



دانشگاه زنجان
دانشکده مهندسی

گروه برق

پایان نامه کارشناسی

گرایش: مخابرات

عنوان:

شناسایی صوت

استاد راهنما: دکتر اصغر ظاهری

نگارش: مسلمی علیرضا

شهر یور 92

1392/06/03

Voice

Recognition

شناسایی صوت

استاد راهنما: دکتر اصغر طاهری

Alireza Moslemi
ZANJAN UNIVERSITY

فهرست

مقدمه

3

فصل اول: زیست‌سنجی و کاربردهای آن در سیستم‌های امنیتی

5

تعریف، ضرورت و کاربردها

5

بررسی عملکرد سیستم‌های موجود

6

اجزای سیستم‌های زیست‌سنجی

7

ارزیابی کارایی سیستم‌های امنیتی مبتنی بر زیست‌سنجی

8

منابع فصل

10

فصل دوم: سیستم‌های امنیتی مبتنی بر تشخیص گوینده

11

تعریف و کاربردها

11

انواع سیستم‌های تشخیص گوینده

11

روش‌های پیاده‌سازی

13

منابع فصل

14

فصل سوم: پردازش صوت پیش‌زمینه‌های تئوری

14

دستگاه شنوایی انسان

14

ویژگی‌های امواج صوتی

19

روش‌های دیجیتالی ذخیره صدا

21

منابع فصل

22

فصل چهارم: پردازش صوت برنامه‌نویسی و پیاده‌سازی

24

ساختار مورد نیاز برای نگهداری ویژگی‌های صدا

24

انجام پردازش صدا به صورت یک رشته مستقل

29

ضبط صدا

37

پخش صدا

46

کتابخانه پردازش صوت

52

منابع فصل

53

فصل پنجم: پردازش صحبت

53

ترکیب و تشخیص صحبت

53

54

مدلی برای توصیف روش تولید صحبت

57

آینده فن آوریهای پردازش صحبت

58

منابع فصل

58

فصل ششم: مدلسازی سیگنال

58

اهمیت مدلسازی سیگنال

59

آشنایی با مدلسازی سیگنال

60

تشخیص الگو

61

الگوریتمهای مدلسازی سیگنال

64

منابع فصل

64

فصل هفتم: روشهای طراحی سیستمهای تشخیص گوینده

64

مقدمه

65

روشهای مبتنی بر چشمپوشی زمانی پویا

66

روشهای مبتنی بر مدلهای نهان مارکف

68

روشهای مبتنی بر مقدارگزینی برداری

70

مقایسه کارایی

72

منابع فصل

72

فصل هشتم: کتابخانه تشخیص گوینده

72

کتابخانه تشخیص صحبت - گوینده جیلانگ هی

73

پایگاه دادههای عبارتهای عبور - شمای کلی

74

عبارت عبور و پردازشهای مربوط به آن

82

چند عبارت عبور برای یک کاربر و پردازشهای مربوط به آن

85

ایجاد یک جدول از کاربران

92

تشکیل و کار با پایگاه دادهها

100

منابع فصل

100

ضمیمه اول

104

ضمیمه دوم

مقدمه

صنعت کامپیوتر که راهبری سیستمهای اطلاعاتی، مالی و امنیتی نوین را بر عهده دارد برای اعمال کنترل صحیح بر نحوه دسترسی به سیستمهای یاد شده توسط کاربران، سالهاست که از سیستم ساده نام کاربر و کلمه عبور استفاده می‌نماید. با وجود آن که سیستم مزبور از نظر پیاده‌سازی بسیار ساده و کم‌هزینه می‌باشد، به این دلیل که تضمین نمی‌کند شخص دارنده یک شناسه مجاز همان شخصی است که شناسه مزبور برای استفاده اختصاصی او در نظر گرفته شده، امنیت مورد انتظار برای سیستمهای مهم را تأمین نمی‌نماید. آمار رو به‌فزون تقلبهای کامپیوتری به صورت گذر غیر مجاز از سیستم یاد شده نشانگر این عدم تواناییست که انگیزه جستجوی جایگزینی مطمئن برای آن را در ذهن متخصصان امر پرورده است. زبان و اسکند مهندسی گروه برق آزمایشگاه پروژه برق دانشگاه زنجان دانشجوئی مطمن برای آن از سوی دیگر تکنیک استفاده از ویژگیهای منحصر به فرد زیستی انسانها برای بازشناسی آنان دانشجوئی گروه برق آزمایشگاه پروژه برق دیرزمانیست که در حیطه جرم‌شناسی به لحاظ باور علمی این مطلب که این ویژگیها غیر قابل تقلید بوده، احتمال مشابهت آنها در افراد صفر یا عددی مشابه آن است به عنوان ابزاری مطمئن و کارآمد مطرح می‌باشد و متخصصان طراحی سیستمهای امنیتی الکترونیکی نیز در دهه‌های اخیر به این فن‌آوری به عنوان مهم‌ترین اساس برای طراحی سیستمهای امنیتی وابسته به فرد خاص نظر داشته‌اند.

اهمیت موضوع یاد شده از یک سو و تازگی آن از سوی دیگر تهیه‌کننده این نوشتار را بر آن داشت موضوع پروژه پایانی خود را مطالعه بر روی این گونه سیستمها و طراحی یک سیستم امنیتی نمونه بر اساس ویژگیهای منحصر به فرد صوتی افراد انتخاب نماید که دستاورد آن یک سیستم تشخیص گوینده وابسته به متن به عنوان نتیجه عملی و این نوشتار به عنوان فراهم آورنده پیش‌زمینه‌های علمی و توصیف‌کننده نحوه عملکرد آن می‌باشد.

پیش از هر چیز تأکید بر این مطلب ضروری به نظر می‌رسد که این نوشتار نه به عنوان یک مرجع برای موضوع طراحی سیستمهای تشخیص گوینده بلکه به عنوان راهنمایی برای افراد علاقمند به ادامه این کار فراهم آمده است و از این لحاظ سعی شده که به جای پوشش تمامی روشهای ممکن و ارائه الگوریتمهایی که در پروژه عملی کاربردی نداشته‌اند از آنها به صورت اشاره‌وار یاد شود و در مهندسی گروه برق آزمایشگاه پروژه برق عوض زمینه‌های تئوری و روشهای برنامه‌نویسی عملاً استفاده شده، به صورت مشروح‌تر بیان شده‌اند. ذکر فهرست منابع هر فصل در پایان آن، توضیح کد توابع استفاده شده و ارائه پیش‌زمینه‌های نظری و روشهای برنامه‌نویسی برای پردازش صوت به صورت ابزارهایی برای تحقق این هدف به کار گرفته

شده‌اند که به نظر می‌رسد نمی‌توانند جایی در بین مطالب یک مرجع برای مطلب یاد شده داشته باشند.

بدنه اصلی نوشتار از سه بخش تشکیل شده است. بخش اول مقدمه‌ای است بر اهمیت و کاربردهای سیستم‌های زیست‌سنجی و به طور خاص سیستم‌های تشخیص گوینده. این بخش شامل دو فصل

است و مطالب آن عمدتاً مقدمات لازم برای بخش سوم نوشتار را فراهم می‌آورد. در بخش دوم پردازش صوت را به صورت نظری مورد بررسی قرار خواهیم داد و سپس مطالب نظری یاد شده را در قالب یک کتابخانه برای پردازش صوت از دیدگاه برنامه‌نویسی پیاده‌سازی خواهیم نمود. این بخش تا

حد زیادی مستقل از دو بخش دیگر نوشتار می‌باشد ولی به لحاظ نوع مخاطب و هدف این نوشتار پرداختن به آن ضروری به نظر می‌رسد. بخش یاد شده نیز شامل دو فصل می‌باشد. بخش آخر که

مرتبط‌ترین بخش نوشتار با سیستم عملی است و در بردارنده چهار فصل است پردازش صحبت را از

دیدگاه تئوری و برنامه‌نویسی مورد بحث قرار می‌دهد. بحث با فصلی در مورد پردازش سیگنال صحبت که در واقع به نوعی می‌تواند ادامه بخش قبل تلقی گردد آغاز می‌گردد و در ادامه مدلسازی

سیگنال به عنوان شیوه‌ای برای استخراج الگوهای قابل مقایسه از سیگنال صحبت مطرح می‌شود و در آخرین فصل نظری این نوشتار روشهای مختلف معمول برای طراحی سیستم‌های تشخیص

گوینده بررسی می‌شوند. نهایتاً در آخرین فصل این نوشتار یک کتابخانه نمونه تشخیص گوینده که در پیاده‌سازی قسمت عملی مورد استفاده قرار گرفته است ارائه می‌گردد.

در ضمیمه اول این نوشتار شیوه نصب و روش استفاده از سیستم پیاده‌سازی شده مطرح گردیده است و ضمیمه دوم آن ارائه‌دهنده فهرست کاملی از منابع متفرقه اینترنتی یافت شده که بسیاری از

آنها به دلایل مختلف در این پروژه مورد استفاده قرار نگرفته‌اند برای خوانندگانی که قصد انجام کاری مشابه با این پروژه را دارند می‌باشند.

با توجه به آن که منابع عمده این نوشتار، عمدتاً انگلیسی زبان می‌باشند ارائه معادلهای مناسب

حتی‌الامکان فارسی برای اصطلاحات علمی که در منابع فارسی معادلی برای آنها یافت نشده سرلوحه کار فراهم آورنده این نوشتار قرار داشته و به منظور انتقال هر چه بهتر مطلب سعی شده با ارائه

پاورقیها یا توضیحات اضافی، مخاطب نوشتار با اصل اصطلاحات آشنا شود. علاوه بر آن در مواردی که ارائه واژه به صورت انگلیسی مناسب‌تر تشخیص داده شده واژه با حروف فارسی در متن اصلی به

کار برده شده تا تناسب متن حفظ گردد. اغلب اصطلاحات مخفف شده به صورت اخیر آورده شده‌اند. البته در فصلهای مرتبط با برنامه‌نویسی به لحاظ نوع مطلب شیوه دیگری متناسب با محتوای آنها به

کار گرفته شده است. در مجموع ضمن آن که تلاش شده مطلب به صورت واضح بیان گردد

فراهم آورنده این نوشتار تلاش برای ارائه اصطلاحات درست فارسی و رعایت قوانین این زبان توانمند را همواره در نظر داشته است.

با وجود آن که طراحی سیستمهای تشخیص گوینده مدتهاست مد نظر کارشناسان قرار دارد و با وجود ارائه نتایج بسیاری از این تحقیقات به صورت سیستمهای تجاری، هنوز سیستمی که بتواند مستقل از محیط و نوع آموزش کاربران عملکرد مناسبی داشته باشد ارائه نشده و این زمینه هنوز هم به عنوان یک افق علمی باز و دارای زمینه تحقیقاتی فراوان مطرح می باشد. با وجود آن که محدودیتهای زمانی و ... مانع از ارائه یک کار دلخواه فراهم آورنده این نوشتار و استاد راهنمای محترم شد امیدواریم این نوشتار بتواند مقدمه ای برای انجام کارهای کامل تر فراهم آورد ضمن آن که علاوه بر موضوع طراحی سیستمهای تشخیص گوینده بسیاری از مطالب این نوشتار می توانند مقدمه ای کاملاً ارضا کننده و مناسب برای افراد علاقمند به طراحی سیستمهای تشخیص صحبت فراهم آورند.

امید است که تلاشهای فراهم آورنده این نوشتار و زحمات بی دریغ استاد راهنمای محترم بتواند پیش زمینه های لازم برای ارائه کارهای کامل تر در نوع خود را فراهم آورد.

فصل اول: زیست سنجی و کاربردهای آن در سیستمهای امنیتی تعریف، ضرورت و کاربردها

زیست سنجی عبارت است از دانش و فن آوری اندازه گیری و تحلیل آماری داده های زیستی. در فن آوری اطلاعات واژه زیست سنجی به مجموعه فن آوریهایی اطلاق می گردد که در آنها از اندازه گیری و تحلیل ویژگیهایی از بدن انسان همچون اثر انگشت، اثر کف دست، شبکیه و عنبیه چشم، الگوهای صوتی، الگوهای مربوط به رخسار، دمانگاری صورت، شکل دست یا گوش، داده های به دست آمده از گام، الگوهای ویدی، دی.ان.ای و یا ویژگیهایی همچون دستخط (امضا) و دینامیک ضربه زدن به صفحه کلید برای تأیید هویت اشخاص استفاده می شود. این فن آوریها در تلاشند تا اندازه گیری و مقایسه ویژگیهای برشمرده شده را به منظور بازشناسی افراد به صورت خودکار درآورند.

فن آوریهای زیستی در ابتدا برای کاربردهای تخصصی نیازمند امنیت بالا پیشنهاد شدند اما اینک به عنوان عناصر کلیدی در توسعه تجارت الکترونیک و سیستمهای برخط و به همان صورت برای سیستمهای امنیتی نابرخط و سیستمهای امنیتی منفرد مطرح می باشند.

این فن‌آوریها اجزاء مهمی را برای تنظیم و نظارت بر نحوه دسترسی و حضور در سیستم فراهم

می‌آورند. محدوده‌های عمده کاربرد این فن‌آوریها عبارتند از: تجارت الکترونیک، نظارت امنیتی،

دسترسی به پایگاه داده‌ها، کنترل مرزها و مهاجرت، تحقیقات قضایی و پزشکی از راه دور.

توسعه فن‌آوریهای زیست‌سنجی فراتر از کاربردهای سنتی نیازمند امنیت بالا، یک اجبار نشأت گرفته

از انگیزه‌های مالی است. امنیت معاملات برای آینده توسعه تجارت الکترونیک یک مسأله حیاتی است

و نگرانیهای فراوانی درباره راه‌های فعلی وجود دارد. مشکل شماره‌های شناسایی شخصی و

شناسه‌های هویتی - مانند کارتها - این است که آنها صحت هویت شخصی را که از آنها استفاده

می‌کند تأیید نمی‌کنند. آمارها میزان زیان ناشی از تقلب را به طور سالانه برای کارتهای اعتباری

بالغ بر چهارصد و پنجاه میلیون دلار و برای خودپردازها حدود سه میلیارد دلار برآورد می‌کنند.

برتری سیستمهای مبتنی بر زیست‌سنجی آن است که به شدت به ویژگیهای فردی اشخاص

وابسته‌اند و به راحتی نمی‌توانند مورد سوء استفاده قرار گیرند.

بررسی عملکرد سیستمهای موجود

فعالیت‌های انجام شده تا به حال منجر به ظهور ماشینهای گران قیمت زیست-سنجی شده است که

علاوه بر قیمت زیاد معمولاً از لحاظ سرعت و عملکرد مناسب نیستند یا حداقل برای دستیابی به

عملکرد مناسب باید محیط استفاده آنها شرایط خاصی را داشته باشد و یا کاربران آنها آموزشهای

گسترده‌ای را گذرانده باشند.

در حالی که بعضی از فن‌آوریهای زیست‌سنجی در قالب تولیدات تجاری به بازار عرضه شده‌اند

بسیاری از این دسته فن‌آوریها در مرحله تحقیق و آزمایش قرار دارند. فن‌آوریهای مزبور نیازمند

کارهای مطالعاتی بیشتر برای افزایش پایداری و بهبود عملکردشان برای استفاده در کاربردهای ویژه

هستند.

پایداری در برابر تقلب، دقت عملکرد، سرعت و تجهیزات مورد نیاز، همخوانی با سخت‌افزار و نرم‌افزار

موجود، هزینه، سادگی استفاده و پذیرش از سوی کاربر از جمله عوامل تعیین‌کننده در موفقیت هر

یک از فن‌آوریهای به کار گرفته شده می‌باشند.

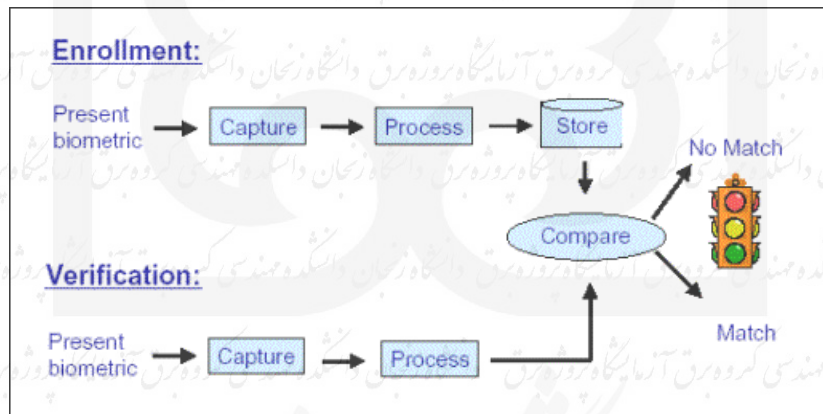
VOICE RECOGNITION

نوع سیستم	دقت عملکرد	سادگی استفاده	میزان پذیرش کاربر
اثر انگشت	بالا	متوسط	پایین
هندسه دست	متوسط	بالا	متوسط
صوت	متوسط	بالا	بالا
شبکیه چشم	بالا	پایین	پایین
چشم مصنوعی	متوسط	متوسط	متوسط
امضا	متوسط	متوسط	بالا
رخسار	پایین	بالا	بالا

جدول شماره 1- مقایسه سیستمهای زیست‌سنجی معمول (منبع شماره 1)

اجزای سیستمهای زیست‌سنجی

عملیات سیستمهای زیست‌سنجی در بر دارنده دو مرحله مجزا می‌باشد: ثبت کاربر و بازشناسی کاربر. در مرحله اول اطلاعات مربوط به کاربر به سیستم وارد می‌شوند و در مرحله دوم اطلاعات ورودی حاضر با اطلاعات ذخیره شده مقایسه می‌گردند.



شکل شماره 1- مراحل لازم عملیاتی در یک سیستم امنیتی مبتنی بر زیست‌سنجی (منبع)

(شماره 3)

مرحله تأیید هویت عبارت است از تطبیق ویژگیهای مورد ادعای یک شخص بر ویژگیهای موجود او گروه برق آزمایشگاه پروژه برق

سیستم‌های امنیتی مبتنی بر زیست‌سنجی بنا به انتخاب به وجود آورنده، به جای مرحله تأیید هویت می‌توانند مرحله دیگری را که بازشناسی نامیده می‌شود جایگزین کنند. در این روش نیاز نیست که درخواست کننده ادعای هویت شخص خاصی را بنماید بلکه سیستم ویژگی‌های او را با تمامی رکوردهای موجود مقایسه می‌کند و در صورت تطابق با یکی از آنها او را به عنوان شخص دارای ویژگی‌های موجود در رکورد یافت شده بازشناسی می‌کند که این فرایند یک پردازش یک به چند را شکل می‌دهد.

سیستم‌های تشخیص هویت زیستی معمولاً شامل اجزای زیر می‌باشند:

(الف) گیرنده اطلاعات: زیرسیستمی است که گرفتن نمونه‌های زیست‌سنجی (صوتی، تصویری و...) را بر عهده دارد. ویژگی‌های خاص استخراج شده از نمونه‌ها قالب‌هایی را برای مقایسه بعدی تشکیل می‌دهند. این فرایند باید سریع و ساده بوده در عین حال قالب‌هایی با کیفیت خوب را تولید کند.

(ب) ذخیره کننده: قالب‌های به دست آمده باید برای مقایسه بعدی ذخیره شوند. این زیر سیستم می‌تواند جزئی از وسیله گیرنده اطلاعات سیستم باشد و یا در یک سرور مرکزی قابل دستیابی توسط یک شبکه جای گیرد. جایگزین دیگر، یک شناسه قابل حمل نظیر یک کارت هوشمند است. هر کدام از انتخاب‌های فوق مزایا و مشکلات خاص خود را دارد.

(ج) مقایسه گر: اگر سیستم زیست‌سنجی در مقام بازشناسی افراد به کار گرفته شود باید هویت شخص با قالب ذخیره شده مورد ادعای او مقایسه شود. در بعضی سیستم‌ها ممکن است امکان بروز آوری خودکار قالب مورد مراجعه پس از هر تطبیق درست وجود داشته باشد. این امر به سیستم توانایی سازگاری با تغییرات تدریجی کوچک در ویژگی‌های کاربر را می‌دهد.

(د) اتصالات: غالباً برای ایجاد ارتباط بین گیرنده اطلاعات، ذخیره کننده و مقایسه گر نیاز به اتصالات لازم وجود دارد. غالباً سیستم‌های زیست‌سنجی نیازمند شبکه و رابط‌های برنامه‌نویسی مورد نیاز برای ایجاد اتصال بین اجزاء می‌باشند. امنیت و کارایی، عناصر کلیدی برای این جزء می‌باشند.

ارزیابی کارایی سیستم‌های امنیتی مبتنی بر زیست‌سنجی

موضوع مهمی که در پذیرش سیستم‌های زیست‌سنجی از اهمیت شایان توجهی برخوردار است تعیین کارایی هر یک از اجزاء و کل سیستم زیست‌سنجی به روشی قابل اعتماد و هدفمند است. برای تعیین کارایی سیستم‌های امنیتی مبتنی بر زیست‌سنجی معیارهای ویژه‌ای به کار گرفته می‌شوند. در این کاربردها تعدادی کاربر (سرویس گیرنده) به سیستم وارد می‌شوند و متقلب به عنوان

شخصی تعریف می شود که مدعی هویت شخص دیگری است. متقلب ممکن است به عنوان کاربر در

سیستم وجود داشته باشد و عمل وی ممکن است عمدی یا غیر عمدی باشد. عمل تأیید هویت باید کاربران را بپذیرد و متقلبان را رد کند.

نرخ پذیرش نادرست (اف. ای. آر) به عنوان نسبت تعداد متقلبانی که به اشتباه توسط سیستم

پذیرفته شده اند به تعداد کل متقلبان آزمایش شده تعریف گردیده، به صورت درصد بیان می شود. نرخ پذیرش نادرست (اف. ای. آر) به عنوان نسبت تعداد کاربران سیستم که به اشتباه توسط این نرخ، احتمال پذیرش متقلبان را توسط سیستم بیان می کند و باید در سیستمهای نیازمند امنیت بالا کمینه شود.

نرخ عدم پذیرش نادرست (اف. آر. آر) به عنوان نسبت تعداد کاربران سیستم که به اشتباه توسط

سیستم پذیرفته نشده اند به تعداد کل کاربران مورد آزمایش قرار گرفته تعریف گردیده، به صورت درصد بیان می شود. این نرخ، احتمال عدم پذیرش کاربران مجاز را توسط سیستم بیان می کند و باید

به صورت ایده آل مخصوصاً در سیستمهایی که در آنها کاربر در صورت عدم پذیرش از دسترسی به سیستم محروم می شود کمینه گردد.

روند تشخیص هویت مبتنی بر زیست سنجی در بردارنده محاسبه فاصله قالب ذخیره شده و نمونه

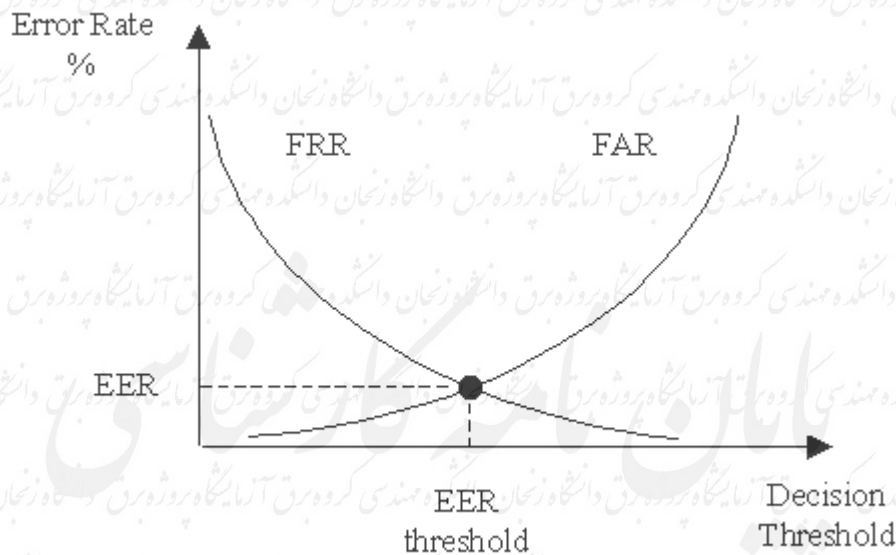
حاضر است. تصمیم برای پذیرش یا رد نمونه حاضر بر اساس یک آستانه از پیش تعریف شده اتخاذ می گردد. بنابراین واضح است که کارایی سیستم به شدت وابسته به انتخاب این آستانه است و این

امر موجب ایجاد یک بدهستان بین نرخ پذیرش نادرست و نرخ عدم پذیرش نادرست می گردد. نرخ

خطای برابر (ای. ای. آر) به صورت آستانه برابری این دو نرخ تعریف می شود و غالباً به عنوان یک

ویژگی نشان دهنده کارایی سیستم مطرح می گردد. شکل شماره 2 نشان دهنده رابطه سه پارامتر

تعریف شده برای یک سیستم نمونه است.



شکل شماره 2- FAR ، FRR و ERR برای یک سیستم نمونه (منبع شماره 1)

پارامتر مهم دیگر کارایی، زمان تشخیص هویت است که به صورت زمان متوسط صرف شده برای

فرایند تشخیص هویت تعریف می‌شود. این زمان شامل زمان لازم برای گرفتن نمونه حاضر نیز می‌باشد.

در حالی که بعضی از عرضه‌کنندگان سیستم‌های امنیتی مبتنی بر زیست‌سنجی برای محصولاتشان پارامترهای کارایی فوق را در شرایط آزمایشگاهی بیان می‌کنند پارامترهای کارایی قابل طرح در

جهان واقعی برای سنجش کارایی واقعی این گونه سیستمها به ندرت وجود دارند. علت این امر این واقعیت است که به حساب آوردن همه پیچیدگیهای ممکن جهان واقعی تأثیر گذار بر سیستمهای

زیست‌سنجی تقریباً غیر ممکن است. به عنوان نمونه زمان واقعی تشخیص هویت به شدت وابسته به زمان آموزش کاربر، محیط عملیاتی و شرایط روانی کاربر همچون میزان فشار روحی اوست. مشخصات ارائه شده توسط عرضه‌کننده را باید به دید راهنماهای نه چندان متناسب با دنیای واقعی

نگریست.

منابع فصل

- 1) Farzin Deravi, University of Kent at Canterbury, Audio-Visual Person Recognition for Security and Access Control, from www.jtap.uk.ic
- 2) The BioAPI Consortium, BioAPI Specification Version 1.1 – March 16, 2001, from www.bioapi.org

دانشجویان محترم:

جهت دسترسی به متن کامل پایان نامه‌ها به کتابخانه دانشکده مهندسی و یا آزمایشگاه پروژه گروه برق مراجعه فرمایید.

1) Steven W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, Chapter 22: Audio Processing, from www.dspguide.com

کتابخانه نامہ کارستانی



منابع فصل

1) Microsoft®, MSDN Library (January 2000 Edition)

2) Thomas Holme, How to play and record sound, from
<http://www.codeproject.com>

فصل پنجم: پردازش صحبت

ترکیب و تشخیص صحبت

کاربردهای نیازمند پردازش صحبت اغلب در دو دسته ترکیب صحبت و تشخیص صحبت مورد

بررسی قرار می گیرند.

ترکیب صحبت عبارت است از فن آوری تولید مصنوعی صحبت به وسیله ماشین و به طور عمده از نگاه زنجان و اسکده مهندسی گروه

پرونده های متنی به عنوان ورودی آن استفاده می گردد. در اینجا باید به یک نکته مهم اشاره شود که بسیاری از تولیدات تجاری که صدای شبیه به صحبت انسان ایجاد می کنند در واقع ترکیب صحبت

انجام نمی دهند بلکه تنها یک تکه ضبط شده به صورت دیجیتالی از صدای انسان را پخش می کنند.

این روش کیفیت صدای بالایی ایجاد می کند اما به واژه ها و عبارات از پیش ضبط شده محدود است.

از کاربردهای عمده ترکیب صحبت می توان به ایجاد ابزارهایی برای افراد دارای ناتوانی بینایی برای

مطلع شدن از آنچه بر روی صفحه کامپیوتر می گذرد اشاره کرد.

تشخیص صحبت عبارت است از تشخیص کامپیوتری صحبت تولید شده توسط انسان و تبدیل آن به

یک سری فرامین یا پرونده های متنی. کاربردهای عمده موجود برای این گونه سیستمها دربرگیرنده

بازه گسترده ای از سیستمها و کاربردها از سیستمهای دیکته کامپیوتری که در سیستمهای آموزشی و

ترکیب صحبت علاوه بر آن که همانند تشخیص صحبت می‌تواند استفاده از کامپیوتر را برای کلیه افراد ناتوان بدنی که دارای تواناییهای شنوایی و گفتاری مناسب هستند آسان‌تر سازد به عنوان یک وسیله خروجی کاربرپسند در محیطهای مختلف می‌تواند با جایگزین کردن بسیاری از علائم دیداری (انواع چراغها و...) و شنوایی (انواع زنگهای اخطار و ...) با گفتارهای بیان کننده کامل پیامها استفاده از و رسیدگی به سیستمهای نیازمند این گونه پیامها را بهینه کند.

در اینجا لازم است به این نکته اشاره شود که پیشرفت در فن آوری تشخیص صحبت (و همچنین تشخیص گوینده) همان قدر که محدوده دی.اس.پی را در بر می‌گیرد نیازمند دانش به دست آمده از محدوده‌های هوش مصنوعی و شبکه‌های عصبی است. شاید این تنوع دانشهای مورد نیاز به عنوان عامل دشواری مطالعه مبحث پردازش صحبت در نظر گرفته شود حال آن که این گونه نیست و این تنوع راهکارها بخت رسیدن به سیستم با کارایی مطلوب را افزایش می‌دهد. برق و دانشگاه زنجان دانشکده مهندسی گروه برق آزمایشگاه پروژه

تواناییهای ابزارهایی که در بخش اول این فصل به آنها اشاره شد امیدواریهای فراوانی را در زمینه موفقیت ابزارهای موجود فراهم می‌آورد و دامنه وسیع شرکتها و مراکز دانشگاهی که در این زمینه فعالیت می‌کنند بر تنوع در قابلیتها و کاربردهای پیاده‌سازی شده این ابزارها می‌افزاید.

منابع فصل

1- Steven W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, Chapter 22: Audio Processing, from www.dspguide.com

فصل ششم: مدلسازی سیگنال

اهمیت مدلسازی سیگنال

تشخیص کامپیوتری صحبت در واقع بر دارنده دو نوع عمل اصلی تشخیص است: تشخیص صحبت و تشخیص گوینده. با تحلیل یک موج صوتی می‌توان خصیصه‌های اندامهای گفتاری گوینده را تخمین زد که این خصیصه‌ها راهکاری برای تشخیص هویت و تصدیق آن به روش زیست‌سنجی فراهم

است بدانیم که در ادبیات تشخیص صحبت خصیصه‌های مربوط به گوینده به عنوان تغییرات غیر داده‌ها حذف می‌گردند ولی سیستمهای تشخیص گوینده نیز از لیفتترینگ استفاده می‌کنند.

هر دو نوع سیستم تشخیص صحبت و تشخیص گوینده اطلاعات موضعی زمان کوتاه را با گرفتن مشتق خصوصیات اولیه نسبت به زمان به دست می‌دهند. به عنوان مثال یک صوت صدادار می‌تواند

با پیدا شدن فرمانتهای آن در طیف تشخیص داده شود، حال آن که یک صوت بی صدا (سایشی) با استفاده از انتقال طیف مدل می‌شود. مقادیر مشتق مرتبه اول خصائص ضرایب دلتا و مقادیر مشتق مرتبه دوم آن شتاب یا ضرایب دلتا-دلتا نامیده می‌شوند. مشتق زمانی با استفاده از یک رابطه

رگرسیون که یک مجموعه فریم را پیش و پس از فریم کنونی می‌کشد تقریب زده می‌شود.

سیستمهای تشخیص گوینده از یک پیمانۀ انتخاب خصیصه نیز در چارچوب تشخیص الگو استفاده می‌کنند. برای تشخیص صحبت تمامی سیگنال باید به یک نمایش متنی نگاشته شود حال آن که

سیستم تشخیص گوینده نیازی به کار تحت این اجبار ندارد. بنابراین پیمانۀ انتخاب خصیصه فقط خصیصه‌ها مربوط به اصوات صدادار را ذخیره می‌کند. اصوات صدادار مستقیماً فرضیات مدلسازی پیشگویانه خطی را برآورده می‌سازند و کمتر تحت تأثیر نویز صوتی قرار می‌گیرند.

منابع فصل

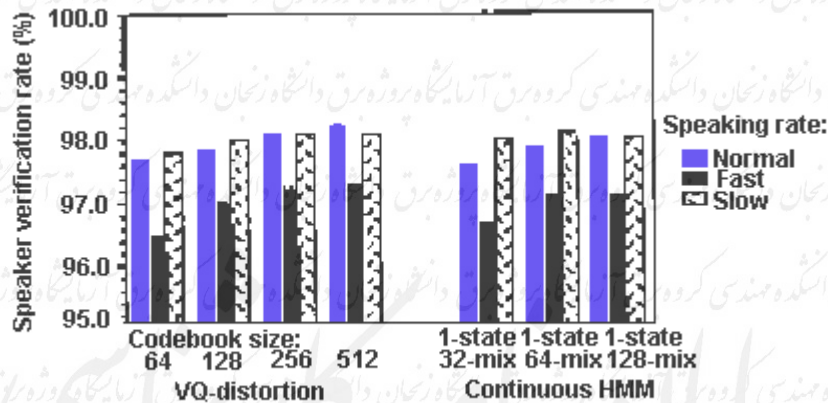
1) Richard Duncan, Mississippi State University, A Description And Comparison Of The Feature Sets Used In Speech Processing

فصل هفتم: روشهای طراحی سیستمهای تشخیص گوینده

مقدمه

همچنان که پیش از این گفته شد سیستمهای تشخیص گوینده در حالت کلی به دو نوع سیستمهای

تأیید هویت گوینده و سیستمهای بازشناسی گوینده تقسیم می‌شوند. تفاوت این دو سیستم در نحوه پذیرش ورودی است: در سیستمهای نوع اول گوینده با ارائه یک شناسه ادعای هویت یک کاربر



شکل شماره 6 - مقایسه سیستمهای مستقل از متن (ماتسوی و فوری 1992)

منابع فصل

- 1) Woon Wei Kian and Yap Wei Wum, Approaches to Speaker Verification Methods (Part of an article titled as Surprise 98 ... reporting on Speaker Verification), from <http://www.iis.ee.ic.ac.uk/~frank/surp98/report/wwyz2/approaches.h>

tm

فصل هشتم: کتابخانه تشخیص گوینده

کتابخانه تشخیص صحبت - گوینده جیلانگ هی

کتابخانه مزبور که از طریق مرجع شماره 1 قابل دستیابی است به زبان ++C نوشته شده و شامل

حدود 20 کلاس است که الگوریتمهای رایج استخراج خصیصهها و تکنیکهای رایج برای تشخیص

صحبت و تأیید هویت گوینده را پیادهسازی می نماید.

```
CreateUsersDB();
```

متد CreatUsersDB فقط یک فایل خالی ایجاد می کند. انتخابها بعداً در این فایل نوشته خواهند شد.

مجموعه کلاسهای فوق یک کتابخانه ایستا به نام HSspeakersDBLib را تشکیل می دهند که

می توان با استفاده از آن یک سیستم تشخیص گوینده وابسته به متن ساده را پیاده سازی نمود.

همچنانکه اشاره شد کتابخانه اصلی مورد استفاده تواناییهای گسترده ای برای کار با انواع روشها و

الگوریتمها دارد که می توان با کار روی آنها سیستمهای تشخیص گوینده با تشخیص صحبت بان

کارایی عملی ایجاد نمود.

فرضیات ما در نحوه استفاده از این کلاس نحوه پیاده سازی این کلاس مؤثر بوده و برای به وجود

آوردن یک مجموعه دارای کاربرد کلی تر باید بیشتر روی این کتابخانه کار شود اما به نظر می رسد

همچنان که از عملکرد برنامه پیاده سازی شده قابل مشاهده است حتی این استفاده ساده از این

کتابخانه نیز می تواند برطرف کننده نیازهای یک برنامه نویس در زمینه طراحی یک سیستم تشخیص

گوینده می باشد.

منابع فصل

- 1) Jialong He, SVLib Library, downloadable from <http://tiger.la.asu.edu/personal.htm>