طراحی و ارزیابی یک سامانهٔ بازسازی گفتار به روش هم‌گذاری واحدهای حساس به بافت نوایی

وحید صادقی
دانشگاه بین المللی امام خمینی

چکیده:

در این مقاله، سه گونهٔ گفتاری متفاوت از کلماتی یکسان از طریق هم‌گذاری هجاها به سه روش متفاوت بازسازی شدند. در روش اوّل، کلمات از طریق هم‌گذاری هجاهای حساس به بافت نوایی بدون پردازش فرکانس پایه و دیرش هجاها ساخته شدند. در روش دوم، ابتدا کلمات با هجاهای غیرنوایی هم‌گذاری و سپس هجاهای درون کلمات هم‌گذاری شده با توجه به جایگاه نوایی کلمات در سطح جمله در حوزهٔ فرکانس پایه و دیرش پردازش شدند. در روش سوم، کلمات بدون هر گونه پردازش نوایی از طریق هم‌گذاری هجاهای غیرنوایی ساخته شدند. کلمات بازسازی‌شده سپس در جایگاه‌های نوایی آغاز و پایان IP، AP و PW در درون سه جملهٔ حامل قرار داده شده و پس از تولید در یک آزمون ادراکی به آزمودنی‌ها ارائه شدند. نتایج آزمون ادراکی نشان داد که کیفیت گفتار بازسازی‌شده نوع اوّل و دوم در تمامی جایگاه‌های نوایی به‌طور معنی‌داری از نوع سوم طبیعی‌تر است. هم‌چنین بر اساس نتایج به دست آمده کیفیت گونه‌های گفتاری نوع اوّل و دوم در حد "طبیعی" است ولی میزان طبیعی بودن گونهٔ گفتاری نوع اوّل در تمامی جایگاه‌ها از نوع دوم بیشتر است و این تفاوت در بعضی از جایگاه‌ها معنی‌دار است. این نتایج نشان می‌دهد واحدهای بازسازی حساس به بافت نوایی، کیفیت گفتار بازسازی‌شده را تا حدّ قابل قبولی بهبود می‌بخشد.

واژگان کلیدی: بازسازی گفتار، واحدهای نوایی، گروه آهنگ، گروه زیروبمی، کلمهٔ واجی، تغییرات زنجیره‌ای و زبرزنجیره‌ای

۱- مقدمه

در برنامه‌های بازسازی گفتار به روش هم‌گذاری، برابرهای آوایی واحدهای بازسازی از پیش ذخیره شده و طی فرایند هم‌گذاری بازخوانی و در کنار قرار می گیرند *(Sato, 1992)*. در این برنامه‌ها بازسازی نوای گفتار در بخشی جداگانه با عنوان واحد پردازش نوایی انجام می‌شود. در این بخش منحنی زیروبمی گفتار طی دو مرحله بازسازی می‌شود. در مرحلهٔ اوّل، الگوی نوایی جملات در دو سطح واجی و آوایی تعیین می‌شود. در سطح واجی الگوی آهنگ به‌صورت توالی از نواخت‌ها مانند نواخت L، H، *L-H و غیره *(Ladd, 1980)* و *(Pierehumbert, 1981)* یا الگوهای زیروبمی کیفی مانند خیزان، افتان *(Ladd, 1980)* بازنویسی می‌شود. در سطح آوایی مقادیر مرجع فرکانس پایه و ترادف زمانی (نسبت زمانی نواخت با زنجیره آوایی) برای هر نواخت یا الگوی زیروبمی کیفی مشخص می‌شود *(Pierehumbert, 1981)*. این مقادیر با محاسبهٔ میانگین مقادیر فرکانس پایه و ترادف زمانی نواخت‌ها از سطح پیکره‌های گفتاری به‌دست می‌آید. در مرحلهٔ دوم زنجیره آوایی هم‌گذاری‌شده متناسب با الگوی نوایی جمله و مطابق با مقادیر مرجع فرکانس پایه و ترادف زمانی در دو حوزهٔ فرکانس و زمان تنظیم می‌شود و به این ترتیب برابرهای آوایی جدیدی برای هر واحد بازسازی متناسب با بافت نوایی جمله مورد نظر ساخته می‌شود.

این روش یکی از رایج‌ترین روش‌های بازسازی آهنگ است و در تهیهٔ بسیاری از برنامه‌های بازسازی مانند فستیوال *(Alan, et al., 1999)* و *(Alan and Lenzo, 2033)* و می تاک *(Allen, et al., 1987)* مورد استفاده قرار گرفته است.

درهمین اواخر روش دیگری در بازسازی نوای گفتار مطرح شده است که بر اهمیت جایگاه و بافت نوایی در تهیهٔ برابرهای آوایی واحدهای بازسازی و تأثیر آن بر کیفیت گفتار بازسازی‌شده تأکید بسیار دارد. این روش مبتنی بر این فرض

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

است که با انتخاب برابرهای آوایی یک واحد بازسازی از بافت‌های نوایی متنوع و ذخیره‌سازی آن‌ها در دادگان برنامه، واحدهای بازسازی با حفظ ویژگی‌های طبیعی زنجیره‌ای و زبرزنجیره‌ای خود در جایگاه‌های نوایی مختلف وارد زنجیرهٔ گفتار بازسازی‌شده می‌شوند که این مسئله باعث می‌شود کیفیت گفتار بهبود یابد *(Yoon, 2004 and 2008)* و *(Hirschberg and Prieto, 1996)*. با وجود واحدهای حساس به بافت نوایی اوّل آن که واحدها با حفظ ویژگی‌های نوایی طبیعی و ذاتی خود در جایگاه‌های نوایی مختلف از جمله دیرش، زیروبمی و شدت انرژی وارد زنجیره گفتار بازسازی‌شده می‌شوند و حجم تغییرات و تنظیمات نوایی لازم برای بازسازی آهنگ گفتار پس از هم‌گذاری واحدها کاهش می‌یابد. دیگر آن که تغییرات زنجیره‌ای پیوستهٔ گفتار طبیعی در گفتار باز سازی‌شده حفظ می‌شود.

مطالعات انجام‌شده بر روی زبان‌های مختلف نشان می‌دهد که برخی ویژگی‌های زنجیره‌ای گفتار به‌صورت تابعی از الگوی نوایی واحدها قابل تغییر است. در *(Pierehumbert and Talkin, 1992)* طی یک آزمایش دربارهٔ رابطهٔ بین میزان فعّالیت هم‌خوان‌های چاکنایی /h/ و /ʔ/ و ساخت نوایی زبان انگلیسی نتیجه‌گیری شده است که عواملی هم‌چون تکیهٔ واژگانی، تکیهٔ زیروبمی و جایگاه این هم‌خوان‌ها نسبت به مرز کلمهٔ واجی[1]، گروه واجی[2] و گروه آهنگ[3] بر میزان فعّالیت این هم‌خوان‌ها اثر می‌گذارد. به‌طور مثال توزیع واک بازداشته در واکه‌های آغازی کلمه منطبق بر مرز گروه آهنگ در مقایسه با مرزهای نوایی کوچک‌تر بیشتر است. در *(Jun, 1993)* با اندازه‌گیری و مقایسه مقدار زمان شروع واک (VOT) برای انسدادی نادمیده /p^h/ در زبان کره‌ای در سه جایگاه نوایی آغاز گروه زیروبمی[4]، آغاز کلمهٔ واجی و میان کلمه واجی نتیجه‌گیری شده است که مقدار این پارامتر در سطوح بالاتر سلسله مراتب نوایی در مقایسه با سطوح پایین‌تر بیشتر است. به‌علاوه مقدار آن در جایگاه‌های آغازی سطوح نوایی در مقایسه با جایگاه‌های میانی بیشتر است. در *(Smith, 1997)* در یک آزمایش تولیدی دربارهٔ واکرفتگی در زبان انگلیسی نشان داده شده است که میزان واکرفتگی این هم‌خوان سایشی بسته به جایگاهش در ساخت سلسله‌مراتبی نوای گفتار متفاوت است به این صورت که واکرفتگی در پایان واحدهای نوایی کوچک‌تر مانند هجا و کلمهٔ واجی، کم‌تر از واحدهای نوایی بزرگ‌تر مانند گروه زیروبمی و گروه آهنگ است. وی این فرضیه را مطرح کرد که فعّالیت الگوهای تولیدی در پایان واحدهای نوایی بزرگ‌تر بیش‌تر از واحدهای کوچک‌تر کاهش می یابد و در مقابل، فعّالیت این الگوها در آغاز واحدهای بزرگ‌تر در مقایسه با واحدهای نوایی کوچک‌تر افزایش می‌یابد. در *(Keating, et al., 1998)* نیز طی یک آزمایش تولیدی نشان داده شده که هم‌خوان های /t/ و /n/ در جایگاه آغازه هجا در سطح واحدهای نوایی مختلف در زبان‌هایی هم‌چون انگلیسی، فرانسه، کره‌ای و تایوانی با الگوهای تولیدی متفاوت تولید می‌شود. هم‌چنین در آزمایش تولیدی بر روی /t, t^h,t^*, n/ در زبان کره‌ای *(Cho and Keating, 2001)* نشان داده شده که این هم‌خوان‌ها در سطوح بالاتر سلسله‌مراتب نوایی تولید قوی‌تری دارند چون هم سطح تماس زبان-کامی در این سطوح بیشتر است و هم دیرش زمانی ناحیهٔ بست طولانی‌تر است. بررسی تغییرات واج‌گونه‌ای پیوسته حساس به بافت نوایی در زبان کره‌ای *(Yoon, 2008)* نشان داد که بخش آغازی واکه /a/ بعد از سایشی دمیده /s^h/ در مقایسه با سایشی سخت /s^*/ نفسی‌تر است و میزان نفسی شدن واک بسته به جایگاه هم‌خوان سایشی در سطوح مختلف ساخت نوایی گفتار مانند کلمهٔ واجی، گروه زیروبمی و گروه آهنگ متفاوت است. این تغییرات پیوستهٔ زنجیره‌ای در زبان فارسی نیز قابل مشاهده است. به‌طور مثال تکیهٔ واژگانی در زبان فارسی بر مقدار زمان شروع واک یا دیرش قطعه رهش انسدادی‌های بی‌واک اثر می‌گذارد (صادقی، ۱۳۸۶ ; نوربخش، ۱۳۸۸). دیرش قطعه رهش انسدادی بی‌واک [t^h] در آغاز هجای تکیه‌بر کلمهٔ "کتاب" در مقایسه با آغاز هجای فاقد تکیه در کلمهٔ "تابان" طولانی‌تر است.

بر این اساس می‌توان گفت که ماهیت بسیاری از تغییرات و تقابل‌های آوایی در ساخت زنجیره‌ای گفتار تنها در قالب الگوی نوایی گفتار معلوم می‌گردد و روشن است که برنامه‌های بازسازی گفتار برای بهبود کیفیت گفتار بازسازی‌شده باید این تغییرات آوایی حساس به بافت نوایی را در برنامه لحاظ کنند. در اکثر برنامه‌های بازسازی گفتار، تمامی تغییرات زنجیره‌ای به غیر از تغییرات هم‌تولیدی در قالب قواعد بازنویسی[5] (برنامهٔ بازسازی می‌تاک[6]) یا قواعد

[1] Phonological word
[2] Phonological phrase
[3] Intonational phrase
[4] Accentual phrase

[5] Rewrite rules
[6] MItalk

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

پساواژگانی[1] (برنامهٔ بازسازی فستیوال) به سبک قواعد واج‌گونه‌ای اس‌پی‌ای در برنامه لحاظ می شود *(AlanBlack, et al., 1999)*. به‌طور مثال، بخش واجی برنامهٔ می‌تاک شامل شش قاعدهٔ بازنویسی زنجیره‌ای است که در مرز کلمه و گروه عمل می‌کنند. یکی از این قواعد، قاعدهٔ زنشی‌شدگی[2] است که بر اساس آن /t/ و /d/ بعد از یک رسای غیرخیشومی و قبل از واکه با تکیهٔ غیراصلی به‌صورت زنشی تولید می‌شوند مثل sat on a chair، *(Allen, et al., 1987)*. بعضی از همین قواعد در برنامهٔ بازسازی فستیوال در بخش پساواژگانی برنامه اعمال می‌شوند *(Alan and Lenzo, 2003)*. امّا همان‌طورکه طراحان این برنامه‌ها نیز پذیرفته‌اند، این بخش برنامه نیاز به تحقیقات بیش‌تر دارد. مسئله این است که تغییرات واج‌گونه‌ای حساس به بافت نوایی را نمی‌توان به‌طور کامل و منظم به این روش در برنامه اعمال کرد. به‌عنوان مثال اگرچه تناوب‌های مقوله‌ای مانند تناوب انسدادی/ زنشی $t \approx r$ در sat on the chair و at all و هم‌چنین دمیده/ نادمیده $k \approx k^h$ در the car و discover را می‌توان با استفاده از قواعد بازنویسی یا پساواژگانی در برنامه لحاظ کرد، ولی این قواعد نمی‌توانند تغییرات پیوستهٔ دمیدگی را در بافت‌های نوایی متفاوت مثل this car، this career و . . . اعمال کنند. برای این که این تغییرات را در برنامه اعمال کنیم، بهترین راه این است که واحدهای بازسازی را از سطح بافت‌های نوایی متفاوت انتخاب کنیم. با به‌کارگیری واحدهای حساس به بافت نوایی، تغییرات نوایی و تغییرات زنجیره‌ای پیوستهٔ گفتار به شکل طبیعی در زنجیرهٔ گفتار بازسازی‌شده اعمال شده و کیفیت گفتار بهبود می‌یابد *(Yoon, 2004 and 2008)*. در *(Yoon, 2008)* برای طراحی و ساخت برنامهٔ بازسازی گفتار کره‌ای به شیوه‌گذاری، به‌طور آزمایشی و در مقیاس محدود از دایفون‌های وابسته به بافت نوایی استفاده شده است. برای این منظور وی ابتدا برابرهای آوایی هر دایفون از مجموع دایفون‌های مورد آزمایش (۳۶ ٪ کل دایفون‌های ممکن زبان کره‌ای) را از سطح چهار جایگاه نوایی از نسخهٔ گفتاری پیکره متنی زبان کره‌ای انتخاب و در یک دادگان جداگانه ذخیره کرد. این جایگاه‌ها عبارت بودند از: سه جایگاه کناری در آغاز گروه آهنگ، گروه زیروبمی و کلمهٔ واجی و یک جایگاه غیرکناری در میان کلمهٔ واجی. سپس با هم‌گذاری دایفون‌ها در جایگاه‌های نوایی مناسب چندین گروه آهنگ را بازسازی کرد. وی سپس به‌ازای هر یک از این گروه‌ها گونهٔ دیگری را نیز ساخت که در آن دایفون‌ها در جایگاه نامناسب نوایی قرار داده شدند. در پایان در یک آزمایش شنیداری از آزمودنی‌ها خواست کیفیت گفتار بازسازی‌شده را ارزیابی کنند. نتایج نشان داد که آزمودنی‌ها گونه‌هایی را که در آن‌ها دایفون‌های وابسته به بافت نوایی به کار رفته بود طبیعی می‌دانند.

۲- طراحی آزمایش

در این بخش با انجام یک آزمایش شنیداری، کیفیت سه گونهٔ گفتار بازسازی‌شده را با مورد مقایسه قرار می‌دهیم. گونهٔ گفتاری نخست، مبتنی بر دادگان واحدهای حساس به بافت نوایی است که تمامی تغییرات طبیعی پیوستهٔ آوایی در سطح واحد بازسازی را در هم به‌لحاظ ویژگی‌های زنجیره‌ای و هم زبرزنجیره‌ای شامل می‌شود و بنابراین زنجیره‌ای که از طریق هم‌گذاری واحدهای حساس به بافت نوایی بازسازی می‌شود، در درون واحدهای بازسازی به کم‌ترین پردازش آوایی نیاز دارند. گونهٔ گفتاری دوم بر پایهٔ دادگانی ساخته می‌شود که در آن به‌ازای هر واحد بازسازی تنها یک برابر آوایی مستقل از بافت نوایی وجود دارد. این برابر آوایی در تمامی جایگاه‌های نوایی آن واحد در زنجیرهٔ گفتار قرار می‌گیرد و پس از هم‌گذاری واحدها، برابر آوایی مورد نظر در حوزهٔ فرکانس و زمان تنظیم می‌شود تا سایر برابرهای آوایی آن واحد به تناسب جایگاه نوایی آن بازسازی شوند، ولی هیچ‌گونه پردازش زنجیره‌ای برای اعمال تغییرات زنجیره‌ای پیوسته در گفتار طبیعی بر روی آن صورت نمی‌گیرد. گونهٔ گفتاری سوم نیز هم‌چون گونهٔ دوم، بر پایهٔ واحدهای مستقل از بافت نوایی ساخته می‌شود با این تفاوت که زنجیرهٔ گفتاری نوع سوم تنها از صورت هم‌گذاری‌شده واحدها به دست می‌آید و هیچ‌گونه پردازش یا تنظیمات آوایی بر روی واحدها انجام نمی‌شود. سؤالات تحقیق عبارتند از: ۱- اعمال (گونه‌های اوّل و دوم) یا عدم اعمال (گونهٔ سوم) تغییرات نوایی در بازسازی گفتار تا چه حدّ بر کیفیت گفتار بازسازی‌شده اثر می‌گذارد.

۲- اعمال (گونهٔ اوّل) تغییرات زنجیره‌ای پیوسته در مقابل عدم اعمال این تغییرات تا چه اندازه بر کیفیت گفتار بازسازی شده اثر می‌گذارد.

۲-۱- دادگان آزمایش

در این آزمایش هجا به‌عنوان واحد بازسازی انتخاب شد. هجا در زبان فارسی دارای سه ساخت CV، CVC و CVCC است

[1] Post-lexical rules
[2] Flapping rule

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

که تعداد هم‌خوان‌ها ۲۳ و تعداد واکه‌ها ۶ عدد است *(ثمره، ۱۳۷۸: ۱۰۸-۱۱۱)*. ساخت سلسله‌مراتبی گفتار فارسی شامل سه واحد نوایی است: گروه آهنگ[1]، گروه زیروبمی[2] و کلمهٔ واجی *(سادات تهرانی، ۲۰۰۷)*. گروه آهنگ بزرگ‌ترین واحد نوایی در سلسله‌مراتبی گفتار فارسی است. هر گروه آهنگ، متناظر با یک جملهٔ ساده است. گروه زیروبمی واحد نوایی کوچک‌تر از گروه آهنگ است که متناظر با یک گروه نحوی مانند گروه اسمی، فعلی یا حرف اضافه است که در آن، هجای پایانی تکیهٔ اصلی واقع می‌شود. در صورت وجود واژه‌بست‌ها هجای پیش از واژه‌بست دارای تکیهٔ اصلی است. کلمهٔ واجی نیز واحد نوایی کوچک‌تر از گروه زیروبمی است که متناظر با یک واژهٔ قاموسی به همراه واژه‌بست‌های آن است. یک هجا مانند /ba/ در زبان فارسی می‌تواند در آغاز (لبهٔ سمت راست) یا پایان (لبهٔ سمت چپ) این سه واحد نوایی متفاوت قرار بگیرد: آغاز یا پایان یک گروه آهنگ (IP)؛ آغاز یا پایان یک گروه زیروبمی (AP) و آغاز یا پایان یک کلمهٔ واجی (PW). به‌طور مثال، /ba/ در *"برنده شدم"* در جایگاه آغاز IP؛ در *"ایران برنده شد"* در جایگاه آغاز AP و در *"برای برنده شدن تلاشی نکرد"* در جایگاه آغاز PW است. برای تشخیص جایگاه‌های نوایی از نشانه‌های مرزی < >، [ ]، { } به ترتیب برای گروه آهنگ، گروه زیروبمی و کلمهٔ واجی استفاده می‌شود. بر این اساس برابرهای آوایی این هجا در آغاز جایگاه‌های نوایی IP، AP و PW به‌ترتیب به‌صورت ba >، ba] و {ba} و در پایان همین جایگاه‌ها به‌صورت <ba، [ba و ba} است. این برابرهای آویی وابسته به بافت نوایی را هجاهای نوایی می‌نامیم. فرض ما بر آن است که یک واحد آوایی مانند b در سطح این جایگاه‌های نوایی به‌صورت متفاوت تظاهر می‌یابد. یک جملهٔ کوتاه مانند *"یک شبکهٔ خبری تأسیس شد"* که شامل یک IP، دو AP، پنج PW و ۱۱ هجا است، دارای ساخت نوایی زیر است:

< [{yek} {ʃa.ba.ke.ye} {χa.ba.ri}] [{taʔ.sis}{ʃod}]>

<[{CVC}{CV.CV.CV.CV}{CV.CV.CV}][{CVC.CVC}{CVC}]>

انسدادی‌های بی واک و واک‌دار زبان فارسی به عنوان همخوان‌های هدف انتخاب شدند. به‌ازای هر یک از انسدادی‌های بی واک /p/، /t/ و /k/ یک کلمه با ساخت هجایی CV(C).CVC(C) انتخاب شد، به‌طوری‌که هم‌خوان‌های مورد نظر در جایگاه آغاز هجای اوّل قرار داشتند. این کلمات عبارتند از پرداخت /par.dɑχt/، ترمیم /tar.mim/ و /kɑr.gar/. این کلمات را کلمات نوع اوّل می‌نامیم. هم‌چنین به‌ازای انسدادی‌های واک‌دار /b/، /d/ و /g/ یک کلمه با ساخت هجایی مشابه انتخاب شد به‌طوری‌که این هم‌خوان‌ها در جایگاه آغاز هجای دوم قرار داشتند: نگار /ne.gɑr/، شدید /ʃa.did/ و دوربرد /durbord/. این کلمات را کلمات نوع دوم می‌نامیم. بر این اساس انسدادی‌های بی‌واک در درون هجاهای فاقد تکیهٔ واژگانی و انسدادی‌های واک‌دار در درون هجاهای دارای تکیهٔ واژگانی قرار داشتند. به‌ازای هر کلمهٔ نوع اوّل، سه جملهٔ حامل کوتاه ساخته شد؛ به‌طوری‌که کلمهٔ هدف هر بار در یکی از جایگاه‌های نوایی آغاز IP، آغاز AP و آغاز PW گرفت. هم‌چنین به‌ازای هر کلمهٔ هدف نوع دوم، سه جملهٔ حامل دیگر ساخته شد؛ به‌طوری‌که کلمهٔ هدف هر بار در یکی از جایگاه‌های نوایی پایان IP، پایان AP و پایان PW گرفت. به این ترتیب به‌ازای هر هم‌خوان انسدادی، سه هجای نوایی ساخته شد؛ یعنی درمجموع هیجده هجای نوایی که از این تعداد ۹ هجا مربوط به آغاز یا لبهٔ سمت چپ واحدهای نوایی و ۹ هجای دیگر مربوط به پایان یا لبهٔ سمت راست واحدها هستند. لازم به ذکر است که برای کنترل اثر متغیر بافت آوایی بر هجاهای هدف، کلمات با ساخت هجایی CVC.CVC(C) طوری انتخاب شدند که هم‌خوان‌های واقع در مرز هجا واک‌دار باشند. مثال‌های ۱-۳ کلمات هدف نوع اوّل را در جایگاه آغاز IP (الف)، AP (ب) و PW (ج) و مثال‌های ۴-۶ کلمات هدف نوع دوم را در جایگاه پایان همان واحدهای نوایی نشان می‌دهد.

۱- "پرداخت" / par.dɑχt /

الف) پرداخت کردم. <[{par.dɑχt} {kar.dam}]>

ب) علی پول را به من پرداخت کرد.

<[{ʔali}] [{pul} {ro}] [{be}{man}] [{par.dɑχt} {card}]>

ج) از پرداخت کردن قسط خسته شدم.

<[{ʔali}] [{pul} {ro}] [{be}{man}] [{par.dɑχt} {card}]>

۲- "ترمیم" /tar.mim/

الف) ترمیم کردم. <[{tar.mim}{kar.dam}]>

ب) بافت آسیب‌دیده را ترمیم کرد.

<[{bɑf.te}{ʔɑ.sib.di.de}{ro}] [{tar.mim} {card}]>

ج) بافت پس از ترمیم کردن بهبود می‌یابد.

<[{bɑft}][{pas.az}{tar.mim}{kar.dan}][{beh.bud}{mi.yɑ.bad}]>

۳- "کارگر" / kɑr.gar /

الف) کارگر است. <[{kɑr.gar}{ʔast}]>

ب) علی کارگر است. <[{ʔali}][{kɑr.gar}{ʔast}]>

[1] Intonational phrase
[2] Accentual phrase

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

ج) از کارگر گرفتن منصرف شد.

<[{?az}{kar.gar}{ge.ref.tan}] [{mon.sa.ref}{ʃod}]>

۴- "نگار" /ne.gar/

الف) ناراحته از نگار. <[{na.ra.ha.te}][{?az}{ ne.gar}]>

ب) پیش نگار برد. <[{pi.ʃe}{ne.gar}][{bord}]>

ج) به نگار گفتن مشکلی را حل نمی‌کنه.

<[{be}{ ne.gar}{gof.tan}][{moʃ.ce.li}{ro}][{hal}{ne.mi.ko.ne}]>

۵- "شدید" /ʃa.did/

الف) (چرا از حال رفت؟) از ترس شدید.

<[{?az}{tar.se}{ʃa.did}]>

ب) از ترس شدید سکته کرد.

<[{?az}{tar.se}{ʃa.did}][{sec.te}{card}] >

ج) شوک شدید دادن بی فایده است.

<[{ʃo.ce}{ʃa.did}{da.dan}][{bi.fa.ye.de}{?ast}]>

۶- "دوربرد" /durbord/

الف) (با چی شلیک کرد؟) با سلاح دوربرد.

<[{ba}{se.la.he}{dur.bord}]>

ب) با سلاح دوربرد شلیک کرد.

<[{ba}{se.la.he}{dur.bord}][{ʃel.lic}{card}]>

ج) دوربرد بودن سلاح کمکی نمی‌کنه.

<[{dur.bord}{bu.da.ne}{se.lah}][{ko.ma.ci}][{ne.mi.ko.ne}]>

۲-۲- روش آزمایش

هیجده جملهٔ ساخته‌شده (۳ واحد آوایی x ۳ جایگاه نوایی x ۲ نوع کلمه) توسط یک گویشور مرد تولید شدند. ضبط جملات در یک اتاق آرام در دانشگاه امام خمینی انجام شد. داده‌ها با استفاده از میکروفون شور با پاسخ بسامدی مناسب بر روی کارت صوتی یک رایانهٔ شخصی ضبط شدند. پس از ضبط جملات مراحل زیر انجام شد:

۲-۲-۱- تقطیع هجا

تمامی هجاهای کلمات هدف تقطیع و استخراج شدند. تعیین دقیق مرز هجاها در فضای آوایی به علت هم‌پوشی و هم‌تولیدی واج‌ها دشوار است. نگارنده برای تعیین مرز هجا در زنجیره‌های آوایی V.C و C.C[1] که تنها مرزهای هجایی ممکن فارسی هستند، از معیارهای ارائه شده در *(Lewis and Tatham, 2001)* پیروی کرده است. بر این اساس، در زنجیرهٔ آوایی V. stop مرز هجا جایی است که دامنهٔ ارتعاش تارآواها برای واکه در هجای اوّل به‌میزان قابل‌توجّه‌ای کاهش می‌یابد. این ناحیه ناظر بر آغاز محدودهٔ بست هم‌خوان انسدادی است. برای هم‌خوان‌های انسدادی بی‌واک تشخیص پایان واکه آسان است، چون به‌دنبال واکهٔ یک دورهٔ بی‌واکی کامل همراه با سکوت و نوفه ظاهر می‌شود ولی برای انسدادی‌های واک‌دار این تشخیص سخت‌تر است؛ چون ارتعاش تارآواها برای واکه در هجای اوّل به‌مدّت ۲۰ تا ۳۰ میلی‌ثانیه پس از شروع هجای دوم ادامه می‌یابد. الگوی تقطیع هجا برای V.fricative مشابه V. stop است با این تفاوت که برای سایش‌های واک‌دار نوفه و واک در پایان واکه به‌طور هم‌زمان تظاهر می‌یابد. در زنجیرهٔ آوایی V. liquid چون تظاهر هم‌خوان‌های روان بر روی طیف‌نگاشت به‌صورت امواج پیچ در پیچ و مدور است، شروع این ناحیه به‌عنوان نقطهٔ شروع هجای دوم در نظر گرفته می‌شود. تعیین لحظهٔ شروع این ناحیه بر روی طیف‌نگاشت آسان نیست. به‌طور معمول کمینهٔ حدّاقل میزان دامنهٔ این ناحیه، محل شروع هجای دوم محسوب می‌شود. برای زنجیره stop. fricative/affricate مرز هجا، پایان قطعه رهش هم‌خوان انسدادی و آغاز قطعه سایش هم‌خوان سایشی یا انسدای‌سایشی است. در زنجیرهٔ liquid. stop مرز هجا ناحیه‌ای است که در آن ارتعاش تارآواها بلافاصله پس رهش هم‌خوان انسدادی شروع شده و یا دامنهٔ آن به‌طور قابل‌توجّه‌ای افزایش یابد. برای زنجیرهٔ fricative. fricative/affricate مرز هجا وسط قطعهٔ سایش این هم‌خوان‌هاست. یکی دیگر از نشانه‌های این مرز هجایی نحوهٔ توزیع انرژی در نواحی مؤثر فرکانسی است. توزیع انرژی در فرکانس‌های مختلف با توجّه به محل تولید این هم‌خوان‌ها متفاوت است. به‌عنوان مثال برای [s] توزیع انرژی روی فرکانس‌های بالای ۴۰۰۰ هرتز است درحالی‌که برای [f] روی فرکانس‌های پایین است. بر این اساس اگر [s] و [f] در مرز هجا در مجاورت ظاهر شوند، مرز آن‌ها را می‌توان از طریق توزیع نابرابر انرژی روی فرکانس‌ها بر روی طیف‌نگاشت تشخیص داد. در زنجیرهٔ آوایی fricative. liquid محل شروع هجای دوم آغاز ارتعاش تارآواها یا افزایش دامنهٔ آن پس از پایان قطعهٔ سایش هم‌خوان سایشی است. برای liquid.stop/fricative/affricate اگر هم‌خوان‌های گرفته بی‌واک باشند، مرز هجا هم‌زمان با لحظهٔ پایان ارتعاش تارآواها برای هم‌خوان روان است. برای هم‌خوان‌های گرفتهٔ واک‌دار ارتعاش در نواحی مرزی ادامه دارد؛ بنابراین لحظه‌ای که دامنهٔ ارتعاش

[1] ساخت هجا در زبان فارسی به‌صورت CV، CVC و یا CVCC است بنابراین مرز هجا در زنجیرهٔ گفتار فارسی به دو صورت است: یکی V.C و دیگری C.C

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

به‌طور ناگهانی کاهش می‌یابد به‌عنوان مرز هجا در نظر گرفته می‌شود. هم‌چنین با توجّه به تأثیر الگوی فرکانسی هم‌خوان‌های خیشومی بر بافت زنجیره‌ای مجاور، مرزهای هجایی شامل هم‌خوان‌های خیشومی ازجمله liquid.nasal، fricative.nasal، stop.nasal و V. nasal برای ساخت کلمات هدف انتخاب نشدند. هم‌چنین چون دیرش قطعه رهش انسدادی‌های بی‌واک بعد از سایشی‌های بی واک /s/ و /ʃ/ به‌طور قابل ملاحظه‌ای کوتاه‌تر می‌شود، هجای دوم در مرز هجایی fricative. stop برابر آوایی مناسبی برای بازسازی دیگر موارد وقوع آن هجا در بافت‌های آوایی دیگر نیست. بنابراین این مرز هجایی نیز در انتخاب کلمات هدف لحاظ نشد. پس از تقطیع و استخراج هجاها، هجاهای هر کلمه به تفکیک جایگاه نوایی در پرونده‌ای جداگانه ذخیره شدند.

۲-۲-۲- بازسازی گونه‌های گفتاری

در مرحلهٔ بازسازی، سه گونهٔ گفتاری متفاوت از کلماتی یکسان از طریق هم‌گذاری هجاهای استخراج‌شده ساخته شدند. برای بازسازی کلمات، هجاهای هدف کلمات نوع اوّل (هجاهای اوّل) با هجاهای هدف کلمات نوع دوم (هجاهای دوم) هم‌گذاری شدند. به این ترتیب سه کلمهٔ با معنی ساخته شد: "پرگار" /pargɑr/ از هم‌گذاری هجای /par/ در "پرداخت" و /gɑr/ در "کارگر"؛ "تردید" /tardid/، از هم‌گذاری /tar/ از "ترمیم" و /did/ از "شدید" و "کاربرد" /kɑrbord/ از هم‌گذاری/kar/ از "کارگر" و /bord/ از "دوربرد" /durbord/. بازسازی این کلمات به سه روش متفاوت انجام شد. در روش اوّل، کلمات هدف از طریق هم‌گذاری هجاهای نوایی ساخته شدند. بر این اساس، برابرهای نوایی هجاهای هدف نوع اول با برابرهای نوایی متناظرشان از هجاهای نوایی نوع دوم هم‌گذاری شدند. به‌طور مثال، برای بازسازی کلمه "پرگار" /pargɑr/ یک بار هجای نوایی par IP> (از کلمه "پرداخت"مثال ۱ الف) با هجای نوایی gɑr> IP (از کلمه "نگار" ۴ الف)؛ بار دیگر هجای نوایی par AP] با هجای نوایی gɑr] AP و در مرحله پایانی هجای نوایی{PW par} با هجای نوایی pw {par هم‌گذاری شدند. همین مراحل برای کلمات "تردید" و "کاربرد" هم تکرار شد. به این ترتیب به ازای هر یک از این کلمات، سه گونه حساس به بافت نوایی با در نظر گرفتن جایگاه نوایی هجاهای هدف ساخته شدند. سپس به‌ازای هر یک از کلمات "پرگار"، "تردید" و "کاربرد" سه جمله حامل ساخته و توسط همان گویشور تولید شد. این کلمات در جملات اوّل در جایگاه نوایی IP؛ در جملات دوم در جایگاه نوایی AP و در جملات سوم در جایگاه PW قرار داشتند. سرانجام، گونه‌های نوایی بازسازی شدهٔ مربوط به هر کلمه در درون این جملات به جای همان کلمات با الگوی نوایی مشابه قرار داده شدند. جملات ۷-۹ کلمات "پرگار"، "تردید" و "کاربرد" را در جملات حامل مربوطه در جایگاه آغاز و پایان سطوح نوایی IP (الف)، AP (ب) وPW (ج) نشان می‌دهد:

۷- "پرگار" /pargɑr/

الف) (چی گفت؟) پرگار. <[{par.gɑr}]>

ب) علی پرگار نمی‌خواهد.

<[{ʔa.li}] [{par.gɑr}][{ne.mi.χɑ.had}]>

ج) از پرگار خریدن خسته شدم.

<[{ʔaz}{par.gɑr}{χa.ri.dan}][{χas.te}{ʃo.dam}] >

۸- "تردید" /tar.did/

الف) (چی گفت؟) تردید <[{tar.did}]>

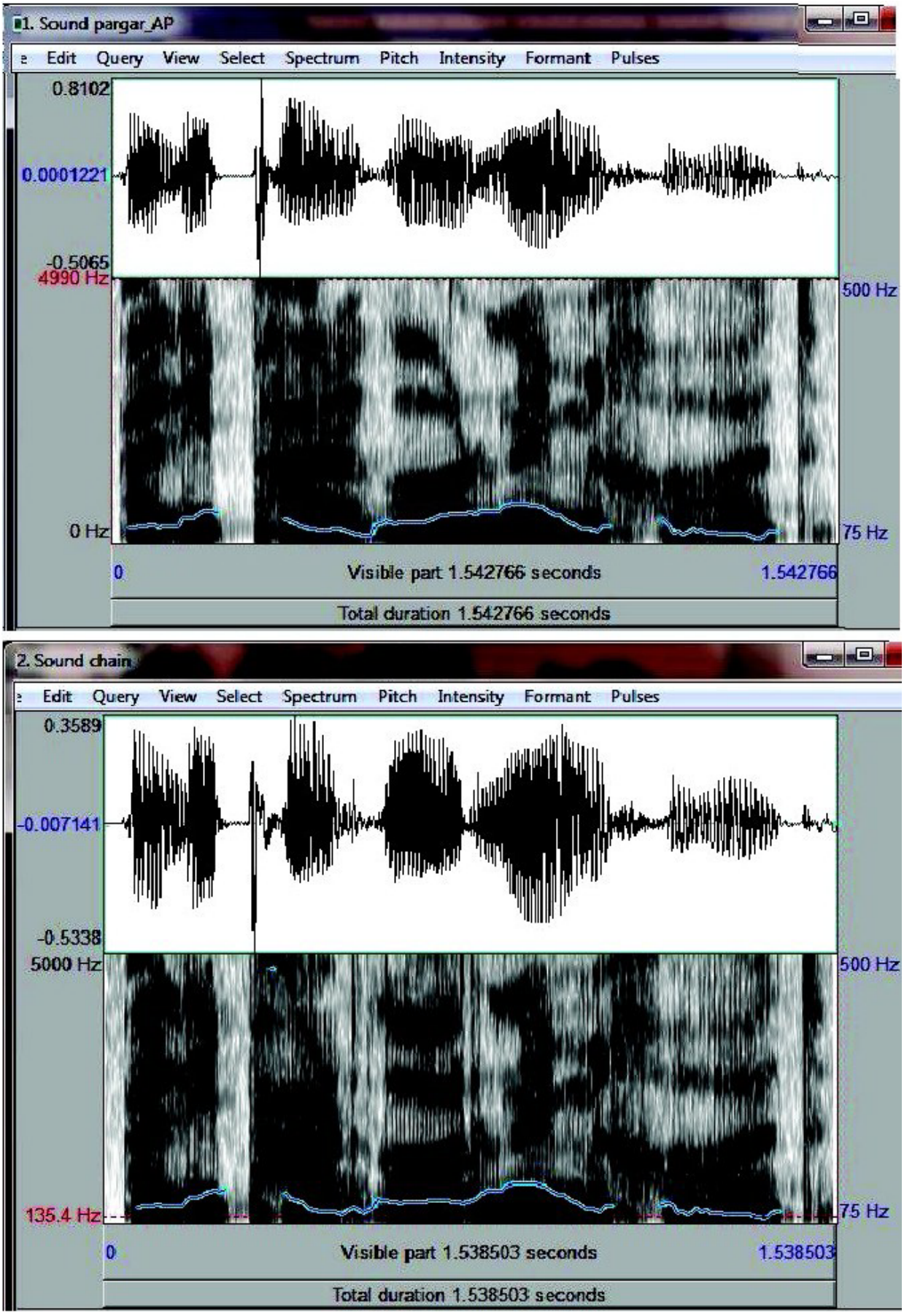
ب) علی تردید نخواهد کرد.

<[{ʔa.li}] [{tar.did}][{na.χɑ.had}{kard}]>

ج) با تردید نگریستن به مسائل.

<[{ba}] [{tar.did}][{ne.ga.ris.tan}{be}{ma.sɑ.ʔel}]>

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]



(شکل ۱): طیف‌نگاشت پاره‌گفتار طبیعی (بالا) و هم‌گذاری شدهٔ (پایین) "علی پرگار نمی‌خواهد"

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

۹- "کاربرد" /kar.bord/

الف) (چی گفت؟) کاربرد. <[{kar.bord}]>

ب) این وسیله کاربرد ندارد.

<[{?in}{va.si.le}][{kar.bord}][{na.dɑ.rad}]>

ج) در صورت کاربرد نداشتن.

<[{dar}{su.ra.te}{kar.bord}{na.dɑʃ.tan}]>

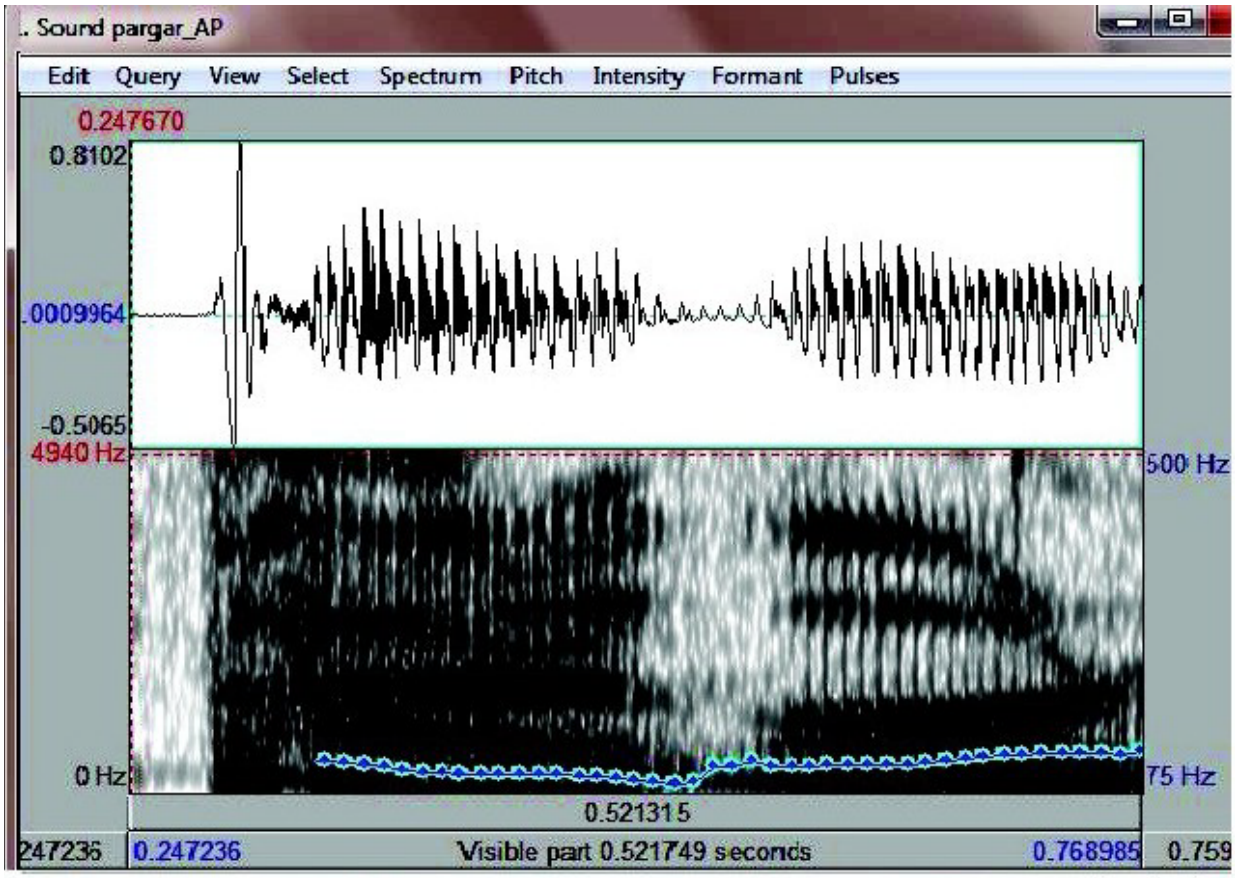
فرض بر آن است کلماتی که به این روش بازسازی می‌شوند ویژگی‌های واج‌گونه‌ای و نوایی گفتار طبیعی را حفظ می‌کنند. روش دوم شامل دو مرحله است. در مرحلهٔ اوّل، با هم‌گذاری برابر نوایی آغاز کلمهٔ واجی به‌عنوان واحد بازسازی هجای هدف نوع اوّل و پایان همین واحد نوایی به‌عنوان واحد بازسازی هجای هدف نوع دوم، یک گونهٔ نوایی کلمهٔ واجی (PW) از کلمات "پرگار"، "تردید" و "کاربرد" ساخته شد. این گونهٔ نوایی سپس در جایگاه‌های نوایی IP و AP در جملات حامل مربوطه (مثال‌های الف و ب در جملات ۷-۹) به‌جای همان کلمات قرار داده شد. هم‌چنین با هم‌گذاری برابرهای نوایی آغاز و پایان گروه زیروبمی یک گونهٔ نوایی گروه زیروبمی (AP) از کلمات مورد نظر ساخته شد. این گونهٔ نوایی سپس در جایگاه نوایی PW در جملات حامل مربوطه (مثال‌های ج در جملات ۷-۹) به جای همان کلمات قرار داده شد. کلماتی که به این روش بازسازی شدند، با الگوی نوایی جملات مربوطه ناسازگار بودند. بر این اساس برابرهای آوایی این کلمات و هجاهای درون آن‌ها گونه‌های غیرنوایی محسوب می‌شوند. در مرحلهٔ دوم، هجاهای غیرنوایی کلمات هم‌گذاری‌شده به روش PSOLA مورد پردازش قرار گرفتند. در این روش فرکانس پایه و دیرش هجاهای خاص در زنجیرهٔ گفتار هم‌گذاری‌شده متناسب با مقادیر این پارامترها در گفتار طبیعی تنظیم می‌شوند و درنتیجه کیفیت گفتار بازسازی شده بهبود می‌یابد. بر این اساس مقادیر فرکانس پایه و دیرش هجاهای اول و دوم در کلمات بازسازی‌شده، متناسب با مقادیر این پارامترها برای همین هجاها در جایگاه‌های نوایی مورد نظر در گفتار طبیعی تنظیم شدند.

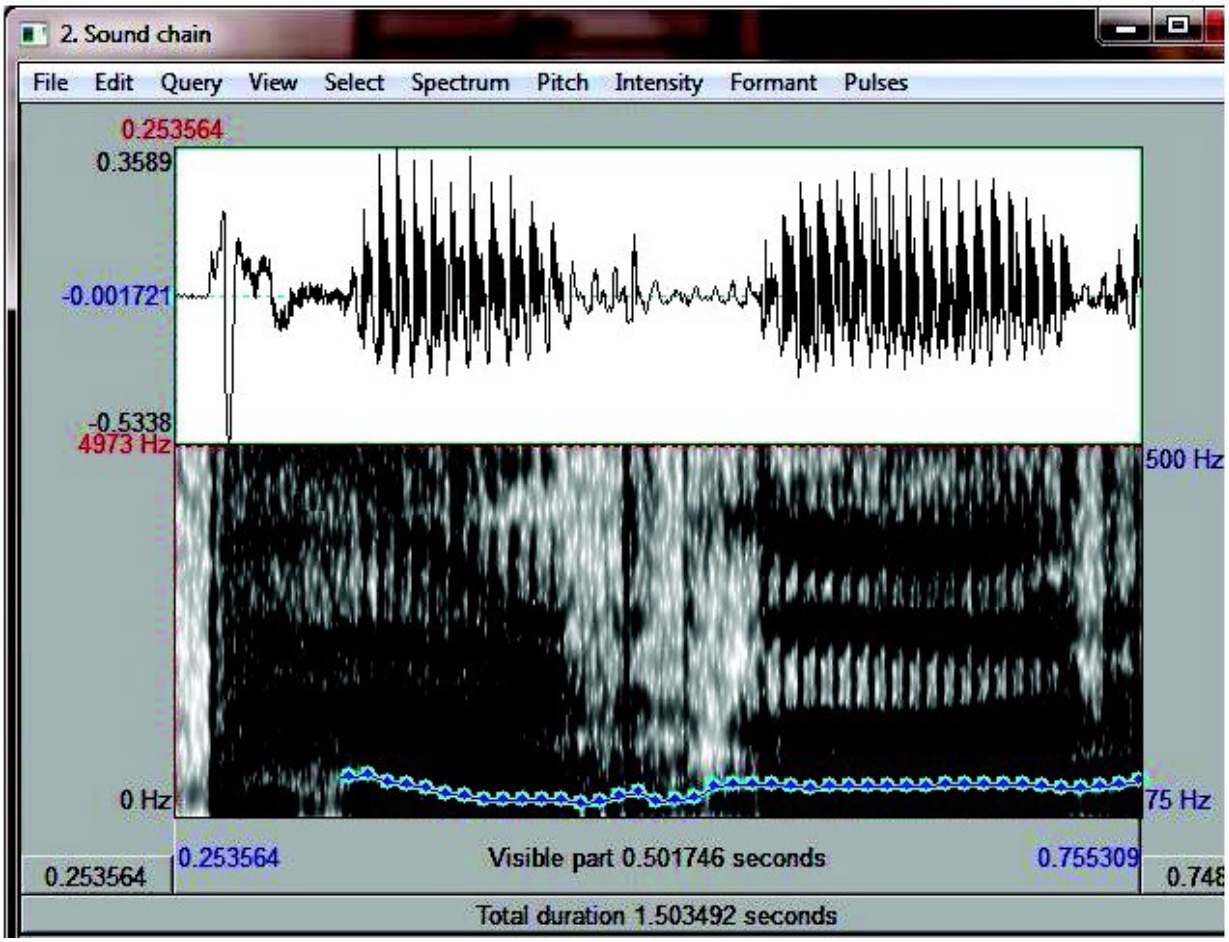
روش سوم مشابه روش دوم است با این تفاوت که زنجیرهٔ آوایی کلمات هدف پس از هم‌گذاری، پردازش نوایی نمی‌شود. در این روش نیز مانند روش دوم، برابر نوایی کلمهٔ واجی برای بازسازی کلمات در جایگاه‌های نوایی IP و AP و برابر نوایی AP برای بازسازی کلمات در جایگاه نوایی PW انتخاب شدند. بعد از هم‌گذاری هجاها و ساخت برابرهای نوایی مورد نظر، هیچ‌گونه پردازشی بر روی هجاهای هدف صورت نگرفت.

برای بازسازی و پردازش نوایی کلمات هدف از نرم افزار پِرَت استفاده شد. برای این منظور برای بازسازی گونه‌های گفتاری نوع اوّل، ابتدا علائم آوایی جملات ۷، ۸ و ۹ در سطح کلمه تقطیع و ذخیره‌سازی شدند. سپس گونه‌های نوایی IP، AP و PW کلمات "پرگار"، "تردید" و "کاربرد" که درقبل از طریق هم‌گذاری هجاهای هدف اول و دوم کلمات دیگر ساخته شده بودند، انتخاب و به جای همان کلمات با الگوی نوایی مشابه در جملات ۷، ۸ و ۹ در کنار دیگر کلمات هم‌گذاری شدند. پس از هم‌گذاری هیچ گونه پردازش نوایی یا زنجیره‌ای بر روی علائم آوایی صورت نگرفت.

(شکل ۱) طیف‌نگاشت گفتار طبیعی (الف) و هم‌گذاری‌شده (ب) جمله "علی پرگار نمی‌خواهد" (مثال ۷ ب) و (شکل ۲) گونهٔ طبیعی (الف) و هم‌گذاری‌شده (ب) کلمه "پرگار" را در همین جمله نشان می‌دهد. چنان‌که مشاهده می‌شود دیرش هجاها و الگوی نوایی گونه‌های طبیعی و بازسازی‌شدهٔ کلمه "پرگار" بسیار به شبیه است. در هر دو گونه، سطح فرکانس پایه برای هجای دوم نسبت به هجای اوّل بالاتر است (f0 هجای اوّل و دوم به‌ترتیب ۱۱۰ و ۱۳۸ هرتز) و این کلمه به‌صورت یک گروه زیروبمی ظاهر شده است. به‌علاوه سطح فرکانس پایه در گفتار هم‌گذاری‌شده علاوه بر بخش پایدار در محدودهٔ گذر هجاها نیز مشابه گفتار طبیعی است. یعنی فرکانس پایه در گفتار هم‌گذاری‌شده، از هجای اوّل به هجای دوم کلمهٔ "پرگار" و از هجای دوم این کلمه به هجای آغازین کلمهٔ "نمی‌خواهد" به‌صورت نرم و پیوسته همانند گفتار طبیعی گذر کرده است. علت این مسئله این است که هجاهای نوایی در گفتار هم‌گذاری‌شده مانند هجاهای زیروبمی par] و [gar در کلمهٔ "پرگار" ویژگی‌های نوایی گفتار طبیعی را علاوه‌بر محدودهٔ پایدار در محدودهٔ گذر هجاها نیز در برمی‌گیرند.

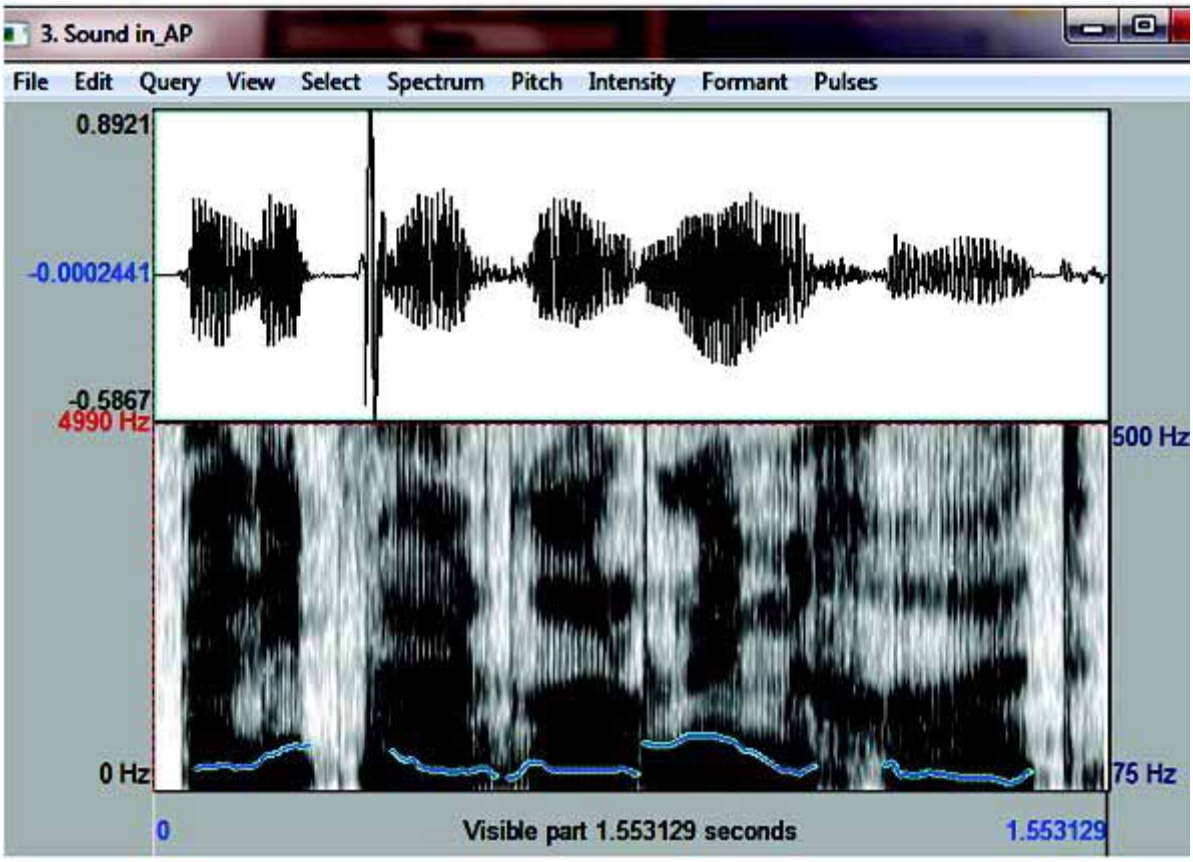
[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]



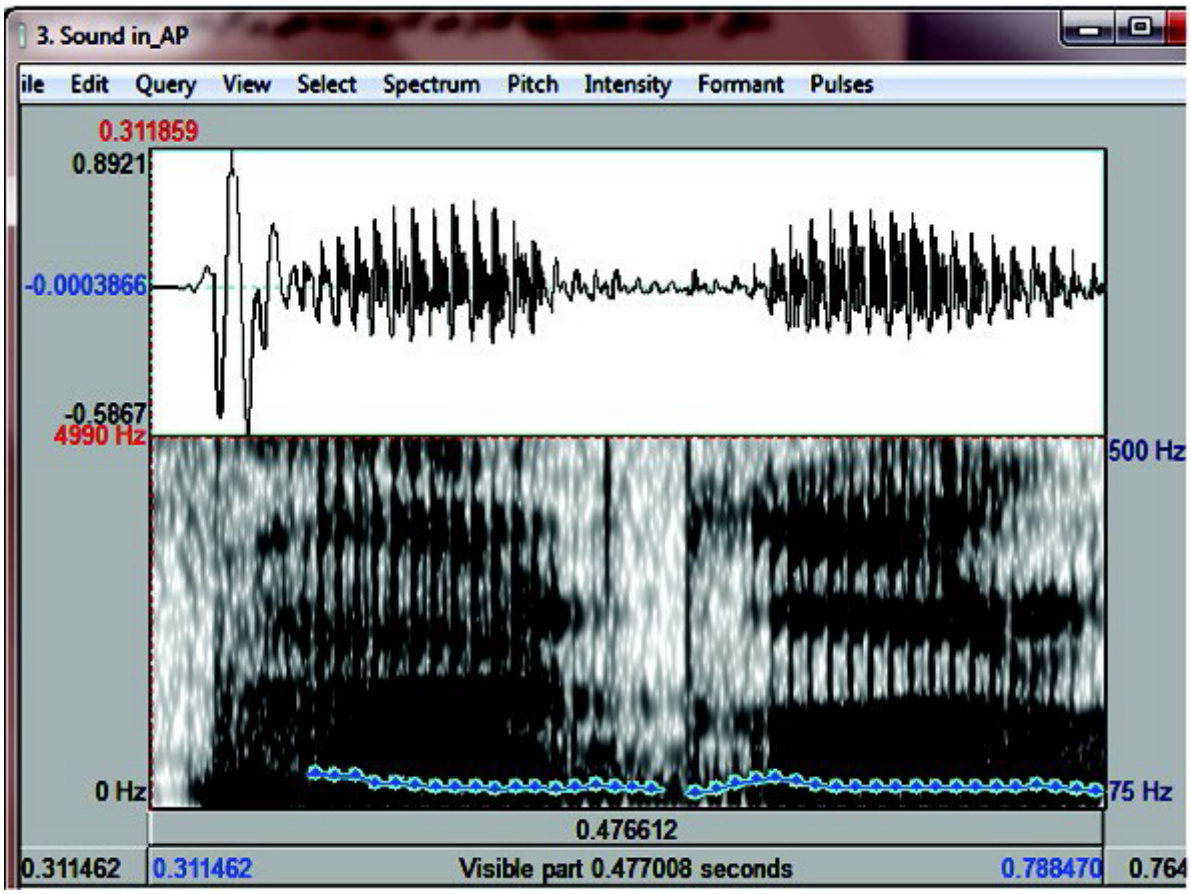


(شکل ۲): طیف‌نگاشت گونهٔ طبیعی (بالا) و هم‌گذاری شدهٔ (پایین) کلمهٔ "پرگار" بر اساس روش بازسازی اوّل

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

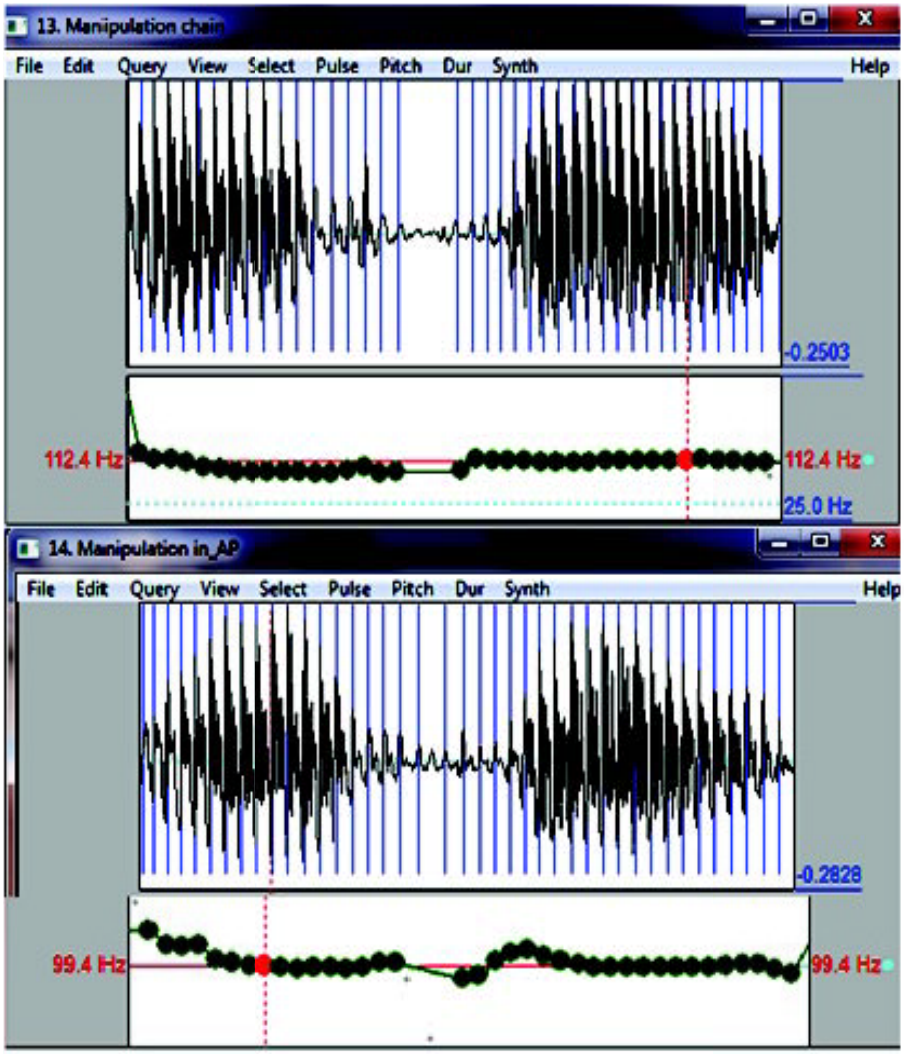


(شکل ۳): طیف‌نگاشت پاره‌گفتار هم‌گذاری شدهٔ "علی پرگار نمی‌خواهد" بر اساس روش بازسازی دوم

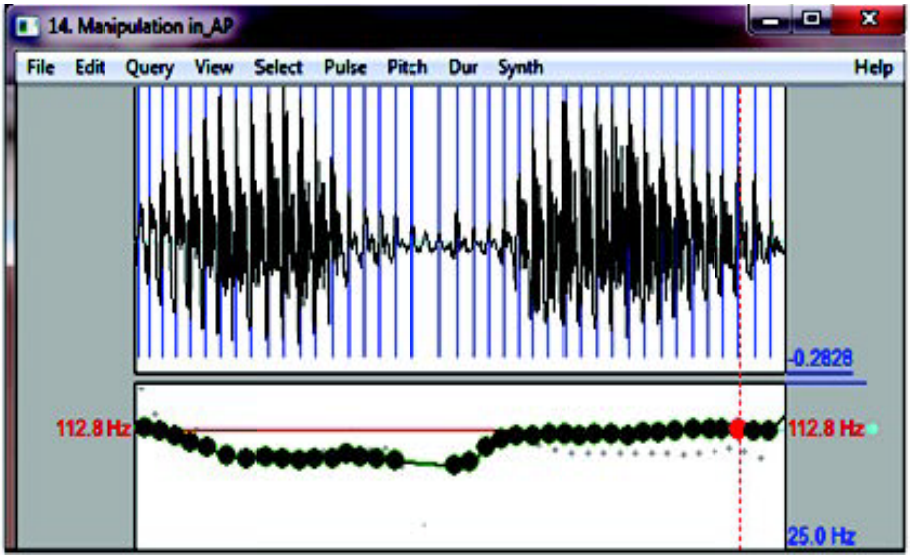


(شکل ۴): طیف‌نگاشت گونهٔ هم‌گذاری شدهٔ کلمهٔ "پرگار" بر اساس روش بازسازی دوم

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]



(شکل ۵): تابع تغییرات فرکانس پالس‌های حنجره در واحد زمان برای گونهٔ طبیعی (بالا) و هم‌گذاری شدهٔ (پایین) کلمهٔ "پرگار" پیش از اعمال تغییرات زیروبمی



(شکل ۶): زیروبمی بازسازی شدهٔ گونهٔ هم‌گذاری شدهٔ کلمهٔ "پرگار"

نکتهٔ دیگر آن که دیرش محدودهٔ بست و قطعهٔ رهش انسدادی‌های [p] و [g] نیز در گفتار هم‌گذاری‌شده ([p] بست: ۳۶ و رهش ۴۴ میلی‌ثانیه؛ [g] بست: ۲۷ و رهش: ۱۴ میلی‌ثانیه) و گفتار طبیعی ([p] بست: ۳۴ و رهش ۴۱

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

میلی‌ثانیه؛ [g] بست: ۲۶ و رهش: ۱۶ میلی‌ثانیه) به‌طور تقریبی به یک اندازه است که نشان می‌دهد ویژگی‌های زنجیره‌ای در سطح هجاهای نوایی در حدّ طبیعی است.

برای بازسازی گونه‌های گفتاری نوع دوم، ابتدا گونهٔ PW کلمات "پرگار"، "تردید" و "کاربرد" در تمامی جایگاه‌های نوایی IP، AP و PW در جملات ۷-۹ به‌جای گونهٔ طبیعی همان کلمات در کنار دیگر کلمات هم‌گذاری شد. این گونه در جایگاه نوایی PW جملات مورد نظر با الگوی نوایی جمله مطابقت داشت. جملات هم‌گذاری‌شده به‌طور دقیق معادل مثال‌های ج) در گونه‌های بازسازی‌شدهٔ نوع اوّل بودند. بنابراین الگوی آهنگ این جملات بدون هرگونه پردازش نوایی، بسیار شبیه به گفتار طبیعی بود. امّا حضور این‌گونه در جایگاه‌های نوایی نامربوط و نامناسب، یعنی جایگاه‌های IP، AP و PW در جملات الف و ب باعث شد الگوی آهنگ جملات در محدودهٔ کلمات هدف از گفتار طبیعی فاصله گرفته و غیرطبیعی شود. در مرحلهٔ بعد، پاره‌گفتارهای هم‌گذاری‌شدهٔ الف و ب مورد پردازش قرار گرفتند؛ به این صورت که مقادیر فرکانس پایه و دیرش هجاهای کلمات هدف در این گونه‌ها متناسب با مقادیر این پارامترها برای همین هجاها در گفتار طبیعی تغییر داده شدند. برای این منظور ابتدا علامت آوایی هر پارهٔ گفتار در نرم‌افزار پرت به بازنمود دوبعدی فرکانس پالس‌های حنجره در واحد زمان تبدیل شد (از طریق گزینهٔ دست‌کاری). از این بازنمود سپس دو خروجی به دست آمد: یکی لایهٔ زیروبمی، برای بازسازی مقادیر فرکانس پایه و دیگری لایهٔ دیرش، برای بازسازی مقادیر دیرش هجاها. بازسازی و تنظیمات نوایی مورد نظر بر روی پاره‌گفتارهای مربوطه سپس در هر یک از این دو لایه به‌طور جداگانه انجام شد.

(شکل ۳) طیف‌نگاشت پاره‌گفتار هم‌گذاری‌شدهٔ جملهٔ "علی پرگار نمی‌خواهد" (مثال ۷ ب) و (شکل ۴) طیف‌نگاشت گونهٔ هم‌گذاری‌شدهٔ کلمهٔ "پرگار" را در همین جمله نشان می‌دهد. مقایسهٔ این طیف‌نگاشت‌ها با طیف‌نگاشت‌های گفتار طبیعی (شکل‌های ۱ و ۲ بالا) نشان می‌دهد دیرش هجاها در گفتار طبیعی به‌طور محسوسی بیش‌تر از گفتار هم‌گذاری‌شده است که علت آن طول بیش‌تر واکه‌ها (اختلاف ۲۰ میلی‌ثانیه‌ای برای واکهٔ اوّل و ۴۵ میلی‌ثانیه‌ای برای واکهٔ دوم) و دیرش بیش‌تر قطعهٔ رهش انسدادی (اختلاف ۲۵ میلی‌ثانیه‌ای) [p] است. هم‌چنین الگوی نوایی گفتار هم‌گذاری‌شده و طبیعی با به‌طور کامل متفاوت است. اوّل آن که میزان فرکانس پایه در سطح هر دو هجا برای کلمهٔ "پرگار" در گفتار هم‌گذاری‌شده به‌طور تقریبی به یک اندازه است (حدود ۱۱۰ هرتز) که علت آن این است که این کلمه در پاره گفتار مورد نظر، گونهٔ هم‌گذاری‌شدهٔ یک کلمهٔ واجی است و تکیهٔ زیروبمی ندارد؛ ولی در گفتار طبیعی سطح فرکانس پایه برای هجای دوم در مقایسه با هجای اوّل به‌طور قابل‌ملاحظه‌ای بیش‌تر است (اختلاف حدود ۳۰ هرتزی) که نشان می‌دهد این کلمه به‌صورت یک گروه زیروبمی مستقل تظاهر یافته است. دوم آن که در گفتار طبیعی فرکانس پایه از بخش پایانی کلمهٔ "پرگار" به بخش آغازی کلمهٔ "نمی‌خواهد" بدون ناپیوستگی و به‌صورت نرم،گذر کرده است که علّت آن حضور نواخت مرزی h در حدّ فاصل میان دو گروه زیروبمی "پرگار" و "نمی‌خواهد" به‌ترتیب با نواخت‌های L+H* و H* است. ولی در گفتار هم‌گذاری‌شدهٔ فرکانس پایه در بخش پایانی هجای دوم کلمهٔ "پرگار" به یک باره قطع و در آغاز هجای [ne] در کلمهٔ "نمی‌خواهد" به‌طور ناگهانی افزایش یافته است؛ این باعث شده در این محدوده ناپیوستگی نوایی قابل‌توجهی ایجاد شود.

بازسازی زیروبمی علامت آوایی به روش PSOLA به صورت افزایش یا کاهش زمانی فاصلهٔ قاب‌ها یا پالس‌های حنجره است. بنابراین هر گونه تغییر در زیروبمی علامت آوایی باعث تغییر در دیرش آن نیز می‌شود. به‌همین علّت زیروبمی قبل از دیرش هجاها بازسازی شد.

(شکل ۵) تابع تغییرات فرکانس پالس‌های حنجره را در واحد زمان برای گونهٔ طبیعی (بالا) و هم‌گذاری‌شدهٔ (پایین) کلمهٔ "پرگار" پیش از اعمال تغییرات زیروبمی نشان می‌دهد که چنان‌که مشاهده می‌شود الگوی تغییرات فرکانس پایه برای گفتار طبیعی و هم‌گذاری‌شده در سطح هر دو هجا با تفاوت دارد. (شکل ۶) بازنمود زیروبمی بازسازی‌شدهٔ گونهٔ هم‌گذاری‌شدهٔ این کلمه را نشان می‌دهد. همان‌طور که مشاهده می‌شود پس از بازسازی زیروبمی، سطح فرکانس پایه در هر دو هجا به سطح مقادیر طبیعی نزدیک شده است. هم‌چنین برای بهبود کیفیت آهنگ گفتار هم‌گذاری‌شده در مرز کلمات هدف و کلمات بعد، یعنی در محدودهٔ گذر از هجای پایانی کلمات هدف به هجای آغازین کلمات بعد، سطح فرکانس پالس‌ها مطابق با سطح فرکانس پالس‌های طبیعی بازسازی شد.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

۲-۲-۳- آزمون ادراکی

فرضیهٔ تحقیق عبارت بود از این‌که ویژگی‌های زنجیره‌ای از بافت نوایی تأثیر می‌پذیرند و این که در بازسازی یک پارهٔ گفتار، برابر آوایی هر واحد بازسازی باید متناسب با جایگاه نوایی آن در سطح زنجیرهٔ گفتار مورد استفاده قرار گیرد تا کیفیت گفتار بازسازی در هر دو سطح زنجیره‌ای و زبرزنجیره‌ای به گفتار طبیعی نزدیک‌تر شود. اگر فرضیه مطرح شده درست باشد، در آن صورت انتظار می‌رود کیفیت پارهٔ گفتارهای بازسازی شدهٔ نوع اوّل در مقایسه با نوع دوم و سوم طبیعی‌تر باشد. برای ارزیابی کیفیت گونه‌های گفتاری بازسازی‌شده و پاسخ به فرضیهٔ تحقیق، یک آزمون ادراکی انجام شد. مراحل انجام این آزمون عبارت بود از:

الف) انتخاب شرکت‌کننده‌ها: شرکت‌کننده‌ها عبارت بودند از ۱۰ گویشور مرد و ۱۰ گویشور زن که همگی به زبان فارسی معیار به‌عنوان زبان اوّل صحبت می‌کردند. آزمودنی‌ها همگی تحصیل کرده بوده و متوسط سن آن‌ها بین ۲۰ تا ۳۵ سال بود. هیچ‌کدام از شرکت‌کننده‌ها سابقهٔ بیماری یا اختلال شنوایی نداشتند. به‌علاوه هیچ‌یک در جریان هدف تحقیق قرار نداشتند.

ب) محرک‌ها و روش آزمایش: ۲۷ پاره‌گفتار بازسازی‌شده (۳ کلمه × ۳ جایگاه نوایی × ۳ روش بازسازی) به‌عنوان محرک‌های شنیداری به‌طور تصادفی به آزمودنی‌ها ارائه شد. از شرکت‌کننده‌ها خواسته شد میزان طبیعی بودن هر محرک را با تمرکز بر روی یک مقیاس ناپیوسته شامل اعداد ۱، به معنای به‌طور کامل غیرطبیعی تا ۵، به معنای به‌طور کامل طبیعی مشخص کنند. فاصلهٔ بین نقاط بر روی مقیاس ۰/۵ در نظر گرفته شد. بلافاصله پس از پخش هر محرک صفحه‌ای حاوی صورت نوشتاری آن و مقیاس مورد نظر بر روی صفحهٔ نمایش‌گر رایانه برای آزمودنی‌ها نمایش داده می‌شد که آن‌ها باید کیفیت هر محرک را با علامت × در کنار عدد موردنظر مشخص می‌کردند. فاصلهٔ بین پاسخ شرکت‌کننده‌ها و پخش محرک بعد ۵ ثانیه در نظر گرفته شد. هم‌چنین درصورت درخواست آزمودنی‌ها، محرک‌ها برای آن‌ها بازپخش می‌شدند. پیش از شروع آزمایش در توضیح به شرکت‌کننده‌ها گفته شد که هر جمله حاوی یکی از کلمات "پرگار"، "تردید" یا "کاربرد" است که نحوهٔ تولید آن‌ها در سطح جملهٔ مورد نظر حائز اهمیت فراوان است.

(جدول ۱): میانگین و پراش میزان طبیعی بودن پاره‌گفتارهای بازسازی‌شده برای روش بازسازی نوع اول، دوم و سوم در سطح جایگاه‌های نوایی IP، AP و PW برای محرک‌های "پرگار"، "تردید" و "کاربرد"

کلمات هدف	جایگاه نوایی / روش بازسازی	IP	AP	PW
پرگار	نوع اول	۳/۹۰ (۱/۰۸)	۴/۰۵ (۰/۸۸)	۳/۸۷ (۱/۰۹)
	نوع دوم	۳/۲۵ (۱/۱.)	۳/۷۲ (۰/۹۳)	۳/۶۷ (۱/۰۱)
	نوع سوم	۱/۷۲ (۰/۹۶)	۱/۵۷ (۱/۰۳)	۱/۸۰ (۱/۸۰)
تردید	نوع اول	۳/۹۷ (۱/۱۰)	۴/۱۰ (۱/۰۲)	۴/۲۰ (۱/۰۸)
	نوع دوم	۳/۶۵ (۰/۹۱)	۳/۶۰ (۰/۹۱)	۳/۶۰ (۰/۹۳)
	نوع سوم	۱/۷۷ (۰/۸۷)	۱/۶۵ (۰/۸۳)	۱/۷۰ (۰/۸۸)
کاربرد	نوع اول	۴/۱۷ (۰/۷۹)	۴/۱۲ (۰/۸۴)	۴/۰۷ (۱/۱۰)
	نوع دوم	۳/۷۲ (۱/۱۱)	۳/۸۷ (۰/۹۵)	۳/۷۰ (۰/۹۳)
	نوع سوم	۱/۸۰ (۰/۸۱)	۱/۵۲ (۰/۷۴)	۳/۶۲ (۰/۷۲)

(جدول ۲): نتایج آزمون تحلیل واریانس یک‌طرفه برای جایگاه‌های نوایی IP، AP و PW برای "پرگار"

منبع تغییرات	جمع مجذورات	درجه آزادی	میانگین مجذورات	میزان F	سطح معنی داری
روش بازسازی (IP)	۵۰/۸۰۸	۲	۲۵/۴۰۴	۴۱/۱۶۷	۰/۰۰۰
روش بازسازی (AP)	۷۲/۳۵۸	۲	۳۶/۱۷۹	۶۸/۰۰۴	۰/۰۰۰
روش بازسازی (PW)	۵۲/۴۰۸	۲	۲۶/۲۰۴	۲۵/۴۱۳	۰/۰۰۰

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

(جدول۳): نتایج آزمون تحلیل واریانس یک‌طرفه برای جایگاه‌های نوایی IP، AP و PW برای "تردید"

منبع تغییرات	جمع مجذورات	درجه آزادی	میانگین مجذورات	میزان F	سطح معنی داری
روش بازسازی (IP)	۵۶/۴۰۸	۲	۲۸/۲۰۴	۳۶/۹۳۶	۰/۰۰۰
روش بازسازی (AP)	۶۷/۰۳۳	۲	۳۳/۵۱۷	۶۰/۳۶۲	۰/۰۰۰
روش بازسازی (PW)	۶۸/۱۳۳	۲	۳۴/۰۶۷	۴۸/۳۰۴	۰/۰۰۰

(جدول ۴): نتایج آزمون تحلیل واریانس یک‌طرفه برای جایگاه‌های نوایی IP، AP و PW برای "کاربرد"

منبع تغییرات	جمع مجذورات	درجه آزادی	میانگین مجذورات	میزان F	سطح معنی داری
روش بازسازی (IP)	۶۳/۶۵۸	۲	۳۱/۸۲۹	۴۱/۶۳۵	۰/۰۰۰
روش بازسازی (AP)	۸۲/۳۰۰	۲	۴۱/۱۵۰	۶۳/۲۰۱	۰/۰۰۰
روش بازسازی (PW)	۶۹/۶۵۸	۲	۳۴/۸۲۹	۴۷/۵۲۳	۰/۰۰۰

ج) تحلیل آماری: برای تحلیل آماری پاسخ‌ها، شامل ۵۴۰ پاسخ (۲۷ پاره گتار ×۲۰ آزمودنی) چند آزمون تحلیل واریانس یک طرفه به‌ازای هر یک از جایگاه‌های نوایی IP، AP و PW به‌طور جداگانه برای هر یک از محرک‌های "پرگار"، "تردید" و "کاربرد" انجام شد که در هر آزمون، روش بازسازی (با سه سطح گونه‌های بازسازی نوع اول، دوم و سوم) به‌عنوان متغیر مستقل و میزان طبیعی بودن متغیر وابسته در نظر گرفته شد.

د) نتایج: (جدول ۱) میانگین و پراش پاسخ‌ها را برای روش بازسازی نوع اوّل، دوم و سوم در سطح جایگاه های نوایی IP، AP و PW به طور جداگانه برای محرک‌های "پرگار"، "تردید" و "کاربرد" نشان می دهد. چنان‌که مشاهده می‌شود، میانگین پاسخ‌ها برای تمامی محرک‌ها و در سطح تمامی جایگاها برای گونه‌های بازسازی نوع اوّل بیش‌ترین و برای گونه‌های بازسازی نوع سوم کم‌ترین است. جداول ۲، ۳ و ۴ نتایج آزمون تحلیل واریانس یک‌طرفه را به‌طور جداگانه برای جایگاه‌های نوایی IP، AP و PW به‌ترتیب برای محرک‌های "پرگار" (۲)، "تردید" (۳) و "کاربرد" (۴) نشان می‌دهد. برای "پرگار"، با توجّه به میزان F به‌دست آمده برای IP، F= ۴۱/۱۶۷؛ AP، F= ۶۸/۰۰۴ و PW، F=۲۵/۴۱۳ در سطح معنی داری α > ۰/۰۰۱ و سطح اطمینان ۰/۹۵ می‌توان گفت که اثر عامل روش بازسازی بر کیفیت گفتار بازسازی‌شده در سطح تمامی جایگاه‌های نوایی معنی‌دار است. نتایج آزمون تعقیبی LSD نشان داد که تفاوت گفتار بازسازی‌شدهٔ نوع سوم در تمامی جایگاه‌ها با نوع اوّل و دوم معنی‌دار است و گونه‌های نوع اوّل و دوم فقط در جایگاه IP با تفاوت معنی‌داری دارند. در جایگاه‌های AP و PW این تفاوت معنی‌دار نیست. برای "تردید" نیز، میزان F به‌دست آمده برای IP، F=۳۶/۹۳۶؛ AP، F= ۶۰/۳۶۲ و PW، F= ۴۸/۳۰۴ در سطح معنی‌داری α > ۰/۰۰۱ و سطح اطمینان ۰/۹۵ نشان می‌دهد که اثر عامل روش بازسازی بر کیفیت گفتار بازسازی‌شده در سطح تمامی جایگاه‌های نوایی معنی‌دار است. نتایج آزمون تعقیبی LSD نشان داد که گونه‌های بازسازی‌شده نوع سوم در تمامی جایگاه‌ها با نوع اوّل و دوم تفاوت معنی‌داری دارند و تفاوت گونه‌های نوع اوّل و دوم فقط در جایگاه IP معنی‌دار نیست. در جایگاه‌های AP و PW این تفاوت معنی‌دار است. هم‌چنین با توجّه میزان F به‌دست آمده برای "کاربرد" در جایگاه‌های IP، F=۴۱/۶۳۵؛ AP، F= ۶۳/۲۰۱ و PW، F= ۴۷/۵۲۳ در سطح معنی‌داری α > ۰/۰۰۱ و سطح اطمینان ۰/۹۵ می‌توان گفت که روش بازسازی در تمامی جایگاه‌های نوایی بر کیفیت گفتار بازسازی‌شده اثر معنی‌داری دارد و مطابق با نتایج آزمون تعقیبی LSD تفاوت گونه‌های نوع اوّل و دوم در تمامی جایگاه‌ها با نوع سوم معنی‌دار است و تفاوت گونه‌های اوّل و دوم در جایگاه PW معنی‌دار است ولی در جایگاه‌های IP و AP معنی‌دار نیست.

۳- نتیجه گیری و بحث

در این تحقیق، فرضیه‌ای مطرح شد مبنی بر این که کیفیت گفتار بازسازی‌شده از واحدهای نوایی در مقایسه با گفتار بازسازی شده از واحدهای غیرنوایی به گفتار طبیعی نزدیک‌تر است چون واحدهای نوایی علاوه‌بر ویژگی‌های زبرزنجیره‌ای، ویژگی‌های زنجیره‌ای گفتار طبیعی را نیز در بر می‌گیرند. برای آزمون، این فرض آزمایشی را انجام دادیم. در این آزمایش، سه گونهٔ گفتاری متفاوت از کلماتی یکسان از طریق هم‌گذاری هجاها به سه روش متفاوت ساخته شدند. در روش اوّل، کلمات از طریق هم‌گذاری هجاهای نوایی بدون پردازش فرکانس پایه و دیرش هجاها ساخته شدند. در روش دوم، ابتدا کلمات با هجاهای غیرنوایی هم‌گذاری شدند و

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]

سپس هجاهای درون کلمات هم‌گذاری شده با توجّه به جایگاه نوایی کلمات در سطح جمله در حوزهٔ فرکانس پایه و دیرش پردازش شدند. در روش سوم، کلمات بدون هر گونه پردازش نوایی از طریق هم‌گذاری هجاهای غیرنوایی ساخته شدند. کلمات بازسازی شده سپس در جایگاه‌های نوایی آغاز و پایان IP، AP و PW در درون سه جمله حامل قرار داده شده و پس از تولید در یک آزمون ادراکی به شرکت‌کننده‌ها ارائه شدند. نتایج آزمون ادراکی نشان داد که کیفیت گفتار بازسازی شدهٔ نوع اوّل و دوم در تمامی جایگاه‌های نوایی به‌طور معنی‌داری از نوع سوم طبیعی‌تر است. درحقیقت، شرکت‌کننده‌ها گفتار بازسازی‌شدهٔ نوع سوم را با توجّه به میانگین به‌دست آمده برای این‌گونهٔ گفتاری در سطح جایگاه نوایی مختلف غیرطبیعی می‌دانند. هم‌چنین، گرچه بر اساس نتایج به‌دست آمده کیفیت گونه‌های گفتاری نوع اوّل و دوم در حدّ "طبیعی" است (میانگین به‌دست آمده برای این گونه‌ها، به‌طور متوسط بین ۳/۵ تا ۴/۵ است)، ولی میزان طبیعی بودن گونهٔ گفتاری نوع اوّل در تمامی جایگاه‌ها از نوع دوم بیش‌تر است و این تفاوت در بعضی از جایگاه‌ها (جایگاه IP برای "پرگار"؛ AP و PW برای "تردید" و PW برای "کاربرد") معنی‌دار است. این نتایج نشان می‌دهد که با به‌کارگیری واحدهای بازسازی حساس به بافت نوایی، کیفیت گفتار بازسازی‌شده تا حدّ قابل قبولی بهبود می‌یابد.

۴- منابع

ثمره، یدالله. ۱۳۶۴. آواشناسی زبان فارسی، تهران، مرکز نشر دانشگاهی.

صادقی، وحید. ۱۳۸۶. تأثیر دمش بر تقبل واکداری- بی واکی انسدادی‌های فارسی. زبان و زبان‌شناسی ۶، ۶۵-۸۱.

نوربخش، ماندانا. ۱۳۸۸. نقش تمایزی زمان شروع واک در هم‌خوان‌های دهانی فارسی معیار، رسالهٔ دکتری، دانشگاه تهران.

سادات تهرانی، ۲۰۰۷.

Allen, J., Hunnicutt, M., Klatt, D. 1987. From Text to Speech. The MITalk System. Cambridge University Press.

Alan, W. Black, Lenzo, K. 2003. Building Synthetic Voices. URL http://www.festvox.org/bsv/.

Alan, W. Black, Paul, T., Richard, C. 1999. The Festival Speech Synthesis System. URL http://www.cstr.ed.ac.uk/projects/festival/manual/festival_toc.html.

Hirschberg and P. Prieto 1996. Training intonational phrasing rules automatically from English and Spanish text-to-speech Speech Communication, 18:281- 290.

Lewis, E., and Tatham, M. 2001. Automatic Segmentation of recorded speech into syllables for speech synthesis. *Proceedings of Eurospeech, 01,* 1703-1707. Aalborg

Pierehumbert, 1981. Synthesizing Intonation. Journal of the Acoustical Society of America. 70, 985-995

Pierehumbert, J. and Talkin, D. 1992. Lenition of /h/ and glottal stop", *Papers in Laboratory Phonology,* Vol II. 92-116

Sadat Tehrani, Nima 2007. The Intonation Grammar of Persian. PhD dissertation, University of Monitoba, Winniepeg.

Cho, T., Keating, P. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. Journal of Phonetics 29, 155-190.

Jun Sun-Ah 1993. The Phonetics and Phonolgy of Korean Prosody. PhD Thesis, The Ohio State University.

Keating, P., Cho, T., et al. 1998. Domain-initial articulatory strengthening in four languages. LabPhon 6, 145-163.

Sato, H. 1992. Speech synthesis using CVC concatenation units and excitation waveform elements. Trans. on Acoustical Society of Japan, 83–69.

Smith, C.L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. Journal of Phonetics 25, 471-500

Yoon, K. 2004. A prosodic phrasing model for a Korean text-to-speech synthesis system, Journal of Phonetics.

Yoon, K. 2008. Design and evaluation of prosodically-sensitive concatenative units for a Korean TTS system. Journal of Computer Speech and Language.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]



وحید صادقی تحصیلات خود را در مقطع کارشناسی رشتهٔ آموزش زبان انگلیسی در سال ۱۳۷۶ به پایان رساند. وی در سال ۱۳۷۹ در رشتهٔ زبان‌شناسی در مقطع کارشناسی ارشد از دانشگاه تهران فارغ التحصیل شد و در سال ۱۳۸۵ نیز در همان رشته مدرک دکترای خود را اخذ نمود. حوزهٔ پژوهشی مورد علاقه وی پردازش گفتار و درک و بازشناسی گفتار است.

نشانی رایانامک ایشان عبارت است از:

vsadeghi5603@gmail.com

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-09]