بسم الله الرحمن الرحیم

دانشکده مهندسی برق و کامپیوتر

پایان‌نامه کارشناسی ارشد مهندسی برق مخابرات

سنتز جملات فارسی با استفاده از قواعدگویشی و کسره‌ی بین کلمات

دانشجو : محمدصادق رستمی

استاد راهنما: : دکتر خشایار یغمایی

بهمن ۱۳۹۱

بسم الله الرحمن الرحیم

دانشکده مهندسی برق و کامپیوتر

تایید دفاع از پایان‌نامه کارشناسی ارشد

پایان‌نامه آقای **محمد صادق رستمی**

برای اخذ درجه **کارشناسی ارشد مهندسی برق- گرایش مخابرات**

با عنوان:

سنتز جملات فارسی با استفاده از قواعدگویشی و کسره‌ی بین کلمات

در تاریخ ..... دفاع شد و مورد تایید قرار گرفت.

تایید کنندگان:

۱) استاد محترم داور..... امضاء

۲) استاد محترم داور..... امضاء

۳) استاد محترم راهنما ..... امضاء

۴) مدیر محترم گروه مخابرات ..... امضاء

اینجانب . . . . . بدین وسیله اظهار می‌دارم که محتوای علمی این نوشتار با عنوان . . . . . که به عنوان پایان‌نامه کارشناسی ارشد مهندسی برق گرایش . . . . . به دانشکده برق و کامپیوتر دانشگاه سمنان ارائه شده است دارای اصالت پژوهشی بوده و حاصل فعالیت علمی این جانب است.

این‌جانب می‌دانم که اگر خلاف ادعای بالا در هر زمانی محرز شود کلیه حقوق مترتب بر این نوشتار از این جانب سلب شده و مراتب قانونی مرتبط با آن نیز از طرف مراجع ذیربط قابل پیگیری است.

نام و نام خانوادگی – شماره دانشجویی

تاریخ و امضاء

با تشکر از زحمات

- پدر و مادر عزیزتر از جانم[۱] که در تمامی مراحل زندگی، رهنمودهای خود را از من دریغ نداشته‌اند و همواره حامی من بوده‌اند؛
- همسر بسیار مهربانم[۲] که بالاترین ارمغان، هم‌دلی و هم‌فکری و هم‌سری و آرامش اوست؛
- دکتر خشایار یغمایی[۳] که راه و روش صحیح زندگی کردن در چندین بعد و درس خواندن و تحقیق کردن و تبدیل دانش تئوری به دانش عملی را به ما آموختند؛
- دکتر سعید مظفری[۴] که علی‌رغم تخصص ایشان در زمینه‌ی پردازش تصویر، بعد از دکتر یغمایی، از هیچ راهنمایی ارزنده‌ی دلسوزانه‌ای مضایقه نکردند و همچنین وقتشان را در اختیار بنده قرار دادند؛
- دکتر پژمان رضایی[۵] که کمک‌های شایانی در تسریع انجام پایان‌نامه نمودند؛
- دکتر بابک ناصر شریف[۶] که مجموعه دادگان صوتی فارس‌دات را در اختیار بنده قرار دادند؛
- کارمندان شرکت توان پرداز کومش[۷] به خصوص مهندس محمد زمان و مهندس محمد نحوی؛
- و تمامی کسانی که بنده را در انجام این پایان‌نامه کمک کردند.

---

[۱] آقای حبیب الله رستمی و خانم زهرا صیادنژاد

[۲] خانم فاطمه عباسی

[۳] دکتر خشایار یغمایی در تاریخ چهاردهم تیرماه نود و یک، در سن ۵۵ سالگی، در شب میلاد حضرت مهدی (عج)، بر اثر عارضه قلبی درگذشتند. روحشان شاد، یادشان گرامی و اثرشان جاوید باد.

[۴] عضو هیت علمی گروه الکترونیک دانشکده برق دانشگاه سمنان

[۵] عضو هیت علمی و مدیر گروه مخابرات دانشکده برق دانشگاه سمنان

[۶] از اساتید هیئت علمی در دانشگاه خواجه نصیرالدین طوسی تهران

[۷] این شرکت در سال ۱۳۸۳ توسط مرحوم دکتر خشایار یغمایی تاسیس شده‌است و نشانه‌ی اینترنتی آن www.tpkumesh.com است.

تقدیم به ...

معمولا رسمه که در این صفحه از پایان‌نامه، اثر رو به فرد یا افرادی تقدیم می‌کنن. اما شخصیتی که می‌خوام در موردش صحبت کنم، بسیار بزرگتر از اینه که بتونم این مختصر پایان‌نامه رو بهش تقدیم کنم.

مرحوم دکتر خشایار یغمایی

ترم چهارم دانشگاه در دوره کارشناسی در دانشگاه سمنان بود که با ایشون آشنا شدم. درس تجزیه و تحلیل سیگنال‌ها و سیستم‌ها. در برخورد اول، فردی بود بسیار با شخصیت، مبادی به آداب اجتماعی و متخلق به اخلاق انسانی. تا اون موقع کمتر کسی رو دیده بودم که سر کلاس یک درس فنی، از سعدی بگه و اشعارش و مثل‌هاش و داستان‌هاش؛ از حافظ بگه و غزلیاتش رو از بر بخونه و معنی کنه. و در زمان درس دادن سعی کنه مثال‌های درسیشو –که فنی بودن- از مسایل اجتماعی بیان کنه؛ به همراه تذکر چند نکته اخلاقی و اجتماعی در لابلای حرفهاش!

این شد که تصمیم گرفتم درس‌های بیشتری رو هم با ایشون بردارم. ترم پنجم هم با ایشون درس دیگه‌ای رو گذروندم و حتی یکی از درس‌های اختیاری رو از یک گرایش دیگه انتخاب کردم که بازهم با ایشون درس داشته باشم. همیشه مفهومی درس می‌دادن و امتحاناشون هم ساده نبود. نمره‌ی بالا گرفتن از ایشون، کار ساده‌ای نبود. امتحاناشون رو کتاب و جزوه باز برگزار می‌کردن. پروژه کارشناسی

رو هم با همراهی دوتا دیگه از دوستانم با ایشون برداشتیم. ایشون همیشه می‌گفت: پروژه‌ی پایانی یک دانشجو در دوره کارشناسی، حتما باید عملی باشه تا دانشجو بفهمه درس‌هایی که خونده کجا و چطوری استفاده میشه!

در همون دوران بود که خوشبختانه متوجه شدیم که ایشون علاوه بر اینکه تجربه‌ی ۲۵ ساله‌ی حضور در دانشگاه رو دارن، شرکتی رو هم تاسیس کردن و در اون شرکت، با کمک برخی از دانشجویان و مهندسین به کارهای فنی و عملی می‌پردازن. این شد که با یکی از بهترین دوستانم – آقای نعیم غفاری- تصمیم گرفتیم، دوره‌ی کارآموزی رو در شرکت ایشون بگذرونیم و از اونجا که پروژه‌ی کارشناسیمون هم با ایشون بود، قرار شد در زمان گذراندن کارآموزی، بر روی پروژه‌مون هم کار کنیم.

محیط شرکت بسیار دوستانه و پویا بود و با حضور ایشون، در هر بار، رونق بیشتری می‌گرفت؛ مخصوصا که هرازچندگاهی در محیط کار هم، برای ما از اشعار حافظ و رباعیات خیام و داستان‌های سعدی میگفتن. پروژه‌های متعددی در اون شرکت[۱] انجام می‌شد که اکثرا ایده‌های خودشون بود. پروژه‌هایی که جنبه‌های آموزشی، فرهنگی و سرگرمی داشتن. که در نهایت منجر به تاسیس موزه‌ی علوم شهرداری سمنان شد[۲].

بعد از گذروندن دوره‌ی کارآموزی و انجام پروژه و دفاع، مجددا در دانشگاه سمنان این‌بار در دوره کارشناسی ارشد پذیرفته شدم و این یعنی این‌که بازهم می‌تونستم با ایشون درس بگذرونم[۳]. بعد از گذروندن دوره‌ی کارآموزی، بنده تمایل زیادی داشتم تا در شرکت توان پرداز مشغول به کار بشم که خوشبختانه مورد توافق ایشون قرار گرفت و با توجه با اینکه دانشجوی کارشناسی ارشد بودم، به صورت پاره‌وقت، مشغول به کار شدم.

تا قبل از این، بیشتر مراودات من و ایشون، بر سر مسایل درسی و دانشگاهی بود؛ اما از این به بعد علاوه بر موارد قبل، بیشتر به ایشون نزدیک شده بودم و می‌تونستم از وجودشون بهره ببرم. روز به روز که شناخت بیشتری از ایشون پیدا می‌کردم، علاقه‌م هم به ایشون بیشتر می‌شد. نه تنها من، بلکه بقیه

---

[۱] توان پرداز کومش

[۲] برای اطلاعات بیشتر می‌توانید به سایت شرکت توان پرداز کومش با آدرس اینترنتی www.tpkumesh.com مراجعه نمایید.

[۳] بنده مجموعا ۲۸ واحد درسی را در دو دوره‌ی کارشناسی و ارشد با ایشان گذرانده‌ام.

کسانی که در شرکت توان پرداز کار می‌کردن، در این‌مورد وضعیتی شبیه من رو داشتن. با خوندن سطرهای بعدی شاید به برخی از این علت‌ها واقف بشید.

یکی از بزرگترین خصوصیات اخلاقی ایشون، این بود که در هر محیطی و هر شرایطی، متناسب با اون شرایط برخورد می‌کردن. با این‌که کارمند شرکت ایشون بودم، اما در دانشگاه و در کلاس درس، همه فقط دانشجوی ایشون بودن و کسی امتیاز دیگه‌ای نداشت مگر این‌که در اون درس، از بقیه پیش می‌افتاد. و اگر درون اتومبیلشون باهاشون هم‌مسیر می‌شدیم، مثه یک دوست با ما بر خورد می‌کردن و برادرانه توصیه‌هایی رو به ما می‌گفتن. و وقتی دختر کوچیکشون رو به محل کار می‌آوردن، احساسات پاک پدرانه درون چشمانشون موج می‌زد؛ و وقتی کسی بیمار می‌شد یا مشکلی براش پیش میومد، بسیار ناراحت می‌شدند و خودشون چندین بار مستقیما ، جویای احوال می‌شدن.

ایشون بسیار متواضع بودن. در زمان افتتاح موزه و پارک علوم کودک و نوجوان سمنان[1]، وقتی که از ایشون تقدیر شد، اسم تک‌تک کارمنداشون رو بردن و بین خودشون و کارمنداشون، تفاوتی نذاشتن. حتی وقتی که برای اولین بار نمایشگاه تابلوهای گویا- که طراحی و ساخت آن در شرکت توان پرداز کومش و تحت نظارت ایشان بود- برپا شده‌بود، در ابتدا حاضر به مصاحبه نشدن و گفتن: کسی که این‌ها را ساخته باید خودش مصاحبه کنه! در حالی که ایده و اعتبار طرح از خود ایشون بود.

ایشون به فرهنگ و ادب و هنر بسیار اهمیت می‌دادن. اشعار فارسی رو بسیار دوست داشتن و اشعار بسیار زیادی رو از حافظ و سعدی و خیام و عطار و ... از حفظ بودن. یکی از محصولات شرکت ایشون، عروسک سخنگویی است که با کنترل از راه دور کار می‌کند و به منظور سرگرمی کودکان، با تاکید بر توسعه‌ی فرهنگ ایرانی از طریق شعر و سرود و داستان، تهیه شده‌است؛ و با تولید و توسعه‌ی دستگاه راهنمای چندزبانه‌ی گردشگری سعی در رونق صنعت گردشگری و در کنار آن، ایران‌شناسی رو داشتن. همچنین تسلط بسیار بالایی بر موسیقی کلاسیک غربی داشتن. طوری که خودشون می‌گفتن فقط بیش از ۴۰۰ کاست موسیقی کلاسیک و بیش از ۱۰۰ کاست موسیقی سنتی ایرانی داشتن. یکی از محصولات شرکت ایشون که مورد توجه همگان قرار می‌گیرد، چنگ لیزری است که قابلیت اجرای همزمان ۱۲ نت را دارد. اساس کار این چنگ بر قطع پرتو نورهای لیزر است.

---

[1] طراحی بیش از ۹۰ درصد تجهیزات، و پیشنهادنامه‌ی آن‌ها توسط دکتر یغمایی نوشته شده‌است.

و نکات بسیار زیادی از زندگی ایشون که این مختصر، مجالی برای عرض اندام آن‌ها نمی‌گذارد. این شخصیت کسی نبود جز مرحوم زنده یاد استاد دکتر ***خشایار یغمایی*** که دیگر در میان ما نیست ...

فقط میتونم بگم که ایشون نه تنها در درس و کار، بلکه در بسیاری از زمینه‌های زندگی، استاد من بود. استادی که با مهربانی، سعی در آموزش شاگردانش داشت و سعی می‌کرد تمام رفتارهایش حساب شده باشد. گفتارش سنجیده بود و کسی را آزرده خاطر نمی‌ساخت. تا جایی که می‌توانست در اعطای نمره به دانشجویانش کمک می‌کرد. او کسی بود که کلمه‌ی **" انسان"** برازنده‌ی او بود.

یادش گرامی، نامش جاوید، راهش پاینده، و روحش قرین رحمت ایزد یکتا و بی‌همتا باد ...

از سمت راست: دکتر یغمایی، نعیم غفاری، محمدصادق رستمی؛ کلاس فیلتر و سنتز؛ روز معلم ۱۳۸۹

چکیده

با پیشرفت چشمگیر علوم رایانه در چند دهه‌ی اخیر و پررنگ‌تر شدن نقش آن در زندگی بشر، علوم مختلفی در کنار آن بوجود آمده و رشد کرده‌اند. یکی از این علوم هوش مصنوعی است و خود نیز به شاخه‌های متعددی تقسیم می‌شود که یکی از آن‌ها پردازش زبان طبیعی است. از شاخه‌های پردازش زبان طبیعی می‌توان به تبدیل متن به گفتار اشاره کرد که بیش از پنج دهه بر روی این موضوع کار شده‌است. در تبدیل متن به گفتار، هدف این است که رایانه قادر باشد تا یک متن را با صدای طبیعی –صدای انسان- بخواند. بدین منظور دو کار باید صورت گیرد: پردازش متن، سنتز گفتار. حدود دو دهه است که زبان فارسی نیز در پردازش رایانه‌ای مورد توجه قرار گرفته است.

برای سنتز گفتار، از سه روش عمده استفاده می‌شود. روش اول که سعی در مدل‌سازی دقیق سیستم تولید گفتار انسان دارد، روش دوم که براساس استفاده از فیلترهای سری-موازی برای مدل‌کردن آواها بنا شده است و روش سوم که از قطعات از پیش ضبط شده استفاده می‌کند. آزمایشات نشان داده است که روش سوم، گفتار طبیعی‌تری تولید می‌کند.

هدف از این پایان‌نامه، تولید گفتار طبیعی در زبان فارسی، با توجه به این سه موضوع است: تکیه‌ی کلمات، آهنگ جملات و کسره‌ی اضافه. در روش پیشنهادی، سنتز گفتار بر اساس استفاده از قطعات از پیش ضبط‌شده به نام دایفون است که دلایل استفاده از آن نیز مطرح شده است. ابتدا با توجه به بحث‌های زبان‌شناسی، نقش کلمات در جملات تعیین می‌گردد و با کمک آن، نوع کلمه مشخص می‌شود و با دانستن نوع کلمه، جایگاه تکیه در آن استخراج می‌گردد. سپس نوع جملات مشخص شده و با توجه به آن آهنگ جملات تعیین می‌شود. همچنین باید جایگاه کسره‌های اضافه در متن نیز مشخص گردد. با توجه به این مطالب، دایفون‌ها در هفت حالت استخراج می‌شوند و در زمان سنتز گفتار، با توجه به تکیه‌ی کلمات و آهنگ جملات و مکان کسره‌ی اضافه کنار هم قرار می‌گیرند.

بررسی نتایج نشان می‌دهد از آن‌جا که آهنگ جملات و تکیه کلمات و پیوستگی در مکان‌های کسره‌ی اضافه، درون قطعات از پیش ضبط‌شده –دایفون‌ها- موجود است، استفاده از این روش نوای گفتار طبیعی‌تری نسبت به برخی سیستم‌های موجود، تولید می‌کند. معیار مقایسه نیز *MOS* بوده است که این مقدار در روش پیشنهادی برای طبیعی بودن و اعمال نوای گفتار در چند حالت مختلفی که بررسی شده‌است، حدود ۴/۵ به‌دست آمده است.

واژگان کلیدی: پردازش زبان طبیعی، تبدیل متن به گفتار، سنتز گفتار، قطعات از پیش ضبط‌شده، دایفون، نوای گفتار، آهنگ جمله، تکیه کلمه، نقش کلمه در جمله، کسره‌ی اضافه

فهرست مطالب

فصل اول: مقدمه ای بر تبدیل متن به گفتار (۱-۶)

۱-۱. خاستگاه تبدیل متن به گفتار ۱

۱-۲. کاربردهای تبدیل متن به گفتار ۲

۱-۳. هدف و بیان پروژه ۳

۱-۴. تاریخچه‌ای مختصر از تبدیل متن به گفتار ۴

۱-۵. مرور مختصر فصل‌ها ۵

فصل دوم: پردازش زبان طبیعی، تبدیل متن به گفتار، ارزیابی (۷-۲۴)

۲-۱. پردازش زبان طبیعی ۷

۲-۱-۱. پردازش زبان طبیعی چیست؟ ۷

۲-۱-۲. تاریخچه‌ی پردازش طبیعی ۸

۲-۱-۳. نیازمندی‌های پردازش زبان طبیعی ۹

۲-۱-۴. مقوله‌های زبان طبیعی ۱۰

۲-۱-۵. کاربردهای پردازش زبان طبیعی ۱۱

۲-۲. تبدیل متن به گفتار ۱۲

۲-۲-۱. واحدهای پردازش دادگان ۱۲

۲-۲-۲. هدف و کاربرد ۱۳

۲-۲-۳. خواص گفتار و نوشتار ۱۴

۲-۲-۴. بیان ساده‌ای از یک سیستم تبدیل متن به گفتار ۱۶

۲-۲-۵. مسائل اساسی در تبدیل متن به گفتار ۱۷

۲-۳. سنجش عملکرد ۱۸

۲-۳-۱. روش‌های ارزیابی سیستم‌های تبدیل متن به گفتار ۱۸

۲-۳-۲. ارزیابی جنبه‌های متفاوت صوتی ۲۰

۲-۳-۳. ارزیابی نوا ۲۲

۲-۳-۴. ارزیابی خوشایند بودن و طبیعی بودن ۲۲

۲-۳-۵. ارزیابی صحت سنتز ۲۳

فصل سوم: پردازش متن، تولید گفتار، محصولات (۲۵-۴۳)

۳-۱. پردازش متن ۲۵

۳-۱-۱. داده‌های مختلف در پردازش متن ۲۵

۳-۱-۲. مسائل و چالش‌های پردازش متن ۲۵

۳-۱-۳. برخی از فعالیت‌های انجام شده در پردازش متن در زبان فارسی ۲۸

۳-۲. سنتز گفتار ۲۹

۳-۲-۱. تولید شکل موج گفتار یا تولید گفتار سطح پایین ۲۹

۳-۲-۲. واحدهای استفاده شده در سنتز گفتار ۳۴

۳-۲-۳. نوای گفتار ۳۵

۳-۲-۴. استفاده از تکیه و آهنگ برای طبیعی‌تر شدن نوای گفتار ۳۹

۳-۲-۵. روش‌های تخمین پارامترها برای استفاده در سیستم تبدیل متن به گفتار ۴۰

۳-۳. معرفی برخی از محصولات ۴۲

فصل چهارم: آهنگ جملات و تکیه‌ی کلمات (۴۴-۶۰)

۴-۱. آهنگ جملات ۴۴

۴-۱-۱. مشخص کردن نوع جمله ۴۴

۴-۱-۲. بررسی آهنگ در انواع جملات ۴۵

۴-۲. تکیه کلمات ۵۱

۴-۲-۱. بررسی تکیه در کلمات ۵۲

۴-۲-۲. مشخص کردن نقش کلمه ۵۵

۴-۳. مشخص کردن مکان کسره‌ی اضافه و نقش آن در آهنگ گفتار و تکیه کلمات ۵۶

۴-۴. رابطه‌ی بین هجا و دایفون ۵۷

فصل پنجم: روش پیشنهادی (۶۱-۸۵)

۵-۱. بیان صریح روش پیشنهادی ۶۱

۵-۲. دلایل استفاده از دایفون ۶۳

۵-۳. بیان فرضیات ۶۳
۵-۴. بیان الگوریتم ۶۴
۵-۵. بلوک دیاگرام کلی سیستم ۶۶
۵-۶. بررسی یک مثال ۶۶
۵-۷. پیاده‌سازی سیستم تبدیل متن به گفتار بر اساس روش پیشنهادی ۶۸
۵-۷-۱. پیش‌پردازش متن ۶۹
۵-۷-۲. تعیین مرز دایفون‌ها ۷۰
۵-۷-۳. استخراج دایفون‌ها ۷۳
۵-۷-۴. جستجو در دایفون‌ها ۷۴
۵-۷-۵. هنجارسازی دایفون‌ها و اتصال آن‌ها ۷۵
۵-۷-۶. محاسبه‌ی حجم اشغالی حافظه در روش پیشنهادی ۷۷
۵-۷-۷. تهیه‌ی پایگاه داده‌ی صوتی برای دایفون‌ها در حالت‌های مختلف ۸۰
۵-۷-۸. تهیه‌ی پایگاه داده‌ی متنی برای کلمات و نقش‌های ممکن و دایفون‌های آن ۸۰
۵-۷-۹. نکات عملی مرتبط با استخراج دایفون‌ها و اتصال آن‌ها ۸۲

فصل ششم: آزمایشات و نتیجه گیری (۸۶-۹۱)

فصل هفتم: پیشنهادها (۹۲-۹۴)

منابع و مراجع (۹۵-۹۹)

واژه‌نامه انگلیسی به فارسی (۱۰۰-۱۰۳)

کلمات اختصاری (۱۰۳)

پیوست (۱۰۴-۱۱۴)

فهرست تصاویر

شکل ۲-۱: پردازش زبان طبیعی در یک نگاه ۷
شکل ۲-۲: بلوک دیاگرام کلی یک سیستم پردازش زبان طبیعی ۸
شکل ۲-۳: تست تورینگ برای تشخیص هوشمندی ۹
شکل ۳-۱: نمایشی از سیستم تولید گفتار انسان ۳۱
شکل ۳-۲: کلیت روش فرمانت در تبدیل متن به گفتار ۳۲
شکل ۳-۳: راست: قسمتی از بیان گفتاری "آ"، چپ: قسمتی از بیان گفتاری "ز" ۳۶
شکل ۳-۴: طیف شکل ۳-۳ (راست: "آ"، چپ: "ز")؛ همان‌طور که مشاهده می‌شود، "آ" شبه‌متناوب است و "ز" فرکانس‌های بالاتری دارد ۳۶
شکل ۳-۵: حرف "گ" در دو موقعیت مختلف؛ بازه‌ی زمانی نیز متفاوت است ۳۷
شکل ۳-۶: سیگنال زمانی بیان کلمه‌ی "گفتار" با دو شدت متفاوت ۳۸
شکل ۳-۷: جهش فاز در محل به هم پیوستن دو قطعه؛ سبب بروز فرکانس‌های بالا ۳۹
شکل ۴-۱: نمایش سیگنال حوزه‌ی زمان (سمت چپ) و اندازه‌ی حوزه‌ی فرکانس (سمت راست) برای فعل "خوانده‌ای" در حالت خبری ساده ۴۶
شکل ۴-۲: نمایش سیگنال حوزه‌ی زمان (سمت چپ) و اندازه‌ی حوزه‌ی فرکانس (سمت راست) برای فعل "خوانده‌ای" در حالت سوالی ساده ۴۷
شکل ۴-۳: نمایش سیگنال حوزه‌ی زمان (سمت چپ) و اندازه‌ی حوزه‌ی فرکانس (سمت راست) برای فعل "خوانده‌ای" در حالت سوالی، شروع با کلمه‌ی پرسش "آیا..." و یا سوالی که جواب آن "بلی" یا "خیر" است ۴۸
شکل ۴-۴: نمایش سیگنال حوزه‌ی زمان (سمت چپ) و اندازه‌ی حوزه‌ی فرکانس (سمت راست) برای فعل "خوانده‌ای" در حالت سوالی با کلمه‌ی پرسش "چه..." ۴۸
شکل ۴-۵: نمایش سیگنال حوزه‌ی زمان (سمت چپ) و اندازه‌ی حوزه‌ی فرکانس (سمت راست) برای فعل "خوانده‌ای" در حالت تعجبی ۴۹
شکل ۴-۶: نمایش سیگنال حوزه‌ی زمان (سمت چپ) و اندازه‌ی حوزه‌ی فرکانس (سمت راست) برای فعل "خوانده‌ای" در حالت تعجبی-سوالی ۴۹

شکل ۴-۷: طیف‌نگار واژه‌ی "خوانده‌ای" در حالت سوالی با واژه‌ی پرسشی "آیا"؛ تصویر بالا: محتویات فرکانسی (محور عمودی، فرکانس بر حسب هرتز) تصویر پایین: شکل موج زمانی سیگنال گفتاری ۵۰
شکل ۴-۸: طیف‌نگار واژه‌ی "خوانده‌ای" در حالت خبری ساده، تصویر بالا: محتویات فرکانسی. تصویر پایین: شکل موج زمانی سیگنال گفتاری ۵۰
شکل ۴-۹: نمایش سیگنال گفتاری در حوزه‌ی زمان (سمت چپ) و اندازه‌ی حوزه‌ی فرکانس (سمت راست) برای کلمه‌ی "کتاب"، در حالتی که هر دو هجا بدون تکیه بیان شده‌اند ۵۱
شکل ۴-۱۰: نمایش سیگنال گفتاری در حوزه‌ی زمان و اندازه‌ی حوزه‌ی فرکانس برای کلمه‌ی "کتاب"، در حالتی که تکیه‌ی کلمه، بر روی هجای دوم قرار گرفته است ۵۲
شکل ۴-۱۱: تکیه در واژه‌های: کتاب، کتاب‌ها، کتاب‌هایش. در هر سه واژه، تکیه در آخرین هجا قرار می‌گیرد. به تغییرات زمانی (شکل پایینی) و فرکانسی سیگنال (شکل بالایی- محور عمودی: هرتز) که با مستطیل مشخص شده‌اند توجه شود ۵۴
شکل ۴-۱۲: نمایش طیف‌نگار (بالا) و شکل زمانی (پایین) عبارت "کتابِ‌زیبا"، در حالی که "ب" و "کسره" و "ز" ، به صورت پیوسته بیان شده‌اند؛ بدین صورت: ب/کسره/ز ۵۸
شکل ۴-۱۳: نمایش طیف‌نگار (بالا) و شکل زمانی (پایین) عبارت "کتاب– زیبا"، در حالی که "ب" و "کسره" ، جدا بیان شده‌اند؛ بدین صورت: ب/سکون/کسره ۵۸
شکل ۴-۱۴: استخراج دایفون‌ها برای واژه‌ی "کتاب‌دار"؛ شماره‌ی ۱، متناظر با دایفون "–ک" و شماره ۹ متناظر با "ر-". باقی دایفون‌ها نیز در متن، مشخص شده است ۵۹
شکل ۵-۱: دایفون سکون/میم در ابتدای واژه‌ی "من"، سمت چپ: نمایش زمانی دایفون، سمت راست: اندازه‌ی محتوای فرکانسی ۶۲
شکل ۵-۲: دایفون سکون/میم در ابتدای واژه‌ی "می‌توانم"، سمت چپ: نمایش زمانی دایفون، سمت راست: اندازه‌ی محتوای فرکانسی ۶۲
شکل ۵-۳: دایفون سکون/میم در ابتدای واژه‌ی "می‌توانی؟"، سمت چپ: نمایش زمانی دایفون، سمت راست: اندازه‌ی محتوای فرکانسی ۶۲
شکل ۵-۴: بلوک دیاگرام کلی سیستم پیشنهادی ۶۷

شکل ۵-۵: مقایسه‌ی دو دایفون سکون/سین (۱) و سین/فتحه (۲)؛ تصویر بالا: طیف‌نگار (محور عمودی، فرکانس بر حسب هرتز است)، تصویر پایین: شکل موج زمانی سیگنال ۷۱
شکل ۵-۶: استخراج دایفون‌ها برای کلمه‌ی "گفتار"؛ این کلمه از ۲ هجا و ۸ دایفون تشکیل شده است. تصویر بالا: طیف‌نگار سیگنال (محور عمودی فرکانس بر حسب هرتز)، تصویر وسط: انرژی متوسط سیگنال، تصویر پایین: شکل موج زمانی سیگنال ۷۲
شکل ۵-۷: جداسازی دایفون‌ها برای "واژه" سلام، با استفاده از طیف‌نگار و شکل موج زمانی سیگنال گفتار و تست شنیداری ۷۳
شکل ۵-۸: اختلاف میانگین زمانی قطعات از پیش ضبط شده، در زمان اتصال ۷۶
شکل ۵-۹: فاصله میان‌کلمه و میان‌هجا در جمله‌ی /به/-/مد//ر//سه/-/می//ر//وم/. همان‌گونه که مشاهده می‌شود فاصله میان‌کلمه بیش از فاصله میان‌هجاست ۷۷
شکل ۵-۱۰: محتویات فرکانسی موجود در یک سیگنال گفتار به مدت ۸ ثانیه. فرکانس نمونه‌برداری این سیگنال، ۸ کیلوهرتز بوده است. همان‌طور که ملاحظه می‌شود، سهم بیشتری از محتویات فرکانسی، بین ۰ تا ۴ کیلوهرتز قرار دارند ۸۳
شکل ۵-۱۱: نمونه‌ی سمت راست نسبت سیگنال به نویز بیشتری دارد و بهتر است ۸۵
شکل ۵-۱۲: تفاوت در سیگنال‌های صوتی با ذخیره سازی در عمق حافظه‌ی متفاوت. خط‌چین: عمق حافظه‌ی ۵ بیتی، خط: عمق حافظه‌ی ۸ بیتی ۸۵
شکل۱-پیوست: نمای کلی نرم‌افزار متمتیکا ۱۰۴
شکل۲-پیوست: اسپکتروم آنالایزر نوعی برای سیگنال گفتار یک کلمه ۱۱۰
شکل۳-پیوست: یک فریم ۱۰۲۴ نمونه‌ای. تصویر چپ: طیف (محور عمودی: فرکانس، محور افقی: اندازه دامنه‌ی طیف)، تصویر راست: شکل حوزه‌ی زمان همان فریم ۱۱۰
شکل۴-پیوست: نمونه‌ای از پیکره‌ی بی‌جن خان. راست: کلمات موجود در پیکره بر اساس چیدمان طبق حروف الفبا. چپ: کلمات موجود در پیکره بر اساس ترتیب نزولی تعداد تکرار ... ۱۱۲
شکل۵-پیوست: نمونه‌ای از برچسب‌های پیکره‌ی بی‌جن خان ۱۱۳
شکل۶-پیوست: قسمتی از پیکره‌ی بی‌جن خان؛ جملات تفکیک شده‌اند و نقش کلمه در جمله، روبروی هر کلمه نوشته شده است) ۱۱۳

فصل اول: مقدمه ای بر تبدیل متن به گفتار

۱-۱. خاستگاه تبدیل متن به گفتار

بیش از چند دهه از حضور رایانه در بین جوامع نگذشته است که با پیشرفت روزافزون آن، کاربردهای مختلفی ارائه شده است. استفاده از رایانه درزندگی، در حال حاضر به قدری گسترش پیدا کرده است که تصور زندگی بدون آن، کمی دشوار است. تقریبا در تمام قسمت‌های زندگی، رایانه، اثری از خود را نشان می‌دهد. اما در جوامع علمی و دانشگاهی، استفاده‌هایی به مراتب بیشتر و متنوع‌تر، در حوزه‌های مختلف از آن صورت می‌گیرد.

یکی از این حوزه‌ها، بحث‌های پردازشی است. با توجه به قدرت محاسباتی بسیار سریع در رایانه، به منظور پردازش سریع اطلاعات از آن استفاده می‌شود. داده‌هایی که به عنوان ورودی وارد یک رایانه می‌شوند، دیجیتال نامیده می‌شوند. بنابراین، پردازش دیجیتال داده‌ها، یکی از کاربردهای اساسی رایانه خواهد بود.

بحث دیگری که در دهه‌های گذشته، رشد چشمگیری داشته و فعالیت‌های بسیار زیادی بر روی آن صورت گرفته است، هوش مصنوعی[۱] است؛ که در واقع هدف از آن، ارائه‌ی سیستم‌های هوشمندی است که توسط انسان ساخته می‌شود و توسعه پیدا می‌کنند [1]؛ در این‌جا نیزاز رایانه کمک گرفته می‌شود. می‌توان این‌گونه گفت که رایانه ابزاری شده است برای هوشمندسازی سیستم‌ها. در کنار بحث هوش مصنوعی، حوزه‌ی دیگری از علوم نیز رشد داشته است که به پردازش زبان طبیعی[۲] موسوم است [1] که هدف نهایی آن، تولید سیستم‌هایی است که به زبان انسان باشد؛ یعنی زبان انسان را بفهمد و زبان انسان را تولید کند. به عبارت دیگر هدف نهایی از بحث هوش مصنوعی، سیستمی است که مانند انسان حرف بزند، بفهمد، فکر کند و بتواند با دیگران ارتباط برقرار کند.

در تعاملات روزمره، برای برقراری ارتباط با دیگران از روش‌های (کانال‌های) مختلفی استفاده می‌شود: متن، گفتار، تصویر و علامت‌های (سیگنال‌های) دیگر که در بین آن‌ها، متن و گفتار، با توجه به ماهیتشان، جایگاه ویژه‌ای دارند. اگر سیستم مورس (تلگراف) را اولین سیستم دیجیتال در نظر بگیریم، اولین پردازش‌های دیجیتال، بر روی متون صورت می‌گرفت [1]؛ که در سال‌های بعد با حضور

---

[۱] Atrificial Intelligence

[۲] Natural Language Processing (NLP)

رایانه، سیگنال‌های بیشتری مورد بررسی قرار گرفت.

سیگنال‌های آنالوگ[1] مانند صوت، قبل از ورود به رایانه، به سیگنال‌های دیجیتال تبدیل می‌شوند. از طرفی، خروجی رایانه نیز باید سیگنالی پیوسته (یا شبه پیوسته) باشد. پردازش بر روی داده‌های دیجیتال صورت گرفته و در نهایت، با تقریب‌هایی، داده‌های دیجیتال به آنالوگ تبدیل شده و به عنوان خروجی رایانه، قابل استفاده‌ی انسان می‌گردد [2].

یکی از کاربردهایی که امروزه، از آن بهره گرفته می‌شود، تبدیل متن به گفتار[2] است. یعنی ورودی رایانه، یک متن بوده و خروجی آن، گفتار. این متن می‌تواند به صورت‌های مختلف در اختیار رایانه قرار گیرد. برای ساده‌تر شدن قضیه فرض کنید این متن، رشته‌ای از کاراکترهاست که پشت سر هم وارد رایانه شده و قرار است به گفتار تبدیل شود؛ مسایل مختلفی وجود دارد و به شاخه‌های متعددی از علوم مختلف نیاز است که در این پایان‌نامه، در حد توان به برخی از آن‌ها اشاره شده است. به عنوان مثال، یکی از مهم‌ترین علومی که در تبدیل متن به گفتار، مورد نیاز است، دانش زبان‌شناسی[3] است.

۱-۲. کاربردهای تبدیل متن به گفتار

کاربردهای متنوعی برای بحث تبدیل متن به گفتار می‌توان در نظر گرفت. یکی از اولین مواردی که به نظر می‌رسد، استفاده از آن برای کمک به نابینایان است. در واقع اگر سیستمی باشد که متن را بخواند، یک فرد نابینا را از خط بریل - برای خواندن متن- یا این‌که فرد دیگری برای او متنی را بخواند، بی نیاز می‌کند. اما برای استفاده در این مورد، باید مسایل بسیار زیادی در نظر گرفته شود و سیستم به قدری بهینه باشد که نیاز یک فرد نابینا را به طور کامل بر طرف نماید.

از کاربردهای دیگر تبدیل متن به گفتار، سیستم‌های گویا را می‌توان نام برد و یا سیستم‌های تلفن بانک یا استفاده در فرودگاه‌ها و یا هر کاربرد دیگری که به صورت موردی تعریف شود [1]. حتی با یک سیستم تبدیل متن به گفتار، می‌توان کتاب‌ها، اخبار، هواشناسی و ... را به جای این‌که بخوانیم، بشنویم. همچنین استفاده از این سیستم، استفاده از رایانه را برای افراد نابینا فراهم می‌آورد و در یک کاربرد گسترده‌تر، امکانات بسیار متنوعی به وسیله‌ی اینترنت، برای کاربرانی که از این سیستم استفاده می‌کنند، فراهم می‌آید. با توجه به اینکه مدت زمانی از طول روز به رفت و آمد می‌گذرد، استفاده از

---

[1] Analog

[2] Text to Speech

[3] Linguistic knowledge

سیستم‌های گویشی، سبب صرفه جویی در وقت نیز می‌شوند.

۱-۳. هدف و بیان پروژه

با این‌که بیش از چند دهه است که بر روی سنتز[۱] گفتار و تبدیل متن به گفتار کار شده است، اما فقط چند سالی است که در زبان فارسی به این موضوع پرداخته شده است. نکته‌ی مورد توجه در این‌جا این است که، با توجه به این‌که نظام آوایی[۲] برای هر زبان، منحصر به همان زبان است و در حالت کلی نمی‌توان به زبان دیگری تعمیم داد و این‌که نحوه تولید نوا معمولا در زبان‌های مختلف، متفاوت است (برای مثال، نظام آوایی و ساختاری زبان مردم شرق آسیا، بیشتر براساس مصوت‌هاست اما زبان فارسی ترکیب بیشتری از صامت‌ها و مصوت‌ها را دارد)، نمی‌توان سیستم‌هایی که برای زبان‌های دیگر ساخته و بهینه شده است را در زبان فارسی عینا استفاده کرد؛ زیرا آن سیستم‌ها با توجه به نظام آوایی همان زبان بهینه شده است [۱]. بنابراین باید مطالعات گسترده‌ای در زبان فارسی صورت گیرد؛ به عبارت دیگر، باید نظام آوایی و ساختاری زبان فارسی، به طور کامل مورد بررسی قرار گیرد و مشخصات آن استخراج شود. به عنوان نمونه یکی از مشکلاتی که در زبان فارسی یا عربی وجود دارد، ننوشته شدن حرکات کوتاه مانند کسره است [3, 4].

گفتار تولید شده توسط ماشین، علاوه بر قابل فهم بودن، باید طبیعی نیز باشد؛ یا به عبارت دیگر مانند روبات نباشد [5]. هرچند که در زبان فارسی، سیستم‌هایی ساخته شده است و به صورت محصولات آماده در اختیار عموم قرار داده‌شده است (که به نمونه‌هایی از این محصولات در انتهای فصل سوم اشاره شده است) که بسیاری از معیارها و استانداردها را برآورده کرده‌اند، اما هنوز فعالیت‌ها و مطالعات بر روی طبیعی‌تر کردن گفتار تولید شده، ادامه دارد.

یکی از خصوصیات نوشتار زبان فارسی، ننوشته‌شدن کسره‌ی نقش نمای اضافه در عبارات وصفی و اضافی است [6]. در عبارت "کتاب سبز" یا "کتاب پرویز" کسره نوشته نشده است بنابراین سیستم تبدیل متن به گفتار باید قادر باشد تا این نوع ترکیب را تشخیص داده و در زمان تولید گفتار، آن را لحاظ کند. هرچند که طی چهار سال گذشته، در این قسمت نیز فعالیت‌هایی انجام شده است، در این پایان‌نامه، با رویکرد دیگری به این امر پرداخته شده است.

جنبه‌ی دیگری از فعالیت‌هایی که در سال‌های اخیر به آن پرداخته شده است، طبیعی‌تر کردن

---

[۱] Synthesis

[۲] Phonetic

گفتار، با توجه به آهنگ[1] جملات و تکیه[2] کلمات است [۲،۳][6, 7, 8, 9, 10]. در این پایان‌نامه، روشی برای اضافه کردن آهنگ و تکیه با استفاده از روش‌های سنتزموجود ارائه شده است.

۱-۴. تاریخچه‌ای مختصر از تبدیل متن به گفتار[3]

پیشینه‌ی پردازش دیجیتال گفتار به بیش از پنج دهه می‌رسد. این شاخه از علم، به سه بخش عمده تقسیم می‌شود: کدینگ گفتار، تشخیص گفتار، تولید گفتار. هدف اصلی از کدینگ گفتار، استفاده‌ی بیشتر از پهنای باند کانال ارتباطی است. البته باید توجه داشت که کاهش اطلاعات به میزانی باشد که صوت منتقل شده کیفیت قابل قبولی داشته باشد. امروزه این سیستم‌ها به میزان زیادی بهینه شده‌اند و در تلفن‌های همراه و انتقال بی‌سیم اطلاعات به وفور مورد استفاده قرار می‌گیرند. در بحث تشخیص گفتار، هدف اصلی این است که یک سیستم پردازشی، گفتار تولید شده را شناسایی کند. به عنوان نمونه یک سیستم تایپ گفتاری را در نظر بگیرید که در آن تایپیست به جای آن‌که متن را به کمک صفحه کلید تایپ کند، می‌خواند و گفتار وی پردازش شده و توسط سیستم به نوشتار تبدیل می‌شود. تولید گفتار، نسبت به کدینگ و تشخیص کمی پیچیده‌تر است زیرا باید در آن ابتدا یک مدل فیزیکی کاملی از سیستم تولید گفتار انسان ارائه شود.

در تولید گفتار، ابتدا سیستم‌هایی ارائه می‌شد که سعی در مدل‌سازی سیستم تولید گفتار انسانی بطور مستقیم داشت. با توجه به پیچیدگی این مدل، در عمل نتوانستند از این مدل‌ها استفاده کنند. روش بعدی که مورد استفاده قرار گرفت، مدل ساده‌ای از سیستم تولید گفتار انسان بود؛ بدین صورت که بر اساس فیلتری مبتنی بر تعدادی قطب کار می‌کند. تولید صوت به کمک این روش موفق بود اما در ابتدا صداهای تولیدی، کاملا روبوتیک بودند و گوش دادن طولانی به آن‌ها خسته کننده بود. اما در ادامه به کمک روش‌های مبتنی بر قانون[4] نوای گفتار نیز شبیه‌سازی شد و به صدای تولیدشده اضافه شد؛ هرچند که از حالت روبوتیک خارج شد اما هنوز طبیعی نبود.

در اواخر دهه‌ی هفتاد میلادی، همزمان با روش‌های قبلی، روش جدیدی ابداع شد که در آن به کمک اتصال قطعات از پیش ضبط شده، گفتار مورد نظر تولید می‌شد. ساده‌ترین مثال برای این روش

---

[1] Intonation

[2] Stress

[3] مطالب این قسمت برگرفته از [5] است.

[4] Rule Based