



سیگنال و سیستم ها در صوت

پروژه‌ی درس تجزیه و تحلیل سیستم ها

استاد درس: دکتر سایه میرزایی

دانشجو: مینو احمدی

810897032

دانشکده علوم مهندسی، دانشگاه تهران

Miinouahmadii@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 صداها 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 ...

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 صداها 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. به طور کلی بیشترین حوزه تشخیص صوت در تشخیص صدای انسان و گفتار است. یک سیستم تشخیص صوت پایه نیاز به 22 مگاهرتز پردازنده، حداقل 64 مگابایت رم، یک میکروفن پایه و یک کارت صدای حداقل 46 بیتی نیاز دارد. افزایش ساینز 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 نایل می آیند. 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. 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 صداها 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 میشوند و تبدیل به نمایش دیجیتالی در می آیند. مدل های 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 لغت های موجود در دایره المعارف می پردازد که بزرگترین 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.

تاریخچه ی توسعه سیستم های تشخیص صوت

پیشرفت تکنولوژی تشخیص صوت در 5 دهه ی گذشته را می توان به شرح زیر سازمان دهی کرد.

نسل اول: اولین تلاش در سال های 1551 و 1541

نسل دوم : فن اوری مبتنی بر قالب در اواخر 1541 و 1591

نسل سوم: فن اوری های مبتنی بر مدل هامی اماری در 1591

آخرین نسل سوم: پیشرفت در 1551 و 2111

قدیمی ترین تلاش:

اولین تلاش برای ابداع سیستم تشخیص صوت در سال های 1551 و 1541 صورت گرفت. زمانی که محققان مختلف به تلاش برای بهره برداری از ایده های اساسی مرتبط با آشناسی اکوستیک 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 مدارهای منطقی استخراج می شد.

سیستم های اولیه :

در سال 1552 و در آزمایشگاه بل در امریکا محققان برای ایزوله سازی تشخیص رقمی را برای یک سخنران، با استفاده از اندازه گیری فرکانس های فونمت و برای هر یک سیستم ساختند. در یک تلاش که در سال 1541 و در یک آزمایشگاه دیگر در امریکا صورت گرفت، محققان به تلاش برای به رسمیت شناختن ده سیلاب یا هجای مجزا در یک رشته گفتار پرداختند .با وارد کردن اطلاعات اماری مربوط به توالی واج مجاز در زبان انگلیسی، آنها دقت تشخیص واج را برای کلمات متشکل از دو یا چند واج افزایش داده اند . این فعالیت اولین استفاده از ترکیب نحوی اماری در تشخیص گفتار بوده است. در سال 1541 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 1541 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 1541 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.

تشخیص گفتار پیوسته:

در اواخر 1541 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.

تشخیص کلمات متصل:

مشکل ایجاد یک سیستم قوی بود تا بتواند رشته ای از کلمات متصل به هم را تشخیص دهد که تمرکز تحقیقات در سال های 1591 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.

مدل سازی اماری :

تحقیقات انجام شده در 1591 با یک تغییر در روش از روش مبتنی بر الگو به سمت مدلسازی اماری شخصیت جدیدی یافت. بسیاری از سیستم های تشخیص صوت امروزی مبتنی بر مدل هاو فریم ورک های اماری هستند که در سال 1591 بنا نهاده و توسعه داده شد. یکی از کلیدی ترین تکنولوژی هایی که در سال 1591 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 فرایندهای تصادفی که دنباله ای از مشاهدات را تولید میکنند، آنرا مشاهده نمود.

الگوریتم های تشخیص صوت

علاوه بر این حداقل نیازمندیهای سخت افزاری، سیستمهای تشخیص صوت نیاز به نرم افزار دارند تا بتواند دادهها را جمع اوری، تحلیل، تفسیر نمایند. نرم افزارهای متفاوت از روشهای متفاوتی به این اهداف نایل 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. 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، قاب بندی، 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 :

اماده 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.

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 معمولاً توسط یک نمونه بردار بنام Hold&Sample انجام میشود.

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.

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.

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 99 % 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 مکرراً 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.

مبنتی 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 اصلیتزین ابزارهای ریاضی برای حل این مسئله مدل های پنهان مارکوف هستند. برای حل این مسئله با استفاده از مدل پنهان مارکوف این مدل های اماری 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 حقیقت 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 مجموعه های 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 درهای 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 میگردد و نهایتاً 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.

مدل های ترکیبی:

از مدل اصلی کانال منبع و یا بخشی و یا نوعی از مدل های اماری مولد معمولاً 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 بلندگوی 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 ال 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 سایر بخشهای سیستم وقف دهند یا سازگار نمایند.

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 بال و عملکرد بال درنگ که در برابر نویزهای محیطی مقاوم است و فرایند آموزش آن زمان بر است .

مدل اکوستیک :

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 تغییرات زمینه، تغییرات سخن گو و تغییرات محیط را برشمرد. .. مدل سازی اکوستیک گفتار به طور معمول برمی گردد به روند ایجاد بازنمود ها یا نمایش های اماری برای دنباله ای از بردار های ویژگی محاسبه شده از شکل موج گفتار مدل سازی اکوستیک 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 اکوستیک هم چنین شامل استفاده از اطلاعات باز خورد از تشخیص دهنده باشد تا بردار های ویژگی های کلام در محیط های پر سر و صدا را تغییر شکل دهد مدل صوتی یا اکوستیک شامل بخش هایی از علم در مورد صدا ها و اکوستیک، فونوتیک یا اوا شناسی، تنوع زیست محیطی و تفاوت های مهم در نوع گویش سخنران است. مدل زبان مربوط به دانش یک سیستم 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 اکوستیک ناشی از لهجه های متفاوت و سبک های صحبت کردن هر فرد، توسط پیچیدگی ها و اختلافات لغوی و گرامری ناشی از تغییرات زبان محاوره ای که در مدل زبانی وجود دارند، تشدید می شود، سیگنال های گفتار در مازول های پردازش سیگنال پردازش می شوند که در این مرحله بردار ویژگی های برجسته برای رمز گشا، استخراج می شود. رمزگشا از هر دو مدل اکوستیک 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 WP در معادله اصلی گفتار است. این مدل یکی از انواع مدل های زبانی گرامر با دستور زبان است که یک ساختار رسمی و مجاز برای زبان به حساب 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 دارای ساختار مختص به خود هستند، امروزه امکان تعمیم دستور زبان اصلی وجود دارد. علاوه بر این روابط احتمالی میان دنباله ها را میتوان به صورت مستقیم بدست آورد.

Corpora یا 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.

در بخش زیر مثال کاربردی از تشخیص صورت را به کمک متلب شبیه سازی کرده ایم فایل های ویس ها = در فایل تکمیلی پروژه آورده شده است.

مثال : استخراج نت ها پیانو

در این مثال میخواهیم با توجه به دانشی که از فرکانس ها داریم نت های پیانو را بررسی کنیم .

الفبای موسیقی نت نام دارد نت های موسیقی شامل ۷ عدد میباشد که به شکل زیر از سمت چپ به راست خوانده میشود .

سی	لا	سل	فا	می	ر	دو
si	la	sol	fa	mi	re	do
B	A	G	F	E	D	C

هر نت موسیقی در حالت تیز : دییز یا شارپ خوانده میشود و نشان دهنده فرکانس بالاست.

هر نت موسیقی در حالت بم : بمل یا فلت خوانده میشود و نشان دهنده فرکانس پایین است.

میخواهیم دستگاهی بسازیم که بتواند در تنظیم و بررسی درستی دکمه های پیانو و تشخیص نت ها به ما کمک کند ما انتظار داریم تا یک رشته متن شامل ادرس فایل صوتی به عنوان ورودی نرم افزار بدهیم و سپس نرم افزار نت خروجی را نشان دهد.

فایل شامل جدول نام و فرکانس نت های موسیقی با نام `notes.csv` و فایل صوتی با نام `test.wav` ورودی های ما خواهند بود .

ابتدا داده های صوتی و فرکانس نمونه برداری را از فایل صوتی با دستور `audioread` می خوانیم سپس داده های کانال های صوتی چپ و راست را باهم جمع و در یک متغیر جداگانه ذخیره میکنیم . نمودار فایل صوتی را با دستور `plot` رسم کرده و در گام بعد با استفاده از تبدیل فوریه نمودار حوزه فرکانس را برای فرکانس های در بازه $[100, fs/2]$ هرتز رسم و فرکانس نت را پیدا میکنیم با استفاده از دستور `readtable` جدول شامل نام نت ها را میخوانیم و با استفاده از آن نام نت مورد نظر را استخراج میکنیم. در واقع برای این کار ابتدا محتوای فایل `test.wav` را میخوانیم و ذخیره می کنیم و نمودار زمانی فایل صوتی را رسم می کنیم .

برای رسم نمودار فرکانس، باید داده نسبت به زمان در فایل صوتی را به کمک محاسبه سری فوریه به داده نسبت به فرکانس تبدیل کنیم و سپس نمودار را با داده ی جدید رسم کنیم .

برای پیدا کردن نت فرکانس صدا، پس از خواندن اطلاعات فایل `notes.csv` با پیمایش تک تک فرکانس ها و با تقریب خوب 0.01 نام نت مورد نظر D7 بدست می آید.

داده های داخل فایل note.csv در زیر آمده است.

notes		
Note	Frequency_Hz	Wavelength_cm
C0	16.35	2109.89
C#0/Db0	17.32	1991.47
D0	18.35	1879.69
D#0/Eb0	19.45	1774.20
E0	20.60	1674.62
F0	21.83	1580.63
F#0/Gb0	23.12	1491.91
G0	24.50	1408.18
G#0/Ab0	25.96	1329.14
A0	27.50	1254.55
A#0/Bb0	29.14	1184.13
B0	30.87	1117.67
C1	32.70	1054.94
C#1/Db1	34.65	995.73
D1	36.71	939.85
D#1/Eb1	38.89	887.10
E1	41.20	837.31

خروجی کد:

note frequency: 2366.389663
sample note is: 'D7'

پس نت مورد نظر D7 تشخیص داده شد.

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. امروزه 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، بسیاری 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 تشخیص 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.

منابع

- [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] مقاله مروری بر سیستم های تشخیص صوت، مفاهیم و روش ها