مراحل و نحوه‌ی تهیه‌ی دادگان‌های صوتی هجایی و دایفونی برای سامانه‌ی تبدیل متن به گفتار فارسی

محرم اسلامی[1]، جواد شیخ‌زادگان[2]، زهرا احمدی‌نیا[3]، علی بهرامی راد[4]
[1]دانشگاه زنجان، [2]پژوهشکده‌ی پردازش هوشمند علائم و دانشکده‌ی امام باقر (ع)
[3]مؤسسه‌ی فرهنگ معاصر، [4]دانشگاه صنعتی تامپره-فنلاند

چکیده:

دادگان‌های صوتی بخشی از سامانه‌های تبدیل متن به گفتار به شیوه‌ی هم‌گذاری را تشکیل می‌دهند. کیفیت دادگان‌ها و نحوه‌ی تهیه و آماده‌سازی آن‌ها نقش مهمی در طبیعی‌تر شدن گفتار بازسازی‌شده ایفا می‌کند. در این مقاله ابتدا به معرفی دو دادگان صوتی هجایی و دایفونی فارسی پرداخته و سپس ضمن برشمردن ویژگی‌های هر کدام از آن‌ها، مراحل تهیه و آماده‌سازی آن‌ها را توضیح می‌دهیم. درنهایت با به‌کارگیری این دادگان‌ها در موتور تبدیل متن به گفتار فارسی "گویا" و نتایج حاصل از انجام آزمون‌های مختلف، تفاوت‌ها و برتری‌های آن‌ها را نسبت به یکدیگر بیان می‌کنیم.

واژگان کلیدی: دادگان‌های صوتی، دادگان هجایی، دادگان دایفونی و سامانه‌ی تبدیل متن به گفتار فارسی

۱- مقدمه

در این مقاله ضمن بیان ضرورت و اهمّیت دادگان‌های صوتی در سامانه‌های تبدیل متن به گفتار، مراحل تهیه و آماده‌سازی دو دادگان صوتی هجایی و دایفونی فارسی را مورد بررسی قرار داده و به بیان ویژگی‌های فنی و مزایا و معایب هر کدام از آن‌ها می‌پردازیم. این دادگان‌ها دربردارنده‌ی واحدهای اولیه‌ی آوایی هستند و در موتور تبدیل متن به گفتار (Text-to-Speech) فارسی *(گویا*[1]) که با هم‌گذاری (concatenation) واحدهای آوایی، متن فارسی را به گفتار تبدیل می‌کند، کاربرد دارند. تشخیص جایگاه مناسب واحدهای آوایی در گفتار طبیعی جهت ایجاد دادگان‌های صوتی و نیز چگونگی ضبط و تقطیع واحدهای آوایی تأثیر به‌سزایی در کیفیت گفتار بازسازی‌شده دارد که در تهیه و آماده‌سازی دادگان‌های صوتی هجایی و دایفونی فارسی به این مهم توجّه شده است. در ادامه با شرح مختصری در مورد برنامه‌های رایانه‌ای بازسازی گفتار (Speech synthesis) به معرّفی دو دادگان صوتی هجایی و دایفونی فارسی مورد استفاده در برنامه‌ی بازسازی گفتار فارسی می‌پردازیم. قابل ذکر است که موتور تبدیل متن به گفتار فارسی "گویا" از سه بخش عمده‌ی "پردازش متن"، "بازسازی گفتار" و "دادگان صوتی" تشکیل یافته است که عمل و کیفیت عمل بخش "بازسازی گفتار" به‌طور کامل وابسته به بخش "دادگان صوتی" است.

برنامه‌های رایانه‌ای بازسازی گفتار با این هدف طراحی می‌شوند که بتوانند نوشتار را به گفتار تبدیل کنند. نزدیک‌تر کردن گفتار بازسازی‌شده به گفتار طبیعی آرمان و هدف پژوهش‌گران در بازسازی گفتار به حساب می‌آید. برنامه‌های تولید گفتار (Speech production) به دو گروه تعاملی (interactive) و غیرتعاملی (non-interactive) تقسیم می‌شوند و میزان رابطه‌ی کاربر با برنامه‌ی تولید گفتار نوع آن را مشخّص می‌کند. اگر کاربر با برنامه‌ی تولید گفتار در تعامل و ارتباط باشد و انتظار داشته باشد که برنامه، هر نوع نوشته‌ی او را به گفتار تبدیل کند، در آن صورت این نوع

۱. "گویا" نام تجاری موتور تبدیل متن به گفتار فارسی است که در سال‌های گذشته توسط پژوهشکدهٔ پردازش هوشمند علائم طراحی و پیاده‌سازی شده است.

برنامه را برنامه‌ی " تولید گفتار تعاملی " می‌نامند. از طرف دیگر اگر هیچ نوع رابطه‌ای بین کاربر و برنامه‌ی تولید گفتار وجود نداشته باشد و برنامه تنها مجموعه جملات و عبارات کلیشه‌ای را تولید کند، به آن برنامهٔ " غیرتعاملی " می‌گویند. بازسازی گفتار از نوع برنامه‌های تعاملی تولید گفتار است و تولید خودکار جمله‌های نو، آن را از دیگر برنامه‌های تولید گفتار متمایز می‌کند.

بازسازی گفتار به دو روش مختلف هم‌گذاری (concatenation) و پارامتری (parametric) می‌تواند صورت گیرد. در روش هم‌گذاری برابرهای آوایی متن ورودی برنامه که واحدهای بازسازی محسوب می‌شوند، با قواعد و اصول خاصی از دادگان صوتی برنامه‌ی بازسازی‌کننده‌ی گفتار احضار و کنار هم چیده می‌شوند و شرایط برای تبدیل به گفتار فراهم می‌شود. در تهیه واحدهای بازسازی، ابتدا پیکره‌ی (Corpus) زبانیِ لازم تهیه می‌شود و سپس پیکره‌ی زبانی مورد نظر از سوی یک یا چند نفر اهل زبان در یک محیط آزمایشگاهی خوانده شده و ضبط می‌شود. بعد از انجام پردازش‌های لازم واحدهای بازسازی از پیکره استخراج و در قالب یک دادگان صوتی در چارچوب برنامه‌ی بازسازی گنجانده می‌شود. در روش پارامتری بازسازی گفتار که به آن روش مبتنی بر قاعده نیز گفته می‌شود، همه‌ی پارامترهای انتزاعی گفتار مانند سایش (frication)، سازه (formant)ها و دامنه‌ی نوسان (amplitude) صداها کنار گذاشته می‌شود و برنامه این پارامترها را در چارچوب قواعد خاص بازسازی می‌کند. در این روش (مانند بازسازی‌کننده‌ی Klatt) ابتدا واحد واجی به چند مشخصه‌ی صوت‌شناختی تقسیم می‌شود و هر کدام از این واحدهای صوت‌شناختی براساس توالی علائم کنترلی (sequence of control signals) بازتولید می‌شوند. در این روش ویژگی‌های آوایی مربوط به هم‌تولیدی و (Coarticulation) گذر از صدایی به صدای دیگر لحاظ می‌شود.

در تقسیم‌بندی دیگر، برنامه‌های بازسازی گفتار را می‌توان به دو گروه عام (unrestricted) و خاص (restricted) تقسیم کرد. بازسازی‌کننده‌های عام با هدف تبدیل هر نوع نوشته‌ای به گفتار طراحی می‌شوند، ولی بازسازی‌کننده‌های خاص، نوشته‌های تخصصی را با هدف محدودتر بازسازی می‌کنند و دارای ساختار ساده‌تر نسبت به بازسازی‌کننده‌های عام هستند. از طرف دیگر برنامه‌ی بازسازی‌کننده می‌تواند یک یا بیش از یک‌زبانه باشد. بازسازی‌کننده‌های بیش از یک زبانه برخلاف برنامه‌های یک‌زبانه می‌توانند به‌طور هم‌زمان نوشته‌های مربوط به چند زبان را در درون نوشته‌ی واحد به گفتار تبدیل کنند *(اسلامی و همکاران ۱۳۸۳)*.

در روش هم‌گذاری، بازسازی گفتار از واحدهای زبانی متفاوتی جهت بازسازی استفاده می‌شود که در زیر به توصیف مختصر هر کدام از آن‌ها می‌پردازیم.در بازسازی‌کننده‌های با کاربرد خاص، از جملات و عبارت‌ها می‌توان به‌عنوان واحدهای بازسازی استفاده کرد که این‌گونه برنامه‌ها مناسب محل‌هایی مانند پایانه‌ها، فرودگاه‌ها، بازارهای بورس و اعلام وضع هوا و غیره است که در آن پیام‌های زبانی به‌صورت تقریباً کلیشه‌ای تولید می‌شوند و ارتباط کاربر با برنامه بازسازی‌کننده بسیار محدود است. کلمه واحد زبانی کوچک‌تر از جمله و عبارت است که می‌تواند به‌عنوان واحد بازسازی گفتار مورد استفاده قرار گیرد *(بوران ۱۹۶۸، چپمن ۱۹۷۱، ایدی و همکاران ۱۹۸۷)*. در طراحی برنامه‌ای جامع برای بازسازی گفتار، کلمه واحد مناسبی نیست؛ چراکه احصا و نیز ذخیره‌کردن اطلاعات دیجیتالی همه‌ی کلمه‌های زبان در عمل غیرممکن است. به همین دلیل پژوهش‌گران در بازسازی گفتار از واحدهای کوچک‌تری استفاده می‌کنند که هم اقتصادی است و هم می‌توان به لحاظ آوایی در استفاده‌ی از آن‌ها به تعمیم‌هایی دست پیدا کرد و کیفیت گفتار بازسازی‌شده را بهبود بخشید. این واحدهای کوچک‌تر عبارتند از: واج(Phoneme)، دایاد (dyad)، دایفون (diphone)، نیم‌هجا (demisyllable) و هجا (Syllable)، که در ادامه به شرح مختصر هرکدام می‌پردازیم.

واج به‌عنوان یک واحد انتزاعی ممیز معنا در زبان به خاطر داشتن فهرست محدود به‌ظاهر مناسب‌ترین واحد بازسازی است؛ ولی همه‌ی تلاش‌ها در هم‌گذاری واج‌ها در بازسازی گفتار با شکست مواجه شده است و علت اصلی این شکست از این واقعیت ناشی می‌شود که مرز و کیفیت آوایی واج‌ها در زنجیره‌ی گفتار با توجّه به بافت آوایی به‌لحاظ صوت‌شناختی متغیر است. واقعیت هم‌تولیدی بین صداهای مجاور در گفتار و واقعیت عروضی (Prosody) در مفهوم فرثی (Firthian) بین صداها باعث می‌شود که استفاده از واج به‌عنوان واحد بازسازی گفتار با موفقیت همراه نباشد. حتی استفاده از واج‌گونه (allophone)ها نیز نمی‌تواند به رفع این مشکل کمک کند، چرا که واج‌گونه‌ها هم مانند واج‌ها فاقد بسیاری از ویژگی‌های هم‌تولیدی هستند. بنابراین گفتار بازسازی‌شده با استفاده از واج‌ها و حتی واج‌گونه‌ها به زحمت قابل درک خواهد بود (سیورتسن ۱۹۶۱). بنابراین پژوهش‌گران به واحدهای دیگری روی آوردند و در انتخاب

واحدها دو ملاک را مدّ نظر دادند؛ اوّل این‌که این واحدها قابل پردازش با رایانه باشند و به‌گونه‌ای انتخاب شوند که استفاده از آن‌ها در رایانه امکان‌پذیر باشد و دوم این‌که این واحدها حائز شرایط هم‌تولیدی و گذر از صدایی به صدای دیگر باشند. داياد، دایفون، نیم‌هجا و هجا حائز شرایط فوق هستند. داياد از بخش پایانی یک واحد (واج) و از بخش آغازی واحد بعدی تشکیل می‌شود. اصطلاح دایفون معادلی برای اصطلاح داياد است؛ اگر چه در مفهوم دقیق‌تر، این دو متفاوت از هم هستند *(کلت ۱۹۸۷)*. در اصل دایفون ناظر بر ناحیه گذر بین صداهای مجاور است، ولی داياد بخش پایدار (steady state) صدا را نیز شامل می‌شود. نیم‌هجا بیشتر شبیه به دایفون است، با این تفاوت که در هجای CVC مرز دایفون وسط واکه است، درحالی‌که در نیم‌هجا پایان ناحیه‌ی گذر و آغاز بخش پایدار واکه، مرز تلقی می‌شود. ترایفون (triphone) واحد دیگری در بازسازی گفتار است که در آن دو گذر واکه به همخوان و همخوان به واکه و یا بالعکس دیده می‌شود (به‌عنوان مثال VCV یا CVC). هجا یکی از واحدهای مهم در بازسازی گفتار به شیوه هم‌گذاری است که در آن فهرست هجاهای بالفعل زبان جمع‌آوری می‌شود و بعد از فراهم کردن پیکرهٔ گفتاری لازم، برابرهای آوایی هر کدام از هجاها از پیکره‌ی استخراج و سپس در دادگان صوتی برنامه قرار می‌گیرد.

برنامه‌ی بازسازی گفتار فارسی در موتور تبدیل متن به گفتار "گویا"، از نوع بازسازی‌کننده‌های عام به شیوه‌ی هم‌گذاری است و از هجا و دایفون به‌عنوان واحد بازسازی استفاده می‌کند که در ادامه به بیان ویژگی‌ها و نحوهٔ تهیهٔ هر کدام از دادگان‌های هجایی و دایفونی و نیز برتری آن‌ها بر هم‌دیگر می‌پردازیم.

این مقاله با احتساب مقدمه به‌عنوان بخش اوّل، از چهار بخش تشکیل شده است. بخش دوم به بیان ویژگی‌های دادگان هجایی اختصاص دارد. در بخش سوم ویژگی‌های دادگان دایفونی را شرح می‌دهیم و در بخش چهارم نتیجه‌گیری آمده است.

۲- دادگان هجایی

وجود ساختمان هجایی ساده و آسان بودن تشخیص مرز هجاها در زبانی مانند زبان فارسی می‌تواند دو عامل تعیین‌کننده در انتخاب هجا به‌عنوان واحد بازسازی گفتار باشد.

۲- ۱- هجا و انواع آن در زبان فارسی

هجا در زبان فارسی عبارت از یک رشته‌ی آوایی پیوسته است که از یک واکه و یک تا سه همخوان تشکیل می‌شود (ثمره ۱۳۷۸). منظور از " رشته‌ی آوایی پیوسته " آن است که اجزای سازنده‌ی هجا طی یک فرآیند تولید بدون مکث تولید می‌شوند. واکه به منزل‌ی مرکز یا هسته یا محور هجاست و همخوان در حکم حاشیه یا دامنه آن است.

در فارسی سه نوع هجا وجود دارد که با قرار دادن C و V به ترتیب به جای همخوان و واکه می‌توان این سه نوع را به صورت CV، CVC و CVCC نشان داد؛ به‌عنوان مثال در سه کلمه‌ی /***gu***/ ، /***gu.***/ و /***gu.t***/[1].

۲- ۲- تعیین مرزهای هجایی در رشته‌های آوایی

از آن‌جا که در گفتار فارسی رسمی، در یک هجا بیش از یک واکه نمی‌تواند وجود داشته باشد، از این‌رو، تعداد هجاها در هر رشته‌ی آوایی می‌تواند با شمارش واکه‌ها مشخص شود. امّا تعیین مرز بین دو هجا بستگی به شمارش همخوان‌های بین دو واکه دارد. حداقل تعداد همخوان بین دو واکه یک و حداکثر سه همخوان است. به این ترتیب آرایش همخوانی بین هر دو واکه در هر رشته‌ی آوایی به یکی از سه شکل زیر خواهد بود:

الف) VCV ب) VCCV ج) VCCCV

مرز هجا در مورد اوّل بین V اول و C است، چون هجا در فارسی می‌تواند با واکه پایان پذیرد، اما نمی‌تواند با واکه آغاز شود. مثال این مورد، رشته آوایی //***dav***/ می‌باشد که به ترتیب گفته شده دو هجای /***da***/ و //***v***/ به دست می‌آید.

در مورد دوم، محل برش هجایی بین دو C خواهد بود زیرا اوّل این‌که واکه در آغاز هجا قرار نمی‌گیرد و دوم این‌که خوشه دو همخوانی در آغاز هجا مجاز نیست. نمونه‌ی این مورد، رشته‌ی آوایی /***doxtar***/ است که در آن دو هجای/***dox***/ و /***tar***/ به دست می‌آید. در مورد سوم، مرز دو هجا با دلایل ذکر شده در بالا، بین C دوم و سوم است. این نقطه برش، رشته‌ی آوایی مذکور را به دوهجای CVCC و CV تقسیم می‌کند. نمونه‌ی این مورد، رشته‌ی آوایی /***angu***/ است که در آن سه همخوان در میان دو واکه

۱- در ضمائم مقاله (ضمیمه‌ی ب)، فهرست علائم قراردادی مورد استفاده برای واج‌های فارسی در این مقاله، آورده شده است.

واقع شده‌اند. در اینجا محل برش هجایی بین /ٔ/ و/g/ می‌باشد و در نتیجه دو هجای /u/ و /ang/ به‌دست می‌آید.

با توجّه به اطّلاعات فوق می‌توان گفت در هر رشته‌ی آوایی، همخوان قبل از هر واکه بیان‌گر ابتدای هجا است. بر مبنای این قانون می‌توان مرزهای هجایی را در هر رشته آوایی به‌آسانی و بدون کم‌ترین شک و ابهامی مشخص کرد . از این امر در تقطیع سیگنال‌های صوتی برای استخراج معادل صوتی هجاهای فارسی استفاده شده است.

۲-۳- نحوه‌ی تعیین هجاهای دادگان هجایی

واجگان زبان فارسی از ۶ واکه و ۲۳ همخوان تشکیل شده است و بدون در نظر گرفتن محدودیت‌های همنشینی واج‌ها در درون هجا، تعداد بالقوه هجاها در زبان فارسی ۷۶۳۱۴ خواهد بود (جهت اطّلاع از نحوه‌ی محاسبه‌ی تعداد بالقوه‌ی هجاها به ضمیمه‌ی الف-۱ مراجعه شود). ولی از این تعداد، کمتر از ۶۰۰۰ هجا فعلیت می‌یابند و بقیه‌ی ساخت‌های هجایی غیرممکن و یا دارای خلاء هستند. لازم به ذکر است که در دادگان هجایی موجود، واکه‌ی مرکّب /ou/ را به‌صورت /ow/ یعنی ترکیبی از یک واکه و همخوان بازنمایی کرده‌ایم. بنابراین کلمه‌ی " جوهر " به‌صورت /owhar/ و نه /ohar/ نمایش داده می‌شود. این مسأله کار شناخت مرز هجاها را نیز راحت‌تر می‌کند. بنابراین تعداد همخوان‌ها در دادگان صوتی هجاهای فارسی (با احتساب /ow/) ۲۴ همخوان است. *[پژوهشکده‌ی پردازش هوشمند علائم، ۱۳۸۴].*

بخش اعظم هجاها از صورت واج‌نویسی‌شده‌ی دادگان بزرگ زبان فارسی گفتاری پژوهشکده‌ی پردازش هوشمند علائم *(پژوهشکده‌ی پردازش هوشمند۱۳۸۱)*، "فارس‌دات بزرگ"، به‌دست آمده است و بخش دیگر با مراجعه به شش جلد فرهنگ فارسی معین و استخراج کلمه‌های تک‌هجایی، جمع‌آوری و به فهرست هجاها اضافه شده‌اند.

۲-۴- شیوه‌ی تعیین جایگاه مناسب استخراج هجاها

جایگاه برابرهای آواییِ هجاها در زنجیره‌ی گفتار و در درون کلمه، حائز اهمیت فراوان است و تأثیر چشم‌گیری در کیفیت گفتار بازسازی شده دارد. برای این منظور جهت تهیه دادگان صوتی هجاهای فارسی، ما ابتدا به‌طور آزمایشی پنج جایگاه مختلف را برای هجاها در گفتار فارسی درنظر گرفتیم و هجاهای یکسانی را در جایگاه‌های پنج‌گانه قرار دادیم و پس از تولید داده‌ها (به‌صورت کلمات گفتاری)، هجاهای مورد نظر را از داده‌های گفتاری استخراج و در پنج گروه جداگانه مرتب کردیم. سپس عبارت‌های یکسانی را با استفاده از هجاهای هر کدام از گروه‌ها تولید کردیم تا بتوانیم کیفیت بازسازی گفتار با استفاده از هر کدام از گروه‌های هجایی در محل‌های پنج‌گانه‌ی مذکور را ارزیابی کنیم. جایگاه هجاها به ترتیب عبارت بودند از:

۱ـ هجای بی‌تکیه از اول کلمهٔ مجزا مانند /pa/ در کلمهٔ "پذیرفته".

۲- هجای بی‌تکیه از اوّل کلمه‌ی دارای تکیه‌ی زیروبمی (pitch accent)؛ مانند /pa/ در جمله‌ی "پذیرفته شدم".

۳- هجای بی‌تکیه از اوّل کلمه‌ی فاقد تکیه‌ی زیروبمی؛ مانند/pa/ در جمله‌ی" پذیرفته‌ی خاص و عام شد ".

۴- هجای تکیه‌بر (stressed) فاقد تکیه‌ی زیروبمی؛ مانند /pe/ در کلمه‌ی /kupe/ در بافت " این کوپه کوچک است".

۵- هجای تکیه‌بر در کلمه مجزا، مانند /pe/ در کلمهٔ /kupe/ .

جمله‌های یکسانی با استفاده از هجاهای هر کدام از گروه‌ها بازسازی شد و بدون هیچ‌گونه پردازشی و تنها صورت هم‌گذاری شده‌ی آن‌ها در معرض قضاوت هفده نفر اهل زبان با تخصص و تحصیلات مختلف قرار گرفت و آن‌ها با دادن نمره‌ی ۱ الی ۵ به هر کدام از گونه‌های بازسازی‌شده نظر خودشان را درباره‌ی کیفیت آن‌ها اعلام کردند. از بین گروه‌های فوق به ترتیب گروه اوّل بیش‌ترین و گروه پنجم کم‌ترین امتیاز را از نظر کیفیت گفتار بازسازی شده به‌دست آوردند و بر پایه همین آزمایش، واحدهای دادگان ۶۰۰۰ هجایی مورد نظر همگی در آغاز کلمات، ولو کلمات بی‌معنی قرار داده شدند. بعد از تولید کلمات طراحی‌شده در یک محیط آزمایشگاهی (اتاقک آکوستیک) و پس از انجام فعالیت‌های لازم (تقطیع و برچسب‌دهی هجایی) برابرهای آوایی هجاها به‌دست آمد و در دادگان صوتی هجاها قرار گرفت.

۲-۵- نحوه‌ی ضبط کلمات حاوی عناصر دادگان و ملزومات سخت‌افزاری و نرم‌افزاری آن

در قدم بعدی، برای تهیه‌ی دادگان مورد نظر، حدود ۶۰۰۰ کلمه که البته بعضی از آن‌ها بی‌معنی هستند، توسط یک گویشور مذکر با لهجه‌ی فارسی تهرانی قرائت و از طریق کارت صوتی Creative Sound Blaster (Live) ضبط گردید. از آن‌جا که در تعدادی از روشهای بازسازی گفتار از پردازش‌های همزمان با گام (pitch) استفاده می‌شود *(ابوطالبی ۱۳۷۷)* و از لوازم این پردازش‌ها، معین بودن محل آغازه‌های دوره‌ی تناوب گام می‌باشد، از این‌رو ضبط سیگنال را به‌صورت دوکاناله انجام داده‌ایم که یکی از آن دو کانال، سیگنال گفتار و دیگری سیگنال حنجره (تکانه‌های تارآواها) است. سخت‌افزار مورد استفاده به‌منظور ضبط سیگنال عبارت است از:

- یک دستگاه رایانه با کارت صوتی Creative Sound Blaster (Live)
- یک میکروفون سونی کاردیوئید (Cardioid) پویا[1] با مقاومت ورودی ۶۰۰ اهم و پاسخ فرکانس ۱۶۰۰۰ـ۸۰ هرتز و سطح خروجی مؤثر ۶۱/۸dBm-
- دستگاه لارینژوگراف به‌همراه گردنبند حس‌گر: این دستگاه دارای حس‌گرهایی است که بر حنجره‌ی گوینده نصب و خروجی آن براساس میزان ارتعاش تارهای صوتی معین می‌گردد. برای ضبط سیگنال خروجی این دستگاه (سیگنال حنجره) آن را به یکی از کانال‌های ورودی کارت صوتی منتقل می‌کنیم.

نکات زیر در هنگام ضبط رعایت گردیده است:

- ضبط کلمات در یک اتاقک ضد صدا صورت گرفته است. عایق‌بندی صوتی اتاقک، از تأثیر نامطلوب صداهای ناخواسته یا منابع نویزی مانند لامپ فلورسنت، رایانه و غیره جلوگیری می‌کند.
- نرخ نمونه‌برداری در هر یک از کانال‌ها، ۲۲۰۵۰ نمونه بر ثانیه و هر یک از نمونه‌ها با ۱۶ بیت ذخیره شده‌اند.
- نرم‌افزار مورد استفاده برای ضبط و ویرایش سیگنال‌ها نرم‌افزار Wave Studio است.
- در هنگام ضبط سعی گردیده فاصله بین دهان گوینده تا میکروفون ثابت بماند و این فاصله حدود ۲۰ سانتی‌متر بوده است.

[1] Dynamic

۲- ۶- مسئله‌ی هم‌زمانی سیگنال گفتار با سیگنال حنجره

سیگنال گفتار در مقایسه با سیگنال حنجره، تأخیر جزئی دارد و دلیل این تأخیر نیز تفاوت فاصله‌ی حدود ۲۸ سانتی‌متری محل نصب حس‌گرهای اندازه‌گیری سیگنال حنجره از محل میکروفون (دریافت‌کننده سیگنال گفتار) می‌باشد. این فاصله با توجّه به سرعت صوت (۳۴۰ m/s) موجب تفاوت زمانی معادل ۰/۸۲۴ میلی‌ثانیه (و یا با توجه به فرکانس نمونه‌برداری ۲۲۰۵۰ هرتز، تأخیری معادل ۱۸ بایت) بین دو سیگنال می‌گردد. این مسأله برای استفاده از سیگنال تارآواها لحاظ شده است.

۲-۷- تقطیع کلمات و ایجاد دادگان هجاها

تقطیع پرونده‌های صوتی مربوط به هجاهای زبان فارسی توسط متخصصان این امر به‌صورت دستی انجام شده است و این کار در محیط نرم‌افزار Wave Studio صورت گرفته است. در تقطیع هجایی نکات زیر رعایت شده‌اند: الف) محل تقطیع به‌گونه‌ای است که عدم پیوستگی و اعوجاج طیفی بین قطعات متوالی به حداقل می‌رسد. برای کاهش میزان ناپیوستگی سعی شده که محل تقطیع در جاهایی باشد که دامنه‌ی شکل موج چندان زیاد نیست. کم بودن دامنه در مرزها تا حدودی، ناپیوستگی در محل‌های اتصال را کاهش می‌دهد. به‌طور خاص در مورد هجاهایی که با آواهای واکدار آغاز شده و یا به پایان می‌رسند، مرز هجاها در محل عبور سیگنال از خط صفر گذشته و دوره‌های تناوب گام به صورت کامل در محدوده هجا قرار می‌گیرد. به‌عبارتی، مرز در ابتدای اولین و انتهای آخرین دوره‌ی تناوب گام واقع گشته است. این نحوه‌ی تعیین ابتدا و انتها باعث می‌گردد هنگامی که در بازسازی گفتار، واج‌های مرزی، واکدار می‌باشند، مسئله‌ی عدم تطابق (phase mismatch) در سیگنال رخ ندهد و مشکلی از این لحاظ در کیفیت پیش نیاید. ب) در مورد هجاهایی که آوای انتهایی آن‌ها واکدار است، نقطه‌ای به‌عنوان پایان هجا در نظر گرفته شده که در آن زمان، در سیگنال ارتعاش تارآواها، حالت واک‌داری به پایان رسیده باشد.

۳- دادگان دایفونی

در این بخش سعی می‌کنیم به‌صورت مختصر مراحل مختلف فعالت‌های مربوط به تهیه‌ی دادگان صوتی دایفونی را شرح دهیم. این فعّالیت‌ها شامل طراحی، ضبط و تقطیع دادگان صوتی دایفونی می‌باشد. قبل از تشریح فعالیت‌های فوق لازم

است به‌طور مختصر دایفون و انواع آن و هم‌چنین تعداد دایفون‌های زبان فارسی را مورد بحث قرار دهیم.

۳- ۱- دایفون و انواع آن

واحد پایه‌ی مورد استفاده در دادگان صوتی برای بازسازی رایانه‌ای گفتار باید به‌گونه‌ای انتخاب شود که اول این‌که حجم حافظه معقولی را اشغال کند، یعنی تعداد واحدهای آوایی مطلوب باشد و دوم این‌که بتوان گذرهای آوایی را در دادگان پوشش داد. از جمله واحدهایی که با هدف تأمین این شرایط، تعریف و مورد استفاده قرار می‌گیرد، دایفون است. دایفون عبارت است از نیمه‌ی پایدار یک آوا تا نیمه‌ی پایدار آوای بعدی. دایفون از دو نیم‌آوای به هم چسبیده تشکیل می‌شود، البته باید درنظر داشت که این ترکیب شامل ترکیب سکوت و نیم‌آوا نیز می‌باشد. با این تعریف انواع دایفون‌های زبان فارسی شامل توالی‌های cc، cv، vc، c-، -c ، -v و v- است، که "-" به معنای سکوت می‌باشد. *[شیخ‌زادگان و همکاران، ۱۳۸۶]*.

۳-۲- تعداد دایفون‌ها در دادگان صوتی

در ابتدا تعداد دایفون‌ها ۹۰۴ دایفون برآورد می‌شد، که این تعداد از ترکیب ۷ واکه (با احتساب واکه‌ی مرکب /ow/) و ۲۳ همخوان به شکلی که در ضمیمه‌ی الف-۲-۱ آمده است، حاصل می‌شد.

امّا در عمل به این نتیجه رسیدیم که لازم است برای واج‌های /k/ و/g/ واج‌گونه‌های آن‌ها یعنی [c] و [ɟ] را نیز لحاظ کنیم تا به کیفیت بهتری برسیم. البته برای دیگر واج‌ها نیز این مطلب صادق است، لیکن یا تأثیر آن‌ها در کیفیت گفتار بازسازی‌شده قابل تشخیص نیست و یا کم است و از طرف دیگر برای حجیم نشدن دادگان از این کار در مورد واج‌های دیگر صرف‌نظر کردیم.

- /k/ در بافت قبل از واکه‌های /i/، /o/ یا /u/ ظاهر می‌شود و در سایر بافت‌ها [c] ظاهر می‌شود.
- /g/ در بافت قبل از واکه‌های /i/، /o/ یا /u/ ظاهر می‌شود و در سایر بافت‌ها [ɟ] ظاهر می‌شود.

با این توضیحات تعداد دایفون‌ها ۹۶۴ تا برآورد شدند که چگونگی به‌دست آمدن این عدد در ضمیمه‌ی الف-۲-۲، مشخص شده است.

۳-۳- تهیه و طراحی پیکره‌ی زبانی لازم جهت استخراج دایفون‌ها

برای طراحی پیکره‌ی زبانی لازم جهت استخراج دایفون‌ها از تک‌کلمه‌ها و کلمات مرکب استفاده شده است. به این منظور ابتدا با مراجعه به واژگان (lexicon) برنامه‌ی بازسازی گفتار فارسی *(گویا)* برای هر یک از دایفون‌ها، سه کلمه را که دارای بیشترین بسامد در واژگان برنامه بوده و شامل دایفون مورد نظر نیز باشند، استخراج کردیم. البته بدیهی است که برای برخی از دایفون‌ها کلمه‌ای پیدا نشد؛ درنتیجه برای آن‌ها از ترکیب کلمات یا حتّی از طراحی کلمات بی‌معنی استفاده شده است. سپس با آزمایش‌هایی که انجام شد و همین‌طور با تجربه‌ای که از قبل در مورد ساختار آوایی (هجایی) کلمات داشتیم، به اصولی برای طراحی دادگان دایفونی رسیدیم. هم‌چنین پس از تطبیق دادن این اصول با کلمات پربسامد استخراج شده از واژگان برنامه ناچار شدیم که برای بعضی از دایفون‌ها دوباره در واژگان به‌دنبال کلمه‌ای بگردیم که اصول مورد نظر را پوشش دهد. این اصول در زیر فهرست شده‌اند:

۱- تمام دایفون‌ها از هجاهای بی‌تکیه استخراج شده‌اند و تنها دایفون‌هایی که بعد از آن‌ها سکوت است، از این قاعده مستثنی هستند.

۲- هیچ یک از دایفون‌ها از داخل کلمات تک‌هجایی استخراج نشده‌اند.

۳- دایفون‌ها به‌گونه‌ای انتخاب شده‌اند که واج اول و دوم هر یک در دو هجای متفاوت قرار داشته باشند. البته اگر واج دوم یک واکه باشد در داخل همان هجای واج اول قرار می‌گیرد ولی در سایر موارد این قاعده صادق است. به‌عنوان مثال برای دایفون /n./ از بین دو کلمه "جشنواره" /a.nv/re/ و "پیشنهاد" /pi.nah/d/ کلمه "پیشنهاد" انتخاب شده است چون‌که در کلمه "پیشنهاد" همخوان‌های /n./ و /n/ در دو هجای متفاوت قرار دارند درحالی‌که در کلمه‌ی " جشنواره " این دو همخوان داخل یک هجا هستند.

۴- در خوشه‌های دوهمخوانی تا حدّ امکان سعی شده است که همخوان اول در داخل هجای cvc باشد و نه داخل هجای cvcc. به‌عنوان مثال برای دایفون /r./ از بین دو کلمه‌ی "فرشچیان" /far.iy/n/ و "خوش‌چهره" /xo.ehre/ کلمه‌ی " خوش‌چهره " انتخاب شده است، چون همان‌طور که مشخص است

در کلمه "فرشچیان" همخوان /ʃ/ داخل هجای cvcc یعنی /ʃar/ قرار دارد درحالی‌که در کلمه‌ی " خوش‌چهره " همخوان /x/ داخل یک هجای cvc یعنی /xo/ است. البته به‌دلیل محدودیت در بعضی از دایفون‌ها، از این قاعده تخطی شده است.

۵- سعی شده است دایفون‌های cv از کلماتی استخراج شوند که در آن کلمات، cv بعد از یک واکه ظاهر شده باشد یا به عبارت دیگر سعی شده است که دایفون‌های cv از هجایی استخراج شوند که هجای قبلی‌اش cv و هجای باز باشد. به‌عنوان مثال برای دایفون /b//، از بین دو کلمه‌ی "مبارزه" /mob/reze/ و "آذربایجان" /n,/zarb/ye/l/ کلمه "مبارزه" انتخاب شده است چونکه در کلمه "آذربایجان" قبل از دایفون /b//، واج /r/ قرار دارد که واکه نیست ولی در کلمه مبارزه قبل از دایفون /b//، واکه /o/ قرار دارد.

۶- علاوه‌بر موارد فوق که جنبه‌ی عمومی دارند، بعضی از ملاحظات خاص نیز صورت گرفته است. به‌عنوان مثال کلمات به‌گونه‌ای طراحی شده‌اند که اگر واج دوم دایفون، واکه‌ی /i/ باشد پس از آن، همخوان /y/ نیاید، یا اگر واج دوم واکه است پس از آن، همخوان /r/ واقع نشود و غیره. در کل برای همخوان‌های روان و لرزشی، از این قبیل ملاحظات، زیاد صورت گرفته است. هم‌چنین برای خوشه‌های دوهمخوانی که با واج /g/ شروع می‌شوند، در بعضی از موارد از قاعده‌ی شماره‌ی ۴ تخطی شده است و علّت، آن است که در این قبیل دایفون‌ها قبل از واج /g/ در بیشتر موارد واج /n/ قرار دارد، درنتیجه به جای این‌که /g/ را از داخل یک هجای cvc استخراج کنیم آن را از داخل cvcc استخراج کرده‌ایم که قبل از /g/ ، /n/ آمده باشد.

۳-۴- نحوه‌ی ضبط کلمات حاوی دایفون‌ها و ملزومات سخت‌افزاری و نرم‌افزاری آن

پس از طراحی پیکره‌ی زبانی لازم، مرحله‌ی ضبط داده‌ها فرامی‌رسد. در این مرحله از گویشوران مرد و زن خواسته شد تا در محیط آزمایشگاهی (اتاقک ضد صدا) پیکره‌ی زبانی را به‌منظور ضبط داده‌ها بخوانند. فرآیند ضبط به‌صورت دوکاناله صورت گرفته است که یکی از کانال‌ها، سیگنال گفتار و دیگری حاوی سیگنال حنجره است. سخت‌افزار مورد استفاده در ضبط سیگنال‌ها عبارت است از:

- یک دستگاه رایانه با کارت صوتی "Creative Sound – Blaster Audigy"
- یک میکروفون "VicBoss BM 707M" با مقاومت ورودی ۶۰۰ اهم
- دستگاه لارینژوگراف به همراه گردنبند حس‌گر

نکات زیر در هنگام ضبط رعایت شده است:

- برای حذف نوفه و صداهای ناخواسته، ضبط کلمات در اتاقک ضد صدا صورت گرفته است.
- نرخ نمونه‌برداری در هر یک از کانال‌ها و نرم‌افزار مورد استفاده برای ضبط سیگنال‌ها، مانند موارد قبل انتخاب شده‌اند.

۳-۵- طراحی اصول استخراج (تقطیع) دایفون‌ها

بعد از ضبط کلمات مرحله‌ی استخراج دایفون‌ها از داخل کلمات ضبط‌شده می‌باشد و برای این کار نیاز به اصول و قواعد خاصّی داشتیم که باید در حین کار مدّنظر قرار می‌گرفت. البته بخشی از این اصول و قواعد از قبل وجود داشت و بخشی در اثر آزمایش‌های گوناگون به‌دست آمده یا تکمیل شد. این اصول به شرح زیر هستند:

۱- دایفون‌هایی که با واکه شروع می‌شدند یا به واکه ختم می‌شدند، از وسط منطقه‌ی پایدار واکه برش زده‌ایم.

۲- دایفون‌هایی که با همخوان سایشی شروع می‌شدند یا به همخوان سایشی ختم می‌شدند، از وسط همخوان برش زده‌ایم.

۳- دایفون‌هایی که با همخوان انفجاری واکدار شروع می‌شدند، قسمت رهش (همراه با سایش) همخوان آغازین دایفون را انتخاب کرده‌ایم.

۴- دایفون‌هایی که با همخوان انفجاری بی‌واک شروع می‌شدند، قسمت رهش (همراه با سایش) و دمش همخوان آغازین دایفون را انتخاب کرده‌ایم.

۵- دایفون‌هایی که به همخوان انفجاری ختم می‌شدند، قسمت بست (سکوت) همخوان را انتخاب کرده‌ایم.

۶- در تمام مراحل فوق، دایفون شامل قسمت گذر بین دو آوای سازنده‌ی خود است.

۷- برش سیگنال‌ها همواره از نقطه‌ی عبور از صفرشان صورت گرفته است.

۴- نتیجه‌گیری

در این مقاله ضمن بیان ضرورت و اهمّیت دادگان‌های صوتی در پیاده‌سازی سامانه‌های تبدیل متن به گفتار، مراحل تهیه و آماده‌سازی دو دادگان صوتی هجایی و دایفونی فارسی را مورد بررسی قرار دادیم و به بیان ویژگی‌های فنّی هر کدام از آن‌ها پرداختیم. این دادگان‌ها در بردارنده‌ی واحدهای اوّلیه‌ی آوایی هستند و در برنامه‌های بازسازی گفتار فارسی که با هم‌گذاری واحدهای آوایی متن فارسی را به گفتار تبدیل می‌کنند، کاربرد دارند. دادگان هجایی دارای دو ویژگی مثبت است: اوّل واضح بودن خروجی گفتار بازسازی‌شده با استفاده از هجاهاست؛ دوم ساده بودن تعیین خودکار مرز هجاهای متن ورودی است و علت این امر برخورداری زبان از یک ساخت هجایی بسیار ساده است که در متن مقاله به آن اشاره کردیم. دادگان هجایی یک مشکل عمده دارد و آن غیرطبیعی بودن (علی‌رغم وضوح) خروجی گفتار بازسازی‌شده با استفاده از هجاهاست. علت این مشکل عدم انطباق ابتدا و انتهای سیگنال‌های صوتی هجاهاست که موجب کوبه‌ای شدن گفتار بازسازی‌شده می‌شود و طبیعی کردن آن نیازمند پردازش سیگنال سنگین است. وجود همین مشکل ما را بر آن داشت که به تهیه‌ی دادگان دایفونی روی بیاوریم که اوّل این‌که دایفون‌ها به خوبی گذرهای آوایی را پوشش می‌هند که این موضوع در عین برخورداری از وضوح، موجب طبیعی‌شدن گفتار بازسازی‌شده نیز می‌شود. از طرف دیگر حجم دادگان دایفونی در مقایسه با دادگان هجایی بسیار کم است.

تجربه‌ی ما نشان می‌هد که جایگاه اخذ واحدهای آوایی (خواه هجا و خواه دایفون) به‌عنوان مثال از کلمه یا جمله، از هجای بی‌تکیه یا تکیه‌دار و نیز چگونگی ضبط و تقطیع داده‌ها، تأثیر به‌سزایی در کیفیت گفتار بازسازی‌شده دارد و حجم پردازش سیگنال را در جهت طبیعی‌کردن گفتار پایین می‌آورد و در تهیه و آماده‌سازی دادگان‌های گفتاری هجایی و دایفونی فارسی به این مهم توجّه شده است.

۵- منابع

ابوطالبی، حمیدرضا. (۱۳۷۷). " بررسی و پیاده‌سازی سنتزکننده‌ای مناسب برای گفتار فارسی "، پایان‌نامه‌ی کارشناسی ارشد مهندسی مخابرات، دانشگاه صنعتی شریف، تهران، ایران.

اسلامی، محرم. و همکاران. (۱۳۸۳). " تبدیل رایانه‌ای متن به گفتار فارسی "، اولین کارگاه پژوهشی زبان فارسی و رایانه، دانشکده‌ی ادبیات و علوم انسانی دانشگاه تهران، تهران، ایران.

پژوهشکده‌ی پردازش هوشمند علائم، ۱۳۸۴، "گزارش طرح" دادگان صوتی هجاهای فارسی، تهران، ایران.

پژوهشکده‌ی پردازش هوشمند علائم. ۱۳۸۱. پیکره‌ی بزرگ گفتاری زبان فارسی (فارس‌دات بزرگ)، تهران، ایران.

ثمره، یداله. (۱۳۷۸). *" آواشناسی زبان فارسی* ". ویرایش دوم، مرکز نشر دانشگاهی، تهران، ایران.

شیخ‌زادگان، جواد و همکاران (۱۳۸۶)، "گزارش طرح صفحه‌خوان شیوا"، پژوهشکده‌ی پردازش هوشمند علائم، تهران، ایران.

Buron, R.H. (1968). "Generation of a 1000-word vocabulary for a pulse-excited vocoder operating as a audio response unit". IEEE Transactions on Audio and Electroacoustics. AU-16, 21-5.

Chapman, W.D. (1971). "Techniques for computer voice response". IEEE International conference Record, 98-90.

Eady,S.J. et al., 1987. "pitch assignment rules for speech synthesis by word concatenation". Proceedings of ICASSP 1987. 1473-6.

Klatt, D.H. (1987). "Review of text-to-speech conversion for English". JASA, 82:3, 636-93.

Keller, Eric. (1994). fundamentals of Speech Synthesis and Speech Recognition, John Wiley & Sons, New York

Sivertsen, E. (1961). "Segment inventories for speech synthesis". Language and Speech, 4, 27-90.

Taylor, Pual. (2009). Text-to-Speech Synthesis, Cambridge University Press, Cambridge

۶- ضمایم

ضمیمه‌ی الف

الف-۱- محاسبه‌ی تعداد بالقوه‌ی هجاهای زبان فارسی

می‌دانیم که در زبان فارسی تنها سه نوع هجای C، CVC و CVCC وجود دارد که در آن C معرّف واج‌های هم‌خوان و Vمعرف واکه‌ها می‌باشد. همین‌طور می‌دانیم که زبان فارسی گفتاری دارای ۲۳ هم‌خوان و ۶ واکه است. بنابراین تعداد بالقوّه‌ی هجاهای زبان فارسی به‌صورت زیر محاسبه می‌شود:

تعداد هجاهای CV = ۲۳×۶=۱۳۸

تعداد هجاهای CVC = ۲۳×۶×۲۳=۳۱۷۴

تعداد هجاهای CVCC = ۲۳×۶×۲۳×۲۳=۷۳۰۰۲

تعداد باقّوه‌ی هجاهای زبان فارسی = ۷۶۳۱۴

الف-۲- محاسبه‌ی تعداد دایفون‌های زبان فارسی

الف-۲-۱- محاسبه‌ی تعداد اوّلیه‌ی دایفون‌ها در زبان فارسی

تعداد "cc" ها = ۲۳×۲۳= ۵۲۹

تعداد "cv" ها = ۲۳×۷= ۱۶۱

تعداد "vc" ها = ۷×۲۳= ۱۶۱

تعداد "c-" ها = ۲۳

تعداد "-c" ها = ۲۳

تعداد "v-" ها = ۷

تعداد "-v" ها =۰

تعداد بالقوه‌ی دایفون‌ها = ۹۰۴

الف-۲-۲- محاسبه‌ی تعداد کل دایفون‌های دادگان صوتی

تعداد کل همخوان‌ها = n_c= ۲۵

تعداد همخوان‌ها بجز g,c,k؛ = n_c -۴= ۲۱

تعداد همخوان‌ها بجز g, k = n_c -۲= ۲۳

تعداد همخوان‌های g, k = n_{c1}= ۲

تعداد همخوان‌های c,؛ = n_{c2}= ۲

تعداد کل واکه‌ها = n_v = ۷

تعداد واکه‌های o,u,ow,/ = n_{v1} = ۴

تعداد واکه‌های a,e,i = n_{v2} = ۳

و نحوه‌ی محاسبه‌ی تعداد دایفون‌ها به‌صورت زیر مشخص می‌باشد:

تعداد "cc" ها = $(n_c - ۴) . n_c + n_{c1} . n_{c1} + n_{c2} . (n_c - ۲)$ = ۵۷۵

تعداد "cv" ها = $(n_c - ۴) . n_v + n_{c1} . n_{v2} + n_{c2} . n_{v2}$ = ۱۵۹

تعداد "vc" ها = $n_v . n_c$= ۱۷۵

تعداد "c-" ها = n_c -۲= ۲۳

تعداد "v-" ها = n_v =۷

تعداد "-c" ها = n_c= ۲۵

تعداد کل دایفون‌های دادگان صوتی = ۹۶۴

ضمیمه‌ی ب

فهرست علائم قراردادی مورد استفاده برای واج‌های فارسی در این مقاله (استاندارد Doulos Sil)

معادل فارسی واج	علائم
آ	/
اَ	a
اُ	o
اِ	e
ب	b
پ	p
ت	t
ج	,
ح و هـ	h
خ	x
د	d
ر	r
ژ	j
او	u
ن	n
م	m
ک (مثل ک در کاهو)	k
ک (مثل ک در کلم)	c
گ (مثل گ در گرگ)	g
گ (مثل گ در گرم)	;
ش	.
ز	z
ـی و ی (مثل یِ در آذربایجان)	y
ای	i
اُو (مثل اُو در جُوهر)	ow
س	s
چ	'
غ	G
ف	f
ق	G

محرم اسلامی در سال ۱۳۷۰ در رشته‌ی مترجمی زبان انگلیسی از دانشگاه علامه طباطبایی فارغ‌التحصیل شد. سپس در سال ۱۳۷۳ در رشته‌ی زبان‌شناسی همگانی دوره‌ی کارشناسی ارشد را دانشگاه تهران به‌پایان رسانید و در سال ۱۳۷۹ در همان رشته و همان دانشگاه موفق به اخذ مدرک دکترا شد. از ایشان بیش از ۳۰ مقاله در نشریات و همایش‌های داخلی و خارجی به‌چاپ رسیده است، و هم‌چنین صاحب تألیفاتی می‌باشند. وی هم‌اکنون عضو هیئت علمی مأمور در دانشگاه صنعتی شریف، گروه زبان‌شناسی رایانه‌ای می‌باشد و زمینه‌های تخصّصی ایشان، زبان‌شناسی همگانی، واج‌شناسی آهنگ و آواشناسی و پردازش رایانه‌ای زبان‌طبیعی (NLP)می‌باشد.
نشانی رایانامک ایشان عبارت است از:
m_eslami@sharif.edu

جواد شیخ‌زادگان مدرک کارشناسی مهندسی مخابرات را در خردادماه سال ۱۳۶۳ از دانشکده‌ی مهندسی برق دانشگاه شریف اخذ کرد و بلافاصله در اولین دوره‌ی کارشناسی ارشد مخابرات در دانشگاه خواجه نصیرالدین طوسی پذیرفته شد و در تیرماه ۱۳۶۷ این دوره را به پایان رساند. ایشان پس از سپری کردن دوره‌ی طرح سربازی خود در مهرماه سال ۱۳۶۹ در اولین دوره‌ی دکتری مخابرات در دانشگاه تربیت مدرس پذیرفته شد و در سال ۱۳۷۴ از این رشته فارغ‌التحصیل شد. نامبرده علاوه بر تدریس در دانشگاه‌های مختلف تهران از اواسط سال ۱۳۶۶ تا حال حاضر، از اوایل سال ۱۳۷۰فعّالیت‌های تحقیقاتی خود را در زمینه‌ی پردازش سیگنال‌های صوتی شروع کرد و در سال ۱۳۷۳ موفق به تأسیس یک مؤسسه‌ی تحقیقاتی گردید که از سال ۱۳۷۷ با مجوز وزارت علوم، تحقیقات و فناوری به نام پژوهشکده‌ی پردازش هوشمند علائم فعّالیت خود را ادامه می‌دهد. ایشان در حال حاضر عضو هیئت علمی، هیئت امناء و هیئت مدیره‌ی پژوهشکده‌ی مذکور، و نیز عضو هیئت مؤسس و شورای اجرایی انجمن رمز ایران از بدو تأسیس (سال ۱۳۷۹) تاکنون هستند. علاوه بر زمینه‌های تحقیقاتی اصلی او که عبارتند از: بازشناسی گوینده، بازشناسی زبان و دادگان‌های گفتاری زبان‌فارسی، موضوعاتی از قبیل: رمزنگاری، پنهان‌نگاری و امنیّت اطّلاعات و ارتباطات نیز از زمینه‌های مورد علاقه‌ی ایشان می‌باشند.
نشانی رایانامک ایشان عبارت است از:
sheikhzadegan@rcisp.ac.ir

زهرا احمدی‌نیا فارغ‌التحصیل مقطع کارشناسی ارشد زبان‌شناسی از دانشگاه تهران است و به مدّت پنج سال در پژوهشکده‌ی پردازش هوشمند علائم در انجام فعّالیت‌های پژوهشی مختلف مربوط به برنامه‌ی تبدیل متن به گفتار فارسی همکاری داشته است. زمینه‌ی تحقیقاتی مورد علاقه‌ی وی واژگان‌نگاری و در حال حاضر پژوهشگر مؤسسه‌ی فرهنگ معاصر است.
نشانی رایانامک ایشان عبارتست از:
za.ahmadinia@ gmail.com

علی بهرامی راد در سال ۱۳۸۱ در رشته‌ی مهندسی برق-الکترونیک از دانشگاه امام حسین (ع) فارغ‌التحصیل شده است. وی در حال حاضر در دوره‌ی کارشناسی ارشد رشته‌ی مهندسی پزشکی در دانشگاه صنعتی تامپره در کشور فنلاند تحصیل می‌کند. زمینه‌ی تحقیقاتی مورد علاقه‌ی وی کاربرد یادگیری ماشینی در پردازش سیگنال‌های صوتی می‌باشد.
نشانی رایانامک ایشان عبارتست از:
abahramir@yahoo.com