

ساختار یک سیستم تبدیل متن به گفتار برای زبان فارسی

بهرنگ قاسمی زاده و مجید نم نبات

QasemiZadeh@digitalclone.net

چکیده

در این مقاله ساختار یک سیستم تبدیل متن به گفتار برای زبان فارسی شرح داده می شود. این سیستم از دو بخش عمده پردازش زبان طبیعی و پردازش سیگنال تشکیل شده است. هر یک از این بخشها از مجموعه ای از پیمانها های مختلف با کارکردهای متفاوت جهت استخراج پارامترهای مهم در سنتز گفتار تشکیل شده اند. پردازشهای لازم توسط این پیمانها ها بر روی متن ورودی به گونه ای در نظر گرفته شده است که علاوه بر تولید خروجی با کیفیت مناسب، سیستم امکان پردازش برخط و تولید گفتار خروجی بصورت همزمان را نیز داشته باشد.¹

کلمات کلیدی: سیستم تبدیل متن به گفتار، پردازش زبان طبیعی، پردازش دیجیتالی سیگنال، سنتز گفتار.

مقدمه

یکی از مسائل مهم و مورد توجه محققان در هوش مصنوعی و زمینه های تحقیقاتی مرتبط با آن، تبدیل خودکار متون الکترونیکی به گفتار² است. امروزه برای اکثر زبانهای زنده دنیا سیستم های تبدیل متن به گفتار وجود دارد، به عنوان نمونه می توان از سیستم های تبدیل متن به گفتار معرفی شده در [1]، [2]، [3] و [4] نام برد. شاید بتوان قدیمی ترین سیستم تبدیل متن به گفتار را MITTalk برای زبان انگلیسی دانست [5]. در این سیستم پارامترهای مهم تولید گفتار به سادگی و به روشی قانون-گرا پیشگویی می شوند و با استفاده از یک موتور سنتز فرمندی گفتار خروجی تولید می گردد. از سوی دیگر شاید بتوان سیستم تبدیل متن به گفتار Festival [6] را یکی از مدرن ترین این سیستم ها نام برد که به شکل کد-باز و برای استفاده محققین با رعایت اصول مهندسی نرم افزار ارائه شده است. این سیستم ابتدا برای زبان انگلیسی ارائه و سپس به مرور زمان، زبانهای دیگری را نیز در بر گرفت. با دانش نویسندگان این مقاله تنها سیستم تبدیل متن به گفتار ارائه شده به شکل تجاری برای زبان فارسی سیستم تبدیل متن به گفتار شرکت هوش مصنوعی رایبوز است، هر چند که نمونه های تحقیقاتی دیگری نیز پیش از این سیستم ارائه شده است.

بطور کلی برای تبدیل متن به گفتار نیاز است تا پارامترهای مهم در تولید گفتار استخراج و استفاده شوند. استخراج بعضی از این پارامترها همچون رشته آوایی نوشتار، کشش زمانی واجها و منحنی پیچ گفتار الزامی و بعضی دیگر همچون سرعت بیان، درنگ بین کلمات و تن صدا از جمله پارامترهای غیر ضروری می باشند که استفاده از آنها منجر به بهبود خروجی سیستم می گردد. این پارامترها را هم می توان به کمک پردازش های مبتنی بر قانون [7][8] و هم به کمک روش های آماری [9][10] استخراج کرد. به عنوان مثال تعیین رشته آوایی یا تبدیل حرف به صدا³ هم به کمک روشهای مبتنی بر آمار و یادگیری ماشین و هم با استفاده از روشهای مبتنی بر قانون امکان پذیر است. روش تعیین هر یک از این پارامترها به شدت وابسته به طبیعت زبانی است که تولید گفتار از متن برای آن صورت می گیرد. به عنوان مثال روش های تبدیل حرف به صدا به کمک روش های آماری برای زبان انگلیسی از دقت مناسبی

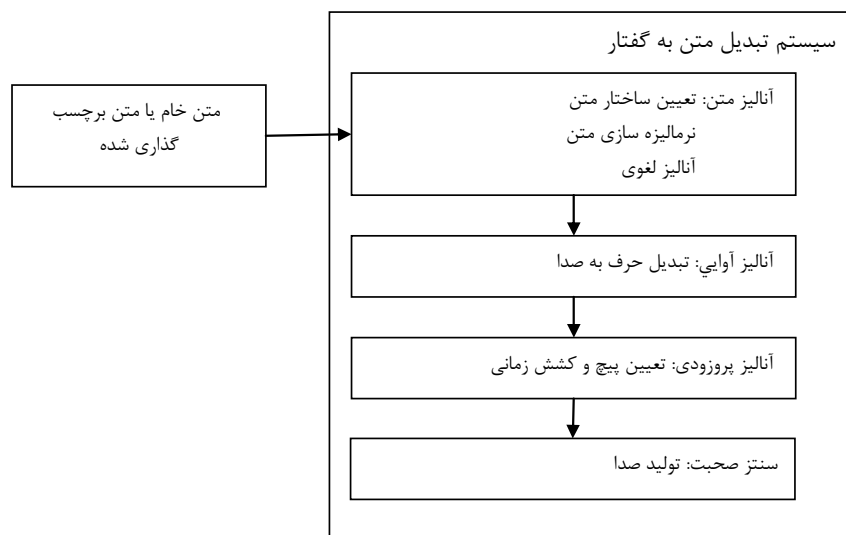
¹ سیستم معرفی شده، در محیط VC++ پیاده سازی شده است.

² Text-To-Speech

³ Letter To Sound (LTS)

برخوردارند اما در ادامه استفاده از چنین روشهایی برای زبان فارسی، دقت مناسبی را نخواهد داشت، این مساله ناشی از شکل نوشتار زبان فارسی و حذف آوایهای صدادر کوتاه از صورت نوشتاری است.

در شکل 1 پیمانها برای تبدیل متن به گفتار در یک سیستم کلاسیک تولید گفتار از متن نشان داده شده است. این پیمانها را می توان به دو بخش عمده بخش پردازش زبان طبیعی⁴ و بخش پردازش دیجیتال سیگنال⁵ تقسیم بندی نمود. در بخش NLP اجزاء سازنده متن ورودی شامل کلمات، گروههای گرامری، جملات و دیگر اطلاعات مرتبط با پردازش زبان که در بخش پردازش سیگنال دیجیتال به عنوان ورودی مورد نیاز است تهیه می شود. بخش DSP با استفاده از نتایج حاصل از بخش NLP، پارامترهای گفتار را شامل رشته آوایی متن و پارامترهای نوایی تعیین و در انتها گفتار خروجی با استفاده از یک موتور سنتز تولید می گردد. در این پژوهش معماری یک سیستم پیشنهادی برای تبدیل متن به گفتار توضیح داده شده است. برای تشریح ساختار این سیستم، روش استخراج هر یک از پارامترهای مهم در تولید گفتار از متن ورودی بیان شده است. با توجه به طبیعت زبان فارسی در استخراج هر یک از این پارامترها هم از روش های مبتنی بر قانون و هم از روش های آماری و یا ترکیبی از این دو استفاده شده است. بنا بر آنچه که گفته شد، مقاله ارائه شده بدین ترتیب سازمان یافته است: در بخش 2 معماری کلی سیستم مختصرا شرح داده شده است. بخش 3 به تشریح پیمانها پردازش زبان طبیعی و ارائه روشهای به کار برده شده می پردازد. در بخش 4 پیمانها پردازش سیگنال و یا تولید گفتار، متدها و روش های استفاده شده در آن، شرح داده شده است. بخش 5 به طرح مسائل مهم در پیاده سازی سیستم پرداخته است. در نهایت و در بخش 6، خلاصه و نتیجه گیری به همراه سمت و سوی تحقیقات آینده ارائه شده است.



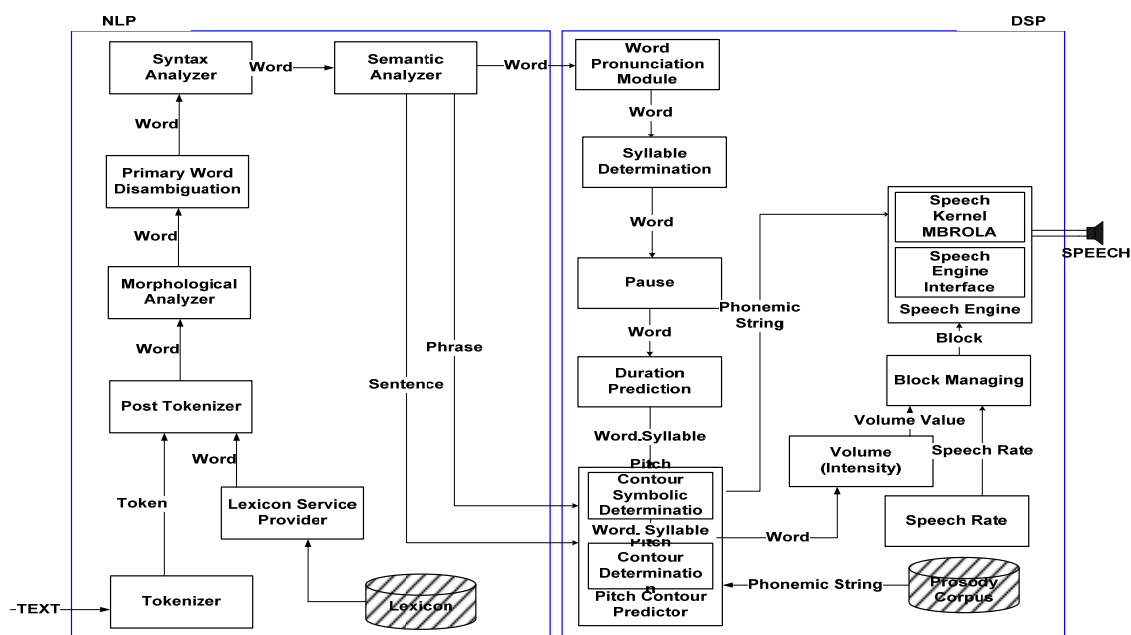
شکل 1: معماری کلی سیستم های تبدیل متن به گفتار

نگاه کلی بر معماری سیستم پیشنهادی

شکل 2 نمایه کلی از معماری سیستم پیشنهادی را نمایش می دهد. در معماری پیشنهادی، سیستم به دو بخش پردازش زبان طبیعی و پردازش سیگنال دیجیتال تقسیم بندی شده است. بخش NLP متن فارسی را از ورودی دریافت می کند و بخش DSP سیگنال گفتار متناسب با آن را برای پخش توسط سیستم ارائه می نماید. بخش NLP از پیمانها های مختلفی همانند تحلیل لغوی گر لغوی، پیمانها رفع ابهام از کلمات، تحلیلگر نحوی و تحلیلگر معنایی ساده تشکیل شده است.

خروجی مهیا شده توسط این بخش شامل ساختار داده چند سطحی است که در پایین ترین سطح از آن کلمات مرتبط با متن ورودی به همراه اطلاعات زبانی مربوط به آن کلمات جای گرفته است. این اطلاعات زبانی شامل رشته آوانگاری واژه، مقوله واژگانی، نقش نحوی و محل آن در گروه های نحوی است. علاوه بر این مرز جملات نیز در این بخش مشخص می شود و ویژگیهای خاص هر جمله به خصوص ویژگیهای آوایی مرتبط با معنی، برای جمله ها استخراج می شود.

بخش DSP از پیمانانه های مختلفی همچون مولد رشته آوایی، تعیین استرس، هجابندی، تعیین کشش زمانی، تعیین درنگ، تعیین منحنی پیچ، بلوک کننده و موتور سنتز تشکیل شده است. این پیمانانه ها بصورت زنجیره ای و بر اساس ترتیب نشان داده شده در شکل 2 داده های ورودی را مورد پردازش قرار می دهند. علاوه بر این پیمانانه ها، پیمانانه های کمکی دیگری همچون دادگان نوایی نیز وجود دارند که برای آموزش مدل های تعیین نوا مورد استفاده قرار می گیرند. بعضی از این پیمانانه ها همچون تعیین منحنی پیچ ممکن است که خود از چند زیرپیمانانه دیگر تشکیل شده باشند. علاوه بر این بعضی از پیمانانه ها همچون تعیین سرعت بیان و شدت صدا⁶ برای توسعه های آتی در نظر گرفته شده و در نسخه کنونی موجود نمی باشند. در ادامه هر یک از بخش های NLP و DSP و نحوه عملکرد آنها شرح داده شده اند.



شکل 2: نمای کلی از معماری سیستم پیشنهادی

بخش پردازش زبان طبیعی

همانطور که پیشتر گفته شد، وظیفه این واحد تعیین نمودن ویژگیهای زبانی متن ورودی، همانند ویژگیهای صرفی و نحوی کلمات متن ورودی است. این کار شامل تعیین مرز کلمات، تخصیص ویژگیهای صرفی، تعیین گروه های نحوی، تعیین مرز جملات و مشخص نمودن نوع جمله شامل پرسشی، تعجبی، امری، خبری به همراه برخی از ویژگیهای معنایی مرتبط با حالت آن مانند خوشحال بودن، عصبانی بودن و ... است.

متن ورودی ابتدا به پیمانانه TOKENIZER وارد می شود. در این پیمانانه به کمک یک ماشین حالت متناهی متن به توکن هایی با انواع مختلف همانند عدد، رشته متنی، علامت سجاوندی و ... تقسیم بندی می شود. این متن توکن بندی شده در اختیار پیمانانه POSTPROCESSING قرار می گیرد. این پیمانانه با استفاده از سرویسهای فراهم آمده توسط واژگان، رشته های متنی را در واژگان جستجو می نماید و در صورت وجود کلمه در واژگان ویژگیهای صرفی به همراه رشته آوانگاری را به رشته متنی نسبت می دهد. علاوه

بر این در این پیمانانه، علامت های سجاوندی دوتایی همانند پرانتزها و روابط میان آنها تشخیص داده می شود و نرمالیزه کردن متن در این پیمانانه تکمیل می شود.

تحلیل گر لغوی وظیفه مشخص نمودن ویژگیهای صرفی و تعیین رشته آوانگاری ریشه کلمات صرف شده را بر عهده دارد. علاوه بر این تحلیل گر لغوی، کلمات همنویسه را مشخص و دیگر گزینش های میسر دیگر را تعیین می نماید. تحلیلگر لغوی به تفصیل در [11] تشریح شده است. توجه به این نکته ضروری است که ممکن است برای هر رشته ورودی چندین کلمه به عنوان گزینش های میسر وجود داشته باشد. این کلمات به شکل یک آرایه برای هر جایگاه در جمله، مشخص می شوند. پس از انجام تحلیل لغوی آرایه کلمات تشخیص داده شده برای هر جمله به پیمانانه Word Sense Disambiguation داده می شود.

در نسخه موجود، وظیفه پیمانانه Word Sense Disambiguation تعیین یکی از گزینش های میسر برای رشته ورودی به عنوان کلمه نهایی است. به عبارت دیگر پس از این پردازش در هر جایگاه تنها یک کلمه به عنوان کلمه نهایی انتخاب می شود. در انجام این پردازش از روشهای ساده قانون مدار به همراه پردازش های آماری استفاده شده است. کلمه مناسب با استفاده از ویژگیهای معنایی ذخیره شده در واژگان به همراه تاریخچه موجود از همنشینی کلمات در کنار یکدیگر گزینش می شود. برای مدل کردن چگونگی هم نشینی کلمات، یک پیکره از متن فارسی تهیه شده است و مدل استفاده شده با کمک این پیکره آموزش داده شده است. پس از گزینش کلمات مناسب، آرایه کلمات مرتبط با هر جمله به پردازشگر نحوی داده می شود.

پردازش نحوی در سیستم پیشنهادی شامل دو مرحله است. در مرحله نخست از این پردازش، ابتدا از یک پردازش مبتنی بر مدل از پیش تعیین شده برای تشخیص ترکیبات اسم و صفت و تعیین کسره های حذف شده از متن استفاده می شود. این کار به کمک یک مدل مبتنی بر درخت تصمیم گیری انجام می شود. این مدل با کمک یک پیکره از جملات فارسی که در آن کلیه ترکیبات اضافی و صفتی مشخص شده اند آموزش داده شده است. پس از تعیین این ترکیبات از یک تحلیل نحوی مبتنی بر قانون کم عمق برای تعیین محدوده گروه های نحوی استفاده می شود. گروههای نحوی تشخیص داده شده سپس به کمک یک تحلیل گر مبتنی بر GPSG⁷ در ساختار درختی مناسب جای می گیرند. روش ارائه شده در اینجا علاوه بر دقت مناسب، از سرعت مناسبی برخوردار است که خود یکی از مسائل مهم در طراحی یک سیستم تبدیل گفتار به متن است.

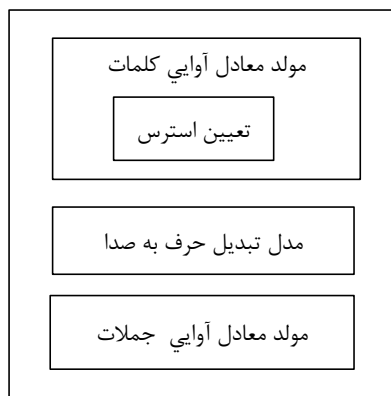
در این نسخه از سیستم پیشنهادی، تحلیلگر معنایی تنها نوع جملات ورودی اعم از پرسشی، سوالی و ... را به همراه برخی از حالت‌های محدود همانند خشم، خوشحالی، تعجب و ... تعیین می نماید. این کار به کمک ویژگیهای معنایی نسبت داده شده به کلمات تشکیل دهنده جمله و ساختار نحوی جملات ورودی انجام شده است. برای تعیین این ویژگی ها از روش های مبتنی بر قانون استفاده شده است.

اطلاعات نهایی بدست آمده به شکل یک آرایه از کلمات تشکیل دهنده متن به بخش DSP تحویل داده می شود. علاوه بر این یک ساختار درختی به صورت مستقل از این آرایه نیز برای نمایش ساختار جملات ورودی تهیه می شود. در این درخت اشاره گرهایی به کلمات موجود در آرایه کلمات وجود دارد. در حقیقت ساختار استفاده شده در اینجا یک Acyclic Directed Graph [12] است. در ادامه عملکرد بخش DSP تشریح شده است.

بخش پردازش دیجیتالی سیگنال

بخش پردازش سیگنال دیجیتالی وظیفه تولید پارامترهای مهم در سنتز گفتار و سپس تولید گفتار خروجی را بر عهده دارد. برای این منظور متن ورودی همراه با اطلاعات حاصل از بخش NLP ابتدا به پیمانانه مولد رشته آوایی برای تولید رشته آوایی نهایی کلمات فرستاده می شود. زیرپیمانانه های موجود در این پیمانانه در شکل 3 نشان داده شده است. این پیمانانه دو زیر پیمانانه اصلی برای تولید معادل آوایی کلمات بنامهای "مولد معادل آوایی کلمات" و "مدل تبدیل حرف به صدا" دارد. به دلیل وجود گستردگی لغات زبان امکان عدم پوشش یک کلمه توسط بخش NLP به منظور تعیین ویژگیهای زبانی از جمله رشته آوایی کلمه وجود دارد. لذا از یک مدل تبدیل حرف به صدا برای پیشگویی رشته آوایی کلمات ناشناخته به کمک مدل‌های یادگیری ماشین استفاده شده است. برای کلماتی که در واژگان پیدا می شوند در مرحله آنالیز تکواژ شناختی و تحلیل صرفی تنها معادل آوایی تکواژهای سازنده کلمات از واژگان استخراج می گردد. برای بیشتر کلمات با متصل کردن این رشته ها به یکدیگر معادل آوایی نهایی کلمه بدست می آید. هر چند موارد استثنایی نیز

وجود دارد. وظیفه زیرپیمانه "مولد معادل آوایی کلمات" در نظر گرفتن این استثنائات و تولید معادل آوایی درست برای تمام کلمات می باشد. اینکار با استفاده از یک ساختار زنجیره ای از قوانین و عملگرها صورت می گیرد. یکی دیگر از کارکردهای این زیرپیمانه، تکمیل فرایند نرمالیزه سازی توکنهای خاص همچون اعداد، زمان، تاریخ، سمبلهای خاص و ... می باشد. در این فرایند رشته آوایی این توکنها با استفاده از نوع تعیین شده برای هر توکن در پیمانه Tokenizer تولید می گردد.



شکل 3: ساختار پیمانه مولد رشته آوایی

در زبان فارسی استرس معمولاً برای اسامی و صفات و قیود بر روی آخرین هجای کلمه بدون در نظر گرفتن پسوندهای تصریفی قرار می گیرد [13]. برای افعال و اعداد نیز قوانین خاصی در زمینه تعیین هجای استرس دار وجود دارد. با توجه به اینکه افعال می توانند مثبت یا منفی، ساده یا پیشوندی یا مرکب و در انتها معلوم یا مجهول باشند و این سه صورت تقسیم بندی می توانند در حالت‌های مختلف با یکدیگر ترکیب گردند، تا به حال مطالعات جامع و کاملی در زمینه تعیین هجای استرس دار برای افعال و حتی دیگر کلمات صورت نگرفته است. در این سیستم برای تعیین استرس، از ترکیب دو روش داده-گرا و قانون-گرا استفاده شده است. در بخش داده-گرا هجای استرس دار کلمات استثناء که از قانون خاصی تبعیت نمی کنند در واژگان مشخص گردیده است. هجای استرس دار بقیه کلمات استثناء نیز بصورت قانون-گرا مشخص می شود. تعیین استرس برای توکنهای خاص دشوار و لذا تعیین استرس برای آنها همزمان با ساخت رشته آوایی برای آنها صورت می گیرد. معادل آوایی نهایی جمله ورودی عموماً با متصل کردن معادل کلمات به یکدیگر می تواند تولید شود، هر چند معمولاً برای سهولت بیان و روانی گفتار، بعضی از کلمات همچون "است" در مواقع خاصی به کلمه قبلی خود متصل می شوند. زیرپیمانه "مولد معادل آوایی جملات" برای در نظر گرفتن این موارد از یکسری قوانین استفاده می نماید. بعد از تولید معادل آوایی متن مورد پردازش، این رشته آوایی در پیمانه هجابند، هجاهای سازنده آن تعیین می گردد.

افراد عموماً به هنگام صحبت کردن تمایل به گروه بندی کلمات با قرار دادن درنگ بین گروهها دارند. به این گروهها اصطلاحاً عبارتهای پروزودیک⁸ گفته می شود. قابل ذکر است گروه بندی کلمات لزوماً با قرار دادن درنگ صورت نمی گیرد، بلکه اینکار می تواند با کشیدن واج آخر گروه و یا تغییر ریتم و ایجاد گسستگی در منحنی پیچ نیز صورت بگیرد. در تئوری پیرهامبرت⁹ برای عبارتهای پروزودیک دو سطح گروه آهنگ و گروههای میانی¹⁰ در نظر گرفته شده است. اسلامی هرگروه گرامری در حالت بیشترین گسترش¹¹ را معادل با یک گروه درنگ در نظر گرفته است. بعلاوه تاکنون برای یافتن ارتباط میان گروههای نحوی و پروزودیک تحقیقات زیادی در دیگر زبانها انجام شده است که منجر به یافتن قوانینی برای محدود کردن مکانهای ممکن بعنوان مرز گروههای پروزودیک شده است. در سیستم کنونی، ما یک سطح برای عبارتهای پروزودیک در نظر گرفته ایم و علاوه بر این هر عبارت پروزودیک را معادل با یک درنگ دانسته ایم و لذا عبارت بندی پروزودیک با استفاده از یک مدل پیشگویی وجود درنگ بین هر دو کلمه صورت می گیرد. بعلاوه از یک سری قوانین برای محدود کردن مکانهای ممکن بعنوان مرز استفاده شده است.

⁸ Intonational and Prosodic Phrase

⁹ Pierrehumbert

¹⁰ Intonational Phrases (IP) and Intermediate Phrases (ip)

¹¹ Maximal Projection

در ادامه با استفاده از مدل‌های یادگیری ماشین، میزان کشش زمانی هر واج و شکل منحنی پیچ متن مورد پردازش مشخص می‌گردد. برای تخمین پیچ، ابتدا شکل منحنی پیچ هر هجا به کمک چندین پارامتر بصورت سمبولیک مشخص می‌گردد و سپس با استفاده از این پارامترها و ترکیب منحنی پیچ هر هجا، کانتور نهایی بدست می‌آید. برای تخمین این پارامترها از مدل‌های یادگیری ماشینی استفاده شده است که برای آموزش این مدلها، حدود 1 ساعت صحبت که متن آن به دقت برچسب گذاری لغوی و نحوی شده است، استفاده شده است. بعد از تعیین پارامترهای نوایی، رشته آوایی همراه با این پارامترها به صف ورودی یک موتور سنتز فرستاده می‌شود. این موتور برای سنتز متون فارسی از یک سنتز پیوندی بر اساس روش MBR-PSOLA [14] استفاده می‌نماید. لازم به ذکر است که مبنای نظری مرتبط با مدل های آوایی زبان فارسی عموماً از [15] و [16] بوده است.

مسائل مرتبط با پیاده سازی

علاوه بر انتخاب روش های مناسب برای استخراج پارامترهای مهم در تولید گفتار، یکی دیگر از مسائل مهم در پیاده سازی و گسترش سیستم های تبدیل متن به گفتار، مسائل مرتبط با پیاده سازی و روش های انتخاب شده برای ارائه داده ها، هماهنگی میان پردازش های مختلف و از همه مهمتر، توانایی سیستم در تولید گفتار در زمان مناسب، زمان واقعی، برای کاربر است. این کار تنها به کمک استفاده از روش های پیچیده Multithreading امکان پذیر نیست.

طراحی سیستم پیشنهادی به گونه ای انجام شده است که در آن تمامی پیمانه های ذکر شده دارای کارکرد مشخص و تعریف شده ای می باشند. در طراحی پیمانه ها همواره به استقلال عملکرد هر پیمانه از دیگر پیمانه ها توجه شده است. علاوه بر این یک ساختار داده ای مشترک ورودی و خروجی برای کل پیمانه ها در نظر گرفته شده است. با ویژگیهای در نظر گرفته شده، هر پیمانه بصورت یک ریسمان¹² مجزا پیاده سازی شده است. علاوه بر پیمانه های مشخص شده در شکل 1، پیمانه کنترل کننده برای هماهنگی میان پیمانه های مختلف و کنترل عملکرد سیستم وجود دارد. برای پاسخ به نیاز کار در زمان واقعی، پس از بافر نمودن مقدار مناسبی از داده کلیه پیمانه های سیستم همواره و در ریسمان های جداگانه مشغول تهیه داده های مناسب برای استفاده در مراحل مختلف پردازشی اند. در حقیقت در پیاده سازی سیستم، پیمانه ها یک Pipeline را تشکیل می دهند به این معنی که هنگامیکه یک متن در حال پخش برای کاربر است قسمت دیگری از متن همزمان در پیمانه های دیگر سیستم در حال پردازش است بدین ترتیب سیستم قادر به تولید پاسخ بر-خط شده است. علاوه بر مطالب ذکر شده، معماری شی گرا در پیاده سازی سیستم، امکان استفاده دوباره از پیمانه ها را فراهم می آورد.

نتیجه گیری و راه آینده

در این مقاله ساختار یک سیستم تبدیل متن به گفتار برای زبان فارسی شرح داده شد. بدین منظور در ابتدا پیمانه های موجود در بخش NLP سیستم و سپس دیگر پیمانه ها موجود در بخش DSP مختصراً شرح داده شدند و متدها و روش های استفاده شده برای استخراج پارامترهای مورد استفاده در سنتز گفتار تشریح شدند.

توسعه واژگان مورد استفاده، گسترش پیکره های موجود و آموزش دوباره مدل ها در بخش NLP یکی از کارهایی است که برای کیفیت بهبود این بخش مورد نظر قرار گرفته است. علاوه بر این گسترش تحلیل های مبتنی بر ویژگیهای معنایی و به خصوص استفاده از هستان شناسی ها و تهیه بانک داده ای از اصطلاحات فارسی یکی دیگر از اهداف آینده برای توسعه بخش NLP سیستم است. پشتیبانی از متون دارای ساختار همانند صفحات وب و تشخیص متون به شکل ساختیافته، تعیین ارتباطات میان کلمات موجود در یک متن سبب فراهم آمدن کیفیت بهتری برای گفتار سنتز شده خواهد بود.

در بخش DSP یکی از مهمترین کارهای آینده، پیاده سازی موتور سنتز گفتار مبتنی بر روش Unit Selection است. برای این منظور در حال حاضر، جمع آوری و گسترش بانک داده ای از گفتار فارسی مد نظر قرار گرفته است. علاوه بر این هم اکنون تحقیقاتی در زمینه تعیین اتوماتیک هجای تکیه دار در زبان فارسی در حال انجام می باشد.

مراجع

- [1] M. Beutnagel, A. Conkie, J. Schroeter, Y. Stylianou and A. Syrdal, "The AT&T Next-Gen TTS System", 137th Acoustical Society of America Meeting, 1999.
- [2] N. Campbell and A. Black, "CHATR: a multi-lingual speech re-sequencing synthesis system", Institute of Electronic, Information and Communication Engineers, 1996.
- [3] W. Hamza1, and R. Donovann, "Data-Driven Segment Preselection in the IBM Trainable Speech Synthesis System", ICSLP, pp. 2609-2612, 2002.
- [4] H. Hon, A. Acero, X. Huang, J. Liu, and M. Plumpe, "Automatic Generation of Synthesis Units for Trainable Text-To-Speech Systems", Microsoft Research, 1998.
- [5] J. Allen, S. Hunnicutt, and D. Klatt, "From Text to Speech: the MITTalk System", MIT Press, Cambridge, Massachusetts, 1987.
- [6] P. Taylor, A. Black, R. Caley, "The Architecture of the Festival Speech Synthesis System", The Third ESCA Workshop in Speech Synthesis, pp. 147-151, 1998.
- [7] C. S. Fordyce, M. Osterdorf, "Prosody Prediction for Speech Synthesis Using Transformational Rule-Based Learning", Spoken Language Processing (ICSLP), pp. 682-685, 1998.
- [8] D. H. Klatt, "Linguistic Uses of Segmental Duration in English: Acoustic and perceptual Evidence", Journal of Acoustic Society of America, vol. 59, pp. 1209-1221, 1976.
- [9] K. McKeown, and S. Pan, "Prosody modelling in concept-to-speech generation: methodological issues", Philosophical Transactions of the Royal Society, pp. 1419-1431, 2000.
- [10] A. K. Syrdal, G. Mohler, K. Dusterhoff, A. Conkie, A. W. Black, "Three Methods of the Intonation Modeling", The Third Workshop on Speech Synthesis, Australia, 1998.
- [11] قاسمی زاده بهرنگ، رحیمی، سعید، نم نبات، مجید، کوچاری، عباس، روشی نوین برای تحلیل صرفی واژه های فارسی، یازدهمین کنفرانس انجمن کامپیوتر ایران، تهران، 1384.
- [12] Jonathan T Gross, Jay Yellen, Graph Theory and its Applications, CRC Press, ISBN: 0849339820, 1998.
- [13] تقی وحیدیان کامیار، نوای گفتار در فارسی، انتشارات دانشگاه فردوسی مشهد، 1379.
- [14] T. Dutoit, and H. Leich, "MBR-PSOLA: Text-to-Speech Synthesis Based on an MBE Resynthesis of the Segments Database", Speech Communication, no 13, pp. 435-440, 1993.
- [15] محرم اسلامی، "نحو و واج شناسی (یک سطح مشترک)", جشن نامه دکتر علی اشرف صادقی، ص 143-176، 1382.
- [16] محرم اسلامی، محمود بی جن خان، "نظام آهنگ زبان فارسی"، پنجمین کنفرانس زبان شناسی، دانشگاه علامه طباطبایی، 1379.