید، لیست‌های پخش آهنگ‌ها را تنظیم نماید یا صدا را کم یا زیاد نماید. تمامی اینکارها تنها با صدای کاربر امکان‌پذیر است. با این وجود، این سرویس از خدمات اصلی پخش موسیقی خارج از Spotify

پشتیبانی نمی‌کند. میکروسافت می‌گوید بلندگو هوشمند همچنین به سؤالات مختلفی پاسخ می‌دهد، و کارهایی دیگر نظیر برقرای تماس اسکایپ و یا بررسی آخرین اخبار هواشناسی را به خوبی انجام می‌دهد.

هسته اصلی فناوری **کاربرد تشخیص گفتار** میکروسافت، تبدیل گفتار به نوشتار است که جریان‌های صوتی را به متن رونویسی می‌کند. میکروسافت می‌گوید این سرویس پایان گفتار را تشخیص می‌دهد و گزینه‌های قالب‌بندی از جمله نقطه‌گذاری (علامت سوال، تعجب، ویرگول، نقطه و غیره) و همچنین ترجمه زبان را ارائه می‌دهد.

کاربرد تشخیص گفتار برنامه‌های موبایل

۴-سیری اپل:

هنگامی که اپل برای اولین بار Siri را در آیفون ۴ در سال ۲۰۱۱ ادغام کرد، دستیار مجازی به تعدادی از سرویس‌های وب متصل شد و دارای قابلیت‌های صوتی مانند سفارش تاکسی را از طریق TaxiMagic، جمع‌آوری جزئیات کنسرت از StubHub، بررسی فیلم از Rotten Tom گوگل و واریسی رستوران‌های خوب از طریق Yelp شد.

امروزه، قابلیت‌های سیری شامل ترجمه، پخش آهنگ، انتقال وجه بین حساب‌های بانکی و غیره است. به گفته اپل، به دلیل قابلیت‌های گسترده یادگیری ماشین، می‌توان دستورات جدیدی را برنامه‌ریزی کرد.

در حالی که سیری پیش از دستیار گوگل و آمازون الکسا راه اندازی شده، هنوز هم نگرانی‌هایی در مورد صحت آن هنگام پاسخ دادن به دستورات یا سوالات در مقایسه با دیگر فن‌آوری‌های موجود در بازار، وجود دارد.

در ژوئن سال ۲۰۱۸، اپل تغییراتی را در Siri انجام داد و میانبرهای ویژه‌ای را راه اندازی کرد. با این تغییرات، اپل ادعا می‌کند کاربران می‌توانند **کاربرد تشخیص گفتار** را از طریق دستور صوتی، متن یا تپ، در سیری گسترش دهد. این برنامه، هم اکنون در iPhone، iPad، Apple watch و قابل دسترسی است. این اقدامات شامل اتصال به دیگر برنامه‌هایی مانند برنامه Tile برای یافتن کلیدواژگان یا به دست آوردن اطلاعات سفر از طریق Kayak است.

گفته می‌شود که میانبرهای سیری قادر به خواندن داده‌های متنی کاربر، مانند رویدادهای تقویم و مکان‌های GPS هستند تا بتوانند میانبرهای جدیدی را ارائه دهند. برای مثال، اگر کاربر برنامه زمانی

برای دیدن یک فیلم در یک تاریخ مشخص دارد، می‌توان از سیری خواسته شد وارد حالت Do Not Disturb شود. زمان کاربر و اطلاعات مکان‌یاب مشخص می‌کند که کاربر واقعاً در داخل تئاتر است یا خیر.

۵- پروژه‌های کاربرد تشخیص گفتار فیسبوک:

پیش بینی می‌شود صنعت کاربرد تشخیص گفتار ۵۵ میلیارد دلاری، با نرخ ۱۱ درصد از سال ۲۰۱۶ تا ۲۰۲۴ رشد کند.

این فناوری در قالب سایر برنامه‌های رونویسی در سایر صنایع در بین بنگاه‌های کوچکتر و کمتر شناخته شده کاربرد خوبی پیدا کرده است. در حال حاضر در مراقبت‌های بهداشتی، متخصصان پزشکی از برنامه‌های رونویسی متن به متن مانند Dolbey برای ایجاد سوابق پزشکی الکترونیکی برای بیماران استفاده می‌کنند.

در بخش‌های اجرای قانون و بخش‌های حقوقی، شرکت‌هایی مانند Nuance از برنامه‌های رونویسی برای تهیه اسناد دقیق و سریع و برای مستندسازی گزارش‌های حادثه استفاده می‌کنند. در رسانه‌ها، روزنامه‌نگاران از برنامه‌های رونویسی مانند Recordly به عنوان ابزاری برای ضبط و انتقال اطلاعات به منظور ثبت گزارش‌های دقیق‌تر اخبار استفاده می‌کنند. در آموزش، Sonix به پژوهشگران کمک می‌کند تا مصاحبه‌های کیفی خود را تغییر و بهبود بخشند.

نمونه فارسی کاربرد تشخیص گفتار، می‌توان به **فارس آوا** اشاره نمود. فارس آوا برنامه‌ای است که اختصاصاً برای زبان فارسی تهیه شده و می‌تواند همانند نسخه خارجی آن، در رسانه‌ها، دادگاه‌ها، استارت‌آپ‌ها، مدارس و دانشگاه‌ها و غیره مورد استفاده قرار گیرد.

مزایا تشخیص گفتار:

- افزایش بهره‌وری کسب و کار؛
- تعامل بین مشاغل و مشتریان را خودکار می‌کند.
- یک سطح امنیتی اضافی اضافه می‌کند.
- گفتار را سریعتر از آنچه انسان می‌تواند تایپ کند ضبط می‌کند.

- به افراد دارای معلولیت کمک می کند؛
- به کنترل دستگاه های خانه شما کمک می کند.
- به رانندگان با سیستم های ASR داخل خودرو و موارد دیگر کمک می کند

معایب تشخیص گفتار:

- اگر گوینده سریع و نه واضح صحبت کند، سیستم ها نمی توانند به طور کامل گفتار را تشخیص دهند.
- برای بهبود دقت تشخیص به واژگان بزرگ نیاز است.
- هر زبان برای ASR نیاز به آموزش جداگانه دارد.
- کسب و کارها می توانند داده های صوتی کاربر را بدون اجازه آنها جمع آوری و استفاده کنند.
- هزینه های زمانی و مالی بالاست.
- نرم افزار ASR حافظه زیادی مصرف می کند و به مقدار زیادی رم نیاز دارد