به نام خدا

دانشکده مهندسی برق و رباتیک
گروه الکترونیک

استخراج ویژگی مبتنی بر پردازش در حوزهٔ اتوکرولیشن جهت بازشناخت گفتار با استفاده از HTK

دانشجو: **سید حمید اخلاق**

استاد راهنما:
دکتر حسین مروی

استاد مشاور:
دکتر امید رضا معروضی

پایان نامه ارشد جهت اخذ درجهٔ کارشناسی ارشد

بهمن ۱۳۸۸

شماره : ۰۷۱۵/آ.ت.ب
تاریخ : ۱۳۸۸/۱۱/۰۵
ویرایش : - - - -

دانشگاه صنعتی شاهرود

مدیریت تحصیلات تکمیلی

فرم شماره (۶)

بسمه تعالی

فرم صورتجلسه دفاع پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای **سید حمید اخلاق** رشته **بـــرق** گرایش **الکترونیک**

تحت عنوان : **استخراج ویژگی مبتنی بر پردازش در حوزه اتوکرولیشن جهت بازشناخت گفتار با استفاده از HTK**

که در تاریخ **۱۳۸۸/۱۱/۰۵** با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است :

قبول (با درجه : عالی امتیاز ۱۹/۵۵) ☐ دفاع مجدد ☐ مردود ☐

۱ـ عالی (۲۰ ـ ۱۹) ۲ـ بسیار خوب (۱۸/۹۹ ـ ۱۸)

۳ـ خوب (۱۷/۹۹ ـ ۱۶) ۴ـ قابل قبول (۱۵/۹۹ ـ ۱۴)

عضو هیأت داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضاء
۱ـ استادراهنما	دکتر مروی	استادیار	
۲ـ استاد مشاور	دکتر معروفی	استادیار	
۳ـ نماینده شورای تحصیلات تکمیلی	دکتر حداد ظریف	استادیار	
۴ـ استاد ممتحن	دکتر زاهدی	استادیار	
۵ ـ استاد ممتحن	دکتر خسروی	استادیار	

تأیید رئیس دانشکده :

تقدیر و تشکر

در ابتدا خدا را به خاطر محبت عظیمش سپاس می‌گویم سپس از زحمات استاد راهنما جناب آقای دکتر حسین مروی و همچنین خانواده و برادرانم تشکر می‌کنم.

تعهد نامه

اینجانب ..... دانشجوی دوره کارشناسی ارشد / دکتری رشته .....

دانشکده ..... دانشگاه صنعتی شاهرود نویسنده پایان نامه/ رساله .....

..... تحت راهنمائی ..... متعهد می شوم .

- تحقیقات در این پایان نامه/ رساله توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه/ رساله تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه / رساله تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه/ رساله رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه/رساله ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه/رساله ، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ :

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه / رساله بدون ذکر مرجع مجاز نمی باشد.

* متن این صفحه نیز باید در ابتدای نسخه های تکثیر شده پایان نامه/ رساله وجود داشته باشد .

چکیده

یکی از مهمترین مسایل مطرح در زمینهٔ سیستم‌های بازشناخت گفتار تأثیر نویز بر سیگنال گفتار و کاهش دقت بازشناسی است، لذا بازشناخت گفتار مقاوم در برابر نویز یکی از موارد مورد مطالعه محققان می‌باشد. در سالهای اخیر تحقیقات بسیاری در این زمینه به عمل آمده و روش‌های مختلفی ارائه شده است. یک دسته از روش‌های موجود در زمینهٔ بازشناخت گفتار مقاوم، استخراج ویژگی‌هایی از سیگنال گفتار می‌باشد که نسبت به نویز مقاوم باشند. در این پایان نامه هدف معرفی روش‌هایی جهت مقاوم سازی ویژگی‌های استخراج شده از سیگنال گفتار در برابر نویز جمع شونده می‌باشد. حوزه‌ای که در این پایان نامه از آن برای استخراج ویژگی مقاوم استفاده شده است، حوزهٔ اتوکرولیشن می‌باشد. برای این منظور در ابتدا بعد از بررسی مهمترین اجزاء تشکیل دهندهٔ سیستم‌های بازشناخت گفتار و مرور بعضی از کارهای انجام شده در زمینهٔ استخراج ویژگی مقاوم در حوزهٔ اتوکرولیشن، روش‌های پیشنهادی معرفی شده‌اند. از روش‌های موجود در حوزهٔ اتوکرولیشن RAS، DAS، AMFCC و PAC می‌باشند. در این پایان نامه ایده‌هایی جدید جهت بهبود روش‌های AMFCC و PAC بر اساس استفاده از مشتق طیف توان و پنجره گذاری مناسب پیشنهاد شده است. نتایج بدست آمده از پیاده سازی روش‌های پیشنهادی بر روی دیتابیس TIMIT بیانگر بهبود در نرخ بازشناخت گفتار پیوسته نسبت به برخی از روش‌های گذشته است.

واژگان کلیدی: بازشناخت گفتار - استخراج ویژگی - اتوکرولیشن - بستهٔ نرم افزاری مدل مخفی مارکوف

لیست مقاله‌های استخراج شده از پایان نامه

1. S. H. Akhlagh and H. Marvi (2009), “Features Derived from Differential Power Spectrum in Autocorrelation Domain for Continuous Speech Recognition”, ICSV16, Kracow, Poland.

2. S. H. Akhlagh and H. Marvi (2010), “Robust Features Derived from Differentiated Phase Autocorrelation Spectrum for Speech Recognition”, 16th National Conference on Communications, Co-sponsored by IEEE Comm. Society, India.

3. S. H. Akhlagh, H. Marvi and O. R. Ma'rouzi, “Model based and cepstral based features obtained from differentiated autocorrelation spectrum for robust speech recognition”, submitted to DSP Elsevier, under review.

فهرست مطالب

فصل اول: مقدمه..... ۱

۱-۱ مقدمه..... ۱

۱-۲ تاریخچه بازشناخت گفتار..... ۴

۱-۲-۱ دهه‌های ۱۹۵۰ و ۱۹۶۰ میلادی..... ۵

۱-۲-۲ دهه‌های ۱۹۷۰ و ۱۹۸۰ میلادی..... ۹

۱-۲-۲-۱ دههٔ ۱۹۷۰..... ۹

۱-۲-۲-۲ دههٔ ۱۹۸۰..... ۱۳

۱-۲-۳ دهه‌های ۱۹۹۰ و ۲۰۰۰ میلادی تاکنون..... ۱۵

۱-۳ بیان مسئله و فصل بندی مطالب..... ۲۰

فصل دوم: معرفی اجزاء سیستم‌های بازشناخت گفتار..... ۲۲

۲-۱ مقدمه..... ۲۲

۲-۲ ابعاد مختلف سیستم‌های بازشناخت گفتار..... ۲۳

۲-۲-۱ بازشناخت کلمات مجزا در مقابل بازشناخت گفتار پیوسته..... ۲۴

۲-۲-۱-۱ بازشناخت کلمات مجزا..... ۲۴

۲-۲-۱-۲ بازشناخت گفتار پیوسته..... ۲۴

۲-۲-۲ اندازهٔ بانک واژگان..... ۲۵

۲-۲-۳ وابسته یا مستقل بودن از گوینده..... ۲۶

۲-۲-۴ محدودیت‌های زبان شناختی..... ۲۷

۲-۲-۵ ابهامات و اشتباهات آکوستیکی..... ۲۸
۲-۲-۶ نویز محیطی..... ۲۸
۲-۳ ساختار سیستم‌های بازشناخت گفتار..... ۲۹
۲-۳-۱ استخراج ویژگی..... ۳۲
۲-۳-۱-۱ پردازش کپسترال..... ۳۳
۲-۳-۱-۲ ضرایب کپسترال پیشگویی خطی (LPCC)..... ۳۵
۲-۳-۱-۳ ضرایب کپسترال فرکانس مل (MFCC)..... ۳۷
۲-۳-۱-۴ پیشگویی خطی مبتنی بر درک انسان (PLP)..... ۴۲
۲-۳-۱-۵ ضرایب مشتق..... ۴۵
۲-۳-۲ مدل زبان..... ۴۵
۲-۳-۳ مدل صوتی..... ۴۷
۲-۴ بازشناخت گفتار مقاوم..... ۴۸
۲-۴-۱ استخراج ویژگی‌های مقاوم..... ۴۹
۲-۴-۲ بهبود سیگنال گفتار..... ۴۹
۲-۴-۳ مقاوم سازی مدل بازشناخت..... ۵۰

فصل سوم: مدل مخفی مارکوف و کاربرد آن در بازشناخت گفتار..... ۵۱

۳-۱ مقدمه..... ۵۱
۳-۲ مدل مخفی مارکوف..... ۵۲
۳-۲-۱ مدل‌های مارکوف مرتبهٔ اول..... ۵۳
۳-۲-۲ مفهوم مدل مخفی مارکوف مرتبهٔ اول..... ۵۵
۳-۲-۳ پارامترهای لازم برای تعریف یک مدل مخفی مارکوف گسسته..... ۵۶
۳-۲-۴ سه مسألهٔ اصلی در مدل‌های مخفی مارکوف..... ۵۷

۳-۲-۴-۱ حل مسألهٔ ارزیابی در مدل مخفی مارکوف..... ۵۹
۳-۲-۴-۲ حل مسألهٔ رمزگشایی در مدل مخفی مارکوف..... ۶۰
۳-۲-۴-۳ حل مسألهٔ یادگیری در مدل مخفی مارکوف..... ۶۲
۳-۲-۵ انواع مدل مخفی مارکوف از لحاظ ساختار..... ۶۴
۳-۲-۶ مدل‌های مخفی مارکوف با چگالی‌های مشاهده‌ایی پیوسته..... ۶۶
۳-۲-۷ اصول بازشناخت گفتار بوسیلهٔ مدل‌های مخفی مارکوف..... ۶۷
۳-۲-۷-۱ بازشناخت کلمه‌های جدا از هم..... ۶۸
۳-۲-۷-۲ بازشناخت گفتار پیوسته..... ۷۱
۳-۲-۷-۲-۱ مدل‌های صوتی بر پایهٔ مدل مخفی مارکوف..... ۷۱
۳-۲-۷-۲-۲ آموزش ادغام شده..... ۷۴
۳-۳ بستهٔ نرم افزاری مدل مخفی مارکوف HTK..... ۷۵
۳-۳-۱ ساختار کلی HTK..... ۷۶
۳-۳-۱-۱ ابزارهای مربوط به آماده‌سازی داده‌ها..... ۷۹
۳-۳-۱-۲ ابزارهای مربوط به آموزش..... ۸۰
۳-۳-۱-۳ ابزارهای مربوط به بازشناخت..... ۸۲
۳-۳-۱-۴ ابزار آنالیز..... ۸۳
۳-۳-۲ انواع فایل‌های مورد استفاده در HTK..... ۸۳
۳-۳-۲-۱ فایل‌های برچسب..... ۸۳
۳-۳-۲-۲ فایل دیکشنری..... ۸۴
۳-۳-۲-۳ فایل‌های مربوط به تعریف مدل‌های مخفی مارکوف..... ۸۵
۳-۳-۲-۴ فایل‌های پیکربندی..... ۸۶

فصل چهارم: مروری بر کارهای انجام شده در حوزهٔ اتوکرولیشن برای استخراج ویژگی..... ۸۷

۴-۱ مقدمه..... ۸۷

۴-۲ بعضی از خواص تابع اتوکرولیشن..... ۸۹

۴-۳ روش رشتهٔ خود همبستگی نسبی (RAS)..... ۹۰

۴-۳-۱ فیلتر کردن رشتهٔ اتوکرولیشن..... ۹۱

۴-۳-۲ محاسبهٔ ضرایب کپسترال از رشتهٔ RAS..... ۹۲

۴-۴ روش مشتق رشتهٔ اتوکرولیشن (DAS)..... ۹۵

۴-۴-۱ بدست آوردن رشتهٔ RAS..... ۹۵

۴-۴-۲ محاسبهٔ مشتق طیف توان (DPS)..... ۹۶

۴-۴-۳ بدست آوردن ضرایب کپسترال..... ۹۷

۴-۵ روش اتوکرولیشن MFCC (AMFCC)..... ۹۸

۴-۶ روش اتوکرولیشن فاز (PAC)..... ۱۰۰

فصل پنجم: روش‌های پیشنهادی..... ۱۰۳

۵-۱ مقدمه..... ۱۰۳

۵-۲ روش پیشنهادی اول: مشتق طیف توان در حوزهٔ اتوکرولیشن (DPSA)..... ۱۰۴

۵-۲-۱ محاسبهٔ تابع اتوکرولیشن سیگنال نویزی..... ۱۰۵

۵-۲-۲ محاسبهٔ طیف توان از روی ضرایب اتوکرولیشن با تأخیر بالا..... ۱۰۶

۵-۲-۳ محاسبهٔ مشتق طیف توان..... ۱۱۰

۵-۲-۴ مرحلهٔ نهایی: محاسبهٔ ضرایب کپسترال (DPSACCs)..... ۱۱۲

۵-۳ روش پیشنهادی دوم: مشتق طیف اتوکرولیشن فاز (DPAS)..... ۱۱۳

۵-۳-۱ بدست آوردن رشتهٔ اتوکرولیشن فاز..... ۱۱۴

۵-۳-۱-۱ اعمال پنجره به رشتهٔ اتوکرولیشن..... ۱۱۵
۵-۳-۱-۲ رشتهٔ اتوکرولیشن فاز..... ۱۱۶
۵-۳-۲ مشتق طیف اتوکرولیشن فاز..... ۱۱۷
۵-۳-۳ مرحلهٔ نهایی: محاسبهٔ ضرایب کپسترال (DPASCCs)..... ۱۲۰
فصل ششم: نتایج آزمایش‌های انجام شده..... ۱۲۲
۶-۱ مقدمه..... ۱۲۲
۶-۲ دیتابیس مورد استفاده..... ۱۲۳
۶-۳ پارامترها و کلاسیفایر مورد استفاده..... ۱۲۴
۶-۴ بررسی کارآیی روش پیشنهادی اول (DPSA)..... ۱۲۵
۶-۴-۱ تنظیم پارامترهای مشتق..... ۱۲۵
۶-۴-۲ مقایسهٔ کارآیی روش DPSA با بعضی از روش‌های دیگر..... ۱۲۶
۶-۵ بررسی کارآیی روش پیشنهادی دوم (DPAS)..... ۱۳۱
۶-۵-۱ تنظیم پارامترهای مشتق..... ۱۳۱
۶-۵-۲ مقایسهٔ کارآیی روش DPAS با بعضی از روش‌های دیگر..... ۱۳۲
۶-۶ نتیجه گیری..... ۱۳۶
۶-۷ پیشنهادها برای کارهای آینده..... ۱۳۷
مراجع..... ۱۳۸

فهرست شکل‌ها

شکل ۱-۱ جمله "Bishop moves to king knight five" به همراه شکل موج مربوط به آن [۱].....۳

شکل ۱-۲ ماشین سخنگو ساخته شده توسط Wheatstone [۲].....۵

شکل ۱-۳ بلوک دیاگرام مدار تشخیص دهنده عدد ساخته شده توسط Davis و همکارانش [۴].....۷

شکل ۱-۴ تصویر فرمنت ۱ در مقابل فرمنت ۲ مربوط به عددهای صفر تا نه [۴].....۸

شکل ۱-۵ پرسپترون چند لایه [۲].....۱۵

شکل ۱-۶ ارزیابی بدست آمده توسط DARPA مربوط به بعضی از ماموریت‌های بازشناخت گفتار [۲].....۱۸

شکل ۲-۱ مدل کلی از فرآیند تولید و بازشناخت گفتار.....۳۰

شکل ۲-۲ ساختار کلی یک سیستم بازشناخت گفتار اتوماتیک [۶۷].....۳۲

شکل ۲-۳ بلوک دیاگرام مربوط به مدل ساده تولید گفتار [۴۲].....۳۳

شکل ۲-۴ بلوک دیاگرام مربوط به استخراج ضرایب کپسترال فرکانس مل.....۳۸

شکل ۲-۵ مقیاس مل [۴۶].....۴۰

شکل ۲-۶ فیلتر بانک مقیاس مل [۴۵].....۴۱

شکل ۲-۷ بلوک دیاگرام مربوط به بدست آوردن ضرایب PLP [۴۹].....۴۲

شکل ۲-۸ مدل مربوط به اثر نویز جمع شونده و اعوجاج کانال بر سیگنال گفتار.....۴۸

شکل ۳-۱ یک نمونه رشتهٔ مارکوف با ۳ حالت [۶۵].....۵۳

شکل ۳-۲ فرآیند پیدا کردن بهترین رشتهٔ حالت به کمک الگوریتم ویتربی [۴۴].....۶۱

شکل ۳-۳ رابطهٔ بین α و β در الگوریتم بام ولچ [۴۴]..... ۶۳
شکل ۳-۴ نمایش سه نوع HMM..... ۶۵
شکل ۳-۵ یک نمونه مدل مخفی مارکوف برای یک کلمه [۲۹]..... ۶۹
شکل ۳-۶ بازشناخت کلمه‌های جدا از هم به کمک مدل مخفی مارکوف [۲۹]..... ۷۰
شکل ۳-۷ یک نمونه مدل مخفی مارکوف مربوط به یک آوا..... ۷۱
شکل ۳-۸ بدست آوردن رشته‌های سه آوایی از روی کلمه‌ها [۶۸]..... ۷۳
شکل ۳-۹ ایجاد پیوند بین حالت‌ها در مدل‌های مربوط به آواهای مشابه [۶۸]..... ۷۳
شکل ۳-۱۰ دو مرحلهٔ اصلی موجود در HTK [۲۹]..... ۷۶
شکل ۳-۱۱ ساختار نرم افزاری HTK [۲۹]..... ۷۷
شکل ۳-۱۲ ابزارهای مختلف موجود در HTK در مراحل مختلف پردازش [۲۹]..... ۷۹
شکل ۳-۱۳ بلوک دیاگرام مربوط به ورودی‌ها و خروجی ابزار HCopy [۲۹]..... ۸۰
شکل ۳-۱۴ مراحل مختلف مربوط به آموزش مدل‌های مخفی مارکوف [۲۹]..... ۸۲

شکل ۴-۱ بلوک دیاگرام مربوط به محاسبهٔ ضرایب RAS-MFCC [۷۳]..... ۹۳
شکل ۴-۲ مقایسه دو روش MFCC و RAS-MFCC در حضور نویز سفید [۷۳]..... ۹۴
شکل ۴-۳ مقایسه دو روش MFCC و RAS-MFCC در حضور نویز رنگی [۷۳]..... ۹۴
شکل ۴-۴ بلوک دیاگرام مربوط به روش DAS برای استخراج ویژگی‌های مقاوم [۷۴]..... ۹۷
شکل ۴-۵ بلوک دیاگرام مربوط به روش AMFCC برای استخراج ویژگی [۷۶]..... ۹۹

شکل ۵-۱ (الف) یک فریم از آوای /iy/ (ب) تابع اتوکرولیشن دو طرفهٔ مربوط به آن..... ۱۰۵
شکل ۵-۲ رشتهٔ اتوکرولیشن مربوط به یک فریم ۳۲ میلی ثانیه‌ایی از چهار نمونه نویز مختلف..... ۱۰۷
شکل ۵-۳ طیف توان و مشتق طیف توان مربوط به یک فریم از آوای /iy/..... ۱۱۱

شکل ۵-۴ بلوک دیاگرام کلی مربوط به روش پیشنهادی اول (DPSACCs).....۱۱۲

شکل ۵-۵ طیف اتوکرولیشن فاز و مشتق طیف اتوکرولیشن فاز مربوط به یک فریم از آوای /aa/...۱۱۹

شکل ۵-۶ بلوک دیاگرام کلی مربوط به روش پیشنهادی دوم (DPASCCs).....۱۲۰

شکل ۶-۱ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی در حضور نویز با توان‌های متفاوت.....۱۳۰

شکل ۶-۲ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی و روش پیشنهادی دوم (DPAS) در حضور نویز با توان‌های متفاوت.....۱۳۵

فهرست جداول

جدول ۶-۱ نرخ بازشناخت مربوط به روش پیشنهادی (DPSA) با سه رابطهٔ مشتق متفاوت.....۱۲۶

جدول ۶-۲ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی در حضور نویز سفید با توان‌های متفاوت.....۱۲۷

جدول ۶-۳ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی در حضور نویز کارخانه با توان‌های متفاوت.....۱۲۷

جدول ۶-۴ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی در حضور نویز همهمه با توان‌های متفاوت.....۱۲۷

جدول ۶-۵ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی در حضور نویز F16 با توان‌های متفاوت.....۱۲۸

جدول ۶-۶ نرخ بازشناخت مربوط به روش پیشنهادی دوم (DPAS) با سه رابطهٔ مشتق متفاوت..۱۳۱

جدول ۶-۷ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی و روش پیشنهادی دوم (DPAS) در حضور نویز سفید با توان‌های متفاوت.....۱۳۲

جدول ۶-۸ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی و روش پیشنهادی دوم (DPAS) در حضور نویز کارخانه با توان‌های متفاوت.....۱۳۳

جدول ۶-۹ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی و روش پیشنهادی دوم (DPAS) در حضور نویز همهمه با توان‌های متفاوت.....۱۳۳

جدول ۶-۱۰ نرخ بازشناخت مربوط به روش‌های مختلف استخراج ویژگی و روش پیشنهادی دوم (DPAS) در حضور نویز F16 با توان‌های متفاوت.....۱۳۳

فصل اول

مقدمه

فصل اول: مقدمه

۱–۱ مقدمه

با پیشرفت علم و ساخت وسایل الکترونیکی پیشرفته برای انجام کارهای مختلف، بشر این نیاز را در خود احساس کرده است که بتواند راحتتر با این وسایل ارتباط برقرار کند. بدون شک یکی از بهترین و آسانترین راههای برقراری ارتباط از طریق صحبت کردن است که در زندگی روزانه در برخورد با دوستان و همکاران از آن برای بیان منظور خود استفاده می‌کنیم. متاسفانه یا خوشبختانه مسیر پیشرفت علم به سمتی پیش می‌رود که این دوستان و همکاران جای خود را به وسایل الکترونیکی از قبیل کامپیوتر و تلفن همراه داده‌اند. افزایش روز افزون استفاده از این وسایل، دانشمندان را به این فکر انداخت که به جای استفاده از دست برای وارد کردن فرمانها و برقراری ارتباط با این وسایل از صحبت کردن معمولی استفاده شود. عملی کردن این فکر مربوط به شاخه‌ای از علم با عنوان

بازشناخت گفتار اتوماتیک[1] می‌شود. در طول ۶۰ سال گذشته تحقیقات زیادی در این حوزه انجام شده است و پیشرفتهای چشم گیری هم بدست آمده است.

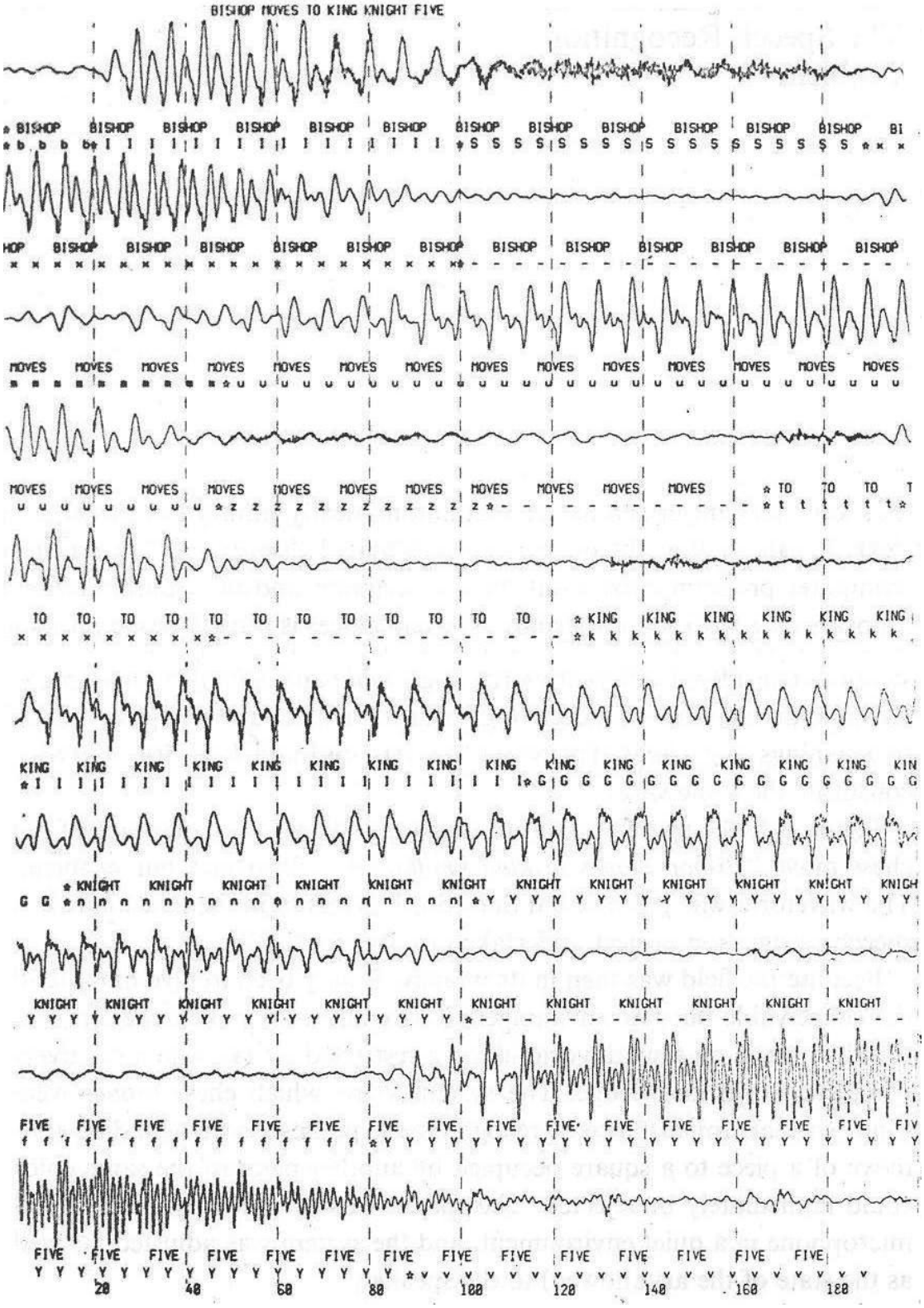
سیستم‌های بازشناخت گفتار اولیه فقط قادر به تشخیص بعضی از حروف یا کلمات محدود و جدا از هم بودند. ولی سیستم‌های بازشناخت گفتار امروزی قادر هستند گفتار پیوسته را با دقت نسبتاً خوبی تشخیص دهند. البته بیان این مطلب نیز لازم است که با وجود تمام پیشرفتهای انجام شده هنوز قادر به ساخت سیستمی که بتواند مستقل از محیط و شخص صحبت کننده کلمات گفته شده را به خوبی انسان تشخیص دهد نیستیم.

مسائل زیادی وجود دارد که ساخت سیستم‌های بازشناخت گفتار را با مشکل مواجه می‌کند. به عنوان مثال شکل موجی که از صوت بدست می‌آید متناسب با شخص گوینده و در شرایط محیطی متفاوت تغییرات زیادی از خود نشان می‌دهد. علاوه بر تغییر شخص و شرایط محیطی بعضی از مواقع حروفی که از یک شخص ثابت و در یک مکان ثابت ادا می‌شوند نیز به خاطر محتوای کلمه ادا شده دارای شکل موجهای متفاوتی هستند. به عبارت دیگر داشتن دو شکل موج یکسان دلیل بر این نمی‌شود که هر دو مربوط به یک صدا می‌شوند. به عنوان مثال همانطور که در شکل ۱-۱ مشاهده می‌شود شکل موج حرف i در کلمه KING بیشتر به شکل موج حرف u در کلمه MOVE شبیه است تا به شکل موج حرف i در کلمه BISHOP [۱].

امروزه سیستم‌های بازشناخت گفتار کاربردهای مختلفی پیدا کرده‌اند. از کاربردهای پزشکی گرفته تا کاربردهای نظامی ولی یکی از مهمترین این کاربردها استفاده از سیستم‌های بازشناخت گفتار برای تبدیل سیگنال صحبت به متن است.

در این فصل در ابتدا تاریخچه مختصری از بازشناخت گفتار آورده شده است سپس هدف از این پایان نامه و ساختار فصلهای مختلف آن بیان شده است.

[1] Automatic Speech Recognition



شکل ۱-۱ جمله "Bishop moves to king knight five" به همراه شکل موج مربوط به آن [۱].