به نام خدا

وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم انسانی و مطالعات فرهنگی
گروه زبان‌شناسی همگانی

مدل‌سازی نوای گفتار در سیستم‌های تبدیل متن به گفتار فارسی

استاد راهنما
دکتر مصطفی عاصی

استاد مشاور اول
دکتر حسین صامتی

استاد مشاور دوم
دکتر محمود بی‌جن‌خان

نگارش: مرتضی طاهری اردلی

رساله برای دریافت درجهٔ دکترای تخصصی
در رشتهٔ زبان‌شناسی همگانی

به:
همهٔ آنهایی که
در راه حقیقت
گام
برمی‌دارند.

سپاسگزاری

در طول دورهٔ دکتری به‌ویژه زمان نگارش رسالهٔ حاضر از حمایت و کمک سخاوتمندانهٔ اساتید، همکاران و دوستان بسیاری بهره برده‌ام که در اینجا برخود لازم می‌دانم از تک‌تک آنها تشکر و قدردانی نمایم.

در ابتدا، از جناب آقای دکتر مصطفی عاصی تشکر می‌کنم که در طول این چند سال در کنار راهنمایی رسالهٔ حاضر، در مقام مدیر گروه زبان‌شناسی و نیز رئیس پژوهشکدهٔ زبان‌شناسی از حمایت همه‌جانبهٔ ایشان بهره برده‌ام. وی نه تنها به صورت حضوری و با خوشرویی تمام آمادهٔ پاسخگویی به سوالات اینجانب بوده‌اند بلکه از راه دور و از طریق رایانامه نیز در اسرع وقت، صمیمانه به سوالات نگارنده پاسخ دادند. همچنین، دست‌نوشته‌ها ارسالی را مطالعه و نکاتی را گوشزد فرمودند. همکاری، همیاری و همدلی ایشان را در طول این چند سال فراموش نخواهم کرد. از ایشان بیش از پیش سپاسگزارم.

جناب آقای دکتر حسین صامتی از دانشکدهٔ مهندسی کامپیوتر دانشگاه صنعتی شریف، در مقام استاد مشاور با حمایت‌های خود از رسالهٔ حاضر مایهٔ دلگرمی و ادامهٔ کار نگارنده بوده‌اند. افتخار آشنایی با این ایشان و نیز دیگر دانش‌آموختگان حوزهٔ مهندسی به اردیبهشت‌ماه سال ۱۳۹۰ و آغاز همکاری نگارنده با شرکت تحقیقاتی عصرگویش‌پرداز برمی‌گردد که جمعی از اساتید و دانشجویان عمدتاً از دانشگاه صنعتی شریف از رشته‌ها و گرایش‌های مختلف مانند هوش مصنوعی، نرم‌افزار و سخت‌افزار کامپیوتر، برق و زبان‌شناسی مشغول تحقیق و توسعه پیرامون بخش‌های مختلف حوزهٔ فناوری گفتار بودند. از ایشان نیز کمال تشکر را دارم.

جناب آقای دکتر محمود بی‌جن‌خان از گروه زبان‌شناسی دانشگاه تهران زحمت مشاور رساله را متقبل شدند و با تمام مشغلهٔ کاری که داشتند با دقت تمام و در سریع‌ترین زمان ممکن، تمامی دست‌نوشته‌های نگارنده را مطالعه و نکات ارزشمندی را متذکر شدند. دقت علمی این استاد فاضل را فراموش نخواهم کرد. از ایشان نیز بی‌نهایت سپاسگزارم.

اما شایسته است از این فرصت استفاده کنم و مراتب قدردانی خود را از اساتید دیگر گروه زبان‌شناسی پژوهشگاه علوم انسانی اعلام دارم. از جناب آقای دکتر یحیی مدرسی و سرکار خانم دکتر آزیتا افراشی کمال سپاس را دارم. این دو استاد فرهیخته در این سال‌ها با فراهم کردن محیطی صمیمی و دوستانه همواره مشوق، و بی‌چشمداشت یاری‌رسان نگارنده بودند. محبت‌های آنها را هیچگاه

فراموش نخواهم کرد و همواره خود را مدیون آنها می‌دانم. همچنین از این افتخار برخوردار بوده‌ام که نخستین تجربهٔ تدریس در حیطهٔ آواشناسی و زبانشناسی را با دانشجویان رشته‌های گفتاردرمانی و شنوایی‌شناسی در مقطع کارشناسی دانشکدهٔ توانبخشی دانشگاه علوم پزشکی ایران داشته باشم. از جناب دکتر مدرسی که این تجربهٔ متفاوت را از سر لطف در اختیار نگارنده قرار دادند کمال تشکر را دارم.

همچنین، در طول این دوره با گذراندن درس‌های مختلف با اساتید فاضل و دانشمند از دانشگاه‌های مختلف بر خود لازم می‌دانم که در اینجا قدردان تک‌تک آنها نیز باشم. از دکتر یدالله ثمره، دکتر محمد دبیرمقدم، دکتر شهین نعمت‌زاده، دکتر ارسلان گلفام، دکتر فرهاد ساسانی، دکتر فرزان سجودی و دکتر احمد پاکتچی نیز سپاسگزارم.

در طول دورهٔ کارشناسی‌ارشد نیز از اساتید فرهیختهٔ دانشگاه علامه طباطبائی درس‌های بسیاری فراگرفتم و انگیزهٔ ادامهٔ تحصیل و پژوهش را از این اساتید دانشمند دارم. از تک‌تک این عزیزان به‌ویژه سرکار خانم دکتر گلناز مدرسی بی‌نهایت سپاسگزارم. اولین آموزه‌های زبان‌شناسی و آواشناسی را از این اساتید دارم. درضمن، لطف و محبت سرکار خانم دکتر مدرسی را در به سرانجام رساندن پایان‌نامهٔ کارشناسی‌ارشد که در همین حوزه یعنی نوای گفتار انجام شد و نیز نگاه دقیق و علمی ایشان را فراموش نخواهم کرد.

در طول تحقیق و نگارش این رساله از راهنمایی‌ها و حمایت‌های سخاوتمندانهٔ افراد بسیار دیگری بهره برده‌ام که قطعاً این پژوهش تحت تاثیر آموزه‌های آنها قرار گرفته است. یکی از بخش‌های فعال در شرکت عصرگویش‌پرداز، گروه سیستم تبدیل متن به گفتار بوده است که شروع همکاری اینجانب با این گروه همزمان با تصمیم گروه برای ارتقای سیستم تبدیل متن به گفتار، به کمک پیاده‌سازی نوای گفتار بوده است. سرآغاز پژوهش و رسالهٔ حاضر به این زمان برمی‌گردد. در ابتدا، از دوست و همکار فاضل و نازنینم دکتر سهیل خرم کمال تشکر را دارم که نخستین درس‌های پیرامون مبحث سیستم‌های تبدیل متن به گفتار را از ایشان دارم و بی‌شک بدون کمک و یاری سخاوتمندانهٔ ایشان این تحقیق بدون سرانجام بود. جناب خرم عزیز در طول این چند سال با صبر و حوصلهٔ تمام، سوالات و کنجکاوی‌های اینجانب را پاسخگو بودند. همچنین روحیهٔ کار تیمی و همکاری گروهی را از ایشان آموختم و همیشه مدیون و سپاسگزار محبت‌هایش هستم.

در طول همکاری با شرکت عصر گویش پرداز این افتخار نصیبم شد که از دانش مهندسی افراد بسیاری بهره ببرم این درحالی بود که حتی با پایه‌ای‌ترین مفاهیم بنیادی در زمینهٔ فناوری گفتار ناآشنا بودم. این افراد همیشه مشوق نگارنده بودند و از آنها نکات بسیاری را فراگرفتم. برخود لازم می‌دانم از دکتر محمد بحرانی، دکتر هادی ویسی، دکتر باقر باباعلی، دکتر ندا موسوی، خسرو حسین‌زاده، یاسر محسنی بهبهانی، سعید صرفجو، امیر سانیان، مرضیه ادراکی، لیلا ضیامجیدی و فهمیه بهمنی‌نژاد از صمیم قلب تشکر نمایم. محبت‌های آن را همیشه به یاد دارم. همچنین، از دکتر ندا موسوی بی‌نهایت سپاسگزارم که سرآغاز آشنایی با دوستان در شرکت عصرگویش‌پرداز بواسطهٔ ایشان بوده است که پیشنهاد همکاری با این گروه را به اینجانب دادند.

نگارنده در طول دورهٔ دکتری از این فرصت نیز برخوردار شد تا با حمایت مالی وزارت علوم، تحقیقات و فناوری ایران از یک دوره فرصت مطالعاتی شش ماه خارج از کشور بهره ببرد. این دوره در دانشگاه رادبود (Radboud University Nijmegen) در کشور هلند و زیر نظر کارلوس گوسنهافن (Carlos Gussenhoven) سپری شد که حاصل آن کسب تجارب ارزشمندی پیرامون نوای گفتار بوده است. خوشبختانه حضور دوست خوب و فاضلم جناب آقای حامد رحمانی در دپارتمان زبان‌شناسی دانشگاه مذکور و بهره بردن از استاد راهنمای مشترک باعث شد تا دوران پرثمر و ارزشمندی را در این مرکز علمی داشته باشم. نکات بسیاری از این عزیزان یاد گرفتم. از آنها نیز کمال تشکر را دارم. همچنین، از این فرصت برخوردار بودم تا در مدرسهٔ تابستانی زبان‌شناسی ۲۰۱۴ هلند که در دانشگاه رادبود برگزار شد شرکت کنم و از کارگاهی که در پنج جلسه و تحت عنوان نوای گفتار کلمه و جمله (word and sentence prosody) برگزار شد بهره ببرم. این کارگاه با مدیریت و تدریس کارلوس گوسنهافن برگزار شد که حاوی نکات بسیار آموزنده‌ای برای نگارنده بوده است.

در طول دورهٔ دکتری، فرصتی دیگر نیز دست داد تا در دو کنفرانس Tonal Aspects of Languages 2014 در دانشگاه رادبود و نیز کنفرانس Speech Prosody 2014 که در دانشگاه ترینیتی کالج دوبلین (Trinity College Dublin) در کشور ایرلند برگزار شد، شرکت کنم. این دو کنفرانس حاوی نکات بسیار ارزشمندی برای نگارنده بود که از آن جمله آگاهی از روش‌شناسی‌های بسیار دقیق در انجام تحقیقات پیرامون نوای گفتار بوده است. درضمن، پیرامون موضوع تحقیق حاضر از گفتگو و مشورت حضوری با سان آ جان (Sun Ah-Jun)، دَنیل هِرست (Daniel Hirst)، هیرویا فوجی‌ساکی (Hiroya Fujisaki)، آگنیشکا وَگنر (Agnieszka Wagner) و آلبرت لی (Albert

Lee) بهره برده‌ام. از این فرهیختگان عرصهٔ نوای گفتار بی‌نهایت سپاسگزارم که با راهنمایی‌های ارزندهٔ خود مسیر را برای انجام تحقیق حاضر هموار کردند. لازم به ذکر است که حضور در کنفرانس نوای گفتار ۲۰۱۴ با حمایت مالی انجمن ایسکا (International Speech Communication Association, ISCA) از پژوهشگران جوان صورت گرفت. از دست‌اندرکاران این انجمن بویژه آلن بِلک (Alan Black) رئیس انجمن مذکور سپاسگزارم.

از یی ژو (Yi Xu) استاد کالج دانشگاهی لندن (University Colleage London) کمال تشکر را دارم که در این سال‌ها نکات بسیاری از وی آموختم. آشنایی نگارنده با ایشان به سال ۱۳۸۹ برمی‌گردد که این آشنایی منجر به همکاری‌های پژوهشی و علمی از سال ۱۳۹۱ تاکنون شده است. در خلال ده‌ها و چه بسا صدها رایانامه‌ای که در این سال‌ها ردوبدل شد ایشان با صبر و حوصله و با سخاوت علمی خاصی که داشتند پاسخگوی تمامی سوال‌های نگارنده بودند و با انعطاف‌پذیری زبانزدی که داشتند از راهی دور نکات بسیاری را به نگارنده گوشزد کردند. حاصل این همکاری ارائهٔ سه مقاله مشترک در کنفرانس‌های مختلف بوده است. این همکاری‌های علمیِ مشترک بخش اصلی از رسالهٔ حاضر را تشکیل داده است. راهنمایی‌های ذی‌قیمت حضوری ایشان در دو کنفرانس فوق‌الذکر نیز مسیر انجام رسالهٔ حاضر را تکمیل کرد. از ایشان بی‌نهایت سپاسگزارم و بخش اعظمی از آموزه‌های خود را در حوزهٔ نوای گفتار همیشه مدیون ایشان هستم.

از جناب آقای دکتر اسلامی نیز سپاسگزارم که در مسیر بخشی از تحقیقِ مدل‌سازی نوای گفتار یعنی برچسب‌گذاری نواییِ نواخت‌ها و فاصله‌نماها از راهنمایی‌های ایشان استفاده کردم و با سخاوت علمی خود پایگاه‌دادگان گفتاری که پیش از این طراحی و تهیه کرده بودند در اختیار نگارنده قرار دادند.

همچنین، در مسیر ساخت پایگاه‌دادگان گفتاری مورد استفاده در رسالهٔ حاضر از کمک و راهنمایی‌های افراد بسیاری بهره برده‌ام. از جناب آقای دکتر محمدمهدی همایون‌پور استاد دانشگاه امیرکبیر تشکر می‌کنم که نگارنده را در انجام چگونگی طراحی و ساخت آن راهنمایی کردند. همچنین، در انجام مراحل مختلف این بخش پژوهشگران و گویندگان بسیاری نیز یاری‌رسان بوده‌اند. لازم می‌دانم مراتب قدرانی خود را به تمامی این عزیزان اعلام دارم: از دکتر هادی ویسی، دکتر محمد بحرانی، محبوبه نعمتی، سلیمان شریفی، محمود کریمی، مشایخی، و سرکار خانم‌ها فهیمهٔ بهمنی‌نژاد،

نازنینی، چاوشی، معصوم‌زاده، رافعی، شهیدثالث، قادری، تعلیم، اکبری، حبیبی، اسحاق‌تبار، صداقتی و دانشیان نیز کمال تشکر را دارم.

درضمن، در بخش انجام آزمون‌های شنیداری رساله، گویشوران فارسی زبان زیادی همکاری و مشارکت کردند. از یکایک آنها سپاسگزارم. همچنین از دوست خوبم جناب آقای سلمان نجاتی تشکر می‌کنم که در فراهم کردن گویشوران مورد نظر، یاری‌رسان نگارنده بودند.

اما تقریباً یک سالی است که افتخار آشنایی با اریک آنونبی (Erik Anonby) استاد دانشگاه کارلتون (Carleton University) نصیب نگارنده شده است. آشنایی با ایشان به واسطهٔ نگارش کتاب ارزشمندشان تحت عنوان "مطالعات بختیاری: واج‌شناسی، متن، واژه‌نامه" بوده و خوشبختانه این آشنایی سرآغاز همکاری با پروژه ماندگار اطلس زبان‌های ایران شده است که پژوهشگران بسیاری با دقت علمی خاص و با مدیریت وی مشغول فعالیت هستند. در آینده نگارنده امیدوار است که در کنار پژوهش و تحقیق در زمینهٔ نوای گفتار، مبحث بسیار ارزشمند و جذاب اطلس زبان‌ها و گویش‌های ایران را نیز دنبال کند.

در خلال سال‌های گذشته لحظات بسیار خوبی با دوستان هم‌دوره‌ای‌ام داشته‌ام که همیشه تلاش کردند تا فضای صمیمی و دوستانه داشته باشیم. از جناب آقای کیومرث جهانگردی و سرکار خانم‌ها بهناز ذوالفقاری و مهتاب نورمحمدی بی‌نهایت سپاسگزارم. برای این عزیزان و دوستانِ همراه، آرزوی بهترین‌ها را دارم. همچنین از دکتر سعید رضایی نازنین، دوست فاضلم کمال سپاس را دارم و لطف و محبت‌هایش را همیشه بیاد خواهم داشت.

سپاس آخر را به اعضای خانواده‌ام به‌ویژه مادرم و خواهرم تقدیم می‌کنم که همواره مشوق من در تمام مسیر زندگی شخصی و دانشگاهی بوده‌اند و با حمایت‌های همیشگی و بی‌دریغ خود مسیر انجام تحقیق و پژوهش را برای این حقیر هموار کردند.

مرتضی طاهری اردلی
مهر ۱۳۹۴

چکیده فارسی

مدل‌سازی نوای گفتار نه تنها نقش مهمی در درک ما از فرایند ارتباط کلامی ایفا می‌کند بلکه می‌تواند در پیشرفت علم فناوری گفتار بویژه سیستم‌های تبدیل متن به گفتار حائز اهمیت باشد. رسالهٔ حاضر درصدد برآمد تا به سه پرسش پیرامون این موضوع پاسخ دهد. نخست اینکه، با توجه به ظرافت‌ها و چالش‌های پیش‌رو در طراحی و ساخت پایگاه‌دادگانِ گفتاریِ مختص سیستم‌های تبدیل متن به گفتار، چگونه می‌توان حالت‌های مختلف نوایی را در این نوع از دادگان لحاظ کرد. پرسش دوم، رویکردِ رمزگذاریِ موازی و تقریبِ هدف (PENTA) به عنوان رویکرد مورد اتخاذ، تا چه میزان در مدل‌سازی نوای گفتار فارسی با استفاده از دادگانِ آزمایشگاهی (نظارت‌شده) موفق عمل می‌کند و پرسش سوم اینکه، در صورت کسب نتایج مطلوب با دادگانِ آزمایشگاهی، بکارگیری این رویکرد تا چه میزان در مدل‌سازی نوای گفتار با دادگان غیرآزمایشگاهی به منظور استفاده در سیستم‌های تبدیل متن به گفتار کارایی دارد. در پاسخ به سوال نخست، تلاش شد تا با استفاده از اطلاعات موجود در متن، بیشترین تفاوت‌های نوایی شناسایی و از این واحدها در طراحی پایگاه‌دادگان استفاده شود. جایگاه‌هایی مورد استفاده عبارتند از هجای ابتدای جمله، هجای قبل از کاما، هجای قبل از نقطه (پایان جمله)، هجای قبل از دونقطه و در نهایت هجای تکیه‌بر. به بیان دیگر، سعی شد واحد انتخاب شده در بخش زنجیری گفتار، در جایگاه‌های مذکور یعنی جایگاه‌هایی که آشکارا بیشترین تفاوت‌های نوایی محرز است نیز انتخاب شود. اما در راستای پاسخگویی به سوال دوم و با اتخاذ رویکرد رمزگذاری موازی و تقریب هدف، از ۱۵۰ پاره‌گفتار خبری که در شرایط آزمایشگاهی تولید شدند استفاده شد. درنهایت، مقایسهٔ منحنی بسامدپایهٔ تولید شده و منحنی طبیعی در جملات خبریِ کانونی و غیرکانونی برپایهٔ دو معیار همبستگی و خطای جذر میانگین مربعات به ترتیب ۰/۸۴ و ۱/۹۴ گزارش شده است. این میزان، بیانگر عملکرد مناسب رویکرد مذکور در پیش‌بینی و تولید منحنی زیروبمی در جملات مذکور است. همین مسیر نیز برای پاسخگویی به سوال سوم طی شد. با استفاده از ۳۰۰ دادهٔ آموزش و ۳۰ دادهٔ آزمون، برای دو معیار عینی همبستگی و خطای جذر میانگین مربعات، به ترتیب، مقادیری برابر ۰/۶۶ و ۳/۳۷ بدست آمد. اما در کنار ارزیابی عینی در دو آزمایش بالا، از آزمون‌های ادراکی نیز استفاده شد. اگرچه نتایج آزمون ادراکی برای سوال دوم تحقیق بسیار خوب ارزیابی شده است اما نتایج حاصل از این آزمون شنیداری برای دادگان غیرآزمایشگاهی یعنی سوال سوم تحقیق، علی‌رغم نتایج نسبتاً قابل قبول در ارزیابی عینی، چندان مطلوب نیست. نتایج حاصل از این آزمون شنیداری دلایل مشخصی را دارد که در فصل چهارم و پنجم به آن پرداخته شده است. لازم به ذکر است که نخستین بار است از این رویکرد برای مدل‌سازی دادگان غیرآزمایشگاهی به منظور بکارگیری در سیستم‌های تبدیل متن به گفتار استفاده شده است.

واژگان کلیدی: تبدیل متن به گفتار، پایگاه‌دادگان، مدل‌سازی، رمزگذاری موازی، تقریب هدف، بسامدپایه

فهرست مطالب

سپاسگزاری ..... ت

چکیده فارسی ..... د

فهرست مطالب ..... ذ

فصل اول: کلیات ۱

مقدمه ۱

۱ـ۱ هدف پژوهش ۹

۱ـ۲ پرسش‌های تحقیق ۱۰

۱ـ۳ روش گردآوری داده‌ها ۱۰

۱ـ۴ آزمودنی‌ها ۱۱

۱ـ۵ مواد پژوهش ۱۲

۱ـ۶ ابزارهای پژوهش ۱۲

۱ـ۷ نحوهٔ اندازه‌گیری ۱۳

۱ـ۸ نوع پژوهش ۱۳

۱ـ۹ مفاهیم اصلی ۱۳

۱ـ۱۰ ساختار پژوهش ۱۵

فصل دوم: مبانی نظری و پیشینهٔ مطالعاتی ۱۷

مقدمه ۱۷

۲ـ۱ سیستم‌های تبدیل متن به گفتار ۱۸

۲ـ۱ـ۱ تحلیل‌های متنی و آوایی ۲۱

۲ـ۱ـ۱ـ۱ پیش‌پردازش ۲۱

۲ـ۱ـ۱ـ۲ تحلیل‌های ساخت‌واژی، نحوی، معنایی و کاربردشناختی ۲۳

۲ـ۱ـ۱ـ۳ تحلیل آوایی ۲۵

۲ـ۱ـ۲ تحلیل و تولید نوای گفتار ۲۵

۲ـ۱ـ۳ بازسازی گفتار ۲۶

۲ـ۱ـ۳ـ۱ سیستم بازسازی انتخاب واحد ۲۷

۲ـ۱ـ۳ـ۲ سیستم بازسازی آماری ـ پارامتری ۳۰
۲ـ۱ـ۳ـ۳ سیستم بازسازی مبتنی بر مدل مخفی مارکوف ۳۲
۲ـ۱ـ۴ واژگان ۳۳
۲ـ۱ـ۵ پیکره‌های متنی ۳۴
۲ـ۱ـ۶ پایگاه‌دادگان گفتاری ۳۵
۲ـ۲ مدل‌سازی نوای گفتار ۳۶
۲ـ۲ـ۱ تاریخچهٔ مطالعات در حوزهٔ نوای گفتار ۳۶
۲ـ۲ـ۲ ارتعاش پرده‌های صوتی، بسامدپایه و زیروبمی گفتار ۳۹
۲ـ۲ـ۳ مطالعات آواشناختی پیرامون منحنی بسامدپایه در فارسی ۴۲
۲ـ۲ـ۴ مدل‌سازی منحنی بسامدپایهٔ گفتار ۴۵
۲ـ۲ـ۴ـ۱ رویکردهای واج‌شناختی به مدل‌سازی نوای گفتار ۴۶
۲ـ۲ـ۴ـ۱ـ۱ رویکرد خودواحد ـ وزنی ۴۷
۲ـ۲ـ۴ـ۱ـ۲ رویکرد ادراکی IPO ۵۲
۲ـ۲ـ۴ـ۲ رویکردهای آواشناختی به مدل‌سازی نوای گفتار ۵۴
۲ـ۲ـ۴ـ۲ـ۱ مدل فوجی‌ساکی ۵۴
۲ـ۲ـ۴ـ۲ـ۲ مدل تیلت (Tilt) ۵۷
۲ـ۲ـ۴ـ۲ـ۳ INTSINT ۵۹
۲ـ۲ـ۴ـ۲ـ۴ رمزگذاری موازی و تقریب هدف (PENTA) ۶۰
۲ـ۲ـ۴ـ۲ـ۴ـ۱ نقش‌های ارتباطی ۶۲
۲ـ۲ـ۴ـ۲ـ۴ـ۲ تقریب هدف ۶۳
۲ـ۲ـ۴ـ۲ـ۴ـ۳ هدف زیروبمی ۶۶
۲ـ۲ـ۴ـ۲ـ۴ـ۳ تحلیل به شیوهٔ بازسازی و بهینه‌سازی تصادفی ۶۶
۲ـ۲ـ۵ کانون نوایی ۶۸
فصل سوم: گردآوری دادگان گفتاری ۷۱
مقدمه ۷۱
۳ـ۱ پیکرهٔ متنی ۷۴

۳ـ۲ پوشش واحدهای آوایی و نوایی ۷۴
۳ـ۳ طراحی و انتخاب خودکار نمونه‌جملات ۷۷
۳ـ۴ اصلاح دستی نمونه‌جملات ۷۸
۳ـ۵ انتخاب گوینده ۷۹
۳ـ۶ ضبط صدا و بررسی فایل‌های صوتی ۷۹
۳ـ۷ برچسب‌گذاری فایل های صوتی ۸۰
۳ـ۸ آمار پوشش حالت‌های آوایی و نوایی ۸۱
۳ـ۹ سخن پایانی ۸۷
فصل چهارم: تحلیل داده‌ها ۸۸
مقدمه ۸۸
۴ـ۱ مدل‌سازی نوای گفتار دادگان کانونی ۸۹
۴ـ۱ـ۱ برچسب‌گذاری نقشی و بازسازی نوای گفتار ۸۹
۴ـ۱ـ۲ نتایج ارزیابی عینی و ذهنی ۹۴
۴ـ۲ مدل‌سازی نوای گفتارِ دادگان غیرآزمایشگاهی ۱۰۱
۴ـ۲ـ۱ برچسب‌گذاری نقشی و بازسازی نوای گفتار ۱۰۲
۴ـ۲ـ۱ نتایج ارزیابی عینی و ذهنی ۱۰۶
۴ـ۲ـ۲ نمونه منحنی‌های بازسازی‌شده ۱۰۸
۴ـ۲ـ۳ میزان تاثیرگذاری لایه‌های مختلف ۱۱۱
۴ـ۲ـ۴ مطالعهٔ موردی ۱۱۴
فصل پنجم: بحث و نتیجه‌گیری ۱۱۶
مقدمه ۱۱۶
۵ـ۱ چگونگی طراحی پایگاه‌دادگان ۱۱۷
۵ـ۲ رویکرد PENTA و مدل‌سازی دادگان آزمایشگاهی ۱۱۹
۵ـ۳ رویکرد PENTA و مدل‌سازی دادگان غیرآزمایشگاهی ۱۲۱
پیشنهادات برای پژوهش‌های آتی ۱۲۶
منابع فارسی ۱۲۷

منابع انگلیسی ۱۲۹
واژه‌نامهٔ فارسی به انگلیسی ۱۳۴
واژه‌نامهٔ انگلیسی به فارسی ۱۳۹
چکیدهٔ انگلیسی ۱۴۴

فصل اول: کلیات

مقدمه

گفتار اصلی‌ترین راه ارتباطیِ انسان‌ها با یکدیگر است و استفاده از آن می‌تواند ارتباط بین انسان و ماشین را آسان‌تر کند. این ارتباط از طریق دو بخش پردازش گفتاری با نام‌های بازشناسی گفتار[۱] و تبدیل متن[۲] به گفتار میسر می‌شود که در کنار ترجمهٔ ماشینی[۳]، مهمترین کاربردهای فناوریِ پردازش زبان طبیعی[۴] محسوب می‌شوند (ژورافسکی و مارتین[۵]، ۲۰۰۷). در ارتباط گفتاریِ بین ماشین و انسان، بخش تبدیل متن به گفتار نقش خروجی را ایفا می‌کند که این خروجی، گفتار بازسازی‌شدهٔ[۶] انسان است. سیستم‌های تبدیل متن به گفتار اساساً از دو بخش پایه‌ای تشکیل شده‌اند؛ بخش نخست، وظیفهٔ

---

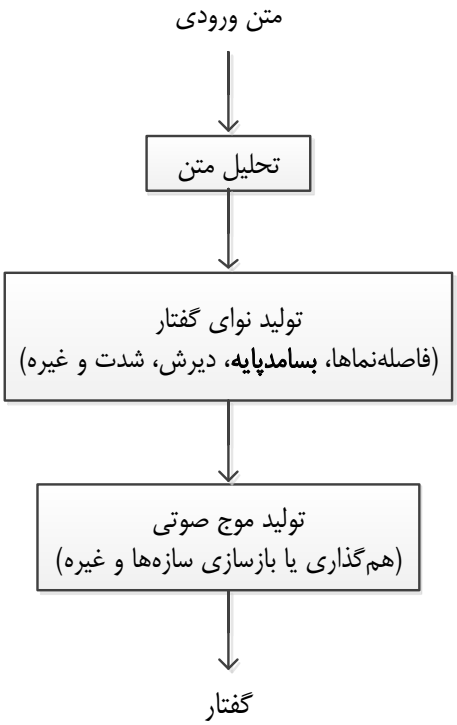
1. speech recognition
2. text-to-speech conversion
3. machine translation
4. Natural Language Processing (NLP)
5. Jurafsky & Martin
6. synthesized

استخراج اطلاعات آوایی و نوایی از متن را برعهده دارد و بخش دوم که بازسازی گفتار[1] نامیده می‌شود در تبدیل اطلاعات آوایی و نوایی به موج گفتاری[2] کاربرد دارد. درحقیقت، تبدیل متن به گفتار تلاش برای تقلید توانایی‌های انسان در خواندن متون است که این متون خود ممکن است از طریق صفحه کلید و یا به صورت یک فایل متنی و یا پس از شناسایی توسط یک سیستم نویسه‌خوان نوری[3] دریافت شوند (همایون‌پور، ۱۳۹۱: ۲۱). از کاربردهای مهمِ سیستم‌های تبدیل متن به گفتار می‌توان به کمک به نابینایان و کم‌بینایان، آموزش زبان، سیستم‌های تلفنی مبتنی بر گفتار[4] و استفاده در ترجمهٔ ماشینیِ گفتار به گفتار اشاره کرد. علاوه بر این کاربردهای عمومی، این سیستم‌ها (به‌ویژه بخش بازسازی گفتار) در پزشکی نیز به کمک بیمارانی آمده است که بر اثر بیماری صدای خود را از دست داده‌اند؛ یعنی بیمارانی که از بیماری‌هایی مانند نورون حرکتی، پارکینسون و یا سرطانِ تارهای صوتی رنج می‌برند، بخشی از تکلم خود را از دست می‌دهند و سیستم‌های مذکور می‌تواند صدایی مشابه صدایِ سالم آنها تولید کرده و آنها را در ارتباط با دیگران یاری کند (خرم، ۱۳۹۳).

از آنجایی که اطلاعات نوایی در کنار اطلاعات آوایی نقش بسزایی در فرایندِ ارتباطِ گفتاری ایفا می‌کند، درکِ صحیح از این عناصر می‌تواند در بهبود کیفیت برنامه‌های کاربردیِ مانند سیستم‌های تبدیل متن به گفتار و بازشناسی گفتار حائز اهمیت باشد. شکل ۱ـ۱ معماریِ کلی یک سیستم تبدیل متن به گفتار را نشان می‌دهد که تولیدِ نوای گفتار در مرحلهٔ میانی بین تحلیل متن[5] و تولید موج صوتی قرار دارد.

---

1. speech synthesis
2. speech wave
3. Optical Character Recognition (OCR)
4. Interactive Voice Response (IVR) system
5. text analysis

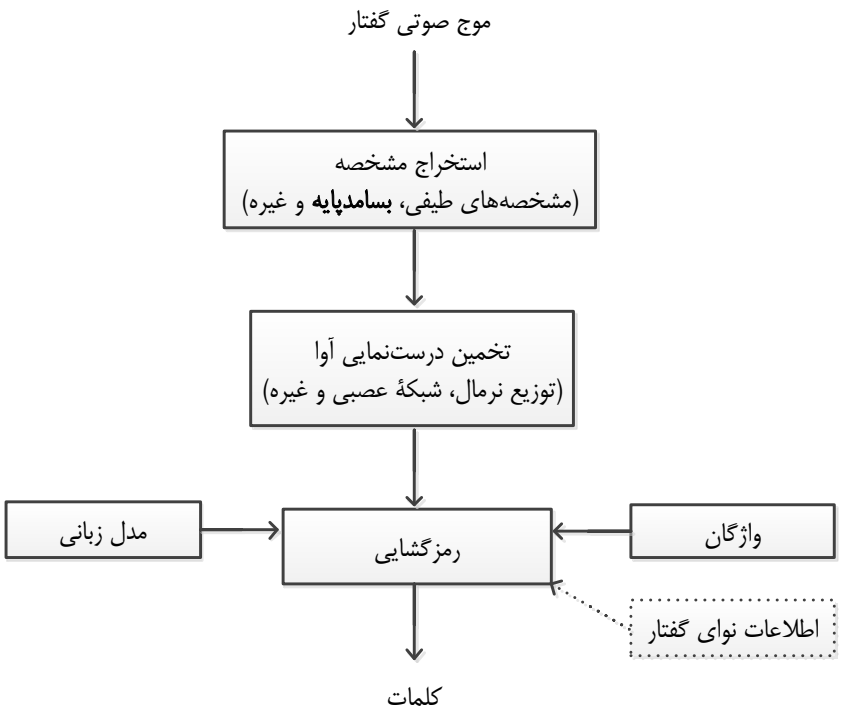


شکل ۱ـ۱: شمای کلی و ساده‌شده از یک سیستم تبدیل متن به گفتار (با قدری تعدیل، برگرفته از سان[1]، ۲۰۰۲)

نوای گفتار همچنین می‌تواند اطلاعات بسیار ارزشمندی برای بهبودِ سیستم‌های بازشناسی گفتار داشته باشد. شکل ۱ـ۲ یکی از روش‌های استفاده از اطلاعات نوای گفتار بویژه بسامدپایه[2] را نشان می‌دهد که در ترکیب با اطلاعات طیفی[3] به منظور شناسایی واحدهای گفتاری مانند واج بکار می‌رود (سان، ۲۰۰۲: ۲). این اطلاعات می‌تواند در مراحل بعدیِ بازشناسی گفتار نیز بکار گرفته شود. برای مثال، می‌توان برخی از رخدادهای نواختی مانند تکیهٔ زیروبمی[4] و نواخت‌های مرزنما[5] را از روی پاره‌گفتار با استفاده از مشخصه‌هایی مانند بسامدپایه، دیرش[6] و شدت[7] پیش‌بینی کرد. یک سیستم

1. Sun
2. fundamental frequency (F_0)
3. spectral
4. pitch accent
5. boundary tone
6. duration
7. intensity

بازشناسی گفتار می‌تواند از این نشانگرهای[1] نوایی برای تعیین برخی اطلاعات معنایی، نحوی و نیز احساسی و عاطفی[2] استفاده کند (همانجا).



شکل ۱ـ۲: شمای کلی و ساده شده از یک سیستم بازشناسی گفتار (با قدری تعدیل، برگرفته از سان، ۲۰۰۲)

لازم به ذکر است که تاکنون بخاطر اهمیت جایگاهِ عناصر زبرزنجیری در ارتقای کیفیت نرم‌افزارهای تبدیل متن به گفتار، پیاده‌سازی نوای گفتار در این سیستم‌ها نسبت به بازشناسی گفتار بیشتر مورد توجه قرار گرفته است.

به طور کلی، بسامدپایه، دیرش و شدت سه مشخصهٔ اصلی تشکیل‌دهندهٔ نوای گفتار هستند (کراتندن[3]، ۱۹۹۷؛ لَد[4]، ۲۰۰۸؛ مدرسی قوامی، ۱۳۹۰) که از بین این سه ویژگی، منحنی بسامدپایه[5]

[1]. marker
[2]. emotional
[3]. Cruttenden
[4]. Ladd
[5]. F_0 contour

مهمترین مولفهٔ شکل‌دهندهٔ نوای گفتار در فارسی است[۱] (ابوالحسنی‌زاده، ۲۰۱۲؛ طاهری اردلی و ژو[۲]، ۲۰۱۲؛ حسینی، ۲۰۱۴؛ رحمانی و همکاران: ۲۰۱۵). این مشخصه یعنی بسامدپایه ماهیتاً یک مفهومی صوت‌شناختی است که به صورت زیروبمی شنیده می‌شود. از طرفی دیگر، زیروبمی به عنوان یک ویژگی شنیداری و بسامدپایه به عنوان یک ویژگی صوت شناختی، با ارتعاش تارآواها به عنوان یک ویژگی تولیدی در تولید گفتار ارتباط مستقیم دارند و هر چه ارتعاش تارآواها بیشتر باشد به همان نسبت بسامدپایه بالاتر و در نتیجه صدا زیرتر خواهد بود و برعکس (اسلامی، ۱۳۸۴: ۳). این فراز و فرودها، نقش‌های زبانیِ گوناگونی در زبان‌های مختلف ایفا می‌کنند که باعث شکل‌گیری اصطلاحات متعددی همچون تکیه[۳]، نواخت واژگانی[۴] و آهنگ[۵] در مطالعاتِ رده‌شناختیِ نوایی شده است. از بین آنها، پرکاربردترین اصطلاح تکیه است که بیشترین تحقیقاتِ انجام‌شده را در این حوزه به خود اختصاص داده است (گوسنهافن[۶]، ۲۰۰۴: ۱۲). بنابر تعریف، تکیه که یک ویژگی ساختاری در زبان است مشخص می‌کند کدام هجا در واژه، قوی‌تر از دیگر هجاهاست (اِسلویتر و فان‌هیوفن[۷]، ۱۹۹۶: ۲۴۷۱). این مفهوم از سه چشم‌اندازِ مختلف یعنی ساختاری (سِلکِرک[۸]، ۱۹۸۰؛ هِیز[۹]، ۱۹۹۵؛ هایمن[۱۰]، ۲۰۰۶ و ۲۰۱۲)، آواشناختی (اسلویتر و فان‌هیوفن، ۱۹۹۶) و ناشنوایی تکیه[۱۱] (پِیرکام و دوپو[۱۲]، ۲۰۰۱؛ ۲۰۰۲؛ ۲۰۰۸ و ۲۰۱۰؛ رحمانی و همکاران، ۲۰۱۵) مورد بررسی قرار گرفته است.

---

[۱]. در پژوهش حاضر اصطلاحات بسامدپایه، زیروبمی، منحنی بسامدپایه و منحنی زیروبمی به صورت جایگزین بکار رفته‌اند. همچنین، از آنجایی که محور اصلی مطالعهٔ این رساله برروی اصلی‌ترین جنبهٔ نوای گفتار یعنی بسامدپایه است، گاه اصطلاح نوای گفتار نیز در این مفهوم بکار گرفته شده است.

[2]. Xu
[3]. stress
[4]. lexical tone
[5]. intonation
[6]. Gussenhoven
[7]. Sluijter & van Heuven
[8]. Selkirk
[9]. Hayes
[10]. Hyman
[11]. stress deafness
[12]. Peperkamp & Dupoux

نواختِ واژگانی یکی دیگر از نقش‌های زبانیِ منحنی بسامدپایه است که در سطح کلمه منجر به تغییر معنا می‌شود. این عملکردِ زبرزنجیریِ گفتار در نیمی از زبان‌های دنیا ایفای نقش می‌کند که عمدهٔ آنها در جنوب شرقی آسیا، ژاپن، افریقا، امریکای شمالی و جنوبی پراکنده شده‌اند (کریستال[1]، ۱۹۸۷؛ یپ[2]، ۲۰۰۲؛ گوسنهافن، ۲۰۰۴: ۲۸).

آهنگ، سومین نقش‌آفرینیِ منحنی بسامدپایه در زبان است که ناظر بر تغییراتِ زیروبمی در سطح پاره‌گفتار است. این زیروبمی‌ها معنایِ واژگانی را تغییر نمی‌دهد بلکه با تغییر سطح و جهت فقط معنایِ بافتی را تغییر می‌دهد (اسلامی، ۱۳۸۴: ۲). زبان فارسی با بکارگیری این مولفه یعنی بسامدپایه، تنها موجب می‌شود تا پاره‌گفتار حالت پرسشی یا خبری به خود بگیرد و کلمهٔ مورد نظر به عنوان کلمهٔ حاملِ اطلاع نو[3] به حساب بیاید یا اینکه در بافت خاصی در تقابل با دیگر عناصر مشابه قرار گیرد (همانجا). این تقابل که از آن تحت عنوان کانون نوایی[4] یاد می‌شود به طور کلی نوای گفتار پاره‌گفتار را دچار تغییرات می‌کند. این تغییرات با افزایش معنی‌دارِ بسامدپایه و دیرش برروی عناصر کانونی و با کاهش چشمگیر بسامدپایه و شدت در عناصر پساکانونی[5] همراه است (طاهری اردلی و ژو، ۲۰۱۲؛ طاهری اردلی و همکاران، ۲۰۱۴؛ طاهری اردلی و ژو، ۲۰۱۵).

اما همانطور که از عنوان رساله یعنی "سیستم‌های تبدیل متن به گفتار" پیداست در این سیستم‌ها هدف دستیابی به سیگنال گفتار از روی متن است؛ بنابراین، تقریباً به جز نشانه‌هایِ مربوط به نقطه‌گذاری که اطلاعاتی پیرامون چگونگی تولید نوای گفتار به ما می‌دهد، هیچ‌گونه اطلاعات دیگری مشخصاً در جمله برای دستیابی به اطلاعات کامل نوایی به مانند آنچه در سطح زنجیری گفتار وجود دارد، رمزگذاری نشده است تا بتوان بر پایهٔ آنها به حالات نواییِ طبیعیِ جمله دست یافت. به نظر

[1]. Crystal
[2]. Yip
[3]. new information
[4]. prosodic focus
[5]. post-focal

می‌رسد از این منظر، تمام نظام‌های نوشتاری دنیا دارای اشتراک هستند (تیلور[1]، ۲۰۰۹: ۳۱). از این رو، تولید مشخصه‌های نوایی از روی متن چالش‌برانگیزترین مسئله‌ای است که پژوهشگران این حوزه با آن روبرو هستند.

در رسالهٔ حاضر، به عنوان اصلی‌ترین پرسش، تلاش خواهیم کرد تا به پیش‌بینی و تولید منحنی بسامدپایه، مهمترین مشخصهٔ نوای گفتار فارسی، با استفاده از اطلاعات موجود در جمله بپردازیم. در سالیان گذشته، در راستای حل مسئله مذکور، زبانشناسان و متخصصین پردازشِ گفتار نظریه‌ها و الگوهای مختلفی ارائه کرده‌اند. در اوایل دههٔ هشتاد از قرن بیستم، رویکرد خودواحد ـ وزنی[2] به عنوان معتبرترین رویکرد زبان‌شناختی به عناصر زبرزنجیری، به معرفی چارچوبی پرداخت که با استفاده از آن بتوان الگویی نظری برای آهنگِ گفتارِ زبان‌های مختلف ارائه داد. نخستین‌بار، پیرهامبرت[3] (۱۹۸۰) چارچوب کلی و مفاهیم نظری این رویکرد را معرفی کرد. این چارچوبِ نظری بعدها دچار تعدیل‌هایی شد اما مبانی اولیه و اصول نظری آن بدون تغییر تا به امروز باقی‌مانده است (لد، ۱۹۹۶ و ۲۰۰۸؛ گوسنهافن، ۲۰۰۴). حاصل تحقیقات گسترده در این زمینه، معرفیِ نظامی برچسب‌گذاری است که نخستین‌بار برای زبان انگلیسیِ امریکایی تحت عنوان برچسب‌گذاری نواخت‌ها و فاصله‌نماها[4] تدوین شده است (بِکمن و هِرشبرگ[5]، ۱۹۹۴). یکی از دلایل معرفی چنین نظامی به منظور بهره‌برداری از آن در مدل‌سازی نوای گفتار در سیستم‌های تبدیل متن به گفتار بوده است. رویکرد فوجی‌ساکی[6] یکی دیگر از چارچوب‌های مطرح است که با نگرشی نسبتاً متفاوت، به صورت جمعی[7] به توصیف دقیقی از

1. Taylor
2. Autosegmental–Metrical
3. Pierrehumbert
4. Tones and Break Indices (ToBI)
5. Beckman & Hirschberg
6. Fujisaki
7. superimpositinoal

منحنیِ بسامدپایه پرداخته است (فوجی‌ساکی و ناگاشیما[1]، ۱۹۶۹؛ فوجی‌ساکی و هیروسه[2]، ۱۹۸۴). این رویکرد با تکیه بر سازوکارِ تولیدیِ بسامدپایه به ارائهٔ مدلی از نوای گفتار پرداخته است که در فصل بعد به تفصیل در مورد آن صحبت خواهیم کرد. تیلت[3]، یکی دیگر از رویکردها با نگرشی مهندسی به نوای گفتار است (تیلور، ۱۹۹۲؛ ۲۰۰۰). این الگو آهنگ گفتار را به عنوان مجموعه‌ای از رخدادهای نواختی در نظر می‌گیرد و در مجموع، شش پارامتر را شامل می‌شود. چهار پارامتر که شکل رخدادها را توصیف می‌کند و دو پارامتر دیگر چگونگی برهم‌نهادگی[4] رخدادها با عناصر زنجیری را نشان می‌دهد. این رویکرد به نوعی الگوگیری از قالب نظریِ واج‌شناختیِ خودواحد ـ وزنی است. اما یکی از رویکردهای متاخر در این زمینه، الگوی رمزگذاریِ موازی و تقریب هدف[5] است که پیاده‌سازی آن را می‌توان در نگاهی کمی یعنی تقریبِ کمّیِ هدف[6] مشاهده کرد (ژو، ۲۰۰۵؛ پروم‌آن[7] و همکاران، ۲۰۰۹؛ ژو و پروم‌آن، ۲۰۱۴). پیش از این، قدرتِ پیش‌بینیِ بالایِ بازسازیِ منحنی بسامدپایه با استفاده از این الگو برای زبان‌های انگلیسی، ماندرین، ژاپنی و تایلندی گزارش شده است. در پژوهش حاضر با بکارگیری رویکرد فوق تلاش خواهیم کرد قدرت پیش‌بینی این الگو را از نوای گفتار فارسی مورد بررسی قرار دهیم و نتایج حاصل از آن را به بحث خواهیم گذاشت.

از طرف دیگر، توسعه و پیاده‌سازی یک سیستم تبدیل متن به گفتار مستلزمِ منابع گوناگون از جمله پایگاه‌دادگان‌های گفتاری[8] است که غالباً زبان‌ویژه‌اند. پایگاه‌دادگان گفتاری در حقیقت یک منبع صوت‌شناختی از عبارات مختلف است که سیستم‌های مذکور با انتخاب قطعاتی از آن و کنار هم قرار

---

1. Nagashima
2. Hirose
3. Tilt
4. alignment
5. Parallel Encoding and Target Approximation (PENTA)
6. quantitative Target Approximation (qTA)
7. Prom-on
8. speech database

دادن آنها موج صوتی نهایی را تولید می‌کنند. اما از آنجایی که نمی‌توان تمام کلمات و جملات ممکن زبان را در چنین مجموعه‌ای فراهم کرد، از مهمترین بخش‌ها، شیوهٔ بهینهٔ آماده‌سازی و تهیهٔ این نوع از دادگان است تا بتوان مناسبترین واحدها را در آن لحاظ کرد. به بیان دیگر، کیفیت خروجیِ یک سیستم به شدت وابسته به چگونگی طراحی پایگاه‌دادگان است به گونه‌ای که هر خطای احتمالی در آن به‌راحتی در صدای بازسازی‌شده منعکس می‌شود. در نتیجه، به منظور دستیابی به صدایی طبیعی، تنها نمی‌توان به ضبط یک نمونه از هجا یا دیگر واحدهای آوایی از زبان بسنده کرد، بلکه لحاظ کردن رخدادهای نوایی در آنها از اهمیت ویژه‌ای برخوردار است. چگونگی لحاظ کردن رخدادهای نوایی یکی دیگر از دغدغه‌های نگارندهٔ رسالهٔ حاضر است که به تفصیل در فصل سوم تحت عنوان "گردآوریِ دادگانِ گفتاری" مورد بحث قرار گرفته است.

۱ـ۱ هدف پژوهش

هدف غایی از انجام پژوهش حاضر ارائهٔ مدلی رایانشی[1] از بسامدپایه به عنوان مهمترین مولفهٔ نوایی گفتار است. این مدل‌سازی به منظور بکارگیری در سیستم‌های تبدیل متن به گفتار است که کمک شایانی به بهبود کیفیت این سیستم‌ها می‌کند. مهم آنکه بخاطر پیچیدگی‌های ذاتی نوای گفتار، پرداختن به این مبحث دارای موانع زیادی است که در این مسیر همفکری و همکاری محققان از زمینه‌های گوناگون از جمله زبانشناسی و هوش مصنوعی می‌تواند راهگشا باشد. همچنین، طراحی و ساخت پایگاه‌دادگانِ گفتاری از دیگر اهداف این تحقیق است که راه را برای ساختن سیستم‌های تبدیل متن به گفتارِ با کیفیت مطلوب‌تر هموار می‌کند. پژوهش در زمینهٔ مدل‌سازی نوای گفتار نه تنها به

[1]. computational

درکِ ماهیتِ زبرزنجیری گفتار و مشکلاتی که زبانشناسان در بررسی آن روبرو هستند کمک می‌کند بلکه کمک شایانی به پژوهشگران در حوزه‌های هوش مصنوعی و فناوریِ گفتار می‌کند.

۱ـ۲ پرسش‌های تحقیق

از آغاز انجام پروژهٔ حاضر تلاش شد تا به سوالات زیر به عنوان اهداف نهایی پاسخ داده شود:

۱. باتوجه به ظرافت‌ها و چالش‌های پیش‌رو در طراحی و ساخت پایگاه‌دادگانِ گفتاریِ مختص سیستم‌های تبدیل متن به گفتار، چگونه می‌توان حالت‌های مختلف نوای گفتار فارسی را در این نوع از دادگانِ گفتاری لحاظ کرد؟

۲. رویکردِ رمزگذاریِ موازی و تقریبِ هدف تا چه میزان در مدل‌سازی نوای گفتار فارسی با استفاده از دادگانِ آزمایشگاهی[1] کارایی دارد؟

۳. در صورت کسب نتایج مطلوب با استفاده از دادگانِ آزمایشگاهی، بکارگیری این رویکرد تا چه میزان در مدل‌سازی نوای گفتار با استفاده از دادگان غیرآزمایشگاهی به منظور استفاده در سیستم‌های تبدیل متن به گفتار کارایی دارد؟

۱ـ۳ روش گردآوری داده‌ها

برای آزمودن کاراییِ رویکرد مورد استفاده، از دو مجموعه دادگان گفتاریِ متفاوت استفاده شد. مجموعهٔ نخست که شامل ۲۸۲۶ پاره‌گفتار است و در فصلی مجزا از رساله به تفصیل بحث شده است. اما در این بخش تنها به مجموعهٔ دوم از دادگان که شامل ۱۵۰ پاره‌گفتارِ تولید شده در شرایط کانونی و غیرکانونی است، می‌پردازیم. برای این مجموعه از پاره‌گفتارها که در شرایط آزمایشگاهی تولید شدند

[1]. experimental data

از پنج گویشور مردِ فارسی‌زبان خواسته شد تا جملات از پیش طراحی شده را تولید کنند. نمونه‌جملاتِ موردنظر برروی کاغذ و روبروی گوینده قرار گرفتند تا آنها را تولید کنند. صدای گوینده‌ها با استفاده از میکروفون که در فاصلهٔ ده سانتیمتری از دهان آن‌ها قرار داشت در حافظهٔ رایانه ضبط شد. فرایند ضبط در محیطی آرام و بدون نوفه انجام گرفت. اما به منظور فعال‌کردنِ شرایط کانونی برروی کلمهٔ مورد نظر در جملات تاکیدی[1]، عبارت دیگری قبل از جملهٔ اصلی (درون پرانتز) قرار داده شد. تفاوت جملهٔ درون پرانتز با جملهٔ اصلی در عنصر کانونی، حضور کلمهٔ "بلکه" در پایان جملهٔ درونِ پرانتز و منفی بودن فعلِ جملهٔ اصلی است. برای مثال:

(۱) (**اونا** بابایِ نیلی رو لندن ندیدن **بلکه**) ماها بابای نیلی رو لندن دیدیم.

لازم به ذکر است که این مجموعه دادگان، پیش از این در چند پژوهش مرتبط دیگر نیز مورد استفاده قرار گرفت (نک: طاهری اردلی، ۱۳۸۹؛ طاهری اردلی و ژو، ۲۰۱۲، طاهری اردلی و همکاران، ۲۰۱۴؛ طاهری اردلی و ژو، ۲۰۱۵؛ طاهری اردلی، ۱۳۹۴).

۱ـ۴ آزمودنی‌ها

پژوهش حاضر شامل پنج بخش تولیدی و ادراکی[2] (ذهنی[3]) مجزا است. در دو بخشِ تولیدی به ترتیب پنج گویشور مرد و یک گویشور زن مشارکت داشتند و در سه آزمونِ ادراکی از پنج شنوندهٔ مرد و پنج شنوندهٔ زن دعوت به همکاری شد. این شرکت‌کنندگانِ فارسی زبان، هیچ‌گونه اختلالِ گفتاری یا شنیداری را به آزمون‌گیرنده گزارش نکردند. درضمن، در سه بخش از پنج بخش، هزینهٔ شرکت به آزمودنی‌ها پرداخت شد.

---

[1]. emphatic
[2]. perceptual
[3]. subjective