مجله پژوهش‌های زبان شناسی
سال دوم، شماره دوم، پاییز و زمستان ۱۳۸۹
دریافت: ۱۳۸۹/۱۰/۱۵ – پذیرش: ۱۳۸۹/۱۲/۲۵
صص ۱۱- ۱

طراحي و پياده سازي دادگان دايفون زبان فارسي براي كاربرد زبانشناسي رايانه‌اي

سیدسعید آیت[۱]

چکیده

دادگان‌های گفتاری نقش مهمی را در تحقیقات و پیاده‌سازی‌های مربوط با زبان‌شناسی رایانه‌ای ایفا می‌کنند. در این مقاله، پس از مطالعه واحدهای آوایی مختلف قابل استفاده برای این منظور، مراحل تهیه یک دادگان دایفون ویژه زبان فارسی ارایه می‌شود. برای این منظور، در ابتدا پایگاه واژگانی که دایفون‌های زبان را شامل شوند، تهیه گردید. سپس نرم‌افزاری طراحی و پیاده‌سازی شد که با گرفتن صورت‌های واجی واژه‌ها، دایفون‌هایی را که قرار است از آن استخراج شوند، مشخص کند. در مرحله بعد سیگنال‌های گفتاری واژه‌ها ضبط گردید و نکات افزایش بررسی گردید. در پایان نیز جداسازی دایفون‌ها و تهیه دادگان مورد نظر صورت پذیرفت. برای افزایش دقت دادگان تهیه شده، مراحل جداسازی دایفون‌ها از سیگنال‌های گفتاری ضبط شده با استفاده از سه روش شنوایی، بررسی سیگنال زمانی و مطالعه طیف‌نگاشت، ارزیابی و از ترکیب هر سه روش برای افزایش دقت دادگان استفاده شد.

واژه‌های کلیدی: دادگان دایفون، زبان‌شناسی رایانه‌ای، واحد آوایی، تقطیع گفتار

[۱] استادیار دانشگاه پیام نور- **گروه علمی** مهندسی کامپیوتر و فناوری اطلاعات Email:dr.ayat@pnu.ac.ir

۱. مقدمه

زبانشناسی رایانه‌ای از دو منظر قابل مطالعه و بررسی است: از یک سو به دلیل نقشی که می‌تواند در مباحث مختلف زبان و زبان‌شناسی ایفا کند و از سوی دیگر به دلیل کاربردی که در سیستم‌های هوشمند گفتاری دارد. امروزه طراحی و ساخت وسایلی که قادر به گفتگو با انسان باشند، جایگاه ویژه‌ای دارند. این امر بشر را به سوی تحقیق و مطالعه پیرامون فرایند تولید و نیز درک گفتار در آدمی و مدل‌سازی آن در دو قالب سیستم‌های تبدیل متن به گفتار(text to speech) و تبدیل گفتار به متن (speech recognition)، و در کنار آن سایر سیستم‌ها و زمینه‌های مرتبط، سوق داده است(آیت، ۱۳۸۶).

زبان‌شناسی رایانه‌ای و پردازش گفتار زمینه‌های مختلفی را در بر می‌گیرد، از جمله تولید گفتار، بازشناسی گفتار، تجزیه و تحلیل گفتار، و دادگان‌های گفتاری(هوانگ۲۰۰۱؛ دلر۲۰۰۰). شایان ذکر است برخلاف بسیاری از فناوری‌های امروزی که زبان و ملیت نقشی در آن ندارند، عملکرد سیستم های گفتاری، تا حدود زیادی مرتبط با زبان است که در طراحی‌ها باید به این امر توجه داشت.

تحقیقات نشان می‌دهد که دادگان گفتاری در اکثر سیستم‌های پردازش گفتار نقش مهمی را به عهده دارد. از این رو در عمل نیاز است برای هر زبان به‌طور مجزا دادگان‌های مختلفی که ممکن است در سیستم‌های پردازش گفتاری استفاده شوند، تهیه کرد. از سوی دیگر، وجود پایگاه‌های گفتاری استاندارد باعث می‌شود بتوان راندمان سیستم‌های مختلف پردازش گفتاری را بر اساس آن مقایسه کرد. برای مثال، در زبان انگلیسی دادگان گفتاری TIMIT و در زبان فارسی دادگان گفتاری فارس‌دات معمولاً استفاده می‌شوند. دادگان فارس‌دات را مرکز تحقیقات علایم هوشمند ایران تهیه کرده و در تهیهٔ آن سعی شده است از گویندگان متنوع بر حسب سن، لهجه، تحصیلات، و جنس استفاده شود(شیخ و بی‌جن‌خان۱۳۸۹:۴۴؛ بی‌جن‌خان ۱۹۹۴).

دادگان‌های تهیه شده برای کاربرد زبان‌شناسی می‌توانند متنی باشند یا گفتاری، که هر دو نقشی اساسی در تحقیقات آن زبان دارند. در زمینه تحقیقات انجام شده دادگان‌های متنی برای زبان فارسی می‌توان مواردی از قبیل عاصی (۱۹۹۷و۱۳۷۳) قیومی، ممتازی و بی‌جن‌خان (۲۰۰۴) را نام برد. در تهیهٔ دادگان نکته شایان ذکر این است که ممکن است برای کاربرد یا سیستمی خاص نیاز باشد دادگان گفتاری ویژهٔ همان سیستم تهیه شود. برای مثال، سیستم بازشناسی گفتاری که قرار است فرمان‌های صوتی چرخ ویژهٔ معلولان را تشخیص دهد، ممکن است تنها نیاز داشته باشد کلمات «برو»، «بایست» و نظایر آن را تشخیص دهد. از سوی دیگر، بسته به کاربرد و محیط استفاده از سیستم گفتاری ممکن است نیاز باشد شرایط خاصی در تهیهٔ دادگان لحاظ شود. برای مثال، در سیستم‌های بازشناسی گفتار تلفنی که قرار است گفتار ورودی را از تلفن دریافت کنند، نیاز است گفتار مورد نیاز از طریق تلفن تلفظ و در دادگان ضبط شود. شایان ذکر است که از دادگان‌های معروف تلفنی در زبان فارسی است (بی‌جن‌خان،۲۰۰۳).

دادگان‌های گفتاری را می‌توان از منظر نوع گفتار ذخیره شده به دو دسته تقسیم کرد دسته اول دادگان‌هایی هستند که در آن جملات مختلف زبان ادا شده‌اند، نظیر فارس دات. این دادگان‌ها معمولا برای سیستم‌هایی نظیر بازشناسی

گفتار استفاده می‌شوند. دسته دوم دادگان‌هایی هستند که واحدهای آوایی خاص آن زبان را شامل می‌شوند؛ مثلاً دادگان واج یا دادگان دایفون(diphone). این دادگان‌ها معمولاً برای سیستم‌های تبدیل متن به گفتار استفاده می‌شوند.

شاخه تبدیل متن به گفتار از مهم‌ترین زمینه‌های زبان‌شناسی رایانه‌ای است. راندمان سیستم‌های تبدیل متن به گفتار، یا گفتارسازها به شدت به نکات در نظر گرفته شده ویژه آن زبان وابسته است. برای مثال، تغیرات فرکانس گام(pitch) در جملات مختلف نظیر خبری یا پرسشی چگونه است و به چه نحو مدل می‌شود؟ از جمله روش‌های مطرح برای پیاده‌سازی سیستم‌های تبدیل متن به گفتار، روش‌های پیوندی (concatenational) یا استفاده کننده از گفتار طبیعی ضبط شده هستند.

در این روش بر اساس ساختار زبان مورد نظر و روش استفاده شده، قطعه‌ای از گفتار، به عنوان واحد پایهٔ صوتی تعیین می‌شود، سپس با استفاده از واحد صوتی تعیین شده، مجموعه‌ای را تشکیل می‌دهند که از صحبت طبیعی استخراج و ذخیره می‌گردد. مجموعهٔ مورد بحث باید به گونه‌ای باشد که به کمک آن بتوان همهٔ ترکیب‌های واجی موجود در آن زبان را پوشش داد. قطعات مورد استفاده ممکن است یکی از موارد کلمه، هجا، واج، نیمه‌هجا، دایفون، و مواردی نظیر این باشد که بسته به مجموعه اصواتی که قرار است گفتارساز تولید کند، همچنین مصالحه میان حافظه و کیفیت بهتر، یکی از آن‌ها انتخاب می‌شود.

در این مقاله به طراحی و پیاده‌سازی یک دادگان دایفون ویژه زبان فارسی پرداخته می‌شود. ساختار مقاله بدین صورت است که در بخش بعد واحدهای آوایی مختلف را مطالعه کرده، دلایل انتخاب دایفون را به عنوان واحد مورد استفاده بیان می‌کنیم. سپس مراحل تهیه دادگان گفتاری ارایه می‌شوند.

۲. واحدهای پردازش دادگان

اگر واحد بزرگی نظیر جمله را کنار بگذاریم، قطعات استفاده شده در تهیهٔ دادگان عبارتند از: واژه، واج، هجا، دایفون و برخی واحدهای دیگر که در ادامه بدانها خواهیم پرداخت.

۲-۱. واژه

بعد از جمله و عبارت، واژه بزرگ‌ترین واحدی است که در تهیهٔ دادگان استفاده می‌شود، ولی باید توجه داشت که تعداد واژه‌های مورد استفاده در یک زبان بسیار زیاد است؛ مثلاً در زبان انگلیسی حداقل ۴۰,۰۰۰ واژه در گفتار روزمره مورد نیاز است. از سویی دیگر، تعداد واژه‌های هر زبان نیز به‌طور پیوسته در حال تغییر است، و حافظهٔ فوق‌العاده زیاد و روزآمدسازی مکرر دادگان را می‌طلبد. از منظری دیگر، بین واژه‌های ادا شده در یک جمله نوعی ارتباط و به عبارتی تأثیر متقابل آواها وجود دارد که در صورت استفاده از دادگان واژگان، تأثیر آواها در انتهای واژه و ابتدای واژهٔ بعد لحاظ نمی‌شود. مشکلات موجود در واحدی نظیر واژه، محققان را بر آن داشت تا واحدهای آوایی کوچک‌تری را انتخاب کنند که ضمن نیاز به حافظهٔ کمتر، کیفیت مناسب‌تری را نیز امکان‌پذیر سازند.

۲-۲. واج

واج عبارت است از واحد اساسی، مجرد، و انتزاعی هر زبان که برای انتقال معانی به‌کار می‌رود. واج‌ها کوچک‌ترین واحدهای آوایی‌اند که تعویض آن‌ها موجب تغییر معنایی واژه می‌گردد؛ مثلاً اگر در واژهٔ «مرد» آوای /d/ را با /z/ عوض کنیم، واژهٔ «مرز» تولید می‌شود که معنای کاملاً متفاوتی دارد. در زبان فارسی ۲۹ واج وجود دارد ودر مقایسه با واژه‌های زبان حافظهٔ کمی برای ذخیرهٔ آن‌ها مورد نیاز است. واج کوچک‌ترین عنصر ممکن در دادگان است و در هر زبانی، واحد آوایی مشخص و با تعداد کاملاً مشخص و محدود محسوب می‌شود. اگر چه واج‌های یک زبان کم‌اند، ولی به دلیل اثر هم‌آوایی میان واج‌ها معمولاً تعیین مرز دقیق واج‌ها میسر نیست. از سوی دیگر، در کاربردی نظیر گفتارساز، قرارگرفتن واجی متفاوت با واجی که هنگام ضبط دادگان استفاده می‌شود، بعد از واج فعلی، اثر متقابل بین واج‌ها را آن گونه که در زبان طبیعی وجود دارد، مدل می‌کند. بنابراین، در عمل در کاربردی نظیر گفتارسازها واج‌ها به ندرت در تهیهٔ دادگان استفاده می‌شوند.

۲-۳. هجا

هجا از آواهایی تشکیل می‌شود که ساخت و ترکیب آن بسته به نوع زبان متفاوت است. هجا رشته آوایی پیوسته‌ای است؛ یعنی اجزای سازندهٔ هجا طی فرایند تولیدی بدون مکث ادا می‌شوند.هجا در زبان فارسی از یک واکه و یک تا سه همخوان تشکیل می‌شود(ثمره ۱۳۷۸). هجای آغازین حتماً همخوان است و نمی‌تواند واکه باشد. در آغاز هجا نیز دو همخوان پشت سر هم نمی‌توانند قراربگیرند. بنابراین، در زبان فارسی سه نوع هجا وجود دارد که با قرار دادن C به جای همخوان و V به جای واکه به سه صورت CV،CVC، CVCC درمی‌آید. مثال‌های این سه نوع عبارت‌اند از /to/، /man/، /gušt/ که معادل واژه‌های «تو»، «من»، و «گوشت» است.

زبان فارسی ۲۳ همخوان و ۶ واکه دارد(مشکوة الدینی ۱۳۷۷). همزه با علامت قراردادی /?/ نیز در این مقاله جزو همخوان‌های زبان فارسی در نظر گرفته شده است. بنابراین، صورت واجی واژه‌هایی نظیر «او» به‌صورت /?u/ و در قالب واجی CV است. از آنجا که در زبان فارسی دو واکه نمی‌تواند در یک هجا قرارگیرد، بنابراین، تعداد هجاها در هر رشتهٔ آوایی با شمارش واکه‌ها مشخص خواهد شد. تعیین مرز هجاها نیز، پس از پیدا نمودن واکه‌ها، کاری ساده است و با توجه به سه ساختار CV، CVC و CVCC، در زبان فارسی، کافی است پس از یافتن محل واکه، همخوان قبل از آن را آغاز هجا در نظر بگیریم. با در نظر گرفتن ۶ واکه و ۲۳ همخوان برای زبان فارسی، تعداد کل هجاهای ممکن در زبان فارسی به‌صورت زیر خواهد بود:

CV=۲۳×۶=۱۳۸

CVC=۲۳×۶×۲۳=۳۱۷۴

CVCC=۲۳×۶×۲۳×۲۳=۷۳۰۰۲

تعدادکل هجاها=۱۳۸+۳۱۷۴+۷۳۰۰۲=۷۶۳۱۴

با وجود این، هجاهای مورد استفاده در زبان فارسی به مراتب کمتر از این است. مثلاً ترکیب‌هایی نظیر «ایژ» /iǰ/، «اوو» /uv/، «ژُ» /ǰo/ و «گی» /gey/ در عمل در زبان فارسی به کار نرفته‌اند. بررسی محدودیت‌های همنشینی، خود یکی از مباحث آواشناسی است.

همچنین، واژه‌هایی مانند «تمبر» /tambr/، با ساخت هجایی CVCCC، و ساختارهای مشابه، ساختارهای مجاز ساخت هجا در زبان فارسی نیستند و از زبان‌های دیگر وارد این زبان شده‌اند. ولی ساخت هجا در زبان انگلیسی به صورت ((((CC)C)V(C(C(C(C)))) است. بدین ترتیب، نزدیک به ۲۰ نوع هجای مختلف را می‌توان در این زبان تولید کرد. استفاده از هجاها به‌عنوان عنصر دادگان، چندان متداول نیست. همان‌طور که گفتیم، فارسی از نظر ساخت آوایی هجاها، جزو زبان‌های ساده است.

۴-۲. دایفون

همان گونه که پیشتر بیان شد، واحد پایه در تهیهٔ دادگان باید به گونه‌ای باشد که اولاً، حجم حافظهٔ معقولی را اشغال کند، به عبارتی، تعداد عناصر دادگان مطلوب باشد؛ و دوماً، بتوان تأثیر متقابل میان آواها را با آن دادگان در نظر گرفت. از جمله واحدهایی که با هدف تأمین این شرایط تعریف و استفاده می‌شوند، واحد دایفون است. این واحد، در واقع برای در نظر گرفتن انتقال از یک واج به واج بعدی ابداع شده است. برای دایفون تعاریف مختلفی بیان شده است، از آن جمله «دایفون عبارت است از نیمهٔ پایدار یک آوا تا نیمهٔ پایدار آوای بعدی» یا «دایفون شامل قسمت آخر یک واج، قسمت اول واج بعد، و گذار میان آن دو است» (آیت ۱۳۷۹).

برای دایفون معادل‌های فارسی مختلفی پیشنهاد شده است، از جمله «دو آوایی»، «دو واجک» و در برخی منابع «دو واج». در این مقاله به دلیل تکثر اصطلاح و آرای صاحب‌نظران، از این معادل‌های فارسی صرف‌نظر شده است.

همان‌طوری‌که پیشتر گفته شد، دایفون از دو نیم‌واج به هم چسبیده تشکیل می‌شود. البته باید در نظر داشت که این ترکیب شامل ترکیب سکون و نیم‌واج نیز می‌شود. با این تعریف، انواع دایفون‌های زبان فارسی در یکی از قالب‌های VC،CV،CC،V-،C- ،C- است. «-» نشانهٔ سکون یا سکوت است. در بررسی اجمالی به نظر می‌رسد تعداد دایفون‌های یک زبان برابر تعداد جای‌گشت‌های ۲ از P است، که P تعداد واج‌های زبان است. در این صورت با در نظر گرفتن مکث، در زبان فارسی ۹۰۰ دایفون خواهیم داشت که در عمل تعداد دایفون‌های زبان فارسی از این تعداد کمتر است. برخی دایفون‌ها، یعنی ترکیبات V- یا VV، اصولاً در زبان فارسی وجود ندارند، چرا که در ساختار هجایی آن به کار نمی‌رود. بنابراین، تعداد کل دایفون‌های ممکن در زبان فارسی به‌صورت زیر است:

C-=۲۳

-C=۲۳

V-=۶

CC=۲۳×۲۳

CV=۲۳×۶

VC=۶×۲۳

۸۵۷ = تعداد کل

برخی دایفون‌ها هم در عمل در زبان فارسی استفاده نمی‌شوند، نظیر «ژُ» /ĵo/. بنابراین از آنجا که این هجا از دو واج تشکیل شده، دایفون نظیر آن هم در زبان موجود نیست.

۵-۲. واحدهای دیگر

واحدهای دیگری نظیر سه‌واجی(triphone)، که می‌تواند شامل نیمهٔ دوم یک واج، یک واج کامل، و نیمهٔ اول واج بعد شود، نیز وجود دارند که بسته به کاربرد ممکن است استفاده شوند. با این وجود، دادگان دایفون از مهمترین دادگان هایی است که به دلیل در نظر گرفتن مرز میان واج‌ها، می‌تواند در گفتارسازهای پیوندی و سایر مطالعات مرتبط زبان‌شناسی استفاده شود. در زمینه دادگان دایفون فارسی کامل‌ترین مورد یافت شده یک پایگاه مقدماتی است که توسط نگارنده ارایه شده است. (آیت، ۱۳۷۹) لذا طراحی یک پایگاه دایفون مناسب‌تر و کامل‌تر برای تحقیقات زبانشناسی و کاربردهای زبان‌شناسی رایانه‌ای زبان فارسی ضروری به نظر می‌رسید که در این مقاله به این مهم پرداخته شده است.

۳. تهیهٔ دادگان گفتاری دایفون

مراحل تهیه یک دادگان گفتاری شامل تهیه پیکره واژگان (corpus)، ضبط گفتار و استخراج واحد آوایی مطلوب است که در ادامه به اجمال بررسی می‌کنیم.

۱-۳. تهیهٔ پایگاه واژگان

در صورتی که واحدهای آوایی انتخاب شده برای دادگان کوچک‌تر از کلمه باشند، معمولاً این مرحله استفاده می‌شود. ابتدا پایگاهی شامل واژه‌های مختلف تهیه می‌شود. واژه‌های انتخابی باید به‌گونه‌ای باشند که تمامی حالات واحد آوایی دلخواه موجود در زبان را دربرگیرند. سپس تعیین شود هر کلمه کدام واحدهای آوایی زبان را دربردارد. لذا برای این کار ابتدا نرم‌افزاری طراحی و پیاده‌سازی شد که صورت واجی واژه را بگیرد و واحدهای آوایی دلخواه موجود در آن کلمه را مشخص کند؛ مثلاً اگر واحد آوایی دلخواه دایفون باشد، واژه‌ای نظیر «سلام» با صورت واجی /salàm/ دارای دایفون‌های/-s/, /sa/, /al/, /là/, /àm/, /m-/ است.

دلیل تهیهٔ پیکره واژگان این است که تا حد ممکن واحدهای آوایی از داخل واژگان استخراج شوند تا از لحاظ مشخصات نوایی وضعیت مناسب‌تری داشته باشند و به‌صورت طبیعی‌تری ادا شده باشند. در صورتی که برای برخی واحدهای آوایی هیچ واژهٔ واقعی در زبان پیدا نشود، می‌توان عبارت نامفهومی که آن واحد آوایی را در برگیرد تهیه و استفاده کرد.

نکته دیگری که باید یادآوری نمود این است که در عمل برخی دایفون‌ها در زبان فارسی استفاده نمی‌شوند و یا اینکه کلمه فارسی که در آن استفاده شده باشند یافت نمی‌شود. جدول ۱، نمونه‌ای از کلمات پایگاه واژگان را به همراه دایفون‌های استخراج شده از آنها نشان می‌دهد.

جدول ۱) نمونه‌ای از کلمات پایگاه واژگان به همراه صورت واجی و دایفون‌های استخراجی

واژه استفاده شده	صورت واجی	دایفون‌های استخراجی
سلام	-salàm-	-s, sa, al, là, àm, m-
صورت	-surat-	-s,su,ur,at
چه	-ce-	-c,ce
خواهد	-xàhad-	-x,ha,xà, àh
زندگی	-zendegi-	-z,ze,gi,eg,en
ژاپن	-2àpon-	-2,po,2à
قبل	-qabl-	-q,bl,qa,ab
کرده	-karde-	-k,de,rd,ka,e-
لازم	-làzem-	-l
نظر	-nazar-	-n,za,az

۲-۳. ضبط گفتار پایگاه واژگان

در این مرحله واژه‌های مشخص شده در بخش قبل را گوینده/گویندگانی(بر حسب کاربرد مورد نظر) در شرایط مشخص و مورد نیاز سیستم می‌خواند/ می‌خوانند و سیگنال صدای مورد نظر ضبط می‌شود.

از آنجا که معمولاً این دادگان یک بار و برای همیشه تهیه می‌شود، سعی بر آن است دادگانی با بهترین کیفیت تهیه شود. برای نیل به این مقصود نکاتی نظیر موارد زیر اهمیت می‌یابد، که در مرحله ضبط دادگان مورد توجه قرار گرفته‌اند:

الف) دستگاه ضبط: بهتر است دستگاه‌های مورد استفاده برای مراحل ضبط گفتار سیگنال‌های مزاحم (noise) کمتری را به گفتار اضافه کنند. این نویز می‌تواند ناشی از مواردی نظیر پاسخ فرکانسی میکروفن، کارت صوتی، تبدیل کننده آنالوگ به دیجیتال(A/D) یا ناشی از خطای چندی کردن (quantization) باشد (واثقی ۲۰۰۷ آیت، و دیگران ۲۰۰۶).

ب) مکان: سیگنال‌های گفتاری در مکانی ضبط شوند که نویزهای محیطی (ambient) کمتری داشته باشد. بسیاری از نویزهای محیطی ممکن است توسط گوش ما شنیده نشوند. برای مثال، نویز فرکانس پایین ناشی از عبور یک کامیون در اطراف محیط ضبط. این نویزها را میکروفن ضبط می‌کند و ممکن است در ویژگی‌های استخراج شده از پایگاه موثر باشند(آیت ۱۳۸۵؛ آیت ۲۰۰۸). تاثیر نویز بر این ویژگی‌ها که می‌تواند مثلاً ضرایب کپستروم استخراج شده از سیگنال‌های گفتاری باشد، ممکن است راندمان سیستم مورد مطالعه را به شدت کاهش دهد. همچنین باید به نویزهای برگشتی (reverberation) دقت کافی داشت. در عمل بهتر است ضبط گفتار در اتاق ضدصوت (soundproof) صورت گیرد.

ج) دقت مورد نظر: در عمل بهتر است داده‌ها با کیفیتی نظیر قالب اندازهٔ ۱۶ بیتی و با فرکانس نمونه‌برداری مثلاً ۲۲۰۵۰ هرتز نمونه‌برداری شوند.

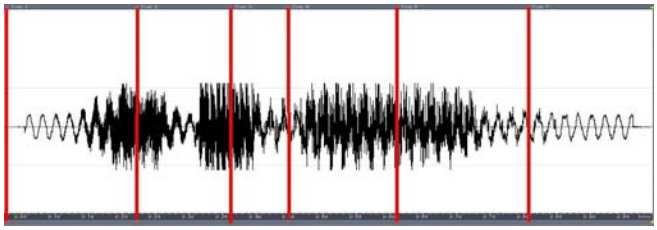
د)گوینده: بسته به سیستم، گوینده یا گویندگان به‌طور مناسب انتخاب شوند. برای مثال، در سیستم سنتز که تبدیل‌کنندهٔ متن به گفتار است و از واحدهای آوایی از پیش ضبط شده برای تولید گفتار استفاده می‌کند، سعی می‌شود واژه‌ها را فردی ادا کند که صدای واضح و دلنشینی دارد. از گوینده خواسته می‌شود واژه‌ها یا جملات را تا حد امکان با حالتی ثابت (monotonic) ادا کند، تا هنگام تبدیل تغییرات نوایی نظیر تفاوت دامنه، به چشم نخورد. چنین سیستمی در بسیاری از لغت‌نامه‌ها استفاده می‌شود، از جمله تلفظ لغات در لغت‌نامهٔ نارسیس (Narsis) یا بابیلون(Babylon). از سوی دیگر، در سیستم بازشناسی‌ای که قرار است گویندگان مختلفی از آن استفاده کنند، سعی می‌شود از تعداد گویندهٔ بیشتری برای ضبط و تهیهٔ دادگان استفاده شود تا در مرحلهٔ عمل سیستم تطابق بیشتری داشته باشد.

ه) نحوه ادا: سعی شود در مرحلهٔ ضبط، واژه‌ها به شکل طبیعی ادا شوند.

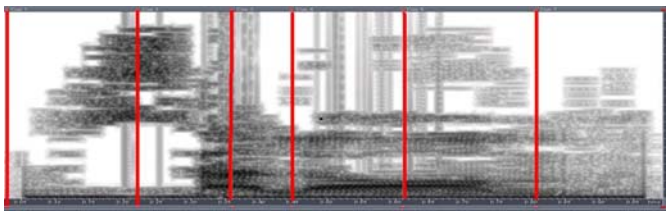
۳-۳. استخراج واحد آوایی دلخواه

در این مرحله، با داشتن فهرست واحدهای آوایی مذکور، می‌دانیم که از هر واژه یا جملهٔ ضبط شده قرار است چه واحدهای آوایی استخراج شوند. این امر که گاه از آن با عنوان تقطیع(segmentation) نیز نام می‌برند، می‌تواند به‌صورت خودکار؛ یعنی با استفاده از نرم‌افزار صورت گیرد یا به صورت دستی؛ یعنی توسط انسان. سیستم‌های کاملاً خودکار در عمل دقت مناسبی ندارند و بهتر است در ترکیب با روش دستی استفاده شوند. برخی نکات بررسی شده پیرامون مساله تقطیع در آثاری نظیر عاصی و حاجی عبدالحسینی (۲۰۰۰)مطالعه شده است.

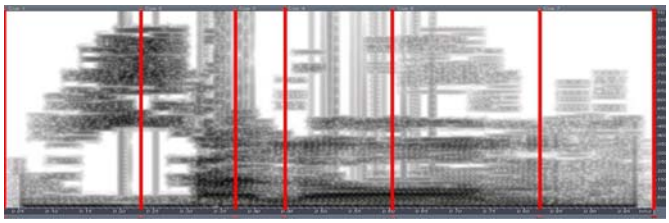
در تهیه پایگاه مورد مطالعه، پیاده سازی‌های نرم‌افزاری در محیط نرم‌افزارهای *Matlab* و *Cool Edit* صورت پذیرفته است و روش دستی برای مرز گذاری واحدها استفاده شده است. برای افزایش کیفیت دادگان نیز از ترکیب چند روش برای تعیین مرز واحد آوایی استفاده شود. برای مثال، ممکن است از سه روش شنیداری، دیدن شکل موج زمانی، و استفاده از طیف‌نگاشت سیگنال که اطلاعات زمان، فرکانس، و انرژی را نشان می‌دهد به‌طور همزمان استفاده کرد تا مرز واحدهای آوایی بهتر و دقیق‌تر تعیین شود. در تهیه دادگان مورد نظر افزایش دقت از هر سه روش به طور همزمان استفاده شده است. برای نمونه، کلمه "سلام" با صورت واجی /salâm/ و دایفون‌های قابل استخراج که در ردیف اول جدول (۱)، آمده است، پس از برچسب گذاری و تقطیع به روش‌های شنیداری و بررسی شکل موج زمانی در شکل (۱)دیده می‌شود. شکل ۲ نیز طیف نگاشت آن را به همراه برچسب‌های زمانی آن نشان می‌دهد. دقت در شکل (۲) ، نشان می‌دهد که خطوط عمودی مشخص کننده مرز دایفون‌ها خیلی دقیق نیست و در زمان پایداری کامل در وسط واج مربوط قرار نگرفته اند. لذا با تغیر مکان مرزهای دایفون ها با استفاده از طیف نگاشت به شکل (۳)دست می‌یابیم. شکل (۴)نیز مرزهای دقیق تعین شده توسط طیف نگاشت را در حوزه زمان نشان می‌دهد.



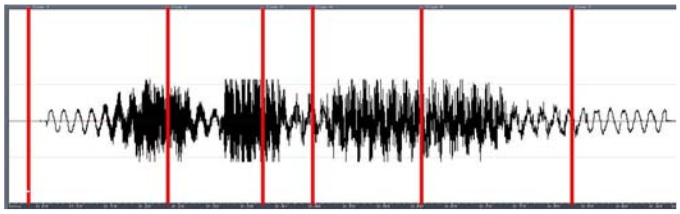
شکل ۱: نمایش زمانی سیگنال گفتار واژه *سلام* و مرزهای دایفون‌ها



شکل ۲: نمایش طیف نگاشت سیگنال گفتار واژه *سلام* و مرزهای دایفون‌ها



شکل ۳: طیف نگاشت سیگنال گفتار واژه *سلام* و مرزهای اصلاح شده دایفون‌ها



شکل ۴: نمایش زمانی سیگنال گفتار واژه *سلام* و مرزهای اصلاح شده دایفون

نکات دیگری که در این مرحله بهتر است در نظر گرفته شود عبارتند از:

الف) نام فایل های صوتی بر اساس نام دایفون نامگذاری شود.

ب) جدولی شامل نام دایفون، نام فایل گفتاری و طول زمانی فایل گفتاری تهیه شود.

ج) مرز بین واج‌ها در هر دایفون نیز به نحوی مشخص یا علامت گذاری شود. این امر می‌تواند در ستونی از جدول مرحله قبل انجام شود.

د) در عمل مناسب‌تر است که هنگام جداسازی سیگنال مربوط به دایفون مورد نظر از کل سیگنال ضبط شده، بازه زمانی تقریبی ۵۰ میلی ثانیه قبل و بعد دایفون هم انتخاب شود.

پس از اینکه این امر برای تمامی واژه‌های موجود انجام شد، پایگاه گفتاری واحد آوایی مورد نظر آماده است.

۴. نتیجه‌گیری

با توجه به اهمیت دادگان‌های گفتاری در تحقیقات و پیاده‌سازی‌های مربوط با زبانشناسی، در این مقاله به تهیه یک دادگان دایفون ویژه زبان فارسی پرداخته شد. برای این منظور در ابتدا اهمیت دادگان های گفتاری مورد مطالعه و سوابق دادگان دایفون زبان فارسی بررسی شد. سپس مراحل تهیه دادگان دایفون با کیفیت مناسب ارایه شد. این مراحل عبارت بودند از: تهیه پایگاه واژگان مناسب، و پیاده‌سازی نرم‌افزاری برای تعیین دایفون‌های مورد نظر از هر واژه. در گام بعدی نکات مربوط به شرایط ضبط سیگنال‌های گفتاری بررسی و سیگنال‌ها ضبط گردید. در نهایت نیز برای افزایش دقت دادگان تهیه شده، مراحل جداسازی دایفون‌ها از سیگنال‌های گفتاری ضبط شده با استفاده از هر سه روش شنوایی، بررسی سیگنال زمانی و مطالعه طیف‌نگاشت، صورت پذیرفت.

قدردانی

این تحقیق با استفاده از اعتبارات دانشگاه پیام نور انجام شده است. در اینجا از معاونت پژوهشی دانشگاه پیام نور تشکر می گردد.

کتابنامه

آیت، سید سعید.(۱۳۸۵). بهسازی گفتار با استفاده از تبدیل موجک و روش‌های ترکیبی، پایان‌نامه دکتری، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف.

آیت، سید سعید.(۱۳۷۹). طراحی و پیاده‌سازی سیستم توید گفتار فارسی با تأکید بر بهبود هر چه بیشتر گفتار تولید شده، پایان‌نامه کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر.

ثمره، یدالله.(۱۳۷۸). *آواشناسی زبان فارسی*، تهران: مرکز نشر دانشگاهی، ویرایش دوم.

شیخ سنگ تجن، شهین، بی جن خان، محمود. (۱۳۸۹)."بررسی کاهش واکه‌ای در زبان فارسی محاوره‌ای" *پژوهش های زبان شناسی*، ش ۱، صص ۳۵-۴۸.

عاصی، مصطفی.(۱۳۷۳)." طرح ایجاد پایگاه داده های زبان فارسی به کمک کامپیوتر"، *مجله اطلاع رسانی نشریه فنی مرکز اطلاعات و مدارک علمی ایران* ، دوره۱۱، ش۱. ص۶ – ۱۰.

مشکوه الدینی، مهدی.(۱۳۷۷). *ساخت آوایی زبان*، مشهد: انتشارات دانشگاه فردوسی.

Assi, S. M. (1997). "Farsi Linguistic Database (FLDB)," *International Journal of Lexicography*. Vol.10, No. 3. 5 - 6.

Assi, M. and Hajiabdolhosseini, M (2000) "Grammatical tagging of a Persian corpus". *International Journal of Corpus Linguistics* , Vol. 5, No. 1, 69-81.

Ayat, S. Manzuri, M. T., and Dianat, R. (2006). "An Improved Wavelet-based Speech Enhancement by Using Speech Signal Features" *International Journal of Computers & Electrical Engineering*, Vol. 32, No. 6, 411-425.

Ayat, S. (2008). "Enhanced Human-Computer Speech Interface Using Wavelet Computing" IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems. Istanbul, Turkey, 37 - 40.

Bijankhan, M. Sheykhzadegan, J, (1994) “FARSDAT: Farsi spoken language database”. In Proceedings of International Conference on Speech Sciences and Technology, Vol. 2: 826-829, Perth, Australia.

Bijankhan, M, Sheykhzadegan, J, Roohani, M. R. Zarrintare, R, Ghasemi, S. Z. Ghasedi M. E. (2003) “TFARSDAT: Telephone Farsi spoken language database "International Conference of EuroSpeech, Geneva, Switzerland, 1525-1528.

Deller, J. R., et. al. (2000). 2nd end. *Discrete-time Processing of Speech Signals*. New York: IEEE Press.

Huang, X. Acero, A and Hon, H (2001). *Spoken Language Processing*, New Jersey, Prentice-Hall.

Ghayoomi, M., Momtazi, S, and Bijankhan, M. (2004) "A Study of Corpus Development for Persian", International *Journal on Asian Language Processing* Vol. 20, No 1, 17-33.

Vaseghi, S. (2007). *Multimedia Signal Processing, Theory and Application in Speech, Music and Communication*. West Sussex: John Wiley Publication.