فصلنامهٔ علمی- پژوهشی زبان‌پژوهی دانشگاه الزهرا(س)
سال ششم، شمارهٔ ۱۳، زمستان ۱۳۹۳

بررسی موردی کاهش آوایی در ارقام تلفنی فارسی

زهرا محمودزاده[۱]
محمد بحرانی[۲]

تاریخ دریافت: ۹۰/۶/۱۵
تاریخ تصویب: ۹۱/۴/۱۹

چکیده

کاهش از جمله فرایندهای آوایی است که برای سیستم‌های بازشناسی گفتار یک چالش محسوب می‌شود. ازجملهٔ این موارد، کاهش ارقام تلفنی مانند "دو-نه"، "هفت-هشت" و "سه-صفر" است. تجزیه و تحلیل آوایی سیگنال‌های تلفنی /se/-/sefr/ نشان می‌دهد که هر چند با حذف /r/ تمایز [se]-[sef] کاهش می‌یابد اما سرنخ‌های آکوستیکی باقی‌مانده در واکهٔ /e/، همانند دیرش واکه و بسامد کانونی سازه‌های F2 و F3، تقابل واجی فوق را انتقال

[۱] استادیار و عضو هیئت علمی پژوهشگاه علوم و فناوری اطلاعات ایران؛ mahmoodzadeh@irandoc.ac.ir
[۲] استادیار و عضو هیئت علمی گروه زبان‌شناسی دانشگاه صنعتی شریف؛ bahrani@ce.sharif.edu

می‌دهند. استفاده از این سرنخ‌های آکوستیکی در بازشناسی گفتار، می‌تواند به بازیابی صورت‌های کاهش‌یافته کمک کند.

واژه‌های کلیدی: کاهش ، تحلیل آکوستیکی، سرنخ آوایی، گفتار تلفنی، بازشناسی گفتار

۱. مقدمه

در گفتار روزمره، کلمات ممکن است کاهش پیدا کنند بدین معنا که همهٔ مشخصه‌های واجی لازم برای تقابل‌های آوایی، در موج صوتی گفتار ظاهر نمی‌شوند. گاه کاهش آوایی[1]، منجربه حذف یک صورت آوایی می‌شود به‌نحوی که در موج صوتی و یا طیف‌نگاشت، تولید آوا مشاهده نمی‌شود. کاهش درجات مختلف، از حذف یک آوا تا تضعیف برخی سرنخ‌های ادراکی را شامل می‌شود. فرایندهایی مانند همگونی، حذف، هم‌تولیدی و کاهش آوایی که سبب تغییر صورت واجی یک کلمه می‌شوند، چالش‌های پیش روی سیستم‌های پردازش گفتار محسوب می‌شوند. مدل‌های متفاوت بازشناسی گفتار، معمولاً براساس فرایند درک گفتار طرح‌ریزی می‌شوند. درک گفتار و بنابراین بازشناسی گفتار، در دو مرحله صورت می‌گیرد: ابتدا درک سرنخ‌های آکوستیکی واحدهای واجی یعنی مشخصه‌های واجی، واج یا هجا و آن‌گاه، در مرحلهٔ بعد، بازشناسی کلمات با استفاده از واحدهای واجی. در صورت‌های آوایی کاهش‌یافته، درک گفتار با استفاده از سرنخ‌های آکوستیکی باقی‌مانده در سیگنال صوتی و بافت واجی روی می‌دهد. هنگامی‌که صورت کاهش‌یافته درگفتار پیوسته و در درون جمله تولید می‌شود، علاوه‌بر اطلاعات آوایی، اطلاعات نحوی و معنایی نیز به بازشناسی صورت کاهش‌یافته کمک می‌کند.

کاهش پهنای باند انتقال اطلاعات برروی خطوط تلفن به‌میزان ۳۱۰۰ تا ۳۴۰۰ کیلوهرتز، سبب افت و کاهش اطلاعات آکوستیکی لازم برای دریافت و تشخیص آواهای

[1] phonetic reduction

زبان حتی از سوی انسان می‌شود زیرا گوش انسان برای بازشناسی آواهای زبانی به اطلاعات آکوستیکی تا حدود ۸ کیلوهرتز نیاز دارد. سیستم‌های بازشناسی گفتار تلفنی نیز به‌دلیلی مشابه یعنی کاهش اطلاعات سیگنال، دچار اختلال می‌شوند و در برخی موارد، قادر به بازشناسی سیگنال‌های مشابه از یکدیگر نمی‌باشند. به‌عنوان مثال، این سیستم‌ها در بازشناسی ارقام «سه و صفر»، «دو و نه»، «هفت و هشت» و نیز ترکیب و باهم‌آیی اعداد، مرز اعداد به‌دلیل هم‌پوشانی آواها از گویشور به گویشور، لهجه به لهجه و نیز فاکتورهایی مانند سن (بالغ، جوان، کودک)، جنسیت (زن، مرد) و غیره دچار اختلال می‌شوند. سیستم‌های بازشناسی گفتار، برای استخراج پارامترهای آوایی، از روش محاسبهٔ ضرایب کپسترال در مقیاس مل[1] استفاده می‌کنند که جانسون[2] (۱۹۹۷: ۱۲۴) معتقد است این روش‌ها نوعی تحلیل شنیداری طیف هستند و بنابراین، مشابهت زیادی به تحلیل طیف در سیستم شنیداری گوش انسان دارند.

از آنجاکه بیشترین میزان خطاها در بازشناسی ارقام تلفنی فارسی، مربوط‌به زوج‌های «سه- صفر» /se/-/sefr/ است، سیگنال این دو رقم، مورد تجزیه و تحلیل آکوستیکی قرار گرفت. فرایند کاهش عدد «صفر»/sefr/ در سیگنال‌های تلفنی سبب از بین رفتن تقابل واجی /sefr-se/ در بازشناسی خودکار می‌گردد. با تجزیه وتحلیل مفصل آکوستیکی و استخراج سرنخ‌های باقی‌مانده در سیگنال، روشن خواهد شد که درک گفتار توسط انسان، براساس چه اطلاعات آوایی صورت می‌گیرد و بنابراین، می‌توان با افزودن این اطلاعات به سیستم‌های بازشناسی گفتار تلفنی، بازیابی این تمایز را بهبود بخشید.

۲. داده‌ها و روش‌شناسی

نمونه‌ها، سیگنال‌های تلفنی ارقام «صفر» /sefr/ و «سه» /se/ هستند که توسط ۱۳ زن و ۱۴ مرد تولید شده‌اند. این نمونه‌ها، از مجموعهٔ دادگان فارس‌دات تلفنی کوچک[3] انتخاب شده‌اند که توسط موسسهٔ پردازش هوشمند علائم تهیه شده است (بی‌جن‌خان و همکاران

[1] mel frequency cepstral coefficients (MFCC)
[2] Johnson
[3] TFarsDat

۲۰۰۳: ۱۵۲۸-۱۵۲۵). نمونه‌ها، از یک رقم منفرد همانند« ۳» تا هفت رقم متوالی همانند «۰۱۳۰۹۸۰» را در برمی‌گیرد و از هر سه جایگاه آغاز، میان و پایان رشتهٔ ارقام، انتخاب شده‌اند. به‌عنوان مثال، در نمونهٔ «۰۳۰»، رقم «سه» در میان رشتهٔ ارقام و رقم «صفر» هم در آغاز و هم در میان این رشته تولید شده است. گویشوران رقم‌های منفرد را به‌صورت یک کلمهٔ مجزا تولید کرده‌اند اما در توالی‌های رقمی، وضعیت یکسان نیست بدین صورت که برخی ارقام، با فاصله و مکث و برخی با سرعت تولید گفتار بالاتر، بدون مکث و متوالی، تولید شده‌اند. بر این اساس، می‌توان داده‌ها را به دو نوع گفتار مجزا[1] و متصل[2] تقسیم نمود. کلمات مجزا، گفتاری است که در آن نمونه‌های گفتاری مورد آزمایش، در درون جمله تولید نمی‌شود به این صورت که فهرستی از کلمات مجزا و مستقل در اختیار گویشوران قرار گرفته و آنها هر کلمه را با اندکی مکث و وقفهٔ کوتاهی از کلمهٔ بعدی تولید می‌کنند و نمونه‌ها ضبط می‌شود (ویبلی و لی[3] ۱۹۹۰: ۱۴۵). بنابراین، در گفتار مجزا، تولید هر کلمه، مجزا و مستقل از کلمهٔ بعدی است و درنتیجه، ویژگی آوایی واج‌ها بیشتر حفظ شده و کمتر دستخوش فرایندهای گفتار پیوسته مانند هم‌تولیدی، همگونی، کاهش و غیره واقع می‌شود. تولید کلمات در گفتار مجزا، با دقت بیشتری نسبت به گفتار پیوسته صورت می‌گیرد. گفتار پیوسته[4]، گفتاری است که در آن کلمات مورد آزمایش، در درون جمله تولید می‌شوند. گفتار پیوسته، به‌دلیل سرعت بالا، دقت کمتر و در نتیجه هم‌پوشانی بیشتر آواها، محل وقوع بسیاری از فرایندهای آوایی نظیر کاهش، حذف، هم‌تولیدی، همگونی و غیره است (فرنتانی[5] ۱۹۹۹: ۳۷۹).

ویبل و لی (۱۹۹۰: ۱۴۵) گفتار متصل را نوعی گفتار خواندنی[6] تعریف می‌کنند که در آن گویشور، فهرستی از کلمات مجزا را با دقت مناسب، اما بدون مکث و توقف، تولید می‌کند. از این رو، در گفتار متصل نیز همانند گفتار مجزا، کلمات در درون جمله تولید

[1] isolated word form
[2] connected speech
[3] Weibel and Lee
[4] continuous speech
[5] Farnetani
[6] read speech

نمی‌شوند بلکه به‌صورت سیاهه‌ای در اختیار گویشور قرار گرفته و گویشور، کلمات را، برخلاف گفتار مجزا، با سرعت بیشتر و بدون مکث و توقف تولید می‌کند.

فرایند کاهش آوایی، در نمونه‌هایی که گویشور رقم /sefr/ را به‌تنهایی به‌صورت یک کلمهٔ منفرد و مجزا تولید کرده است، نیز دیده می‌شود. در نمونه های متوالی ارقام، هنگامی که سرعت گفتار بیشتر است، طبیعی است که کاهش بیشتر روی داده است. از آنجا که در این آزمایش، نمونه‌های کاهش‌یافته‌ای مورد بررسی قرار گرفته‌اند که در آنها همخوان /r/ به‌طور کامل حذف شده است، این نمونه‌ها، به‌صورت [seF] آوانویسی شده‌اند و علامت [F] بزرگ، برای آوانویسی همخوان /f/ در نمونه‌های کاهش‌یافته به کار رفته است. تعداد نمونه‌های مورد بررسی (۱۸۱ نمونه)، شامل نمونه‌های [seF]، [se] و [sefɹ] و آواهای [F]، [f] و [ɹ] برحسب جنسیت در جدول ۱ ارائه شده است.

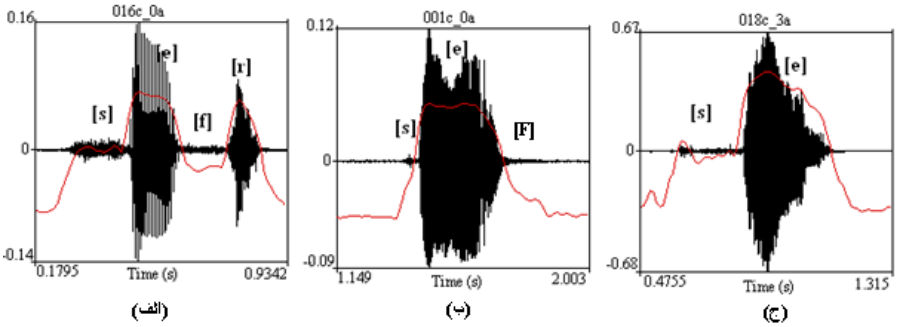
جدول۱: نمونه‌های مورد مطالعه و تعداد آنها برحسب جنسیت

جنسیت	نمونه						کل
	[ɹ]	[f]	[F]	[sefɹ]	[se]	[seF]	
مرد	۱۶	۱۶	۱۶	۱۶	۱۵	۱۶	۹۵
زن	۱۶	۱۶	۱۲	۱۶	۱۴	۱۲	۸۶
کل	۳۲	۳۲	۲۸	۳۲	۲۹	۲۸	۱۸۱

تشخیص نمونه‌هایی که در آنها /sefr/ تحت‌تأثیر فرایندهای آوایی، به‌طور ناقص و کاهش‌یافته تولید شده است، براساس موج صوتی و نیز گوش‌دادن به این نمونه‌ها صورت گرفته است. موج صوتی نمونه‌هایی که کاهش یافته‌اند با نمونه‌هایی که به‌طور کامل تولید شده‌اند، تفاوت فاحشی دارد (شکل ۱ الف و ب). در نمونه‌های کاهش‌یافته (شکل ۱ب)، منحنی شدت انرژی[1] نوفهٔ سایش همخوان /r/، دیده نمی‌شود و نوفهٔ سایش /f/، با انرژی و دیرش کم مشاهده می‌شود. گاهی نیز تنها رهش بسیار کوتاه و جزئی سایش /f/ دیده

[1] intensity contour

می‌شود که نشان می‌دهد در دستگاه گفتار، گرفتگی سایش تنها برای چند لحظه صورت گرفته است. در بسیاری از نمونه‌های کاهش‌یافته، تولید گرفتگی /f/ شنیده می‌شود؛ اما منحنی انرژی، تفاوت چندانی بین /sefr/-/se/ نشان نمی‌دهد. شکل ۱ (ب)، سیگنال و منحنی شدت انرژی در یک نمونهٔ کاهش‌یافتهٔ [seF] و شکل ۱(ج)، سیگنال و منحنی شدت انرژی در یک نمونهٔ [se] را نشان می‌دهد. همان‌طور که در شکل‌ها نیز مشخص است، تفاوتی بین سیگنال و منحنی شدت انرژی [seF] و [se] مشاهده نمی‌شود. در این پژوهش، پارامتر منحنی شدت انرژی، معیاری برای تولید یا حذف /r/ در نظر گرفته شده است. میترر و ارنستس[1] (۲۰۰۶) نیز پارامتر شدت انرژی را به‌عنوان معیاری برای حذف یا عدم حذف /t/ در خوشه‌های پایانی به کار بردند.



شکل ۱: موج صوتی و منحنی شدت انرژی در نمونهٔ
(الف) [sefɪ]، (ب) [seF] و (ج) [se]

ثمره (۱۳۸۵: ۶۷) همخوان /r/ را یک همخوان لرزشی یا غلتان در زبان فارسی تعریف می‌کند که دارای سه واج‌گونهٔ زنشی [ɾ]، سایشی [ɪ] و واک‌گونه است. وی (۱۳۸۵: ۶۹) محل وقوع گونهٔ سایشی [ɪ] را در پایان واژه، بعد از واکه‌های پیشین و نیز بعد از همخوان می‌داند و معتقد است که [ɪ] معمولاً و به‌ویژه پس از همخوان‌های بی‌واک، واک‌رفته می‌شود. در این پژوهش، همخوان /r/ در پایان کلمه و پس از همخوان سایشی بیواک /f/، در کلمهٔ «صفر» /sefr/، تولید شده است. براساس دیدگاه ثمره (۱۳۸۵: ۶۹)، گونهٔ

[1] Mitterer and Ernestus

تولیدشده در این کلمه، از نوع سایشی و واک‌رفتهٔ [ɹ] است زیرا جایگاه تولید این همخوان در پایان کلمه، پس از واکهٔ پیشین /e/ و نیز پس از همخوان بیواک /f/ است. شکل۱(الف)، موج صوتی [ɹ] را به‌صورت نوفهٔ سایشی بیواکی نشان می‌دهد که دامنهٔ انرژی آن نیز کم است. این موج صوتی متفاوت از موج همخوان لرزشی /r/ است که در آن ارتعاش تارهای صوتی مشاهده می‌شود.

در رشته‌هایی از ارقام که رقم [sefɹ] به‌طور متوالی تولید شده، به‌عنوان مثال، در توالی «۳۴۰۰۰۱»، کاهش بیشتر روی داده است زیرا تکرار، از جمله عوامل بسیار مهم در کاهش و حذف محسوب می‌شود. در واقع، بین بسامد تکرار یک تکواژ و طول آن، رابطهٔ عکس وجود دارد به‌نحوی که هرچه بسامد تولید یک واحد آوایی بیشتر باشد، تضعیف آن بیشتر است (یوسیشکین و ودل[1] ۲۰۰۹: ۲۷۴).

۲. ۱. پارامترهای آکوستیکی

عمده‌ترین تفاوت آوایی [se] – [seF]، پس‌از حذف [ɹ]، همخوان سایشی [f] است. چو و مک کویین[2] (۲۰۱۱: ۲۵۳) با انجام آزمایش‌های ادراکی بر روی خوشه‌های کاهش‌یافته و صورت‌های آوایی حذف‌شده دریافتند که شنوندگان، با استفاده از اطلاعات بافتی و معنایی جمله، اطلاعات آکوستیکی آواها و نیز اطلاعات نوایی[3] گفتار در مرز کلمه و گروه آهنگ[4]، صورت‌های کاهش‌یافته را بازشناسی و درک می‌کنند. آنها هم‌چنین دریافتند که بازیابی آواهای حذف‌شده‌ای که فاقد هم سرنخ‌های آکوستیکی و هم اطلاعات بافتی هستند، نشان می‌دهد که دانش واجی خاص هر زبان، که مبتنی بر اطلاعات دقیق و مفصل آوایی آن زبان است، توسط شنونده برای درک این صورت‌ها به کار گرفته می‌شود. از آنجا که پس‌از حذف /r/، تنها عامل تمایز واجی /se/-/sefr/، واج /f/ است، بنابراین سیگنال آوایی این همخوان و واکهٔ /e/، مورد تجزیه و تحلیل آکوستیکی قرار گرفت.

[1] Ussishkin and Wedel
[2] Cho and McQueen
[3] prosodic information
[4] intonational phrase

هدریک و اهده[1] (۱۹۹۳) معتقدند که پارامترهای نوفهٔ سایش وگذرهای سازه‌ای واکه‌ها، حاوی سرنخ‌های ادراکی سایشی‌ها هستند. پارامترهای نوفهٔ سایش عبارتند از: دیرش سایش[2]، مرکز ثقل طیف[3] (COG)، بسامد قلهٔ طیفی[4] و شدت انرژی قلهٔ طیفی[5] سایش. پارامترهای واکه عبارتند از دیرش واکه و بسامد کانونی[6] هریک از گذرهای سازه‌ای[7] F1، F2 و F3. اندازه‌گیری و محاسبهٔ پارامترها در نرم‌افزار پرت[8] صورت گرفت.

۲. ۱. ۱. دیرش سایش

دیرش سایش عبارت است از مدت زمانی که نوفهٔ سایش هر یک از همخوان‌های [f] و [ɪ] طول می‌کشد و واحد آن میلی‌ثانیه (ms) است. اندازه‌گیری دیرش سایش [ɪ] بر اساس موج‌صوتی و طیف‌نگاشت صورت گرفت. در نمونه‌های [seF]، افزایش میزان نوفهٔ سایش برروی سیگنال و نیز مشاهدهٔ نواحی نوفه‌ای با شدت انرژی بیشتر در بسامدهای بالاتر از ۲۰۰۰ هرتز بر روی طیف‌نگاشت، ناحیهٔ سایش [ɪ] را نشان می‌دهد. این ناحیه در مقایسه با ناحیهٔ سایش [f] که انرژی نوفه‌ای کمتر و پراکنده برروی همهٔ بسامدها دارد، شدت انرژی بیشتری داشته و بنابراین منحنی شدت انرژی، افزایش انرژی طیفی در این همخوان را نشان می‌دهد. اندازه‌گیری دیرش سایش [f] نیز براساس سیگنال و طیف‌نگاشت صورت گرفت. همان‌طور که گفته شد، شدت نوفهٔ این همخوان در مقایسه با [ɪ]، بسیار کم است و منحنی انرژی، تنها افت انرژی را نشان می‌دهد و هیچ قلهٔ طیفی در نوفه دیده نمی‌شود. گاهی به‌دلیل افزایش نوفه، حین ضبط داده‌ها، تشخیص محل پایان دیرش سایش [f] و آغاز [ɪ] قابل‌تشخیص نیست. در این موارد، زمان شروع افزایش منحنی انرژی به‌عنوان نقطهٔ شروع سایش [ɪ] در نظر گرفته شده است.

---

[1] Hedrick and Ohde
[2] frication duration
[3] center of gravity (COG)
[4] spectral peak frequency
[5] spectral peak intensity
[6] locus offset frequency
[7] formant transition
[8] praat

۲. ۱. ۲. مرکز ثقل طیف

گشتار مرکز ثقل طیف، عبارت است از میانگین بسامدی طیف. در واقع COG به بسامدی اطلاق می‌شود که طیف انرژی را به دو بخش تقسیم می‌کند به نحوی که در این بسامد، میزان انرژی بسامدهای بالا با میزان انرژی بسامدهای پایین برابر می‌شود و بنابراین به آن مرکز ثقل طیف می‌گویند. مرکز ثقل طیف، پارامتری است که از طیف سایش استخراج می‌شود و واحد آن هرتز (Hz) است.

پارامتر COG با سایز حفرهٔ جلویی دهان ارتباط دارد بدین معنا که مقدار آن، در همخوان‌های پیشین یا قدامی، بیشتر از همخوان‌های پسین است (گوردون و همکاران[1] ۲۰۰۲). آواهایی که انرژی بسامدهای بالای آنها زیاد است، بسامد COG آنها نیز برروی بسامدهای بالا قرار دارد. ونسان و پولز[2] (۱۹۹۷) کاهش آکوستیکی COG به‌همراه کم‌شدن دیرش همخوان را ازجمله نشانه‌های جهانی کاهش آوایی بر می‌شمارند بدین معنا که کاهش مقدار بسامدی COG، نشان می‌دهد که انرژی بسامدهای بالا کاهش یافته است و بنابراین، اطلاعات تولیدی کمتری در سیگنال وجود دارد. از سویی، نتایج مطالعات ناکامورا[3] (۲۰۰۷) نیز نشان می‌دهد که کاهش طیفی، یکی از دلایل عمدهٔ کاهش دقت بازشناسی گفتار پیوسته است.

محاسبهٔ این پارامتر در نرم‌افزار پرت امکان‌پذیر است بدین صورت که می‌توان مقدار این پارامتر را پس‌از محاسبهٔ طیف سایش در این نرم‌افزار محاسبه کرد. در این پژوهش، مقدار COG، براساس کل دیرش سایش محاسبه شد. این پارامتر روشن می‌کند که طیف نوفه، بیشتر تحت تسلط کدام نواحی بسامدی است. برای حذف انرژی واک و نیز انرژی نوفه‌ای جریان برق، بسامدهای زیر ۱۰۰۰ هرتز در کل نمونه‌ها فیلتر شد زیرا در برخی نمونه‌ها، مقداری ارتعاش واک همراه با نوفهٔ سایش [f] و یا ارتعاش بسیار ضعیف تارهای صوتی هنگام تولید [ɪ] قابل‌مشاهده است. لازم به ذکر است پیش‌تنظیمات پرت برای محاسبهٔ مقادیر COG، بدون تغییر مورد استفاده قرار گرفت.

---

[1] Gordon, et al.
[2] Van Son and Pols
[3] Nakamura

۲. ۱. ۳. بسامد و شدت انرژی قلهٔ طیف سایش

همخوان سایشی [f] یک همخوان لبی-دندانی است که از نزدیک‌شدن لب پایین با دندان‌های بالا تولید می‌شود. از آنجا که درمقابل گرفتگی لبی-دندانی، تقریباً هیچ فیلتری در دستگاه گفتار نیست، بنابراین طیف این همخوان هیچ قلهٔ طیفی ندارد و انرژی برروی همهٔ بسامدها پراکنده است. از این رو، در این پژوهش، فقط بسامد وشدت انرژی قلهٔ طیفی [ɪ]، پس از محاسبهٔ منحنی شدت انرژی آن در نرم‌افزار پرت، به‌طور دستی اندازه‌گیری شد.

۲. ۱. ٤. دیرش واکهٔ [e] و گذرهای سازه‌ای

دیرش واکه، عبارت است از مدت زمان تولید واکهٔ [e] و واحد آن میلی‌ثانیه (ms) است. دیرش واکه، از اولین پالس تا آخرین پالس، اندازه‌گیری شد. سازه‌ها، وضعیت دستگاه گفتار حین تولید یک آوا را نشان می‌دهند و به شکل دستگاه گفتار بستگی دارند. بنابراین، موج حنجره تأثیری در مقادیر سازه‌ای یک آوا ندارد. سازه‌ها، درواقع، فیلترهای بسامدی هستند که حاصل وضعیت دستگاه گفتار هستند و بنابراین، برحسب آواها متفاوتند. این فیلترها که در حفرهٔ دهان، حلق و بینی شکل می‌گیرند، برخی بسامدها را تضعیف و برخی را تقویت نموده، سبب به وجودآمدن باندهای بسامدی انرژی، یا به عبارت دیگر، سازه‌ها برروی طیف‌نگاشت می‌شوند. در پژوهش‌های آواشناسی، معمولاً بسامدهای اول تا سوم F1، F2 و F3 مورد مطالعه قرار می‌گیرند (ریتز و جونگمن[1] ۲۰۰۹: ۱۷۷). فلمینگ[2] (۲۰۰۲: ۲۲) معتقد است که گذرهای همخوان‌های لبی و کامی، برحسب واکهٔ بعدی تغییر می‌کند و به‌طورکلی، همخوان‌های لبی، معمولاً F2 و F3 پایین‌تری نسبت به واکهٔ مجاور خود دارند زیرا گرفتگی لبی، معمولاً سبب افت همهٔ گذرهای واکه‌ای می‌شود اما، در همخوان‌های لثوی، F2 بسامد بالاتری دارد.

صادقی (۱۳۸۴: ۵۱۶) منحنی تغییرات بسامدهای بازخوانی اول تا چهارم واکه‌های پیشین را به‌عنوان تابعی از طول حفرهٔ حلق (۶ تا ۱۱ سانتی متر) محاسبه می‌کند که با

[1] Reetz and Jongman
[2] Flemming

استفاده از آن، می‌توان بسامد سازهٔ اول /e/ را از ۵۰۰ تا ۸۰۰ هرتز (طول کمتر به بیشتر)، سازهٔ دوم بین ۱۵۰۰ تا ۱۸۰۰ هرتز و سازهٔ سوم بین ۲۵۰۰ تا ۲۹۰۰ هرتز تخمین زد. انصارین (۲۰۰۴) مقادیر سازه‌ای واکه‌های فارسی را با انجام آزمایش آکوستیکی برروی ۱۲ گویشور فارسی‌زبان و ارائهٔ فضای واکه‌ای به دست آورده است. وی میانگین مقادیر سازهٔ اول ودوم واکهٔ /e/ را به‌صورت $F1= ۶۴۴_{HZ}$ و $F2= ۲۱۱۵_{HZ}$ در کل داده‌ها به دست آورده است.

در این پژوهش، اندازه‌گیری سازه‌ها درمیان و پایان واکهٔ /e/، پس‌از استخراج الگوی سازه‌ای در نرم‌افزار پرت صورت گرفت. از آنجا که حداکثر باند بسامدی سیگنال‌های تلفنی تا ۳۴۰۰ هرتز است، بنابراین در تنظیمات پرت، تعداد سازه‌ها برابر با سه سازه در نظر گرفته شد.

۳. نتایج و تحلیل آماری داده‌ها

تجزیه و تحلیل آماری داده‌ها در نرم‌افزار SPSS 18 و با استفاده از آزمون‌های تحلیل واریانس یک‌طرفه و دوطرفه و نیز آزمون‌های تعقیبی[1] LSD و تامهن[2] صورت گرفت. در جدول ۲، آمار توصیفی پارامترهای آوایی آمده است.

۳. ۱. دیرش کلمه و دیرش سایش

مقایسهٔ میانگین دیرش [seF] با [se] نشان می‌دهد که به چه میزان، کاهش سبب شباهت واژگانی این دو رقم شده است. میزان شباهت واژگانی، تعیین‌کنندهٔ دایرهٔ تراکم واژگانی[3] حول یک واژه است که خود، عاملی بسیار مهم در درک و بازیابی واژگانی[4] است. هرچه دایرهٔ تراکم واژگانی حول یک واژه بیشتر باشد، سرنخ‌های اکوستیکی بیشتر برای بازیابی آن واژه مورد نیاز است (یوسیشکین و ودل ۲۰۰۹: ۲۸۷-۲۶۷).

[1] post hoc test
[2] Tamhane
[3] lexical density neighborhood
[4] lexical access

جدول ۲: آمار توصیفی پارامترهای آوایی COG (هرتز)، دیرش سایش، دیرش کلمه و دیرش واکهٔ /e/ (میلی‌ثانیه)

نمونه	پارامتر	میانگین	تعداد	انحراف	مینیمم	ماکزیمم
[ɪ]	دیرش سایش	۱۱۰/۰۹	۳۲	۴۱/۲۷۰	۴۲	۲۴۲
	COG	۲۲۶۳/۱۳۸		۳۴۲/۴۰۶	۱۴۲۱/۴۵	۲۷۹۴/۹۰
[f]	دیرش سایش	۷۹/۹۱	۳۲	۳۷/۴۹۷	۲۶	۱۶۹
	COG	۲۱۱۵/۰۰۰		۲۶۹/۰۹۳	۱۶۷۹/۰۰	۲۶۷۶/۸۰
[F]	دیرش سایش	۸۶/۳۹	۲۸	۳۷/۲۴۷	۲۸	۱۵۵
	COG	۲۰۴۰/۴۹۹		۱۸۴/۸۷۵	۱۶۸۸/۰۵	۲۳۷۲/۰۸
[sefɪ]	دیرش کلمه	۴۹۶/۴۴	۳۲	۹۹/۰۲۳	۲۹۳	۷۱۱
	دیرش واکه	۱۸۷/۱۹		۵۴/۰۳۶	۱۰۰	۳۳۰
[se]	دیرش کلمه	۳۴۸/۳۴	۲۹	۱۰۲/۰۲۱	۲۰۰	۶۲۳
	دیرش واکه	۲۳۰/۹۳		۹۲/۸۹۶	۱۰۹	۴۶۶
[seF]	دیرش کلمه	۳۶۶/۱۸	۲۸	۹۰/۲۹۸	۲۳۵	۶۳۹
	دیرش واکه	۱۸۱/۰۴		۴۴/۲۸۹	۱۲۴	۳۱۴

تحلیل واریانس یک‌طرفه، نشان می‌دهد که میانگین دیرش کلمه (جدول ۲) در نمونه‌های [sefɹ]، [se] و [seF] تفاوت معناداری با یکدیگر دارد (F(2, 86)=21.216, P=0.000). برطبق آزمون تعقیبی LSD، میانگین دیرش [sefɹ]، به‌طور معناداری بیشتر از دو نمونهٔ دیگر است (P=0.000)؛ اما نمونه‌های [se] و [seF]، تفاوت معناداری با یکدیگر ندارند (p>0.05). عمده‌ترین دلیل آن، حذف /r/ است. افزایش شباهت واژگانی /sefr/-/se/، بازیابی نمونه‌های کاهش‌یافته و تمایز [se]-[seF] را دشوار می‌سازد. دیرش سایش [f] در نمونه‌هایی که تضعیف یافته‌اند، تقریباً ۷ میلی‌ثانیه بیشتر از نمونه‌هایی است که در آن [sefɹ] به‌طور کامل‌تری تولید شده است. این تفاوت، به لحاظ آماری، معنادار نیست. ثمره (۱۳۸۵: ۵۸ و ۵۹) اظهار می‌کند که سایش [f] قبل از همخوان‌های سایشی، کمتر از جاهای دیگر به گوش می‌رسد و تولید آن در جایگاه قبل از همخوان‌های دیگر به‌جز انفجاری‌ها، ملایم و غیرکشیده است. نتایج این پژوهش نیز نشان می‌دهد که دیرش سایش [f]، در جایگاه قبل‌از یک همخوان سایشی نظیر [ɹ]، کمتر از حتی در پایان کلمه [seF] است. به‌طور کلی، سایشی‌های لبی، کم‌انرژی‌تر و به‌طور طبیعی کوتاهتر از سایر سایشی‌ها هستند زیرا به‌دلیل کوتاه‌شدن طول دستگاه گفتار در جلوی محل گرفتگی، هیچ‌گونه فیلتری بر روی آنها صورت نمی‌گیرد.

تولید [ɹ] مستقل از [f] است. زبان در تولید [f] نقشی ندارد و از این رو، در مرحلهٔ آمادگی برای آوای بعدی قرار دارد (ثمره ۱۳۸۵: ۵۸). می‌توان افزایش دیرش سایش را در نمونه‌های [seF] چنین توجیه کرد که در [sefɹ]، گرفتگی لبی-دندانی، با سرعت بیشتری، از وضعیت سایش خود خارج شده است زیرا دستگاه گفتار، به سمت تولید همخوان [ɹ] رفته و بنابراین، دیرش همخوان سایشی کاهش یافته است. اما در [seF]، دستگاه گفتار، در وضعیت تولید سایش، اندکی بیشتر باقی می‌ماند زیرا این آوا، در پایان گفتار قرار دارد و افزایش دیرش پایانی[1]، یکی از فرایندهای آوایی است که در پایان کلمه روی می‌دهد (کیتینگ[2] ۲۰۰۳: ۱۲۰).

[1] word-final lengthening
[2] Keating

۳. ۲. مرکز ثقل طیف

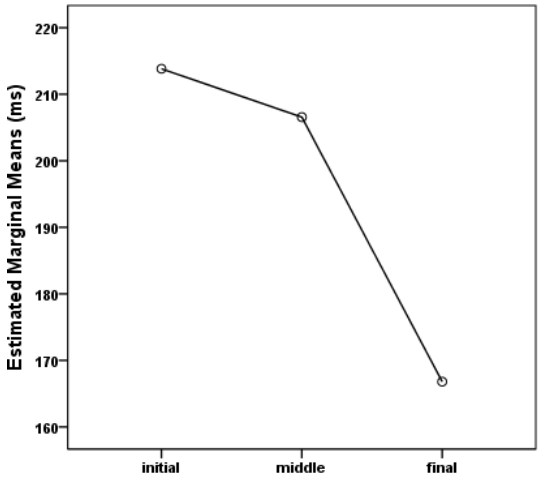
میانگین مقادیر COG (جدول ۲) نشان می‌دهد که مقدار این پارامتر در نمونه‌های تضعیف‌شده، کاهش یافته است. آزمون واریانس یک‌طرفه، نشان می‌دهد که این تفاوت معنادار نیست ($F(1, 58)= 1.518, P>0.05$). از آنجا که دو پارامتر دیرش سایش و COG تفاوت معناداری بین [sefɹ] و [seF] نشان نمی‌دهند، می‌توان چنین نتیجه گرفت که کاهش /sefr/، تنها سبب حذف همخوان /r/ شده است، اما تأثیر چندانی بر پارامترهای نوفهٔ سایش /f/ نداشته است. ازاین‌رو، به نظر می‌رسد این دو پارامتر، به‌عنوان سرنخ‌های آکوستیکی عمل می‌کنند که سبب درک صورت‌های کاهش‌یافتهٔ /sefr/ می‌شوند و تمایز شنیداری و ادرکی آنها را از واژهٔ /se/ سبب می‌شوند. در سیستم‌های بازشناسی گفتار نیز می‌توان با به‌کارگیری پارامترهای نوفهٔ سایش، بازیابی تقابل واجی /se/ - /seF/ را بهبود بخشید.

۳. ۳. دیرش واکه

برخلاف دیدگاه ثمره (۱۳۸۵: ۸۹)، میانگین دیرش [e] (جدول ۲) در پایان کلمهٔ [se]، بیشتر از مقدار آن در جایگاه ماقبل خوشه در کلمهٔ [sefɹ] است. دلیل آن را می‌توان خارج شدن سریع‌تر اندام‌های گفتاری از وضعیت واکه برای تولید همخوان بعدی در [sefɹ] دانست. آزمون تحلیل واریانس یک‌طرفه، نشان می‌دهد که دیرش واکه در این نمونه‌ها با یکدیگر تفاوت معنادار دارد ($F(2, 80)= 5.578, P= 0.005$). آزمون تعقیبی LSD، نشان می‌دهد که میانگین دیرش [e] در [se]، به‌طور معناداری بیشتر از دو نمونهٔ دیگر است ($P<0.05$). اما تفاوت معناداری، بر این اساس، بین [sefɹ] و [seF] مشاهده نمی‌شود ($P>0.0.5$). بنابراین این پارامتر آکوستیکی، تمایز ادراکی /se/-/sefr/ را حتی در صورت‌های کاهش‌یافته نشان می‌دهد و از این‌رو، می‌تواند به‌عنوان یک سرنخ آکوستیکی مهم، در بازشناسی گفتار به کار رود. ثمره (۱۳۸۵: ۸۹) معتقد است میزان دیرش این واکه در جایگاه قبل از خوشه‌های همخوانی پایانی نظیر [fɹ] در کلمهٔ [sefɹ]، بیشتر از تولید این آوا در پایان کلمهٔ [se] است؛ اما نتایج این پژوهش، نشان می‌دهد که میزان دیرش آن در

پایان واژه بسیار بیشتر، بیش از ۴۰ میلی‌ثانیه، از تولید آن در جایگاه قبل از خوشه‌های پایانی است. کیتینگ (۲۰۰۳: ۱۲۱) اظهار می‌کند که افزایش دیرش واکهٔ /o/ در پایان کلمه در زبان انگلیسی، به دلیل فرایند افزایش دیرش پایانی روی می‌دهد.

جایگاه، تأثیر معناداری بر دیرش واکه دارد ($F(2, 80)= 3.850, P< 0.05$). میانگین دیرش واکهٔ /e/ (نمودار۱)، هنگامی‌که رقم تولیدشده در پایان رشتهٔ ارقام تولید شود، به‌طور معناداری، کمتر از آغاز و میان رشتهٔ ارقام است ($P>0.05$). اما بین آغاز و میان کلمه، تفاوت معناداری مشاهده نمی‌شود ($P> 0.05$).



شکل ۲: تأثیر جایگاه رقم (آغاز، میان یا پایان رشتهٔ ارقام) بر میانگین دیرش واکهٔ [e]

۳. ٤. گذرهای سازه‌ای و بسامد کانونی

مقادیر سازه‌ای F1، F2 و F3 در وضعیت ایستان در میان واکه و پایان واکهٔ [e]، در هر یک از سه نمونه، در جدول ۳ آمده است (حرف m برای اشاره به بسامد سازه در میان واکه و حرف f برای اشاره به بسامد سازه در پایان واکه به کار رفته است. لذا F1m یعنی مقدار بسامدی سازهٔ اول در میان واکه و F1f یعنی مقدار بسامدی سازهٔ اول در انتهای واکه).

جدول ۳: میانگین مقادیر سازه‌ای (F1, F2, F3) برحسب هرتز در میان و انتهای واکهٔ [e]

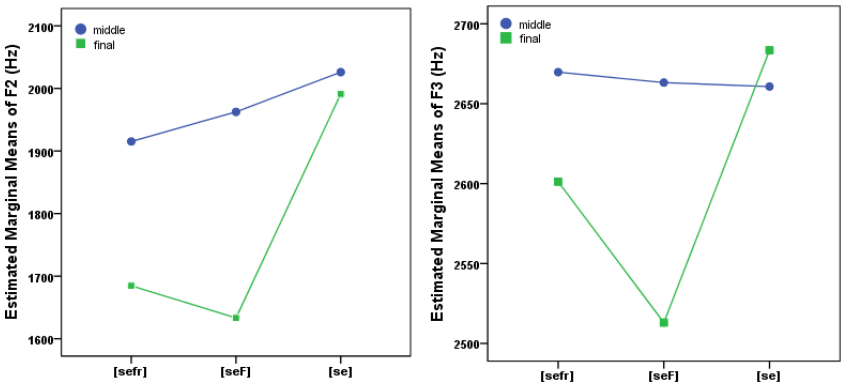
نمونه	تعداد	آمار توصیفی	F1m	F1f	F2m	F2f	F3m	F3f
[sef.ɪ]	۳۲	میانگین	۵۴۲/۴۷	۴۹۴/۴۴	۱۹۱۵/۲۲	۱۶۸۴/۶۶	۲۶۶۹/۷۵	۲۶۰۱/۲۲
		انحراف معیار	۸۵/۲۵۶	۸۴/۸۹۸	۱۶۹/۳۱۳	۱۲۶/۳۸۴	۱۶۵/۹۶۸	۱۹۶/۵۵۴
[se]	۲۹	میانگین	۵۴۷/۷۶	۵۱۹/۰۷	۲۰۲۶/۰۰	۱۹۹۱/۱۴	۲۶۶۰/۷۲	۲۶۸۳/۴۵
		انحراف معیار	۷۲/۰۳۸	۱۱۳/۸۳۱	۱۷۲/۹۳۹	۱۹۳/۷۶۱	۱۳۸/۵۹۷	۱۶۱/۹۶۶
[seF]	۲۸	میانگین	۵۵۷/۳۶	۵۳۰/۳۹	۱۹۶۲/۵۴	۱۶۳۳/۲۱	۲۶۶۳/۲۱	۲۵۱۳/۰۰
		انحراف معیار	۷۱/۵۲۶	۱۰۱/۵۳۹	۱۶۲/۴۵۸	۱۵۶/۲۱۷	۱۱۸/۱۶۴	۱۴۴/۹۰۵
کل	۸۹	میانگین	۵۴۸/۸۸	۵۱۳/۷۸	۱۹۶۶/۲۰	۱۷۶۸/۳۴	۲۶۶۴/۷۵	۲۶۰۰/۲۶
		انحراف معیار	۷۶/۲۸۶	۱۰۰/۲۹۱	۱۷۲/۷۴۳	۲۲۳/۱۰۲	۱۴۱/۸۲۶	۱۸۱/۹۴۱

مقادیر گذر هر سه سازهٔ واکهٔ [e]، از بخش ایستان واکه به انتهای آن در کلمهٔ [se]، تغییر چندانی نداشته است و مقادیر آن، به میزان زیادی، ثابت باقی مانده است. اما مقادیر سازه‌های F2 و F3، از بخش ایستان واکه به‌تدریج به سمت انتهای آن در کلمات [sef.ɪ] و

[se] کاهش یافته است. آزمون واریانس دوطرفه، نشان می‌دهد که این تفاوت معنادار است ($F(2, 83)= 51.569$ & 10.323, $P= 0.000$). کاهش مقادیر سازه‌های F2 و F3 در انتهای واکهٔ [e] در نمونه‌های [sefɹ] و [seF]، در پیِ تولید گرفتگی سایش [f] روی داده است. در واقع، تغییر وضعیت دستگاه گفتار، از واکه به همخوان، سبب بروز تغییر در الگوی بسامدی سازه‌های دوم و سوم شده است. شیب گذر این سازه‌ها، حاوی سرنخ‌های آکوستیکی لازم برای درک همخوان‌های سایشی است. مقادیر سازه‌ای F1، F2 و F3 در انتهای واکه بسامدهای کانونی همخوان محسوب می‌شوند.

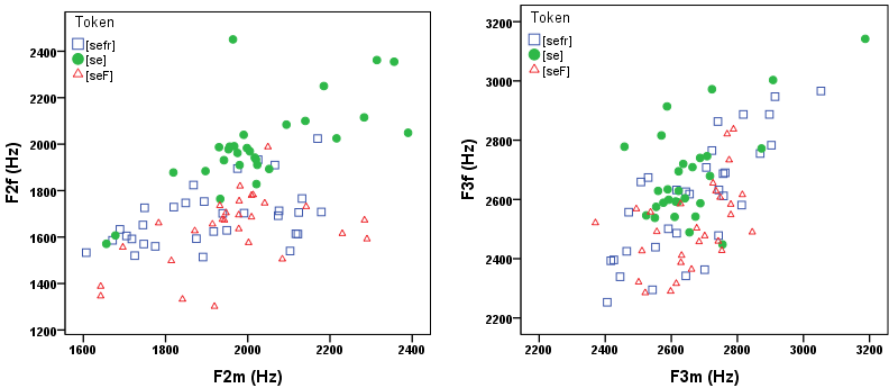
نتایج تست LSD، نشان می‌دهد که مقدار بسامدی گذر F2f در پایان واکه و پیش از گرفتگی [f] در هر یک از نمونه‌های [sefɹ] و [seF]t به‌طور معناداری، کمتر از مقدار بسامدی همین گذر در پایان واکه در کلمهٔ [se] است ($P=0.000$). اما تفاوت معناداری، بر این اساس، بین [sefɹ] و [seF] مشاهده نشد ($P>0.05$). همان‌طور که گفته شد، دلیل این افت بسامدی، به میزان ۲۰۰-۳۰۰ هرتز، تغییر وضعیت دهان از واکه به همخوان [f] است.

نتایج تست تامهن نشان می‌دهد که میانگین بسامد سازهٔ سوم در پایان واکه در [seF]، به‌طور معناداری، کمتر از [se] است ($P=0.000$). اما میانگین بسامد سازهٔ سوم در پایان واکه در [sefɹ]، تفاوت معناداری با [se] و [seF] ندارد ($P> 0.05$). کاهش بسامد سازهٔ سوم در پایانهٔ واکه در [sefɹ] و [seF]، به ترتیب ۶۸ و ۱۵۰ هرتز، به دلیل خارج شدن دهان از وضعیت واک به همخوان است اما میانگین آن در پایانهٔ واکه در [seF]، کاهش بیشتری نسبت به [sefɹ] نشان می‌دهد زیرا هنگامی که لب‌ها برای ایجاد گرفتگی لبی-دندانی به دندان‌های بالا نزدیک می‌شوند، به نظر می‌رسد، هم‌زمان، تیغهٔ زبان هم به سمت لثه برای تولید سایش [ɹ] حرکت می‌کند و از آنجا که گذر F3 در همخوان لثوی افزایش پیدا می‌کند، به نظر می‌رسد سبب شده تا بسامد این سازه در [sefɹ] کاهش کمتری نسبت به [seF] پیدا کند (نمودار ۳ و ۴).



شکل ۳ و ٤: تغییر بسامدهای F2 و F3 از میان
به پایان واکهٔ /e/ در هر یک از نمونه‌ها

نمودارهای ۵ و ۶ نشان می‌دهد که تغییرات F2، از میان به انتهای واکه، تمایز بهتری بین نمونه‌های [sefɪ] و [seF] با [se] ایجاد می‌کند زیرا بسامد سازهٔ دوم در نمونهٔ [se]، بالاتر از دو نمونهٔ اول است. اما برحسب تغییرات F3، نمی‌توان چنین تمایزی بین داده‌ها قائل شد. بر این اساس، به نظر می‌رسد که الگوی گذر سازهٔ دوم، اطلاعات آکوستیکی مطمئن‌تری، حاکی از تولید سایشی [f]، را در سیگنال گفتار انتقال می‌دهد.



شکل ۵ و ٦: پراکندگی داده‌ها برحسب تغییرات
F2 و F3 از میان به انتهای واکهٔ /e/

در جدول ۴، میانگین بسامد سازه‌های واکه در نمونهٔ [se] ، برحسب جایگاه رقم در رشتهٔ ارقام، ارائه شده است.

جدول ٤: میانگین بسامد سازه‌های واکهٔ [e] در نمونهٔ [se] برحسب هرتز بر اساس جایگاه این رقم در رشتهٔ ارقام

جایگاه رقم [se]	تعداد	آمار توصیفی	F1m	F1f	F2m	F2f	F3m	F3f
آغاز	۸	میانگین	۵۲۳/۲۵	۴۵۳/۱۳	۲۱۰۸/۸۸	۲۰۵۹/۳۸	۲۶۸۹/۱۳	۲۷۶۷/۱۳
		انحراف معیار	۵۴/۴۲۶	۸۶/۲۲۱	۱۵۹/۸۰۵	۲۰۷/۹۷۰	۱۳۸/۸۲۹	۱۸۰/۵۲۵
میان	۱۲	میانگین	۵۴۳/۹۲	۵۰۴/۳۳	۱۹۶۷/۴۲	۱۹۱۹/۷۵	۲۶۰۶/۷۵	۲۶۲۱/۰۸
		انحراف معیار	۴۸/۲۵۶	۱۰۵/۹۲۱	۱۸۰/۵۸۶	۲۰۷/۱۴۲	۶۹/۹۳۴	۱۰۲/۲۲۵
پایان	۹	میانگین	۵۷۴/۶۷	۵۹۷/۳۳	۲۰۳۰/۴۴	۲۰۲۵/۶۷	۲۷۰۷/۴۴	۲۶۹۲/۲۲
		انحراف معیار	۱۰۴/۵۵۶	۱۰۸/۸۷۷	۱۵۹/۵۴۲	۱۴۴/۶۲۳	۱۸۹/۷۶۶	۱۸۹/۷۶۰

تحلیل واریانس دوطرفه ($F(2,26)= 1.11, 4.45, 1.68, 1.50, 1.66 \& 2.12, P>0.05$)، نشان می‌دهد که مقادیر سازه‌ای [se] (جدول ۴) تفاوت معناداری، به‌جز در بسامد F1f

برحسب جایگاه آغاز، میان یا در پایان رشتهٔ ارقام، با یکدیگر ندارد بدین معنا که الگوی بسامدی واکهٔ [e]، تحت‌تأثیر واج‌های بعدی، حین تولید رشتهٔ ارقام قرار نگرفته است. همان‌طور که نتایج نشان داد (جدول ۳)، تغییر بسامد سازهٔ اول F1، تأثیری در تمایز نمونه‌ها ندارد. به نظر می‌رسد از آنجا که گویشور، قصد داشته تا هر رقم را مستقل از رقم دیگر تولید کند، هرچند رقم یا رقم‌های بعدی را بلافاصله، به‌صورت متصل، تولید کرده اما انفصال و استقلال آوایی و آکوستیکی واج‌ها، تا حدی، حفظ شده است و بازتاب آن در سیگنال گفتار، در مشخصه‌های آکوستیکی، تظاهر نموده است به‌نحوی که مقادیر سازه‌ای واکه در [se]، در رشته‌های متوالی ارقام، تحت تأثیر واج آغازین کلمهٔ بعدی، قرار نگرفته است (نمودار ۲ و ۳).

در این پژوهش، میانگین مقدار سازهٔ اول ۵۴۸ هرتز، سازهٔ دوم ۱۹۶۶ هرتز و سازهٔ سوم ۲۶۶۴ هرتز برای واکهٔ [e] به دست آمد. مقادیر به‌دست‌آمده با انصارین (۲۰۰۴) و صادقی (۱۳۸۴: ۵۱۶) به استثنای سازهٔ دوم تطابق دارد. صادقی، بسامد سازهٔ دوم را بین ۱۵۰۰–۱۸۰۰ هرتز به دست آورد اما در این پژوهش، مقدار مینیمم F2 ۱۶۰۷ هرتز و مقدار ماکزیمم آن ۲۳۹۰ هرتز است.

٤. پارامترهای صوتی /r/ در بافت /sefr/

مقادیر دیرش سایش و COG همخوان سایشی [ɹ]، در جدول ۲ ارائه شده است. مقادیر این دو پارامتر، تفاوت معناداری با مقادیر پارامترهای [f] و [F] دارد ($F(2, 89)= 5. 327 \& 5.110, P< 0.05$). دیرش سایش [ɹ]، به‌طور معناداری، بیشتر از [f] و [F] است (LSD, $P<0.05$) اما میانگین COG آن تنها، به‌طور معناداری، بیشتر از [F] است (Tamhane, $P<0.05$). مقادیر دو پارامتر دیگری که برای این همخوان اندازه‌گیری شد، بسامد قلهٔ طیف سایش و شدت انرژی آن است که میانگین آنها به ترتیب ۲۳۲۶_{Hz} و $۴۶/۵۸_{dB}$ است.

۵. بحث و نتیجه‌گیری

نتایج آزمایش، نشان می‌دهد که سرنخ‌های آکوستیکی لازم برای بازشناسی صورت‌های کاهش‌یافتهٔ /sefr/ در دو همخوان [f] و واکهٔ [e] قرار دارد. پارامترهای نوفهٔ سایش، شامل دیرش سایش و مرکز ثقل طیف و نیز پارامترهای واکه، شامل دیرش واکه، و کانون بسامدی F2، همگی، سرنخ‌های آکوستیکی لازم برای ایجاد تمایز شنیداری [se] – [seF] و بنابراین، ادراک صورت‌های کاهش‌یافتهٔ /sefr/ را فراهم می‌آورند.

عمده‌ترین تفاوت آوایی [se] – [sefɹ]، پس از حذف [ɹ]، همخوان سایشی [f] است. اما از آنجا که شدت انرژی سایش این همخوان لبی–دندانی، بسیار کم است و درواقع، در طی تولید این همخوان، منحنی انرژی همواره سیر نزولی دارد، بنابراین، بازشناسی سایش از سکوت برای سیستم‌های بازشناسی گفتار و حتی برای گوش، به‌سختی، صورت می‌گیرد. در این سیستم‌ها نیز همانند گوش، نوعی تحلیل شنیداری طیف صورت می‌گیرد (جانسون ۱۹۹۷: ۱۲۴).

کاواساکی[1] (۱۹۹۲)، دو عامل آکوستیکی را دلیل درک یا عدم درک دنباله‌های آوایی[2] بر می‌شمارد: ۱) شدت تغییر آکوستیکی[3] در درون یک دنباله ۲) میزان و درجهٔ تفاوت آکوستیکی بین دنباله‌ها. عامل اول، براساس این فرضیه است که شدت تغییر آکوستیکی، در یک دنبالهٔ آوایی، مستقیماً به برجستگی ادراکی[4] آن ارتباط دارد بدین معنا که تغییر آکوستیکی، به لحاظ ادراکی، برجسته‌تر از وضعیت سکون است و بنابراین، دنباله‌هایی که تغییر طیفی کمی دارند، اولویت ادراکی ندارند. عامل دوم، منوط به این استلزام است که تمایز تقابل‌ها باید به حداکثر خود برسد. بنابراین، تقابل‌هایی که به‌لحاظ شنیداری متمایز نیستند، به لحاظ ادراکی نیز اولویت ندارند. براساس اولین عامل آکوستیکی مطرح‌شده ازسوی کاواساکی، درک [f] در این بافت آوایی [sefɹ]، به دشواری صورت می‌گیرد زیرا شدت تغییر طیفی در درون این واحد آوایی، بسیار کم است و درنتیجه، برجستگی ادارکی آن، با وضعیت سکون (یعنی وضعیت Ø که هیچ آوایی

---

[1] Kawasaki
[2] phonetic sequences
[3] magnitude of acoustic modulation
[4] perceptual saliency

تولید نشده است) برابر است و بنابراین، به‌لحاظ شنیداری، درک تمایز آن در صورت‌های کاهش‌یافته از [se] دشوار است.

آزمایش‌های ادراکی فیجو و همکاران[1] (۱۹۹۸) نیز نشان می‌دهد که درک /f/، وابسته با بافت است به‌طوری که درک این همخوان سایشی، در بافت واکه های پسین، بیشتر (در حدود ۸۵ درصد) از بافت واکه‌های پیشین (حدوداً ۶۰ درصد) است. آنها دلیل این امر را ایجاد قلهٔ طیفی حاصل از جفت‌شدگی[2] آکوستیکی حفرهٔ جلو و عقب گرفتگی در نتیجهٔ گرد شدن لب‌ها و هم‌تولیدی با واکهٔ پسین ذکر می‌نمایند.

از آنجا که در صورت‌هایی که گویشوران [ɪ] را تولید کرده‌اند، مهم‌ترین عامل تقابل واجی /sefr - se/، انرژی نوفهٔ سایش همخوان [ɪ] است، بنابراین مهم‌ترین پارامتری که سرنخ‌های آکوستیکی لازم برای تمایز شنیداری /sefr - se/ را فراهم می‌کند، شدت انرژی قلهٔ بسامدی نوفهٔ سایش [ɪ] است و پارامترهای نوفهٔ سایش [f]، شامل دیرش و طیف آن، اطلاعات چندانی را برای بازشناسی این آواها، به‌ویژه در سیستم‌های بازشناسی گفتار فراهم نمی‌کنند. انرژی نوفهٔ سایش [ɪ] بر روی بسامدهای حدود ۱۵۰۰ تا ۳۰۰۰ هرتز متمرکز است. قلهٔ بسامدی انرژی سایش این همخوان، در حدود ۲۳۰۰ هرتز با میانگین شدت انرژی تقریبی ۴۷_{dB} است. سرنخ‌های آکوستیکی دیگری که تمایز شنیداری /sefr - se/ را سبب می‌شوند، پارامترهای واکه، شامل دیرش و بسامد کانونی پایانهٔ گذرهای واکه‌ای F2 و F3 هستند که حاوی اطلاعات محل تولید همخوان هستند. مقادیر F2 و F3، به‌طور متوسط، به‌ترتیب ۲۰۰ و ۱۰۰ هرتز در پایانهٔ واکهٔ [e] و قبل از همخوان سایشی [f] افت می‌کند. به نظر می‌رسد مهم‌ترین سرنخ آکوستیکی که می‌تواند تمایز شنیداری /se/-/sefr/ را انتقال دهد، این دو پارامتر هستند. با عملیاتی‌کردن این دو پارامتر در سیستم بازشناسی ارقام تلفنی، انتظار می‌رود بازیابی صورت‌های کاهش‌یافتهٔ /sefr/، به میزان زیادی بهبود یابد.

[1] Feijoo
[2] acoustic coupling

منابع

ثمره، یدالله (۱۳۸۵). *آواشناسی زبان فارسی: آواها و ساخت آوایی هجا*. تهران: مرکز نشر دانشگاهی.

صادقی، وحید (۱۳۸۴). *بازشناسی واجی کلمات فارسی: رویکردی مبتنی بر نظریه بهینگی*. دانشگاه تهران: پایان‌نامهٔ دکتری.

Ansarin, A. A (2004). "An acoustic analysis of Modern Persian Vowels". *SPECOM: 9th conference: Speech and Computer*. Russia: St. Petersburg. PP. 315-318.

Bijankhan, M, J. Sheykhzadegan, M.R. Roohani, R. Zarrintare, S.Z. Ghasemi and M.E Ghasedi (2003). "TFarsDat- The Telephone Farsi Speech Database". *Proceedings of EuroSpeech*. Switzerland: Geneva. PP. 1525-1528.

Cho, T, J. McQueen (2011)." Perceptual recovery from consonant-cluster simplification in Korean using language-specific phonological knowledge". *Journal of Psycholinguistic Research*. PP. 40, 253-274.

Farnetani, E (1999). "Coarticulation and connected speech processes". Edited by W. J. Hardcastle and J. Laver. *The handbook of phonetic sciences*. Oxford: Blackwell.

Feijoo, S, S. Fernandez, R. Balsa (1998). "Context effects in the auditory identification of the Spanish fricatives /f/ and /Θ/. hyper and hypo speech". *Proceedings of the ICA/ASA. 98*. PP. 2025-2026.

Flemming, E(2002). "Auditory representations in phonology". Edited by L. Horn; *outstanding dissertations in Linguistics*. New York: Routledge.

Gordon, M, P. Barthmaier K. Sands (2002). "A cross-linguistic acoustic study of voiceless fricatives". *Journal of the International phonetic association*; 32 (2). PP. 141-174.

Hedrick, M. S, Ralph N. Ohde (1993). "Effect of relative amplitude of frication on perception of place of articulation". *Journal of Acoustical Society of America*, 94 (4). PP. 2005-2026.

Johnson, K (1997). *Acoustic and auditory phonetics*. Oxford: Blackwell

Kawasaki, H (1992). "An acoustical basis for universal phonotactic constraints". *Language and Speech*. PP. 35, 73-86.

Keating; P. A (2003). "Phonetic encoding of prosodic structure". *Proceedings of the 6th international seminar on speech production*. Australia. Sydney. PP. 119-124.

Mitterer; H, M. Ernestus (2006). "Listeners recover /t/s that speakers reduce. evidence from /t/-lenition in Dutch". *Journal of phonetics*. PP. 34; 73- 103.

Nakamura, M, k. Iwano, s. furui (2007). "Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance". *Computer Speech and Language*. PP. 22. 171-184.

Reetz, H, A. Jongman (2009). *Phonetics. transcription, production, acoustics, and perception*. UK: Wiley-Blackwell.

Ussishkin, A, A. Wedel (2009)." Lexical access, effective contrast, and patterns in the lexicon. Edited by Boersma. Paul., and Silke Hamann". *Phonology in Perception*. Germany: Mouton de Gruyter.

Van Son, R.J.J.H, L.C.W. Pols (1997). "The correlation between consonant identification and the amount of acoustic consonant reduction". http.//www.fon.hum.uva.nl/IFA-publications/Eurospee-ch97/A0301/A0301.html

Weibel, A, K.F. Lee (1990). *Readings in speech recognition*. Burlington :Morgan Kaufmann.