طراحی و پیاده‌سازی سیستم تبدیل متن به گفتار برای زبان کردی و بررسی کیفی آن

وفا بارخدا، انور بهرام‌پور، فردین اخلاقیان و هشام فیلی

چکیده: در این مقاله اولین سیستم تبدیل متن به گفتار پیاده‌سازی شده برای زبان کردی معرفی شده است. زبان کردی دارای دو رسم‌الخط رایج بر اساس الفبای عربی و لاتین است. در قسمت تحلیل متن، علاوه بر رفع ابهامات رایج در متون مختلف، مشکلات مربوط به هر دو رسم‌الخط بررسی شده است. همچنین نمادهای استانداردی تعریف شده‌اند که سیستم قادر است متن ورودی به هر یک از رسم‌الخط‌های فوق را به رشته‌ای از نمادهای استاندارد تبدیل نماید. همچنین منحنی‌های تغییرات گام برای انواع مختلف جمله‌ها در این زبان برای اولین بار بررسی شده است. در قسمت تولید گفتار، سه سیستم مختلف بر اساس واحدهای واج‌گونه، هجا و دایفون پیاده‌سازی شده است. برای بررسی کیفیت این سیستم‌ها و مقایسه آنها با همدیگر از چهار آزمون MOS، قابلیت فهم، DRT و MRT استفاده شده است. نتایج این آزمون‌ها نشان‌دهنده قابلیت فهم بالای این سیستم‌ها و به‌ویژه سیستم مبتنی بر دایفون است.

کلید واژه: **تحلیل نوایی، دایفون، زبان کردی، سیستم تبدیل متن به گفتار، سیستم سنتز اتصالی، هجا، واج‌گونه.**

۱- مقدمه

در بسیاری از مناطق جهان نجات زبان‌های در شرف انقراض در دستور کار دولت‌ها، نهادهای مدنی، دانشمندان و پژوهشگران قرار دارد. به‌عنوان مثال زبان مانچو که در زمان سلسله Qing به‌عنوان زبان ملی کشور چین اعلام شده بود، امروزه با خطر نابودی مواجه است [۱]. این زبان اکنون تحت حمایت دولت چین قرار گرفته است تا از نابودی آن جلوگیری به عمل آید. نمونه‌هایی از این دست را می‌توان در سراسر جهان یافت. در دنیای امروز یکی از موارد مهم برای بقای هر زبان رشد و سازگاری آن با فناوری روز است.

سیستم سنتز گفتار (تبدیل متن به گفتار) سیستمی است که قادر است متن نوشته‌شده را به سیگنال صحبت تبدیل نماید. ساختن این سیستم‌ها از حدود ۲۵۰ سال پیش مد نظر دانشمندان بوده است. با این حال بیشترین پیشرفت در اوایل قرن گذشته و با روی کار آمدن سیستم‌های الکترونیکی حاصل شد. این سیستم‌ها دارای کاربردهای فراوان و متنوعی هستند. از جمله مهم‌ترین کاربرد آنها، استفاده در سیستم‌هایی است که

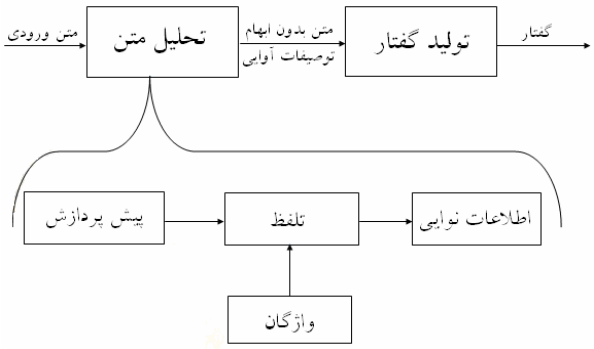
این مقاله در تاریخ ۲۱ دی ماه ۱۳۸۸ دریافت و در تاریخ ۳ مرداد ماه ۱۳۸۹ بازنگری شد.

وفا بارخدا، گروه کامپیوتر و فناوری اطلاعات، دانشکده مهندسی، دانشگاه کردستان، پاسداران، سنندج (email: w.barkhoda@ieee.org).

انور بهرام‌پور، گروه فناوری اطلاعات، دانشگاه آزاد اسلامی – واحد سنندج (email: bahrampour@iausdj.ac.ir).

فردین اخلاقیان، گروه کامپیوتر و فناوری اطلاعات، دانشکده مهندسی، دانشگاه کردستان، پاسداران، سنندج (email: f.akhlaqian@uok.ac.ir).

هشام فیلی، گروه مهندسی برق و کامپیوتر، دانشکده فنی، دانشگاه تهران، تهران، ایران (email: hfaili@ut.ac.ir).



شکل ۱: بلوک دیاگرام سیستم‌های تبدیل متن به گفتار.

امکان برقراری ارتباط بین ماشین و انسان را به شکل گفتاری فراهم می‌نمایند. همچنین کمک به افراد نابینا جهت خواندن متون، استفاده در مراکز خدمات ارتباطی همچون تلفن‌های گویا، کمک به آموزش و یادگیری زبان توسط کودکان و همچنین یادگیری زبان‌های بیگانه، دسترسی به بانک‌های اطلاعاتی از طریق گفتار، کاهش حجم اطلاعات صوتی و غیره از دیگر کاربردهای سیستم‌های تبدیل متن به گفتار خواهند بود.

در بسیاری از زبان‌های دنیا تحقیقات گسترده‌ای بر روی سیستم‌های تبدیل متن به گفتار انجام گرفته و حتی نمونه‌های تجاری آن نیز به بازار عرضه شده است. از این قبیل برای زبان انگلیسی می‌توان به سیستم‌های تبدیل متن به گفتار CHATR [۲] و [۳] و AT&T NEXT GEN [۴] اشاره کرد. همچنین در زبان‌های مختلف همچون فرانسوی [۵] و [۶]، عربی [۷] تا [۱۱]، نروژی [۱۲]، کره‌ای [۱۳]، یونانی [۱۴] و غیره هم در این زمینه کارهای وسیعی صورت گرفته است. حتی در بسیاری از این زبان‌ها سیستم‌هایی با توانایی استفاده از لهجه‌های مختلف آن زبان هم طراحی شده است. در سال‌های اخیر در ایران هم تلاش‌هایی برای تولید سیستم‌های تبدیل متن به گفتار صورت گرفته است. البته بیشتر این تلاش‌ها مربوط به زبان فارسی بوده است [۱۵] تا [۲۱] و در مورد بقیه زبان‌های ایرانی تا به حال کار چندانی صورت نگرفته و به‌طور خاص در مورد زبان کردی در این زمینه تاکنون هیچ کار علمی انجام نشده است.

زبان کردی یکی از زبان‌های زیرشاخه ایرانی است که این زیرشاخه هم به نوبه خود متعلق به شاخه بزرگ هند و اروپایی می‌باشد [۲۲]. این زبان دارای ۲۷ هم‌خوان (صامت)، دو واج نیمه‌هم‌خوان و هشت واکه (مصوت) است. به این ترتیب تعداد کل واج‌های زبان کردی برابر ۳۷ است [۲۲] تا [۲۵]. همچنین این زبان دارای دو رسم‌الخط رایج است که بر اساس رسم‌الخط‌های عربی و لاتین هستند.

سیستم‌های تبدیل متن به گفتار دارای دو مرحله اصلی تحلیل متن و تولید گفتار هستند. شکل ۱ شمای کلی این سیستم‌ها را نشان می‌دهد. بخش اول متن را به‌عنوان ورودی گرفته و بعد از ابهام‌زدایی، آن را به‌عنوان ورودی بخش دوم آماده می‌کند. همچنین تحلیل‌های نوایی هم

Archive of SID

www.SID.ir

در این مرحله صورت می‌گیرد. از بین بردن ابهامات موجود در متن و استخراج درست اطلاعات نوایی کار بسیار پیچیده و سختی است و هنوز یکی از موضوعات جذاب در تحقیقات به شمار می‌رود.

در بخش دوم این سیستم‌ها، گفتار مورد نظر تولید شده و اطلاعات نوایی مطلوب به آن اضافه شود. روش‌های مورد استفاده در این بخش به سه دسته کلی تولیدی، مبتنی بر فرمنت و اتصالی تقسیم می‌شوند. همچنین برای یکنواخت‌کردن سیگنال و اضافه‌کردن اطلاعات نوایی هم روش‌های زیادی پیشنهاد شده است که در این میان دو روش TD - PSOLA و HNM کاربرد و محبوبیت بیشتری یافته‌اند. در [۲۶] تا [۳۰] این روش‌ها با جزئیات کامل مورد بررسی قرار گرفته‌اند.

در این مقاله ما اولین سیستم تبدیل متن به گفتار را برای زبان کردی معرفی می‌کنیم. با توجه به این که زبان کردی دارای دو رسم‌الخط عربی و لاتین است که هر کدام از نمادهای ویژه‌ای استفاده می‌کنند، ما در بخش تحلیل متن به رفع این مشکل پرداخته‌ایم. در واقع با تعریف‌کردن نمادهای میانی و نگاشت هر دو رسم‌الخط به آن، سیستم توانایی استفاده از هر دو رسم‌الخط را پیدا کرده است. همچنین علاوه بر مشکلات معمول در متن‌ها، هر کدام از این رسم‌الخط‌ها مشکلات خاص خود را دارند که ما در بخش تحلیل متن به آن پرداخته‌ایم. به‌علاوه در این مقاله برای اولین بار منحنی‌های تغییرات گام را برای انواع مهم جملات کردی تعیین کرده‌ایم.

همچنین در قسمت تولید گفتار از سه واحد زبانی واج‌گونه، هجا و دایفون استفاده شده است. در واقع برای هر کدام از این واحدها پایگاه داده جداگانه‌ای طراحی و در نهایت کیفیت آنها با هم مقایسه شده است. برای سنجش کیفیت از چهار آزمون MOS، قابلیت فهم، DRT و MRT استفاده کرده‌ایم. همچنین کیفیت سیستم‌های طراحی‌شده را با چند سیستم در زبان انگلیسی و یک سیستم در زبان فارسی مقایسه کرده‌ایم.

ساختار مقاله به این صورت است که در بخش بعد مرحله تحلیل متن در زبان کردی را بررسی می‌کنیم. همچنین در این بخش منحنی‌های تغییرات گام را برای انواع مختلف جملات کردی بیان می‌کنیم. در بخش سوم مرحله تولید گفتار و سه سیستم طراحی‌شده مبتنی بر واج‌گونه، هجا و دایفون را بررسی خواهیم کرد. سپس در بخش چهارم به نتایج آزمون‌های کیفیت و مقایسه سیستم‌ها می‌پردازیم و نهایتاً در بخش پنجم جمع‌بندی و کارهای آینده مطرح خواهند شد.

۲- تحلیل متنی و نوایی در زبان کردی

ساختار سیستم‌های تبدیل متن به گفتار تا حد زیادی وابسته به ساختار زبان مورد نظر بوده و در هر یک از مراحل آن ملاحظاتی وجود دارد که با توجه به آن زبان تعیین می‌شود. هدف نهایی بخش تحلیل متن تبدیل متن مبهم در ورودی به متنی صریح و بدون ابهام و همچنین استخراج اطلاعات نوایی است. در این بخش، ما مرحله تحلیل متن از سیستم‌های تبدیل متن به گفتار را برای زبان کردی مورد مطالعه قرار می‌دهیم. همچنین مشکلات مربوط به زبان کردی را بررسی کرده و برای آنها راهکارهای مناسب ارائه می‌دهیم.

یکی از مهم‌ترین مسائل در بخش تحلیل متن، استخراج اطلاعات نوایی است. در واقع این قسمت یکی از سخت‌ترین و چالش برانگیزترین قسمت‌های طراحی یک سیستم تبدیل متن به گفتار می‌باشد. از مهم‌ترین اطلاعات نوایی که ریتم و آهنگ گفتار را تعیین می‌کند، منحنی تغییرات گام است که تأثیر زیادی در کیفیت گفتار و میزان طبیعی‌بودن آن دارد. با توجه به این که این الگوها برای زبان کردی تا به حال استخراج نشده‌اند، ما در این مقاله با استفاده از نرم‌افزار پرات، مهم‌ترین الگوهای زبان کردی را استخراج کرده‌ایم. برای این کار از کمک چند زبان‌شناس نیز بهره برده‌ایم.

۲-۱ پیش‌پردازش

مهم‌ترین هدف در قسمت پیش‌پردازش، از بین بردن ابهام متن ورودی است. به‌طور کلی مشکلاتی که باید در این مرحله برطرف شوند به دو دسته مستقل از زبان و وابسته به زبان مورد نظر تقسیم می‌شوند. گروه اول مشکلاتی هستند که در همه زبان‌ها مشترک بوده و به‌عبارتی مخصوص یک زبان خاص نمی‌باشند. برای حل این مشکلات روش‌های کلی وجود دارد که صرف نظر از زبان مورد نظر می‌توان کم و بیش از آنها بهره گرفت. مهم‌ترین نمونه از این مشکلات، اعداد هستند که به شکل‌های مختلفی ظاهر شده و باید با توجه به زمینه متن، نوع آنها تشخیص داده شوند. به‌عنوان مثال عدد ۳۲۵۴۵۶۷ اگر عدد ریاضی باشد به‌صورت سه میلیون و دویست و پنجاه و چهار هزار و پانصد و شصت و هفت تلفظ می‌شود در حالی که اگر این عدد شماره تلفن باشد به‌صورت سیصد و بیست و پنج، چهل و پنج، شصت و هفت خوانده می‌شود. همچنین می‌توان به اینها تاریخ، زمان، کد پستی، شماره حساب بانکی و غیره را افزود که همگی عدد هستند ولی به‌صورت‌های مختلفی تلفظ می‌شوند.

برخلاف زبان‌هایی همچون انگلیسی و فارسی، در زبان کردی گفتار و نوشتار تا حد بسیار زیادی بر هم منطبق هستند؛ به‌عبارت دیگر متن به همان صورتی که نوشته می‌شود، خوانده می‌شود؛ بنابراین مشکلاتی نظیر کسره اضافه و هم‌نویسه‌ها در این زبان وجود ندارد و همچنین تعداد عبارات مخفف و نمادهای پیچیده در آن نادر است. با توجه به این موارد می‌توان گفت که بخش پیش‌پردازش در زبان کردی به نسبت سایر زبان‌ها ساده‌تر است. البته این بدان معنا نیست که هیچ مشکلی در این قسمت وجود ندارد؛ چرا که برخی از مسایل موجود در این زبان همچون انواع مختلف اعداد، مستقل از زبان بوده و باید به نوعی پردازش شوند. در این سیستم پردازش انواع مختلف اعداد نظیر تاریخ، ساعت، اعداد ریاضی و غیره پوشش داده شده است.

گروه دوم، مشکلاتی هستند که خاص زبان مورد نظر بوده و بنابراین برای آنها نمی‌توان از روش‌های کلی استفاده کرد. خوشبختانه همان‌طور که بیان شد، این مشکلات در زبان کردی کمتر از زبان‌های دیگر است که این به دلیل نزدیکی نوشتار و گفتار زبان می‌باشد. اما این زبان هم دارای مشکلات خاص خود است. به‌عنوان مثال در زبان کردی دو نوع واج /ر/ پرحجم و /ر/ کم‌حجم وجود دارد. این دو نوع /ر/ با هم متفاوت بوده و در واقع تفاوت معنایی ایجاد می‌کنند، از این رو نمی‌توان آنها را به‌جای هم به کار برد.

مشکل دیگری که مربوط به زبان کردی است، وجود چند رسم‌الخط برای زبان می‌باشد که در این میان دو رسم‌الخط عربی و لاتین دارای اهمیت بیشتری می‌باشند. معمولاً کردزبانان هر کدام فقط با یکی از این رسم‌الخط‌ها آشنایی دارند و نمی‌توانند از متون نوشته‌شده به رسم‌الخط دیگر استفاده کنند. در این میان باید یکی از رسم‌الخط‌ها را به‌عنوان معیار انتخاب کنیم و یا این که سیستم تبدیل متن به گفتار را طوری طراحی کنیم که توانایی استفاده از هر دو رسم‌الخط را دارا باشد. در این صورت نیاز به یک مرحله پردازش اضافی داریم که بتواند هر کدام از رسم‌الخط‌ها را که در ورودی ظاهر شدند، به تعدادی نماد از پیش تعریف شده (رسم‌الخط استاندارد) تبدیل کند به‌گونه‌ای که در ادامه سیستم با این نمادهای تعریف‌شده کار کند. ما برای این کار تعدادی نماد استاندارد

Archive of SID
www.SID.ir

جدول ۱: نمادهای مورد استفاده در رسم‌الخط استاندارد تعریف‌شده.

Arabic	غ	ع	ش	س	ژ	ز	ڕ	ر	د	خ	ح	چ	ج	ت	پ	ب	ا	ئ	
Latin	–	–	Ş	s	j	z	rr	r	d	x	–	ç	c	t	p	b	a	–	
Standard	X	G	S	s	j	z	R	R	d	x	H	C	c	t	p	b	a	A	
Arabic	ی	ی	ه	–	ێ	ھ	وو	ۆ	و	و	ن	م	ڵ	ل	گ	ک	ق	ڤ	ف
Latin	y	î	e	i	ê	h	û	o	U	w	n	m	ll	l	g	k	q	v	f
Standard	y	I	e	i	Y	h	U	o	U	w	n	m	L	l	g	k	q	v	f

تعریف کرده‌ایم و سیستم قادر است که هر کدام از رسم‌الخط‌ها را به آن نگاشت دهد. در جدول ۱ این رسم‌الخط استاندارد و کاراکترهای نظیر آن در هر کدام از دو رسم‌الخط دیگر آمده‌اند.

هر کدام از این دو رسم‌الخط مشکلات مخصوص به خود را دارند. در رسم‌الخط عربی واکه /بزروکه/ (کسره کوتاه معادل /i/ در رسم‌الخط لاتین) نوشته نمی‌شود، برای دو واج /و/ واکه (w) و /و/ هم‌خوان (u) یک نماد وجود دارد و هر دو به‌صورت /و/ نوشته می‌شوند. همچنین برای /و/ کشیده (û) هم از نماد /وو/ استفاده می‌شود. در رسم‌الخط لاتین هم نماد /همزه/ نوشته نمی‌شود و برای واج‌های /ع/، /ح/، /ل/ پرحجم و /ر/ پرحجم نماد استانداردی وجود ندارد.

۲-۲ تلفظ

دسته دیگری از کلمات وجود دارند که تلفظ آنها از ساختار رایج در زبان کردی تبعیت نمی‌کند که به‌عنوان مثال می‌توان به برخی اسامی خاص اشاره کرد. در بخش تلفظ به حل مشکل این کلمات پرداخته‌ایم. برای حل این مسئله از یک دادگان جداگانه استفاده می‌کنیم که در آن به‌ازای هر استثنا در زبان، یک مدخل وجود دارد که در این مدخل خود کلمه و تلفظ درست آن نوشته شده است. با استفاده از این دادگان، هرگاه که در متن به این کلمه خاص برخورد کردیم، تلفظ درست آن را جایگزین کرده و در فاز تولید گفتار از این فرم استفاده می‌کنیم.

تنها مسئله‌ای که در این رابطه مطرح است، اندازه دادگان می‌باشد. برای زبانی همچون انگلیسی که تعداد استثناها زیاد هستند، تعداد مدخل‌ها می‌تواند بسیار زیاد باشد. این تعداد در واقع در حدود چند ده هزار کلمه و یا بیشتر می‌باشد و به این ترتیب کار گردآوری آنها بسیار سخت و زمان‌بر است. در زبان کردی تعداد این کلمات کمتر بوده و اندازه دادگان در حدود ۱۰۰۰ مدخل خواهد بود.

به‌عنوان مثال اسم خاص /عبدالله/ در این دسته قرار می‌گیرد، چرا که انتظار داریم این کلمه به‌صورت /عەبدوللا/ نوشته شود؛ اما با توجه به این که این کلمه اسم خاص بوده و در اصل از زبان عربی وارد زبان کردی شده است، پس به همان فرم زبان عربی نوشته می‌شود. همان‌طور که می‌بینیم تعدادی از واکه‌ها (همانند فتحه /ع/) نوشته نشده است. توجه شود که در کردی همه کلمات دخیل از زبان‌های دیگر و اسامی خاص این مشکل را ندارند. در واقع در این زبان سعی می‌شود تمام کلمات با ساختار زبان منطبق شوند و فقط در موارد خاصی که کاربرد کلمه بسیار گسترده باشد، شکل اصلی آن در زبان مبدأ نوشته می‌شود. به‌عنوان مثال می‌توان کلمه computer را مثال زد که به‌صورت /کومپیوتەر/ نوشته می‌شود.

علاوه بر اینها کلمات دیگری در خود زبان کردی وجود دارد که بسیار پرکاربرد هستند و دارای تلفظی متفاوت با نوشتار خود می‌باشند. به‌عنوان مثال عبارت /هەتا دوایی/ به معنی /غیره/ به‌صورت /هتد/ نوشته می‌شود و بسیار هم پرکاربرد است. می‌توان این نوع کلمات را شناسایی کرده و آنها را به دادگان اضافه کرد. با توجه به پرکاربرد بودن اکثر مدخل‌ها، وجود این دادگان بر روی کیفیت سیستم تأثیر زیادی خواهد داشت.

۲-۳ آهنگ جمله

یکی دیگر از مفاهیم مهم در پردازش سیگنال گفتار و استخراج اطلاعات نوایی، لحن و یا آهنگ جمله است. در واقع انسان‌ها در حین صحبت‌کردن، جملات را با استفاده از یک ریتم خاص بیان می‌کنند که این امر منجر به بیان بهتر مفهوم مورد نظر می‌شود. آهنگ گفتار با استفاده از تغییر در مقادیر گام ایجاد می‌شود؛ به این مفهوم که با تغییردادن مقادیر گام در حین صحبت‌کردن، نوعی بالا و پایین‌رفتن در ریتم گفتار حاصل می‌شود. این آهنگ در زبان‌های مختلف به‌صورت تصادفی نبوده و از الگوهای خاصی پیروی می‌کند و همچنین از زبانی به زبان دیگر متفاوت است.

استخراج الگوهای تغییر گام برای انواع مختلف جمله در هر زبان بسیار مهم است. در بسیاری از زبان‌ها این کار از مدت‌ها قبل مورد توجه قرار گرفته و انواع مهم الگوها استخراج شده‌اند. متأسفانه در زبان کردی از قبل این کار نه در حوزه زبان‌شناسی و نه در حوزه پردازش و سنتز گفتار صورت نگرفته است. برای استخراج الگوهای گام از یک زبان‌شناس کرد به‌عنوان مشاور کمک گرفتیم. اولین مرحله کار تقسیم‌بندی مناسب انواع جمله‌های زبان کردی است که در ادامه توضیح داده می‌شود. سپس برای هر کدام از گروه جمله‌های زبان، تعداد پنج جمله مناسب در نظر گرفته و با استفاده از نرم‌افزار Praat ۵.۱.۰۷ منحنی تغییرات گام را تولید نمودیم. بعد از استخراج منحنی تغییرات گام برای تمامی جملات، با کمک زبان‌شناس، الگوهای انتزاعی تغییرات گام در هر گروه را تولید کردیم.

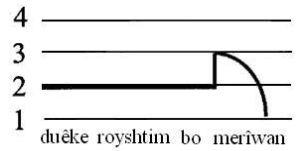
ما جمله‌های زبان کردی را به دو گروه عمده تقسیم کرده‌ایم: ساده و مرکب. گروه جملات ساده شامل جملات خبری، امری، پرسشی و همچنین حالت‌های مختلف آنهاست. همچنین جملات استفهام انکاری هم در این دسته قرار گرفته‌اند. گروه مرکب هم شامل جملات هم‌پایه و جملات شرطی است. در ادامه این گروه‌ها به تفصیل بیان می‌شوند.

۲-۳-۱ جملات ساده

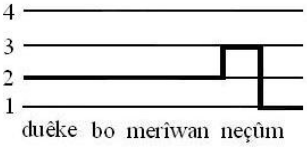
ما جملات ساده را شامل حالاتی در نظر گرفته‌ایم که فقط دارای یک فعل بوده و به عبارتی فقط از یک تک‌جمله تشکیل شده باشند. برای این تقسیم‌بندی یک استثنا هم قائل شده‌ایم که عبارتند از جملات استفهام انکاری. در واقع ما این نوع جملات را در دسته پرسشی قرار داده‌ایم. همان‌طور که بیان شد، انواع مختلف جملات ساده که در این مقاله مورد بررسی قرار گرفته‌اند عبارتند از جملات خبری، امری و پرسشی. همچنین بعضی از این جملات خود دارای حالات مختلفی همچون جملات مثبت و یا منفی هستند. در برخی موارد همچون جملات پرسشی هم امکان دارد که کلمه پرسش در طول جمله جابه‌جا شود و یا این که با استفاده از تغییر لحن، پرسش صورت گیرد. ما سعی کرده‌ایم که حتی‌الامکان این موارد را هم مورد بررسی قرار دهیم.

Archive of SID

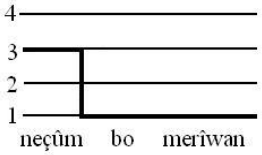
www.SID.ir



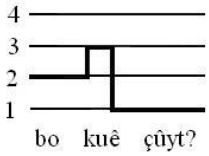
شکل ۲: الگوی جمله خبری مثبت، معنای جمله: /دیروز به مریوان رفتم/.



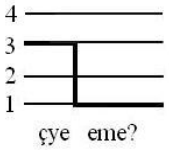
شکل ۳: الگوی جمله خبری منفی، معنای جمله: /دیروز به مریوان نرفتم/.



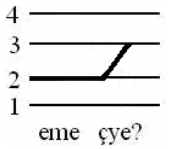
شکل ۴: الگوی جمله خبری منفی، معنای جمله: /نرفتم به مریوان/.



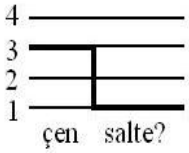
شکل ۵: جمله سؤالی با استفاده از کلمات پرسشی، معنای جمله: /کجا رفتی؟/.



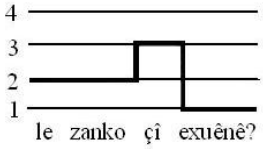
شکل ۶: جمله سؤالی با استفاده از کلمات پرسشی، معنای جمله: /چیه این؟/.



شکل ۷: جمله سؤالی با استفاده از کلمات پرسشی، معنای جمله: /این چیه؟/.



شکل ۸: جمله سؤالی با استفاده از کلمات پرسشی، معنای جمله: /چند سال داری؟/.



شکل ۹: جمله سؤالی با استفاده از کلمات پرسشی، معنای جمله: /تو دانشگاه چی می‌خونه؟/.

جملات خبری

در حالت کلی الگوی کلی این نوع جملات به‌صورت افتان است و در نهایت به سطح ۱ می‌رسد. اما با توجه به این که در جمله تکیه‌ای رخ بدهد یا نه، ممکن است که تا حدودی در الگو تغییر حاصل شود. در حالت کلی می‌توان دو الگوی مثبت و منفی را برای جملات خبری در نظر گرفت. در شکل‌های ۲ و ۳ این دو الگو را ملاحظه می‌کنید.

در مورد جملات خبری منفی همان‌طور که مشاهده می‌شود الگو اندکی تغییر می‌کند. نکته مهم در مورد این جمله‌ها این است که تکیه جمله بر روی فعل منفی به‌کار رفته است. همان‌طور که در شکل ۴ نشان داده شده است، اگر این فعل به ابتدای جمله منتقل شود، تکیه هم همراه آن به ابتدای جمله منتقل می‌شود.

جملات پرسشی

سؤالات در زبان کردی همانند بیشتر زبان‌ها دارای دو الگوی متفاوت می‌باشد. سؤال‌هایی که در آنها از یکی از کلمات پرسشی زبان استفاده شده و در جواب می‌بایست توضیحی کامل ارائه شود و سؤال‌های بله/ خیر که در پاسخ به سادگی می‌توان یکی از دو کلمه بله یا خیر را به‌کار برد. در گروه اول می‌توان از کلمات پرسشی همچون / چی– What/، / کی– Who/، / بوچی– Why/، / چون– How/، / که‌ی– When/، / کوی– Where/، / کام– Which/، / چه‌نی– How many/ و ... استفاده کرد. در حالت کلی تکیه جمله بر روی کلمه پرسش قرار می‌گیرد. در شکل‌های ۵ تا ۹ نمونه‌هایی از این الگوها را ملاحظه می‌کنید.

سؤالات بله/خیر در زبان انگلیسی معمولاً با تغییردادن محل کلمات ایجاد می‌شود؛ برای مثال سؤال /?are you coming/ که با جابه‌جاکردن محل دو کلمه are و you تولید شده است. در زبان کردی این اتفاق نمی‌افتد و با استفاده از تغییر لحن، جمله پرسشی تولید می‌شود. همچنین در موارد زیادی از سؤال کوتاه در انتهای جمله استفاده می‌شود که اصطلاحاً به آنها استفهام انکاری گفته می‌شود. به‌عنوان مثال /رفتی خونه، مگه نه؟/ که الگوی دوقسمتی و مخصوص به خود را دارد. در شکل‌های ۱۰ تا ۱۴ همه این الگوها نمایش داده شده‌اند.

جملات امری

در زبان کردی دو نوع جمله امری مثبت و منفی قابل بررسی هستند. جملات امری مثبت غالباً با اضافه‌شدن حرف /ب/ و جملات منفی هم غالباً با اضافه‌شدن حروف /مه/ به بن مضارع ساخته می‌شوند. به‌عنوان مثال بن مضارع فعل /خوابیدن/ در زبان کردی به‌صورت /نوو/ است. افعال امر مثبت و منفی این فعل به‌صورت /بنوو/ و /مه‌نوو/ ساخته می‌شوند. همچنین در جملات امری، کلمات امر می‌توانند در ابتدا یا انتهای جمله بیایند. در شکل‌های ۱۵ تا ۱۸ تمامی حالات بررسی شده‌اند.

۲-۳-۲ جملات مرکب

منظور از جملات مرکب، جملاتی است که دارای الگوی دوقسمتی باشند. در این میان ما دو نوع جملات هم‌پایه و جملات شرطی و همچنین حالات مختلف آنها را بررسی کرده‌ایم.

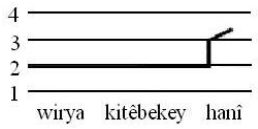
جملات هم‌پایه

جملات هم‌پایه با حرف عطف /و/ به هم وصل می‌شوند. جملات عطفی در زبان کردی به این صورت تلفظ می‌شوند که حرف عطف به جمله اول متصل می‌شود. یعنی در واقع حرف عطف به‌گونه‌ای تلفظ می‌شود که انگار قسمتی از جمله اول است. در این صورت الگوی دو جمله به‌صورت /۲۳۱|۲۳۱/ خواهد بود. جمله اول ناگهان به سطح سه صعود کرده و به‌صورت خمیده به سطح یک می‌آید، جمله دوم از سطح دو شروع شده ناگهانی به سطح سه رفته و ناگهانی به سطح یک می‌آید. شکل ۱۹ این الگو را نشان می‌دهد.

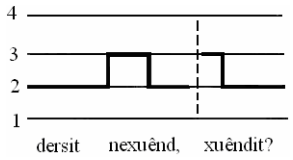
جملات شرطی

این جملات به دو دسته کلی تقسیم می‌شوند. حالت اول وقتی اتفاق می‌افتد که جمله شرط اول بیاید. در این صورت با توجه به این که جمله

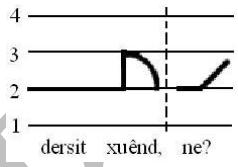
Archive of SID
www.SID.ir



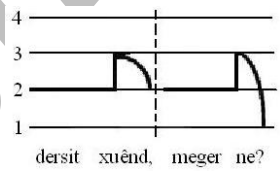
شکل ۱۰: الگوی جمله سؤالی با استفاده از تغییر لحن، معنای جمله: /وریا کتاب رو آورد؟/.



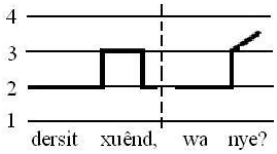
شکل ۱۱: الگوی جمله استفهام انکاری، معنای جمله: /درس نخوندی، خوندی؟/.



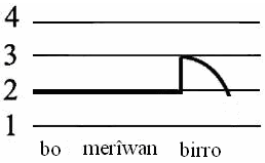
شکل ۱۲: الگوی جمله استفهام انکاری، معنای جمله: /درس خوندی، نه؟/.



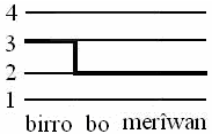
شکل ۱۳: الگوی جمله استفهام انکاری، معنای جمله: /درس خوندی، مگه نه؟/.



شکل ۱۴: الگوی جمله استفهام انکاری، معنای جمله: /درس خوندی، این‌طور نیست؟/.



شکل ۱۵: الگوی جمله امری مثبت با کلمه امر در انتهای جمله، معنای جمله: /به مریوان برو/.



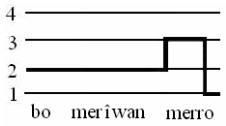
شکل ۱۶: الگوی جمله امری مثبت با کلمه امر در ابتدای جمله، معنای جمله: /برو به مریوان/.

دوم مثبت باشد یا منفی، الگو اندکی تغییر می‌کند. حذف کلمه شرط / اگر/ تغییری در الگوی جمله ایجاد نمی‌کند (شکل‌های ۲۰ و ۲۱).

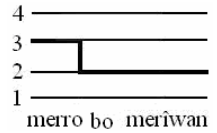
در زبان کردی کمتر اتفاق می‌افتد که جمله شرط دوم بیاید. در این حالت هم با توجه به این که جمله دیگر مثبت یا منفی باشد، الگو اندکی تغییر می‌کند. الگوی مربوط به حالات بیان‌شده در شکل‌های ۲۲ و ۲۳ نمایش داده شده‌اند.

۳- تولید گفتار در زبان کردی

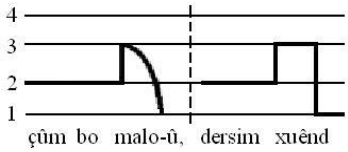
در بخش قبل مرحله تحلیل متن در زبان کردی را بررسی کردیم. در



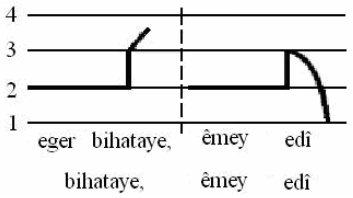
شکل ۱۷: الگوی جمله امری منفی با کلمه امر در انتهای جمله، معنای جمله: /به مریوان نرو/.



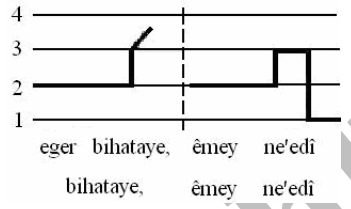
شکل ۱۸: الگوی جمله امری منفی با کلمه امر در ابتدای جمله، معنای جمله: /نرو به مریوان/.



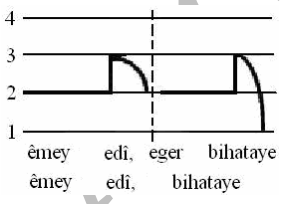
شکل ۱۹: الگوی جمله مرکب هم‌پایه، معنای جمله: /رفتم خونه و درس خوندم/.



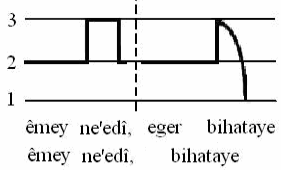
شکل ۲۰: الگوی جمله شرطی، اگر عبارت شرط در ابتدا بیاید و جمله دوم مثبت باشد، معنای جمله‌ها: /اگر می‌آمد، ما را می‌دید/، /می‌آمد، ما را می‌دید/.



شکل ۲۱: الگوی جمله شرطی، اگر عبارت شرط در ابتدا بیاید و جمله دوم منفی باشد، معنای جمله‌ها: /اگر می‌آمد، ما را نمی‌دید/، /می‌آمد، ما را نمی‌دید/.



شکل ۲۲: الگوی جمله شرطی، اگر عبارت شرط در انتها بیاید و جمله اول مثبت باشد، معنای جمله‌ها: /ما را می‌دید، اگر می‌آمد/، /ما را می‌دید، می‌آمد/.

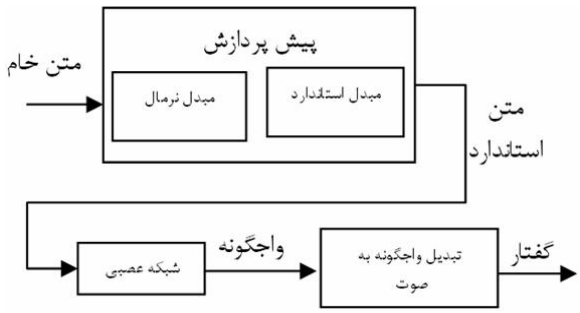


شکل ۲۳: الگوی جمله شرطی، اگر عبارت شرط در انتها بیاید و جمله اول منفی باشد، معنای جمله‌ها: /ما را نمی‌دید، اگر می‌آمد/، /ما را نمی‌دید، می‌آمد/.

این بخش مرحله دوم سیستم‌های تبدیل متن به گفتار را بررسی می‌کنیم که مربوط به تولید گفتار در این زبان می‌شود. ما برای تولید گفتار از سه واحد واج‌گونه، هجا و دایفون استفاده کرده و در واقع سه سیستم تبدیل

Archive of SID

www.SID.ir



شکل ۲۴: معماری کلی سیستم سنتز مبتنی بر واجگونه.

متن به گفتار مجزا را طراحی کرده‌ایم. در این میان اصول طراحی سیستم‌های مبتنی بر هجا و دایفون مشابه هم می‌باشد. یکی از مهم‌ترین و زمان‌برترین مراحل در طراحی این سیستم‌ها، ایجاد پایگاه داده مناسب بود.

روش طراحی سیستم مبتنی بر واجگونه قدری متفاوت بوده و بخش تحلیل متن آن به‌صورت جداگانه انجام گرفته است؛ بنابراین در این بخش توضیح مختصری در مورد آن ارائه می‌دهیم. برای استخراج واجگونه‌ها از یک شبکه عصبی سه‌لایه استفاده شده است که با دقت بسیار خوبی این کار را انجام می‌دهد. پایگاه داده مورد نیاز در این سیستم کوچک بوده و در زمان نسبتاً کمی ساخته شد. لازم به ذکر است که در همه سیستم‌ها برای یکنواخت‌سازی و تغییردادن گام از الگوریتم TD - PSOLA استفاده کرده‌ایم.

۳-۱ سیستم سنتز مبتنی بر واجگونه

گونه‌های مختلف تلفظ یک واج با توجه به موقعیت آن در کلمه را واجگونه گویند [۲۷]. به‌عنوان مثال در زبان کردی، واج /س/ در کلمه /سه‌د/ کاملاً پرحجم و مشابه واج /ص/ در زبان عربی و در /سنه/ کاملاً کم‌حجم و مشابه واج /س/ در زبان فارسی تلفظ می‌شود.

همان‌طور که ملاحظه می‌شود در بین تلفظ دو واجگونه تفاوت زیادی وجود دارد که این تفاوت زیاد، مهم‌ترین دلیل ما برای انتخاب واجگونه‌ها در این سیستم است [۲۸]. البته بایستی به این نکته توجه شود که در مورد همه واجگونه‌ها تفاوت خیلی زیاد نبوده و به‌صورت عادی قابل تشخیص نمی‌باشند؛ ولیکن در عمل این تفاوت بر روی کیفیت سیگنال خروجی تأثیر زیادی دارد که ما در عمل با آن مواجه شدیم.

تعدادی از واجگونه‌ها از قوانین ثابت و مشخصی پیروی می‌کنند؛ به‌عنوان مثال در زبان کردی و در مورد اکثر واج‌ها، اگر واج مورد نظر در پایان کلمه بیاید، واک خود را از دست داده و آن واجگونه را اصطلاحاً نمونه واک‌رفته یا بی‌رهش آن واج می‌نامند [۲۵]. همان‌طور که ملاحظه می‌شود تشخیص این‌گونه واجگونه‌ها ساده است؛ اما مشکل کار اینجاست که بیشتر واجگونه‌ها از قانون مشخصی تبعیت نمی‌کنند و به همین دلیل تشخیص واجگونه بسیار سخت بوده و از عهده زبان‌شناسان خبره بر می‌آید.

با توجه به مطالب بالا و این که تشخیص و استخراج واجگونه‌ها با استفاده از روش‌های مبتنی بر قاعده به سادگی امکان‌پذیر نیست، ما تصمیم گرفتیم که از قدرت یادگیری شبکه‌های عصبی استفاده کنیم. شبکه‌های عصبی پس از آموزش، می‌توانند واجگونه‌ها را با دقت بالایی تشخیص دهند. بنابراین پایگاه داده مناسبی را به همراه واجگونه‌های مطلوب طراحی کرده و مورد استفاده قرار دادیم که در قسمت‌های بعدی به تفصیل مورد بررسی قرار خواهد گرفت. سیستم طراحی‌شده دارای سه مرحله اصلی می‌باشد که در شکل ۲۴ نمایش داده شده است.

در مرحله اول متن خام به‌عنوان ورودی گرفته شده و در دو مرحله عملیات پیش‌پردازش روی آن صورت می‌گیرد، و نهایتاً متن استاندارد به‌عنوان خروجی حاصل می‌شود. سپس در مرحله دوم شبکه عصبی واجگونه‌های متن استاندارد را یافته و سیگنال از پیش ضبط شده برای این واجگونه‌ها در مرحله سوم به هم متصل می‌شوند. در ادامه هر کدام از این مراحل را با جزئیات بیشتر مورد بررسی قرار می‌دهیم.

۳-۱-۱ پیش‌پردازش

همان‌طور که بیان شد، برای زبان کردی دو رسم‌الخط عربی و لاتین به‌صورت رسمی وجود دارد. هر یک از این دو، مشکلات خاص خود را دارند که باید به‌صورت جداگانه بررسی شوند. برای حل این مشکل نمادهای استانداردی برای سیستم در نظر گرفته شده است و متن ورودی در هر کدام از رسم‌الخط‌ها به این رسم‌الخط میانی نگاشت داده می‌شوند. این کار در مرحله مبدل نرمال انجام می‌گیرد. در قسمت مبدل استاندارد هم بقیه مشکلات عمومی متن، همچون مشکلات مربوط به اعداد، عبارات مخفف، نمادها و ... برطرف می‌گردد و نهایتاً در قسمت خروجی این مرحله، متن استانداردی حاصل می‌شود که حتی‌الامکان عاری از ابهام و مشکلات مربوط به نوشتار است.

۳-۱-۲ شبکه عصبی

همان‌گونه که قبلاً اشاره شد ما باید واجگونه‌های زبان کردی را از متن ورودی استخراج کنیم. تعداد واجگونه‌های هر واج می‌تواند خیلی متغیر باشد، به‌عنوان مثال در زبان کردی تعداد واجگونه‌های واج /پ/ و یا /ب/ خیلی زیاد و واجگونه‌های واج‌های /ش/ و یا /چ/ کمتر است. به‌صورت تجربی می‌توان گفت که تعداد واجگونه‌های زبان کردی در حدود ۲۰۰ عدد است [۲۵]. از این تعداد برخی از آنها با یکدیگر تفاوت زیادی دارند اما برخی دیگر خیلی به هم نزدیک بوده و تفاوت آنها را به سختی می‌توان تشخیص داد. به همین منظور لازم نیست که همه واجگونه‌های زبان را در یک سیستم تبدیل متن به گفتار در نظر بگیریم. ما در سیستم طراحی‌شده ۶۶ مورد از مهم‌ترین و واضح‌ترین واجگونه‌ها را در نظر گرفته‌ایم. البته این تعداد به‌طور مساوی بین واج‌ها تقسیم نشده است. به‌عنوان مثال برای واج /پ/ پنج واجگونه در نظر گرفته شده است در حالی که برای /ر/ فقط یک مورد در نظر گرفته شده است. باید به این موضوع اشاره داشته باشیم که ساختار سیستم کاملاً قابل انعطاف بوده و به سادگی می‌توان تعداد واجگونه‌های مورد نظر را در صورت نیاز افزایش داد. لیست واجگونه‌های مورد استفاده در سیستم را در جدول ۲ ملاحظه می‌کنید. نکته مهم تهیه‌کردن پایگاه داده برای واجگونه‌های مورد نظر و به‌منظور آموزش شبکه می‌باشد که در ادامه توضیح داده خواهد شد.

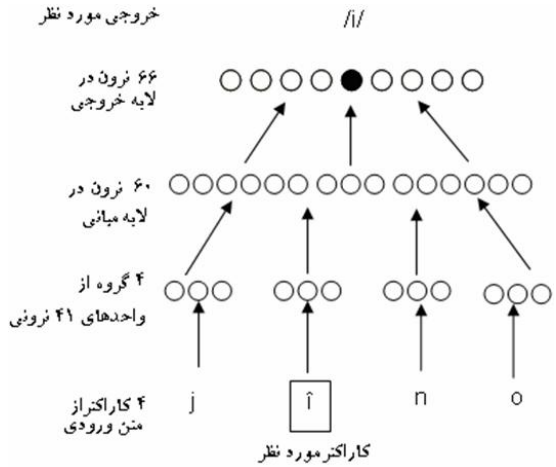
ساختار شبکه

برای استخراج واجگونه‌ها از یک شبکه عصبی سه‌لایه پیشرونده با الگوریتم پس‌انتشار خطا استفاده کردیم. همان‌طور که بیان شد برای تشخیص واجگونه‌ها در زبان کردی قانون و قاعده ثابت و کلی وجود ندارد و دقیقاً به‌همین منظور از شبکه عصبی استفاده کردیم؛ اما پس از انجام مطالعات زبان‌شناسی به این نتیجه رسیدیم که بیشترین میزان تأثیر بر روی تلفظ واج‌ها در زبان کردی (واجگونه‌ها) حداکثر به یک واج قبل و دو واج بعدی وابسته است. اصولاً نحوه تأثیرگذاری واج‌ها بر تلفظ همدیگر به ساختار زبان وابسته است. در مورد زبان انگلیسی، تلفظ هر واج به سه واج قبل و سه واج بعدی وابسته می‌باشد [۲۹]. با توجه به مطالب بیان‌شده، برای لایه ورودی شبکه چهار گروه نرون استفاده شد که در هر گروه تعداد ۴۱ نرون وجود دارد. توجه شود که علاوه بر ۳۸ نماد استاندارد در نظر

Archive of SID
www.SID.ir

جدول ۲: لیست واج‌گونه‌های مورد استفاده در سیستم مبتنی بر واج‌گونه.

واج	p	b	t	d	k	g	q	f	s	Sh	z	J	G	A
واج‌گونه	PpO*&	bEB	t@T	d!D	k?K	G%g	Qq	FVf	s	$S	zZ>	Jj	^	A
واج	c	h	H	m	x	X	n	v	y	l	r	L	R	Y
واج‌گونه	Cc	h	H	mWM	x	X	nN	v	y	l	r	L	R	Y
واج	e	a	N	U	u	o	w	I	i	C	Ü	.	,	
واج‌گونه	e	a	#	U	u	o	w	I	i	~	–	.	,	



شکل ۲۵: ساختار شبکه عصبی به‌کار رفته در سیستم مبتنی بر واج‌گونه.

گرفته شده، علایم نگارشی نقطه، ویرگول و فاصله هم در پیاده‌سازی مورد توجه بوده‌اند. با این حساب در لایه ورودی کلاً ۱۶۴ نرون وجود دارد. لایه خروجی هم متشکل از ۶۶ نرون است که مقادیر آنها در بازه [۰,۱] قرار دارد. هر نرون به یکی از واج‌گونه‌های زبان اشاره دارد و خروجی شبکه برابر خروجی نرونی قرار می‌گیرد که بزرگ‌ترین مقدار ممکن را داشته باشد. ساختار نهایی شبکه طراحی‌شده در شکل ۲۵ نمایش داده شده است.

نحوه عملکرد سیستم به این صورت است که یک پنجره چهارتایی بر روی متن استاندارد (خروجی مرحله پیش‌پردازش) می‌لغزد به‌گونه‌ای که در هر نوبت واج مورد نظر در موقعیت دوم قرار داشته باشد. سپس شبکه عصبی با توجه به خود واج و یک واج قبل و دو واج بعد، واج‌گونه مورد نظر را شناسایی می‌کند. نرخ نهایی تشخیص درست واج‌گونه‌ها توسط شبکه عصبی بسیار خوب و برابر ۹۸٪ بود. لازم به ذکر است که برای آموزش شبکه پایگاه داده‌ای با اندازه ۲۱۰۰ نمونه و همچنین برای ارزیابی سیستم از ۴۰۰ نمونه جدید استفاده کردیم. جزئیات بیشتر در [۲۸] بیان شده است.

۳-۲ سیستم سنتز مبتنی بر هجا

یکی از واحدهای زبانی که می‌توان برای توسعه یک سیستم تبدیل متن به گفتار از آن استفاده کرد، هجا است. در زبان‌های مختلف الگوهای متفاوتی برای هجاها وجود دارد. با توجه به این که معمولاً در اکثر زبان‌ها تعداد این الگوها زیاد است، تعداد کل هجاها هم زیاد بوده و در نتیجه برای استفاده در سیستم‌های همه‌منظوره قابل استفاده نمی‌باشند. به‌عنوان مثال تعداد هجاهای زبان انگلیسی بیش از ۱۵۰۰۰ مورد است [۳۰] و ساختن پایگاه داده با این اندازه زمان‌بر بوده و همچنین نیاز به حافظه نسبتاً زیادی دارد.

اما در برخی از زبان‌ها با توجه به این که الگوی هجاها محدودتر است، تعداد نهایی هجاها هم در حد معقول بوده و می‌توان از آن در سیستم‌های

جدول ۳: الگوهای هجایی در زبان کردی.

		سووک	پر	گران
ئاسایی	الگوی هجا	CV	CVC	CVCC
(قانون‌مند)	مثال	ده، تو	واز، لیخ	کورت، به‌رد
لیکدراو	الگوی هجا	CCV	CCVC	CCVCC
(به‌هم پیوسته)	مثال	برو، چیا	بژار، بزووت	خواست، بنیشت
ناته‌واو	الگوی هجا	V	VC	VCC
(ناتمام)	مثال	– ی	– ان	– اند

همه‌منظوره استفاده کرد. به‌عنوان مثال در زبان هندی برای هجاها چهار الگوی کلی CV، CCV، VC و CVC وجود دارد و تعداد کل هجاها در این زبان کمتر از ۱۰۰۰۰ مورد خواهد بود. در [۳۱] از واحدهای مشابه هجا استفاده شده است که تعداد کل آنها ۱۲۴۲ مورد است. همان‌طور که ملاحظه می‌شود این تعداد کاملاً معقول بوده و ساخت پایگاه داده‌ای با این اندازه خیلی سخت نیست.

همچنین در زبان فارسی هم از این واحد برای توسعه سیستم تبدیل متن به گفتار استفاده شده است. با توجه به این که الگوهای هجا در زبان فارسی محدود به سه حالت CV، CVC و CVCC است. تعداد کل هجاهای فارسی در حدود ۴۰۰۰ مورد بوده و می‌توان در یک سیستم همه‌منظوره از آن استفاده کرد [۱۶].

در زبان کردی سه گروه اصلی هجا وجود دارد که عبارتند از ئاسایی (قانون‌مند)، لیکدراو (به‌هم پیوسته) و ناته‌واو (ناتمام) [۲۲]. گروه ئاسایی مهم‌ترین و اصلی‌ترین گروه هجایی در زبان کردی است و تقریباً اکثر هجاهای کردی در این گروه قرار می‌گیرند. در گروه لیکدراو هجاهایی قرار می‌گیرند که در آنها دو واج هم‌خوان پشت سر هم و در ابتدای کلمه ظاهر می‌شوند. دسته ناته‌واو هم مربوط به برخی پسوندهای خاص مانند /ان/ در مصدر / سووتان/ است. تعداد این هجاها هم خیلی زیاد نیست و در برخی حالات خاص رخ می‌دهد. هر کدام از گروه‌های هجایی بالا خود به سه دسته سووک (سبک)، پر (پر) و گران (گران) تقسیم می‌شوند [۲۲]. جدول ۳ این تقسیم‌بندی‌ها را به‌صورت کامل و همراه با مثال نمایش می‌دهد.

همان‌طور که در جدول نمایش داده شده است تعداد کل الگوهای هجایی در زبان کردی نه عدد است که البته در این میان سه الگوی CV، CVC و CVCC الگوهای اصلی زبان محسوب می‌شوند. با توجه به این که در عمل گروه ئاسایی بسیار پرکاربرد است، می‌توان در پیاده‌سازی‌ها فقط این سه الگو را به‌کار برد و به این ترتیب تعداد هجاها کاهش می‌یابد.

۳-۳ سیستم سنتز مبتنی بر دایفون

در اغلب سیستم‌های امروزی از دایفون به‌عنوان واحد اصلی زبانی استفاده می‌شود. با توجه به این که دایفون‌ها حاوی قسمت انتقال یک واج به واج بعدی هستند، پس نسبت به واحدهای دیگر کیفیت مطلوب‌تری دارند. البته ترکیب استفاده از این واحد و روش‌هایی همچون انتخاب واحد

Archive of SID

www.SID.ir

جدول ۴: نمرات کسب‌شده توسط هر سیستم در آزمون MOS.

	تلاش شنیداری	میزان درک	نحوه ادا	کیفیت تلفظ	خوشایندی	نظر کلی
سیستم مبتنی بر واج‌گونه	۲٫۸۰	۲٫۹۱	۲٫۶۷	۲٫۳۲	۲٫۰۸	۲٫۲۱
سیستم مبتنی بر هجا	۳٫۰۵	۳٫۲۱	۳٫۰۱	۲٫۷۷	۲٫۶۱	۳٫۰۲
سیستم مبتنی بر دایفون	۳٫۸۵	۳٫۹۳	۳٫۷۲	۳٫۳۲	۳٫۳۷	۳٫۵۷
جملات طبیعی	۴٫۸۰	۴٫۹۱	۴٫۷۷	۴٫۶۱	۴٫۷۱	۴٫۸۵

هم در برخی سیستم‌ها با موفقیت مورد استفاده قرار گرفته است. با توجه به نتایج گزارش‌شده در سیستم‌های مشابه، ما هم تصمیم به استفاده از دایفون‌ها به‌عنوان واحد اصلی سیستم گرفتیم.

با توجه به این که دایفون عبارت است از نیمه انتهایی یک واج به علاوه نیمه ابتدایی واج بعدی، و با توجه به این که تعداد واج‌های زبان کردی ۳۷ مورد می‌باشد، پس در بدترین شرایط تعداد دایفون‌های زبان کردی برابر با ۱۳۳۲ مورد خواهد بود. البته باید به این نکته توجه داشت که همه این دایفون‌ها معتبر نبوده و برخی از آنها عملاً هیچ وقت در زبان کردی مورد استفاده قرار نمی‌گیرند. به‌عنوان مثال هیچ‌گاه دو واج /ع/ و /غ/ به دنبال هم استفاده نمی‌شوند و یا همین‌طور واج‌های /خ/ و /ح/. همچنین هیچ‌گاه دو واکه به دنبال هم استفاده نخواهد شد و با توجه به این که تعداد واکه‌ها در زبان کردی برابر شش است، پس تمامی جایگشت‌های دوتایی آنها هم نامعتبر خواهد بود. به همین دلیل تعداد واقعی دایفون‌های زبان در حدود ۹۰۰ مورد است.

بعد از انتخاب واحد مورد نظر، بایستی برای واحدها نمونه‌های مناسبی در نظر گرفته شوند. ما برای ساخت پایگاه داده از یک متن بزرگ استفاده کردیم. برای ضبط، دو گوینده حرفه‌ای زن و مرد را انتخاب کرده و از آنها خواستیم که متن را در شرایط خنثی بخوانند. مدت زمان خواندن در حدود یک ساعت بود. سپس با استفاده از ابزار Cool Edit واحدهای مناسب را جدا کرده و در پایگاه داده ذخیره کردیم [۳۲].

٤- آزمون‌های کیفیت

در این مقاله، ما سه سیستم مختلف تبدیل متن به گفتار را برای زبان کردی پیاده‌سازی کردیم. در این سیستم‌ها به‌ترتیب از واج‌گونه، هجا و دایفون به‌عنوان واحدهای اصلی استفاده شده است. میزان هزینه و زمان طراحی این سیستم‌ها بسیار متفاوت بود. به‌عنوان مثال طراحی پایگاه داده برای واج‌گونه‌ها ساده‌تر و سریع‌تر از دایفون صورت گرفت. اما در این میان کیفیت صدای تولیدشده در این سیستم‌ها با هم برابر نبود. به همین دلیل برای سنجش کیفیت این سیستم‌ها از آزمون‌هایی استفاده کرده‌ایم که در ادامه توضیح داده خواهند شد.

از میان آزمون‌های انجام‌گرفته، سه آزمون MOS، DRT و MRT آزمون‌هایی استاندارد و شناخته‌شده هستند که در اغلب سیستم‌های طراحی‌شده در زبان‌های دیگر هم مورد استفاده قرار می‌گیرند. با توجه به این موضوع می‌توان نتایج حاصل از این آزمون‌ها را با نتایج حاصل از سیستم‌های دیگر زبان‌ها مقایسه کرد. هر کدام از این آزمون‌ها سنجش کیفیت گفتار را بر اساس تعدادی از پارامترهای خاص دنبال می‌کنند. البته در این میان آزمون MOS بیشتر ملاک‌های طبیعی‌بودن و خوشایندی گفتار را ارزیابی می‌کند در حالی که دو آزمون دیگر بیشتر جنبه قابلیت فهم را مورد سنجش قرار می‌دهند. در این میان ما از یک آزمون قابلیت فهم هم استفاده کرده‌ایم که کمتر در سیستم‌های دیگر استفاده شده است. در این آزمون نرخ تشخیص درست کلمات و هجاها توسط سیستم بررسی شده‌اند.

٤-١ آزمون MOS

اولین آزمون برگزارشده MOS است که دارای شش ملاک برای ارزیابی کیفیت گفتار خروجی می‌باشد. این ملاک‌ها به‌ترتیب عبارتند از: نظر کلی درباره کیفیت صدا، تلاش شنیداری برای درک پیام، میزان درک (مشکل در فهم کلمات)، نحوه ادا (میزان تمایز بین صداها)، کیفیت تلفظ و خوشایندی صدا [۳۱]. روش آزمون به این صورت است که چند جمله را با استفاده از سیستم مذکور تولید کرده و از تعدادی شنونده می‌خواهیم که به هر کدام از ملاک‌های ذکرشده نمره‌ای بین یک تا پنج بدهد. نمره ۱ به معنی بسیار ضعیف و نمره ۵ به معنای بسیار خوب است. در نهایت میانگین نمرات به‌دست آمده برای هر ملاک توسط همه شنونده‌ها را به‌عنوان نمره نهایی آن ملاک در نظر می‌گیریم.

برای انجام این آزمون از تعداد ۱۹ نفر (۱۳ مرد و ۶ زن) خواسته شد که به هفت جمله تولیدشده توسط هر سیستم گوش داده و به معیارهای ذکرشده امتیاز بدهند. با توجه به این که تعداد معیارها زیاد بود، هر جمله برای شنوندگان دو بار پخش شد. لازم به ذکر است که تمامی افراد مورد نظر کرد بوده و همچنین مشکل شنوایی بارزی نداشتند. شنوندگان از قبل از محتوای جملات خبر نداشته و خود باید آن را تشخیص می‌دادند. همچنین برای این که میزان سخت‌گیری شنوندگان را بسنجیم، علاوه بر جملات تولیدشده توسط سیستم‌های مذکور، همان جملات را با صدای طبیعی ضبط کرده و در حین آزمون از آنها خواسته شد که به این جملات هم امتیاز بدهند. ترتیب پخش جملات برای هر شنونده کاملاً تصادفی بوده و سعی شده است که جملات تولیدشده توسط همه سیستم‌ها و همچنین جملات طبیعی به‌صورت درهم (ترتیب تصادفی) به شنوندگان ارائه شوند. جدول ۴ نمرات نهایی برای تمامی سیستم‌ها را نشان می‌دهد.

همان‌طور که از جدول پیداست، سیستم مبتنی بر دایفون بهترین کیفیت را داراست و پس از آن سیستم مبتنی بر هجا قرار دارد. همچنین سیستم مبتنی بر واج‌گونه پایین‌ترین کیفیت را داراست. از میان پارامترهای مورد ارزیابی، میزان درک بالاترین نمره‌ها را به خود اختصاص داده است. برای سنجش بهتر و دقیق‌تر قابلیت فهم آزمون‌های دیگری ترتیب داده شده‌اند که در ادامه بیان خواهند شد.

٤-٢ آزمون قابلیت فهم

آزمون بعدی که بر روی سیستم‌ها انجام گرفت برای سنجش میزان قابلیت فهم آنان بود. در واقع آزمون MOS بیشتر برای سنجش میزان طبیعی‌بودن به کار می‌رود. از آنجا که قابلیت فهم یکی از مهم‌ترین ملاک‌های ارزیابی سیستم‌های تبدیل متن به گفتار است، پس لازم است که به‌صورت جداگانه مورد ارزیابی قرار گیرد. آزمون انجام‌گرفته به این صورت است که از شنوندگان خواسته می‌شود به جملات گوش داده و هر آنچه را که می‌فهمند، بنویسند. سپس از روی این نوشته‌ها دو ملاک نرخ تشخیص درست کلمات و نرخ تشخیص درست هجاها به‌دست می‌آید.

Archive of SID

www.SID.ir

جدول ۵: نتایج حاصل از آزمون قابلیت فهم.

	نرخ تشخیص درست کلمات	نرخ تشخیص درست هجاها
سیستم مبتنی بر واج‌گونه	۷۷٫۶	۸۵٫۴
سیستم مبتنی بر هجا	۸۳٫۹	۸۸٫۵
سیستم مبتنی بر دایفون	۹۲٫۷	۹۶٫۹
جملات طبیعی	۱۰۰	۱۰۰

جدول ۶: نتایج نهایی حاصل از آزمون DRT.

	واکدار/ بی‌واک	خیشومی	نگه‌داری	سایش	سختی	فشردگی	میانگین
سیستم مبتنی بر واج‌گونه	۹۶٫۸۷	۹۷٫۳۹	۹۴٫۲۷	۹۵٫۸۳	۹۵٫۳۱	۹۷٫۹۱	۹۶٫۲۶
سیستم مبتنی بر هجا	----	----	----	----	----	----	----
سیستم مبتنی بر دایفون	۹۷٫۹۱	۹۸٫۹۵	۹۵٫۳۱	۹۶٫۳۵	۹۷٫۳۹	۹۸٫۴۳	۹۷٫۳۹
کلمات طبیعی	۱۰۰	۱۰۰	۹۹٫۲۱	۱۰۰	۹۸٫۸۷	۱۰۰	۹۹٫۶۸

جدول ۷: نتایج حاصل از آزمون DRT برای افراد غیر کردزبان.

	واکدار/ بی‌واک	خیشومی	نگه‌داری	سایش	سختی	فشردگی	میانگین
سیستم مبتنی بر واج‌گونه	۹۵٫۰۰	۹۶٫۲۵	۹۵٫۰۰	۹۳٫۷۵	۹۳٫۷۵	۹۵٫۰۰	۹۴٫۷۹
سیستم مبتنی بر هجا	----	----	----	----	----	----	----
سیستم مبتنی بر دایفون	۹۷٫۵۰	۹۵٫۰۰	۹۳٫۷۵	۹۵٫۰۰	۹۷٫۵۰	۹۶٫۲۵	۹۵٫۸۳
کلمات طبیعی	۱۰۰	۹۷٫۵۰	۹۷٫۵۰	۹۵٫۰۰	۹۶٫۲۵	۹۷٫۵۰	۹۷٫۲۹

در این آزمون از همان گروه ۱۹ نفره قبلی استفاده شد. در واقع این آزمون هم‌زمان با آزمون MOS انجام گرفت به این ترتیب که در حین امتیازدادن به جملات، از آنها خواسته شد که متن فهمیده‌شده را بنویسند. ملاک‌های ذکرشده با استفاده از (۱) محاسبه شدند

$$WR = \frac{Correct\ Word\ \ Number}{Total\ Word\ \ Number}$$
$$SR = \frac{Correct\ \ Syllable\ \ Number}{Total\ \ Syllable\ \ Number} \quad (۱)$$

پس از انجام آزمون نتایج مربوط به تمام افراد را برای تمامی جملات ثبت کرده و از آنها میانگین گرفتیم. جدول ۵ نتایج این آزمون را نمایش می‌دهد. همان‌طور که از جدول پیداست، میزان قابلیت فهم کلیه سیستم‌ها به خصوص سیستم مبتنی بر دایفون، بسیار بالاست. به‌طور طبیعی نرخ تشخیص درست هجاها بالاتر از کلمات بوده است. در واقع با توجه به این که هجاها واحدهای کوچک‌تر و ساده‌تری نسبت به کلمات هستند، این نتایج قابل پیش‌بینی بود.

٤-۳ آزمون DRT

سومین آزمون DRT بود که برای سنجش شش پارامتر استاندارد سیستم‌های تبدیل متن به گفتار طراحی شده است [۳۰]. با استفاده از این پارامترها می‌توان تمامی واج‌های زبان را از هم جدا کرد. لازم به ذکر است که این پارامترها برای زبان انگلیسی استفاده می‌شوند ولی چون به لحاظ زبان‌شناسی، پارامترهای دقیق برای زبان کردی استخراج نشده است، پس به ناچار ما هم از همین پارامترها استفاده کرده‌ایم.

روش آزمون به این صورت است که برای هر یک از این معیارها ۱۶ جفت کلمه (در مجموع ۹۶ جفت) طراحی می‌شوند به‌گونه‌ای که این کلمات تک‌هجایی بوده و همچنین در هر جفت، فقط واج اول کلمات با هم متفاوت باشند. در واقع هر جفت کلمه به این صورت انتخاب می‌شود که واج اول آنها در آن پارامتر با هم متفاوت بوده و بقیه واج‌ها کاملاً یکسان باشند. آزمون به این صورت است که از هر جفت یکی از کلمات به‌صورت تصادفی برای شنونده پخش شده و از او خواسته می‌شود تا تشخیص دهد که کدام یک از دو کلمه را شنیده است. به این ترتیب با توجه به تعداد کلمات درست حدس زده شده، معیار مورد نظر سنجیده می‌شود. لیست کامل کلمات مورد استفاده در این آزمون در [۲۶] ارائه شده است.

برای انجام این آزمون علاوه بر ۱۹ نفر مورد استفاده در دو آزمون قبل، از پنج نفر غیر کرد زبان مرد استفاده شده است. میانگین نمرات کسب‌شده برای هر سیستم در جدول ۶ آمده است. با توجه به این که در این آزمون از کلمات تک‌هجایی استفاده می‌شود، سنجش سیستم مبتنی بر هجا بی‌معنی خواهد بود. لازم به ذکر است که ما میانگین نمرات مربوط به افراد غیر کرد زبان را هم به‌صورت جداگانه محاسبه کرده‌ایم که در جدول ۷ نمایش داده شده است. همان‌طور که ملاحظه می‌شود، میانگین نمرات اندکی پایین‌تر از نمرات کلی می‌باشد.

همان‌طور که ملاحظه می‌شود نرخ تشخیص درست کلمات برای هر دو سیستم بسیار بالاست و به‌ویژه سیستم مبتنی بر دایفون دارای نرخ بسیار خوبیست. نکته جالب توجه در این آزمون این است که میانگین نرخ تشخیص درست برای جملات طبیعی هم ۱۰۰٪ نبوده و چند مورد تشخیص غلط هم داشته است. با این اوصاف می‌توان گفت که نتایج به‌دست آمده برای هر دو سیستم قابل قبول است.

٤-٤ آزمون MRT

آخرین آزمون انجام‌گرفته، آزمون MRT بود که در واقع حالت توسعه‌یافته‌ای از آزمون DRT می‌باشد. آزمون به این صورت است که ۵۰ گروه شش‌کلمه‌ای در نظر گرفته می‌شود با این شرط که اولاً همه کلمات تک‌هجایی بوده و ثانیاً در ۲۵ گروه اول فقط واج آخر کلمات با هم متفاوت باشد و در ۲۵ گروه دوم، فقط واج اول کلمات با هم فرق داشته باشند. انجام آزمون به این صورت است که همانند آزمون DRT از هر گروه یک کلمه به‌صورت تصادفی برای شنونده پخش شده و از وی

Archive of SID

www.SID.ir

جدول ۸: نتایج نهایی به‌دست آمده از آزمون MRT.

	موقعیت ابتدایی	موقعیت پایانی	میانگین
سیستم مبتنی بر واج‌گونه	۷۹٫۳۰	۸۲٫۸۰	۸۱٫۰۵
سیستم مبتنی بر هجا	----	----	----
سیستم مبتنی بر دایفون	۹۱٫۲۰	۸۷٫۴۰	۸۹٫۳۰
کلمات طبیعی	۹۷٫۳۰	۹۶٫۴۵	۹۶٫۸۷۵

جدول ۹: نتایج به‌دست آمده از آزمون MRT برای افراد غیر کردزبان.

	موقعیت ابتدایی	موقعیت پایانی	میانگین
سیستم مبتنی بر واج‌گونه	۷۶٫۰۰	۷۶٫۸۰	۷۶٫۴۰
سیستم مبتنی بر هجا	----	----	----
سیستم مبتنی بر دایفون	۸۲٫۴۰	۸۴٫۰۰	۸۳٫۲۰
کلمات طبیعی	۹۲٫۸۰	۹۲٫۰۰	۹۲٫۴۰

جدول ۱۰: مقایسه سیستم‌های طراحی‌شده در زبان کردی با سیستم‌های سنتز در دیگر زبان‌ها.

	EL	LT	AK	RS	SS	ATT	FTTS	KA	KS	KD
تلاش شنیداری	۲٫۲۸	۲٫۸۱	۲٫۸۳	۳٫۴۳	۳٫۷۰	۳٫۷۲	۳٫۶۵	۲٫۸۰	۳٫۰۵	۳٫۸۵
میزان درک	۲٫۴۷	۲٫۹۷	۳٫۰۲	۳٫۶۳	۳٫۸۸	۳٫۹۱	۳٫۹۷	۲٫۹۱	۳٫۲۱	۳٫۹۳
نحوه ادا	۲٫۲۱	۲٫۶۵	۲٫۶۷	۳٫۳۴	۳٫۶۹	۳٫۸۱	۳٫۷۴	۲٫۶۷	۳٫۰۱	۳٫۷۲
کیفیت تلفظ	۲٫۰۴	۲٫۶۰	۲٫۶۰	۳٫۱۷	۳٫۴۹	۳٫۳۰	۳٫۳۷	۲٫۳۲	۲٫۷۷	۳٫۳۲
خوشایندی	۲٫۰۱	۲٫۴۹	۲٫۴۹	۳٫۰۵	۳٫۶۰	۳٫۶۵	۳٫۱۲	۲٫۰۸	۲٫۶۱	۳٫۳۷
نظر کلی	۱٫۷۳	۲٫۱۹	۲٫۲۸	۳٫۱۳	۳٫۶۰	۳٫۷۴	۳٫۶۸	۲٫۲۱	۳٫۰۲	۳٫۵۷

خواسته می‌شود تا حدس بزند کدام کلمه را شنیده است. در نهایت میانگین تعداد کلمات درست را برای دو گروه ابتدایی و انتهایی به‌صورت جداگانه محاسبه می‌کنند. لیست کامل کلمات مورد استفاده در این آزمون در [۲۶] ارائه شده است.

ما برای انجام این آزمون از همان افراد شرکت‌کننده در آزمون DRT استفاده کردیم. نتایج به‌دست آمده از این آزمون و همچنین نتایج مربوط به افراد غیر کردزبان در جدول‌های ۸ و ۹ نمایش داده شده است. لازم به توضیح است که در اینجا هم چون کلمات تک‌هجایی هستند، پس سنجش سیستم مبتنی بر هجا بی‌معنی خواهد بود.

با توجه به نتایج به‌دست آمده می‌بینیم که نرخ تشخیص درست نسبت به آزمون DRT اندکی افت داشته است. در واقع با توجه به این که در آزمون DRT فقط دو کلمه مورد آزمون بودند ولی در اینجا شش گزینه برای شنوندگان مطرح است، پس طبیعی است که نرخ تشخیص درست پایین‌تر باشد. همان‌طور که ملاحظه می‌شود نرخ تشخیص کلمات طبیعی هم ۱۰۰٪ نبوده و در اینجا هم حدود ۳٪ خطا وجود دارد.

با توجه به نتایج به‌دست آمده از سه آزمون اخیر، می‌توان ادعا کرد که قابلیت فهم هر سه سیستم در حد قابل قبول بوده و به خصوص در این میان سیستم مبتنی بر دایفون دارای قابلیت فهم بسیار مطلوبی می‌باشد. به خصوص نتایج دو آزمون DRT و MRT برای دو سیستم مبتنی بر واج‌گونه و دایفون بسیار خوب است. هر چند که با استفاده از این دو آزمون نمی‌توان سیستم مبتنی بر هجا را مورد ارزیابی قرار داد اما با توجه به نتایج آزمون قابلیت فهم، می‌توان دید که سیستم مبتنی بر هجا هم دارای قابلیت فهم بینابینی می‌باشد.

٤-۵ مقایسه سیستم‌های طراحی‌شده با سیستم‌های دیگر زبان‌ها

با توجه به این که سیستم‌های طراحی‌شده اولین موارد در نوع خود هستند، پس نمی‌توان کیفیت آنها را با سیستم دیگری در زبان کردی مقایسه کرد. بنابراین ما مجبور شدیم که کیفیت این سیستم‌ها را با سیستم‌های مربوط به زبان‌های دیگر مقایسه کنیم. البته باید این نکته را مد نظر داشته باشیم که هر زبان شرایط خاص خود را دارد و همچنین شرایط مورد آزمون هم در تمامی سیستم‌ها کاملاً یکسان نبوده است. از همین رو به این نوع مقایسه تا حدودی ایراد وارد است. اما در شرایط فعلی که سیستم دیگری برای زبان کردی وجود ندارد، ما ناگزیر از انجام این مقایسه بودیم. به هر حال حتی‌الامکان سعی شده است که شرایط آزمون را مشابه با آنچه که در مورد سیستم‌های دیگر صورت گرفته است انتخاب کنیم. نتایج این مقایسه می‌تواند جایگاه سیستم‌های طراحی‌شده را در میان سیستم‌های دیگر زبان‌ها تا حدودی روشن سازد.

برای انجام این مقایسه ما از شش سیستم برای زبان انگلیسی با عناوین EL، LT، AK، RS، SS و ATT و همچنین سیستم FTTS برای زبان فارسی استفاده کرده‌ایم [۳۳]. نتایج این مقایسه در جدول ۱۰ آمده است. لازم به ذکر است که در این جدول سیستم‌های پیشنهادی مبتنی بر واج‌گونه با KA، مبتنی بر هجا با KS و مبتنی بر دایفون هم با نماد KD نمایش داده شده است.

همان‌گونه که ملاحظه می‌شود سیستم‌های طراحی‌شده برای زبان کردی به نسبت دیگر زبان‌ها دارای موقعیت متفاوتی هستند. اگر بخواهیم سیستم‌های مطرح‌شده را به سه دسته تقسیم کنیم، سیستم مبتنی بر واج‌گونه در دسته اول و با کم‌ترین امتیاز قرار می‌گیرد؛ سیستم مبتنی بر هجا در دسته دوم و سیستم مبتنی بر دایفون به همراه سه سیستم SS، ATT و FTTS در دسته سوم و با بالاترین کیفیت قرار خواهد گرفت.

۵- جمع‌بندی و کارهای آینده

در این مقاله طراحی و پیاده‌سازی سیستم تبدیل متن به گفتار برای زبان کردی بررسی شد. همان‌گونه که بیان شد، از قبل کار علمی قابل ملاحظه‌ای در این زمینه صورت نگرفته بود و این مقاله اولین مورد در نوع خود برای زبان کردی است. ما در این تحقیق دو بخش تحلیل متن و تولید گفتار را به‌صورت جداگانه مورد بررسی قرار داده و آنها را برای زبان کردی پیاده‌سازی کردیم.

در بخش تحلیل متن علاوه بر ابهام‌زدایی از متن، برای اولین بار منحنی تغییرات گام برای انواع مختلف جمله‌ها هم استخراج گردید. یکی از مشکلات در زبان کردی وجود دو رسم‌الخط استاندارد برای این زبان بود که برای حل این مشکل، ما نمادهای استانداردی را تعریف کرده و از آنها به‌عنوان یک رسم‌الخط میانی استفاده کردیم. همچنین هر کدام از این رسم‌الخط‌ها هم دارای مشکلات خاص خود بودند؛ از این رو می‌بایست مشکلات هر کدام را جداگانه بررسی کرده و حل کنیم.

در مرحله تولید گفتار، ما سه سیستم تبدیل متن به گفتار مختلف را

Archive of SID
www.SID.ir

بر مبنای واج‌گونه، هجا و دایفون پیاده‌سازی کردیم. هر کدام از این سیستم‌ها دارای ویژگی‌های مختص به خود بوده و کیفیت سیگنال خروجی آنها نیز متفاوت است. در واقع هدف ما از طراحی این سیستم‌ها، انجام مقایسه بین آنها بود. در این میان زمان‌برترین قسمت کار ساختن پایگاه داده برای هر کدام از این سیستم‌ها بود.

ما برای تشخیص واج‌گونه‌ها در متن ورودی، از یک شبکه عصبی سه‌لایه کمک گرفتیم. دقت نهایی شبکه عصبی بسیار خوب و برابر ۹۸٪ بود که قابل قبول می‌باشد. به‌علاوه در مواردی هم که شبکه اشتباه حدس می‌زد، باز واج‌گونه خروجی بسیار نزدیک به واج‌گونه هدف تشخیص داده می‌شد که باعث می‌شود نتایج بسیار قابل قبول‌تر باشد. برای سیستم مبتنی بر هجا یک پایگاه داده در مقیاس آزمایشگاهی و برای سیستم مبتنی بر دایفون پایگاه داده کامل با صدای زن و مرد طراحی شد.

برای سنجش کیفیت سیستم‌های طراحی‌شده چهار آزمون مختلف برگزار گردید. در اولین آزمون که آزمون استاندارد MOS می‌باشد، بیشتر جنبه‌های طبیعی‌بودن سیستم‌ها را مورد ارزیابی قرار دادیم. سه آزمون دیگر بیشتر جنبه قابلیت فهم سیستم‌ها را ارزیابی می‌کنند. در آزمون اول نرخ تشخیص درست کلمات و نرخ تشخیص درست هجاها را مورد مقایسه قرار داده‌ایم. دو آزمون دیگر، آزمون‌های استاندارد DRT و MRT هستند که البته فقط سیستم‌های مبتنی بر واج‌گونه و مبتنی بر دایفون را ارزیابی می‌کنند. همان‌طور که نتایج این سه آزمون نشان می‌دهد قابلیت فهم سیستم‌ها خوب بوده و به‌ویژه در این میان سیستم مبتنی بر دایفون قابلیت فهم بسیار خوبی را داراست.

در بخش اطلاعات نوایی، علاوه بر منحنی تغییرات گام، دو مؤلفه تکیه و دیرش هم مطرح هستند که با توجه به پیچیدگی زیاد آنها و به‌ویژه مقوله تکیه، ما در این مرحله از آنها صرف نظر کرده و آن را به آینده موکول کرده‌ایم. همچنین برای ساخت سیستم‌های تبدیل متن به گفتار روش‌های دیگری هم پیشنهاد شده است که می‌توانند در کارهای آینده مورد توجه قرار گیرند. روش انتخاب واحد یکی از روش‌هایی است که در دهه اخیر بسیار مورد توجه قرار گرفته است. مهم‌ترین چالش این روش پایگاه داده بسیار بزرگ آن است که کار طراحی و ساخت آن بسیار زمان‌بر و طاقت‌فرساست. با توجه به مزایای زیاد این روش، می‌توان در ادامه از آن استفاده کرد. همچنین در راستای ساخت سیستم‌های سنتز ابزارهای جدیدی به‌کار گرفته شده‌اند که بازدهی بسیار خوبی از خود نشان داده‌اند. مدل مخفی مارکوف (HMM) یکی از این ابزارهاست که به خصوص از سال ۲۰۰۷ به بعد بسیار مورد توجه بوده است و در بسیاری از تحقیقات چند سال اخیر مورد استفاده واقع شده است.

مراجع

[1] E. S. Rawski, *The Last Emperors: a Social History of Qing Imperial Institutions*, Berkeley and Los Angeles: University of California Press, ISBN 0520212894, 1998.

[2] A. Black, CHATR Version 0.8: *A Generic Speech Synthesis, System Documentation*, ATR-Interpreting Telecommunications Laboratories, Kyoto, Japan, 1996.

[3] A. Hunt and A. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP 96*, vol. 1, pp. 373-376, Atlanta, Georgia, 7-10 May 1996.

[4] M. Beutnagel, A. Conkie, J. Schroeter, Y. Stylianou, and A. Syrdal, "The AT&T NEXT-GEN TTS System," *Joint Meeting of ASA, EAA, and DAGA*, 1999.

[5] T. Dutoit, *High Quality Text - to - Speech Synthesis of the French Language*, Ph.D. Dissertation, the Faculté Polytechnique de Mons, 1993.

[6] T. Dutoit, F. Bataille, V. Pagel, N. Pierret and O. van der Vreken, "The MBROLA project: towards a set of high quality speech synthesizers free of use of non commercial purposes," in *Proc. Fourth Int. Conf. on Spoken Language Processing*, vol. 3, pp. 1393-1396, Philadelphia, US, 3-6 Oct. 1996.

[7] W. Hamza, *Arabic Speech Synthesis Using Large Speech Database*, Ph.D. Thesis, Cairo University, Electronics and Communications Engineering Department, 2000.

[8] A. Youssef and O. Emam, "An Arabic TTS system based on the IBM trainable speech synthesizer," Le traitement automatique de l arabe, JEP TALN 2004, 2004.

[9] F. Chouireb, M. Guerti, M. Naïl, and Y. Dimeh, "Development of a prosodic database for standard Arabic," *Arabian J. for Science and Engineering*, vol. 32, no. 2B, pp. 251-262, Oct. 2007.

[10] A. Ramsay and H. Mansour, "Towards including prosody in a text-to-speech system for modern standard Arabic," *Computer Speech and Language*, vol. 22, no. 1, pp. 84-103, Jan. 2008.

[11] H. Al-Muhtaseb, M. Elshafei, and M. Al-Ghamdi, "Techniques for High Quality Arabic Speech Synthesis," Information Sciences, vol. 140, pp. 255-267, 2002.

[12] I. Amdal and T. Svendsen, "A speech synthesis corpus for Norwegian," in *Proc. Fifth Int. Conf. on Language Resources and Evaluation (LREC'06)*, pp. 1373-1376, Genova, Italy, 2006.

[13] K. Yoon, "A prosodic phrasing model for a Korean text-to-speech synthesis system," *Computer Speech & Language*, vol. 20, no. 1, pp. 69-79, Jan. 2006.

[14] P. Zervas, I. Potamitis, N. Fakotakis, and G. Kokkinakis, "A Greek TTS based on non uniform unit concatenation and the utilization of festival architecture," in *Proc. First Balkan Conf. on Informatics*, pp. 662-668, Thessalonica, Greece, 21-23 Nov. 2003.

[15] A. Farrokhi, S. Ghaemmaghami, and M. Sheikhan, "Estimation of prosodic information for Persian text-to-speech system using a recurrent neural network," in *Proc. Speech Prosody 2004*, Nara, Japan, 23-26 Mar. 2004.

[16] H. R. Abutalebi and M. Bijankhan, "Implementation of a text-to - speech system for Farsi language," in *Proc. Sixth Int. Conf. on Spoken Language Processing*, vol. 1, pp. 661-664, Beijing, China, Oct. 2000.

[17] F. Hendessi, A. Ghayoori, and T. A. Gulliver, "A speech synthesizer for Persian text using a neural network with a smooth ergodic HMM," *ACM Trans. on Asian Language Information Processing (TALIP)*, vol. 4, no. 1, pp. 38-52, Mar. 2005.

[18] A. Koochari, M. Namnabat, S. M. Kasaeiyan, and A. Niazade, "Duration modeling for Persian text-to-speech system by neural network," in *Proc. Int. Conf. on Multidisciplinary Information Sciences & Technologies, InSciT2006*, Mirida, Spain, 25-28 Oct. 2006.

[19] M. Namnabat and A. Koochari, "Generating F0 contours for speech synthesis in Persian language using classification and regression tree," in *Proc. 12th Int. Computer Society of Iran Computer Conf., CSICC'07*, Tehran, Iran, 20-22 Feb. 2007.

[20] M. M. Homayounpour and M. Namnabat, "FarsBayan: a unit selection based Farsi speech synthesizer," in *Proc. Nineth Int. Conf. on Spoken Language Processing, InterSpeech 2006-ICSLP*, Pittsburgh, US, 17-21 Sep. 2006.

[21] M. Namnabat and M. M. Homayounpour, "A letter to sound system for Farsi Language using neural networks," in *Proc. Int. Conf. on Signal Processing, ICSP2006*, vol. 1, Beijing, China, 16-20 Nov. 2006.

[22] S. Baban, *Phonology and Syllabication in Kurdish Language*, Kurdish Academy Press, First Edition, Arbil, 2005. (in Kurdish)

[23] W. M. Thackston, *Sorani Kurdish: a Reference Grammar with Selected Reading*, Harvard: Iranian Studies at Harvard University, 2006.

[۲۴] ع. رخزادی، *آواشناسی و دستور زبان کردی*، انتشارات ترفند، تهران، ۱۳۸۰.

[۲۵] م. کاوه، *زبان‌شناسی و دستور زبان کردی (لهجه سقزی)*، انتشارات احسان، ویرایش اول، تهران، ۱۳۸۵.

[۲۶] و. بارخدا، *طراحی و پیاده‌سازی سیستم تبدیل متن به گفتار در زبان کردی*، پایان‌نامه کارشناسی ارشد، گروه کامپیوتر و فناوری اطلاعات، دانشگاه کردستان، ۱۳۸۸.

[27] R. J. Deller Jr., J. G. Proakis, and J. H. Hansen, *Discrete Time Processing of Speech Signals*, John Wiley and Sons, 2000.

[28] F. Daneshfar, W. Barkhoda, and B. ZahirAzami, "Implementation of a Text-to-Speech System for Kurdish Language," in *Proc. Fourth Int. Conf. on Digital Telecommunications, ICDT'09*, pp. 117-120, Colmar, France, 20-25 Jul. 2009.

Archive of SID

www.SID.ir

[29] J. T. Sejnowski and R. Rosenberg, *Parallel Networks that Learn to Pronounce English Text*, the Johns Hopkins University, Complex Systems Inc, pp. 145-168, 1987.

[30] S. Lemmetty, *Review of Speech Synthesis Technology*, M. Sc Thesis, Helsinki University of Technology, 1999.

[31] M. N. Rao, S. Thomas, T. Nagarajan, and H. A. Murthy, "Text-to-speech synthesis using syllable-like units," in *Proc. of National Conf. on Communications*, pp. 277-280, IIT Kharagpur, India, Jan. 2005.

[32] W. Barkhoda, B. ZahirAzami, O. Shahryari, and A. Bahrampour, "A comparison between allophone, syllable, and diphone based TTS systems for Kurdish language," *IEEE Int. Symp. on Signal Processing and Information Technology*, *ISSPIT'09*, pp. 557-562, Ajman, UAE, 14-17 Dec. 2009.

[۳۳] م. شیخان، م. نصیرزاده و ع. دفتریان، "طراحی و پیاده‌سازی سیستم تبدیل متن به گفتار طبیعی برای زبان فارسی،" *مجله علمی پژوهشی دانشکده مهندسی دانشگاه فردوسی مشهد*، سال ۱۷، شماره ۲، صص. ۴۸-۳۱، ۱۳۸۴.

وفا بارخدا مقطع کارشناسی خود را در سال ۱۳۸۶ و در دانشگاه تربیت معلم تهران به پایان رسانده و دوره کارشناسی ارشد خود را در سال ۱۳۸۸ در دانشگاه کردستان و با عنوان پایان‌نامه طراحی و پیاده‌سازی سیستم تبدیل متن به گفتار در زبان کردی به اتمام رساند. از وی تا کنون ۲۷ مقاله در کنفرانس‌ها و مجلات بین‌المللی به چاپ رسیده است. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: شناسایی افراد بر اساس ویژگی‌های بیومتریک، پردازش تصویر و بینایی ماشین، پردازش زبان طبیعی، منطق فازی و پردازش سیگنال.

انور بهرام‌پور دوره کارشناسی مهندسی کامپیوتر گرایش نرم‌افزار را در دانشگاه تربیت معلم تهران در سال ۱۳۸۲ و در سال ۱۳۸۴ در همان رشته-گرایش از دانشگاه علم و صنعت ایران در مقطع کارشناسی ارشد و با عنوان پایان‌نامه بروز رسانی تدریجی و برخط نرم‌افزارهای مبتنی بر مؤلفه فارغ‌التحصیل گردیده است. ایشان از سال ۱۳۸۵ تا کنون به عنوان عضو هیأت علمی تمام وقت در گروه مهندسی فناوری اطلاعات دانشگاه آزاد اسلامی واحد سنندج مشغول به خدمت می‌باشد. زمینه‌های مورد علاقه ایشان عبارتند از: سیستم‌های توزیع شده، پردازش زبانهای طبیعی، سیستم‌های تبدیل متن به گفتار و الگوریتم‌های تکاملی.

فردین اخلاقیان مدرک کارشناسی مهندسی الکترونیک خود را در سال ۱۳۶۹ از دانشگاه صنعتی اصفهان و مدرک کارشناسی ارشد مهندسی برق گرایش پردازش سیگنال و سیستم را در سال ۱۳۷۲ از دانشگاه تربیت مدرس تهران دریافت نمود. ایشان در سال‌های ۱۳۷۳ لغایت ۱۳۷۹ در گروه کامپیوتر دانشگاه کردستان بعنوان عضو هیأت علمی فعالیت نموده و در سال ۱۳۸۰ دوره دکترای خود را در دانشگاه ولنگنگ استرالیا شروع نمود. نامبرده مدرک دکترای خود را در سال ۱۳۸۴ اخذ نموده و از آن تاریخ تاکنون بعنوان استادیار گروه کامپیوتر دانشگاه کردستان فعالیت می‌نماید. زمینه‌های علمی مورد علاقه وی شامل شناخت الگو، پردازش تصویر و ماشین-بینایی و پردازش‌های هوشمند می‌باشد.

هشام فیلی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد نرم‌افزار و دکتری هوش مصنوعی را بهمترتیب در سالهای ۱۳۷۶ و ۱۳۷۸ و ۱۳۸۵ در دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف پایان رسانده است و از سال ۱۳۸۷ تا کنون عضو هیأت علمی دانشکده مهندسی برق و کامپیوتر دانشکده فنی دانشگاه تهران است. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش هوشمند متن و گفتار، داده‌کاوی، بازیابی اطلاعات و شبکه‌های اجتماعی.

Archive of SID

www.SID.ir