



وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران

گزارش طرح

بهبود سیستم بازشناسی اعداد تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی

مجری: زهرا محمودزاده

شهریور ۱۳۹۰

بہ نام خدا

چکیده

استفاده از دستاوردهای آواشناسی یکی از راههایی است که می‌تواند به بهبود دقت بازشناسی گفتار کمک کند. در گفتار تلفنی عوامل متعددی سبب از بین رفتن تمایزات واجی و در نتیجه ابهام آوایی می‌شوند. اما این تمایزات واجی توسط برخی از پارامترهای اکوستیکی باقیمانده در سیگنال گفتار، انتقال پیدا می‌کنند. تحلیل ۱۸۱ نمونه آوایی که توسط ۱۳ زن و ۱۴ مرد در بافتهای مختلف آوایی تولید شده‌اند، در مرحله اول این پژوهش نشان می‌دهد که پارامترهای نوفه سایش شامل دیرش سایش و مرکز ثقل طیف و نیز اطلاعات آوایی واکه شامل دیرش واکه، بسامد پایانه F2 و شیب گذر F2، تمایز واجی [se]-[sefɪ] را حتی در نمونه‌های کاهش یافته‌ای مانند [seF] در سیگنال گفتار انتقال می‌دهند.

از آنجا که پارامترهای واکه سرنخهای دقیقتر و بهتری برای انتقال تمایز واجی [se]-[sefɪ] در سیگنال تلفنی هستند، بنابراین سه پارامتر دیرش واکه، بسامد پایانه F2 و شیب گذر F2 در ۸۱۱ نمونه آوایی (۴ زن و ۹ مرد) مورد اندازه‌گیری قرار گرفت و به عنوان داده‌های تست برای افزایش دقت سیستم بازشناسی مورد استفاده قرار گرفت. در این تحقیق برای افزایش دقت بازشناسی از روش تلفیق دو شیوه بازشناسی بر اساس مدل مخفی مارکوف و ضرایب مل-کپستروم، و نیز بازشناسی بر اساس پارامترهای استخراج شده از واکه صورت گرفت. یکی از راه‌های ترکیب نتایج حاصل از دو روش فوق، استفاده از معیار اطمینان می‌باشد. یکی از روش‌هایی که می‌توان برای سیستم مبتنی بر پارامتر، معیار اطمینان تعریف کرد استفاده از درجه شباهت نرمال شده‌ای است که با استفاده از توزیع گوسی به دست می‌آید. مشاهده می‌شود که دقت حدود ۱٪ بیشتر از دقت سیستم بازشناسی مبتنی بر HMM (۹۳٫۳۳٪) است و به ۹۴٫۳۱٪ می‌رسد. بنابراین می‌توان گفت روش ترکیبی مبتنی بر معیار اطمینان خطای بازشناسی ارقام "سه" و "صفر" را به‌طور نسبی حدود ۱۵٪ کاهش داده است.

کلیدواژه‌ها: بازشناسی ارقام تلفنی بر اساس پارامترهای اکوستیکی، آواهای [e]، [f] و [ɪ]، دیرش واکه، بسامد F2، COG نوفه سایش

فهرست مطالب

عنوان	شماره صفحه
۱. مقدمه	۲
۱-۱. بیان مسئله	۳
۲-۱. پیشینه پژوهش	۴
۳-۱. سؤالهای پژوهش	۴
۴-۱. فرضیه‌ها	۵
۵-۱. مراحل اجرایی پژوهش	۵
۲. مقدمه	۷
۲-۱. کاهش آوایی	۸
۲-۲. تأثیر کاهش بر پارامترهای اکوستیکی	۹
۲-۲-۱. مشخصات واجی	۱۰
۲-۲-۳. روش شناسی	۱۲
۲-۳-۱. داده‌ها	۱۲
۲-۳-۲. نرخ و سبک گفتار	۱۷
۲-۳-۳. پارامترهای اکوستیکی	۱۸
۲-۳-۳-۱. پارامترهای نوفه سایش	۱۹
۲-۳-۳-۲. پارامترهای واکه	۲۲

۲۶.....	۴-۲. تجزیه و تحلیل آماری داده‌ها
۲۶.....	۲-۴-۱. دیرش کلمه
۲۹.....	۲-۴-۲. دیرش سایش [f]
۳۰.....	۲-۴-۳. مرکز ثقل طیف COG
۳۲.....	۲-۴-۴. دیرش واکه [e]
۳۴.....	۲-۴-۵. تأثیر جایگاه رقم در رشته ارقام بر دیرش واک
۳۶.....	۲-۴-۶. گذرهای سازه‌ای و بسامد مکانی (locus frequency)
۳۹.....	۲-۴-۶-۱. گذر و بسامد مکانی F2
۴۱.....	۲-۴-۶-۲. گذر و بسامد مکانی F3
۴۵.....	۲-۴-۶-۳. تأثیر جایگاه کلمه [sef.i] در رشته ارقام بر مقادیر سازه‌ای واکه
۵۰.....	۲-۴-۶-۴. تأثیر جنسیت بر پارامترهای سازه‌ای
۵۲.....	۲-۵. همخوان سایشی [i]
۵۶.....	۲-۶. بحث و تحلیل نتایج
۶۰.....	۳-۱. مقدمه
۶۰.....	۳-۲. روش شناسی
۶۰.....	۳-۲-۱. نمونه‌ها
۶۰.....	۳-۲-۲. شناسایی ارقام
۶۰.....	۳-۲-۲-۱. شناسایی ارقام "سه" و "صفر" بر اساس پارامترهای واکه
	۳-۲-۲-۲. شناسایی ارقام "سه" و "صفر" با استفاده از سیستم بازشناسی گفتار مبتنی بر مدل مخفی مارکوف
۶۲.....	
۶۳.....	۳-۲-۲-۳. ترکیب نتایج بازشناسی حاصل از دو روش فوق
۶۴.....	۳-۳. تجزیه و تحلیل آماری پارامترها
۶۴.....	۳-۱-۳. دیرش واکه
۶۷.....	۳-۲-۳. بسامد پایانه F2
۷۰.....	۳-۳-۳. گذر F2
۷۱.....	۳-۴. نتیجه‌گیری
۷۳.....	منابع
۷۶.....	پیوست

فهرست جدولها، شکلها و نمودارها

عنوان	شماره صفحه
جدول ۱-۲: نمونه‌های مورد مطالعه و تعداد آنها	۱۳
شکل ۱-۲: سیگنال و منحنی انرژی [sef.I]	۱۴
شکل ۲-۲: طیف‌نگاشت و منحنی انرژی [sef.I]	۱۴
شکل ۳-۲: سیگنال و منحنی انرژی در نمونه‌های کاهش‌یافته [seF]	۱۵
شکل ۴-۲: طیف‌نگاشت و منحنی انرژی در نمونه‌های کاهش‌یافته [seF]	۱۶
شکل ۵-۲: سیگنال و منحنی انرژی [se]	۱۶
شکل ۶-۲: طیف‌نگاشت و منحنی انرژی [se]	۱۷
شکل ۷-۲: ناحیه سایش [f] و [I] بر روی سیگنال	۲۰
شکل ۸-۲: ناحیه سایش [f] و [I] بر روی طیف‌نگاشت	۲۰
شکل ۹-۲: بسامدهای سازه‌ای، افزایش F1 و کاهش دو سازه F2 و F3 در انتهای واکه	۲۴
شکل ۱۰-۲: کاهش هر سه سازه F1، F2 و F3 در انتهای واکه	۲۴
شکل ۱۱-۲: الگوی گذر F1 از میان به انتهای واکه تغییر بسامدی نیافته، اما F2 و F3 در انتهای واکه کاهش یافته‌اند.	۲۵
جدول ۲-۲: آمار توصیفی پارامتر دیرش زمانی (ms) آواها و کلمات	۲۶

جدول ۲-۳: نتایج آزمون واریانس یک طرفه، مقایسه پارامتر میانگین دیرش کلمات	۲۷
جدول ۲-۴: نتایج آزمون LSD، مقایسه میانگین دیرش زمانی کلمات با یکدیگر	۲۷
نمودار ۲-۱: تفاوت دیرش زمانی نمونه‌ها	۲۸
جدول ۲-۵: آمار توصیفی پارامتر دیرش سایش [f] در نمونه‌های کاهش یافته و نیافته	۲۹
جدول ۲-۶: آمار توصیفی پارامتر مرکز ثقل طیف سایش (COG) در نمونه‌های کاهش یافته و نیافته	۳۰
نمودار ۲-۷: طیف سایش [f] و بسامد COG	۳۰
جدول ۲-۷: نتایج آزمون واریانس یک طرفه، مقایسه پارامتر میانگین مرکز ثقل طیف	۳۱
نمودار ۲-۳: تفاوت میانگین COG سایش در نمونه‌های کاهش یافته با نیافته	۳۱
جدول ۲-۸: آمار توصیفی پارامتر دیرش واکه [e]	۳۲
جدول ۲-۹: نتایج آزمون واریانس یک طرفه، مقایسه پارامتر میانگین دیرش واکه [e] و نیز تأثیر جایگاه (place) بر آن	۳۲
جدول ۲-۱۰: نتایج آزمون LSD، مقایسه میانگین دیرش واکه در نمونه‌ها با یکدیگر	۳۳
نمودار ۲-۴: میانگین دیرش واکه در نمونه‌ها	۳۳
جدول ۲-۱۱: آمار توصیفی پارامتر دیرش واکه [e] برحسب جایگاه تولید رقم در رشته ارقام	۳۴
جدول ۲-۱۲: نتایج آزمون LSD، مقایسه میانگین دیرش واکه [e] برحسب جایگاه رقم در رشته ارقام	۳۵
نمودار ۲-۵: دیرش واکه برحسب جایگاه رقم در رشته ارقام	۳۶
نمودار ۲-۶: تأثیر جایگاه رقم بر میزان دیرش واکه آن در هر نمونه	۳۶
جدول ۲-۱۳: آمار توصیفی پارامترهای سازه‌ای واکه	۳۷
جدول ۲-۱۴: تفاوت میانگین بسامدی سازه‌ها از میان به انتهای واکه در هر یک از نمونه‌ها	۳۷
جدول ۲-۱۵: نتایج آزمون واریانس دو طرفه، مقایسه میانگین بسامدی سازه‌ها در نمونه‌ها و تأثیر جنسیت (sex) بر آن	۳۸

جدول ۲-۱۶: نتایج آزمون LSD، مقایسه میانگین مقادیر سازه F2 در نمونه‌ها	۳۹
نمودار ۲-۷: تفاوت بسامد گذر F2 در پایان واکه در هر یک از نمونه‌ها	۴۰
نمودار ۲-۸: تفاوت بسامد گذر F2 از میان به پایان واکه	۴۱
جدول ۲-۱۷: نتایج آزمون Tamhane، مقایسه میانگین مقادیر F3 در نمونه‌ها	۴۱
نمودار ۲-۹: تفاوت بسامد گذر F3 در پایان واکه در هر یک از نمونه‌ها	۴۲
نمودار ۲-۱۰: تفاوت بسامد گذر F3 از میان به پایان واکه در هر یک از نمونه‌ها	۴۳
نمودار ۲-۱۱: تغییر بسامد سازه‌ها از میان به پایان واکه	۴۴
نمودار ۲-۱۲: پراکندگی داده‌ها برحسب تغییر F2 از میان به پایان واکه	۴۴
نمودار ۲-۱۳: پراکندگی داده‌ها برحسب تغییر F3 از میان به پایان واکه	۴۵
جدول ۲-۱۸: آمار توصیفی مقادیر سازه‌ای واکه برحسب جایگاه رقم (place) در رشته ارقام	۴۵
جدول ۲-۱۹: نتایج آزمون تحلیل واریانس دوطرفه، تأثیر جایگاه بر پارامترهای سازه‌ای واکه در [se]	۴۶
نمودار ۲-۱۴: بسامد F2f برحسب جایگاه نمونه در رشته ارقام	۴۸
نمودار ۲-۱۵: بسامد F3f برحسب جایگاه نمونه در رشته ارقام	۴۹
نمودار ۲-۱۶: تغییر بسامد هر سه سازه در جایگاه آغازین رشته ارقام برحسب نمونه	۴۹
نمودار ۲-۱۷: تغییرات هر سه سازه در جایگاه میانی رشته ارقام برحسب نمونه	۵۰
نمودار ۲-۱۸: تغییرات هر سه سازه در جایگاه پایانی رشته ارقام برحسب نمونه	۵۰
جدول ۲-۲۰: آمار توصیفی مقادیر سازه‌ای برحسب جنسیت (sex)	۵۰
نمودار ۲-۱۹: تأثیر جنسیت بر F1 از میان به پایان واکه	۵۱
نمودار ۲-۲۰: تأثیر جنسیت بر F2 از میان به پایان واکه	۵۱
نمودار ۲-۲۱: تأثیر جنسیت بر F3 از میان به پایان واکه	۵۲
جدول ۲-۲۱: آمار توصیفی پارامترهای همخوان [J]	۵۲

- جدول ۲-۲۲: نتایج آزمون واریانس دوطرفه، مقایسه میانگین دیرش و COG همخوان [I] با [f] در نمونه‌های کاهش یافته و نیافته ۵۳
- جدول ۲-۲۳: نتایج آزمون LSD و Tamhane، مقایسه میانگین پارامترهای دیرش و COG در نمونه‌ها ۵۳
- نمودار ۲-۲۲: تفاوت دیرش سایش [I] با [f] در نمونه‌های کاهش یافته و نیافته ۵۴
- نمودار ۲-۲۳: تفاوت COG سایش [I] با [f] در نمونه‌های کاهش یافته و نیافته ۵۵
- جدول ۳-۱: دقت بازشناسی سیگنال‌های "سه" و "صفر" بر اساس پارامترهای مختلف ۶۱
- جدول ۳-۲: نتایج نهایی بازشناسی ارقام "سه" و "صفر" با استفاده از روش رأی‌گیری بین نتایج حاصل از پارامترها ۶۳
- جدول ۳-۳: میانگین مقادیر دیرش واکه برحسب جایگاه ۶۴
- جدول ۳-۴: نتایج آزمون تحلیل واریانس دو طرفه، تأثیر جایگاه و جنسیت بر دیرش واکه ۶۴
- جدول ۳-۵: نتایج آزمون تعقیبی LSD، تفاوت دیرش واکه برحسب جایگاه ۶۵
- نمودار ۳-۱: تفاوت دیرش واکه در نمونه‌ها برحسب جایگاه ۶۵
- جدول ۳-۶: میانگین دیرش واکه برحسب جنسیت در هر یک از نمونه‌ها ۶۶
- نمودار ۳-۲: تفاوت میانگین دیرش واکه برحسب جنسیت در هر یک از نمونه‌ها ۶۶
- جدول ۳-۷: بسامد افزایش و کاهش F2 از میان به انتهای واکه ۶۷
- جدول ۳-۸: آمار توصیفی مقادیر F2 از میان به انتهای واکه برحسب جایگاه ۶۸
- جدول ۳-۹: میانگین مقادیر گذر F2 برحسب جنسیت ۶۸
- جدول ۳-۱۰: نتایج آزمون تحلیل واریانس دوطرفه ۶۹
- جدول ۳-۱۱: میانگین کاهش سازه دوم برحسب جایگاه ۶۹
- جدول ۳-۱۲: میانگین افزایش سازه دوم برحسب جایگاه ۷۰

فصل اول

معرفی طرح پژوهشی

۱. مقدمه

از جمله اهداف استراتژیک پژوهشگاه، گسترش فناوری اطلاعات در دو حوزه نرم‌افزاری و سخت‌افزاری است. از سویی ارائه و انتقال اطلاعات به زبان فارسی، دسترسی فارسی‌زبانان بیشتری را به منابع اطلاعاتی میسر می‌سازد. بنابراین تولید و تهیه نرم‌افزارها و سخت‌افزارهایی که قادر به پردازش اطلاعات در حوزه زبان فارسی باشند، در راستای گسترش فناوری اطلاعات قرار دارد.

همانطور که گفته شد، گسترش علوم و فناوری اطلاعات در ایران مستلزم گسترش کاربرد زبان فارسی به وسیله سیستمهای انتقال اطلاعات است. این سیستمها برای انتقال اطلاعات در حوزه زبان فارسی باید قادر به پردازش زبان فارسی در دو حوزه متن و گفتار (text and speech) باشند. پژوهش حاضر در حوزه پردازش گفتار تلفنی فارسی (Persian telephonic speech processing) صورت می‌گیرد.

در سیستمهای پردازش گفتار تلفنی هدف بر آن است تا گفتار گوینده، بر روی خط تلفن، به طور کامل توسط سیستم بازشناسی شده و در برخی موارد به صورت خروجی نوشتاری ارائه شود. در سیستم بازشناسی ارقام تلفنی، هدف بر آن است که ارقام فارسی، به صورت شماره حساب بانکی، شماره کد ملی، شماره تلفن، شناسنامه، شناسه قبض و پرداخت و غیره توسط کاربر از طریق تلفن در اختیار سیستم قرار گرفته و سیستم با بازشناسی این ارقام، به صورت خروجی ارائه کند. بنابراین بهبود سیستمهای بازشناسی ارقام تلفنی، گامی در جهت بهبود سیستمهای فناوری اطلاعات است.

۱-۱. بیان مسئله

یکی از مدل‌های بازشناسی گفتار، مدل‌های مبتنی بر آواشناسی آکوستیک یا (Acoustic-phonetic feature based models) است. در این مدل‌ها سعی می‌شود تا بر اساس مشخصه‌های آکوستیکی و آوایی سیگنال گفتار، واجها^۱ گفتار استخراج شوند. از آنجا که بر روی خطوط تلفن، میزان اطلاعات سیگنال گفتاری به میزان قابل توجهی کاهش می‌یابد، استخراج واجها به وسیله سیستم‌های بازشناسی گفتار تلفنی دچار مشکل و اختلال می‌شود و بنابراین تحلیل و استخراج آواها نیازمند بررسی و تحلیل دقیق مشخصه‌های آکوستیکی با استفاده از دست‌یافته‌های علم آواشناسی و زبانشناسی است.

آواهای زبان بر اساس تفاوت‌های تولیدی و مشخصه‌های آکوستیکی به چند طبقه به صورت زیر تقسیم می‌شوند: ۱. انفجاری^۲، ۲. سایشی^۳، ۳. انفجاری-سایشی یا انسایشی^۴، ۴. خیشومی^۵ و ۵. ناسوده^۶.

سیستم شنیداری گوش انسان هر یک از این طبقات آوایی را بر اساس مشخصه‌ها و مقادیر آکوستیکی متفاوت درک می‌کند. در سیستم‌های بازشناسی گفتار که مبتنی بر مشخصه‌های آکوستیکی هستند، نیز فرایند بازشناسی بر اساس مدل گوش انسان با استخراج خودکار این مشخصه‌ها صورت می‌گیرد. بنابراین سیستم برای استخراج و بازشناسی دقیق آواها نیازمند اطلاعات آکوستیکی دقیق و تفصیلی برای هر یک از آواها می‌باشد.

کاهش پهنای باند انتقال اطلاعات بر روی خطوط تلفن به میزان ۳۱۰۰ تا ۳۴۰۰ هرتز، سبب افت و کاهش اطلاعات آکوستیکی لازم برای دریافت و تشخیص آواهای زبان حتی از سوی انسان می‌شود زیرا گوش انسان برای بازشناسی آواهای زبان به اطلاعات آکوستیکی تا حدود ۸ کیلوهرتز نیاز دارد. سیستم‌های بازشناسی گفتار تلفنی نیز به دلیلی مشابه یعنی کاهش اطلاعات سیگنال دچار اختلال می‌شوند و در برخی موارد قادر به بازشناسی سیگنال‌های مشابه از یکدیگر نمی‌باشند. به عنوان مثال این سیستم‌ها در بازشناسی ارقام "سه و صفر"، "دو و نه"، "هفت و هشت" و نیز ترکیب و باهم آیی ارقام، مرز

^۱. phonemes

^۲. stops

^۳. fricatives

^۴. affricates

^۵. nasals

^۶. approximants

ارقام به دلیل هم‌پوشانی آواها از گویشور به گویشور، لهجه به لهجه و نیز فاکتورهای ماند سن (بالغ، کودک)، جنسیت (زن، مرد) و غیره دچار اختلال می‌شوند.

۲-۱. پیشینه پژوهش

در زمینه آواشناسی زبان فارسی، مطالعات مترقم اکوستیکی صورت گرفته است. اما ادبیات موجود در زمینه تجزیه و تحلیل اکوستیکی موج صوتی تلفنی بسیار کم و محدود است. از جمله تحقیقاتی که بر روی سیگنالهای تلفنی صورت گرفته است، پایان‌نامه دکترى خانم شهین شیخ سنگ تجن (۱۳۸۹) تحت عنوان "بررسی صوت شناختی کاهش واکه‌های زبان فارسی در مکالمات تلفنی" (شیخ سنگ تجن ۱۳۸۹) است که تضعیف واکه‌های زبان فارسی بر روی سیگنالهای تلفنی را تجزیه و تحلیل آواشناسی می‌کند. نتایج این نشان می‌دهد که در زبان فارسی واکه‌های پایدار /i/ و /u/ در هجاهای بدون تکیه تغییر چندانی نمی‌یابند، در حالیکه واکه‌های ناپایدار به ویژه واکه‌های /e/ و /a/ در چنین هجاهایی به نسبت بسیار زیادی به یکدیگر و به واکه‌های پایدار گرایش می‌یابند. از این رو کاهش واکه‌ای در فارسی از یک رویکرد تلفیقی (Hybrid) تبعیت می‌کند (شیخ سنگ تجن ۱۳۸۹).

کاربرد دستاوردهای علم آواشناسی در سایر علوم و از جمله بازشناسی گفتار سابقه دیرینه دارد. از جمله این آثار "افزایش دقت بازشناسی گفتار واجهای زبان فارسی با استفاده از خصوصیات نوایی گفتار در سطح واجها" است. نتایج این پژوهش نشان می‌دهد که افزودن ویژگیهای نوایی در سطح واجها سبب افزایش یک و نیم درصد در صحت بازشناسی واجها می‌شود (الماس گنج و دیگران ۱۳۸۹).

۳-۱. سؤالهای پژوهش

در این پژوهش سعی بر آن است که به سؤالهای زیر پاسخ داده شود.

۱. چه پارامترهای اکوستیکی، آواهای فارسی را در هر یک از بافتهای متفاوت آوایی از یکدیگر متمایز می‌سازد؟

۲. مقادیر کمی هر پارامتر اکوستیکی در بافتهای مختلف چقدر است؟

۳. چه پارامترهای اکوستیکی آواهای فارسی را بر روی سیگنالهای تلفنی از یکدیگر متمایز می‌سازد و مقادیر این پارامترها چقدر است؟

۴. بهبود سیستم بازشناسی ارقام تلفنی با استفاده از چه پارامترها و مقادیر اکوستیکی صورت می‌گیرد؟

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی £ ۵

۴-۱. فرضیه‌ها

فرضیات این پژوهش عبارتند از:

۱. مدل اکوستیکی-آوایی بازشناسی گفتار با استخراج مشخصه‌های اکوستیکی هر واج زبان فارسی امکان بازشناسی ارقام را در سیستم بازشناسی ارقام تلفنی میسر می‌سازد.
۲. استخراج واجهای ارقام تلفنی در سیستم بازشناسی با استفاده از مدل زبانی امکان‌پذیر است.
۳. مقادیر اکوستیکی آواهای گفتار بر روی خطوط تلفن تضعیف می‌شود.

۵-۱. مراحل اجرایی پژوهش

پژوهش شامل مراحل زیر است.

- ۱- مطالعه و پژوهش در زمینه خصوصیات و مشخصه‌های اکوستیکی آواهای زبان فارسی و استخراج مقادیر پارامترهای صوت‌شناختی
- ۲- مطالعه و پژوهش بر روی سیگنالهای تلفنی ارقام و استخراج مشخصه‌های اکوستیکی
- ۳- برنامه‌نویسی جهت استخراج و بازشناسی واجهای فارسی در سیستم بازشناسی ارقام تلفنی

فصل دوم

تجزیه و تحلیل اکوستیکی

۲. مقدمه

در گفتار روزمره، کلمات ممکن است کاهش پیدا کنند. بدین معنا که همه مشخصه‌های واجی لازم برای تمایزات آوایی در سیگنال گفتار ظاهر نشوند. گاهی کاهش آوایی^۷ منجر به حذف یک صورت آوایی می‌شود، به نحوی که در سیگنال و یا طیف‌نگاشت^۸، تولید آوا مشاهده نمی‌شود. بدین ترتیب کاهش درجات مختلف از حذف یک آوا تا تضعیف برخی سرنخهای ادراکی را شامل می‌شود. این چنین فرایندهایی که سبب تغییر صورت واجی یک کلمه می‌شوند، چالشی هستند برای پردازش گفتار. از جمله این فرایندها همگونی، حذف، هم‌تولیدی^۹ و کاهش آوایی را می‌توان نام برد.

مدلهای متفاوت بازشناسی گفتار معمولاً بر اساس درک گفتار طرح‌ریزی می‌شوند. درک گفتار و بنابراین بازشناسی گفتار در دو مرحله صورت می‌گیرد. ابتدا درک سرنخهای اکوستیکی واحدهای واجی که می‌تواند شامل مشخصه‌های واجی، واج یا هجا باشد و سپس بازشناسی کلمات با استفاده از واحدهای واجی. در صورت‌های کاهش یافته، درک گفتار با استفاده از سرنخهای اکوستیکی باقیمانده در سیگنال صوتی و بافت واجی روی می‌دهد. هنگامی که صورت کاهش یافته در گفتار پیوسته و در درون جمله تولید می‌شود، علاوه بر اطلاعات آوایی، اطلاعات نحوی و معنایی نیز به بازیابی صورت کاهش یافته کمک می‌کند.

^۷. phonetic reduction

^۸. spectrogram

^۹. coarticulation

در این پژوهش نیز سعی بر آن است تا فرایند کاهش رقم صفر /sefr/ در گفتار خواندنی^{۱۰} و تلفنی مورد تجزیه و تحلیل صوت‌شناختی قرار گیرد زیرا کاهش آوایی این رقم سبب از بین رفتن تمایزات واجی /sefr-se/ در بازشناسی خود کار می‌گردد. با تجزیه و تحلیل مفصل اکوستیکی و استخراج سرنخهای اکوستیکی باقیمانده در سیگنال روشن خواهد شد که درک گفتار توسط انسان بر اساس چه اطلاعات آوایی صورت می‌گیرد و بنابراین می‌توان با افزودن این اطلاعات به سیستمهای بازشناسی گفتار، بازشناسی این تمایز را بهبود بخشید.

۲-۱. کاهش آوایی

کاهش تمایز واجی /sefr-/se/ را می‌توان به عوامل مختلفی نسبت داد. اولین عاملی که نقش بسیار عمده‌ای در حذف این تمایز دارد، ماهیت تلفنی سیگنال است زیرا در سیگنالهای تلفنی به علت حذف اطلاعات صوتی و کاهش فرکانس انتقال به ۳۱۰۰ تا ۳۴۰۰ هرتز بسیاری از اطلاعات آوایی حذف می‌شود. علاوه بر حذف فرکانسهای بالا، عامل دیگر، نویز جریان برق در خطوط تلفن است که خود سبب تغییر برخی مقادیر صوتی نظیر طیف و شدت انرژی آوا شده و بنابراین سبب ایجاد ابهام شنیداری و نیز دشواری بازشناسی خود کار می‌شود. علاوه بر موارد بالا، عوامل بافتی و آوایی نیز سبب کاهش این تمایز می‌شوند.

تمایز واجی /sefr-/se/ شامل دو همخوان /f, r/ است. عدم بازشناسی این دو آوا، سبب حذف این تمایز می‌شود. عامل بافتی کاهش این دو همخوان، جایگاه تولید آنهاست که در خوشه پایانی کلمه است. اسکور (۱۹۷۷: ۵۹) به محدودیت جهانی زنجیره‌های واجی یعنی اصل "ساخت هجایی مرجح"^{۱۱} به صورت CV(CV) در زبانها اشاره می‌کند و معتقد است که پیوستار بافتی به این اولویت وابسته است. بدین صورت که خوشه‌ها و همخوانهای منفرد در جایگاه پایانی با این الگو تطابق ندارند و بنابراین کاهش پیدا می‌کنند. در دیدگاه اسکور (۱۹۷۷: ۶۲)، تضعیف در درجه اول به توالی واحدهای آوایی بستگی دارد: فرایند تضعیف بیشتر در خوشه‌ها و جایگاه پایانی رخ می‌دهد و سپس در جایگاه بین‌واکه‌ای.

علاوه بر جایگاه تولید سرعت گفتار و سبک گفتاری گویشور نیز متغیرهایی هستند که بر میزان و شدت کاهش تأثیرگذار هستند. کرچنر (۱۹۹۸: ۱۷۹) در زمینه عوامل مؤثر بر وقوع کاهش، دو تعمیم اصلی ارائه می‌کند: وی معتقد است در شرایط مساوی اولاً کاهش، بیشتر، زمانی روی می‌دهد که

^{۱۰}. read speech

^{۱۱}. preferred syllable structure

واحدهای آوایی اطراف یک همخوان بازتر باشند؛ بافت بین واکه‌ای نمونه‌ای از این تعمیم است و ثانیاً کاهش، بیشتر، موقعی روی می‌دهد که گفتار سریعتر باشد و یا سبک گفتار عامیانه‌تر^{۱۲} باشد. کرچنر^{۱۳} (۱۹۹۸: ۱۷۹) این دو تعمیم را جلوه‌هایی از بافتهای مبتنی بر کوشش می‌داند. بدین معنا که کاهش در این بافتها راحتتر رخ می‌دهد زیرا ایجاد گرفتگی در این بافتها مستلزم کوشش و انرژی بیشتری است.

در واژه /sefr/، توالی /fr/ در جایگاه پایانی خوشه است. علاوه بر آن بافت اطراف همخوان سایشی /f/ نیز یک بافت باز /e-r/ محسوب می‌شود و بنابراین احتمال کاهش آن بیشتر می‌شود.

۲-۲. تأثیر کاهش بر پارامترهای اکوستیکی

میزان قوت یک همخوان به بخشهای آوایی آن شامل دیرش، شدت، واکداری و ساختار سازه‌ای بستگی دارد (Lavoie, 2001, 8). عوامل بروز تضعیف که به آنها اشاره شد، پیامدهای آوایی متفاوتی دارند و تأثیرات متفاوتی بر پارامترهای مختلف صوت‌شناختی و تولیدی دارند. لاووا (همان: ۱۵۸) ویژگیهای یک همخوان ضعیف را بدین صورت ارائه می‌کند: کاهش دیرش، کاهش تماس زبانی کامی^{۱۴}، ساختار سازه‌ای وسیعتر، انسداد ناقص نوفه‌دار، فقدان انفجار رهش یا انفجار رهش ضعیف، افزایش انرژی منظم (واکداری)، کاهش انرژی نامنظم شامل شدت سایش یا دمش.

لاووا^{۱۵} (۲۰۰۱: ۱۵۹) معتقد است اولین معادل صوت‌شناختی کاهش، دیرش کوتاهتر است. وی پیش‌بینی می‌کند که انسدادیهای ضعیفتر دارای بستهاهای کوتاهتر و نسبتاً ناقص هستند به نحوی که نوفه سایش به درون بست نفوذ می‌کند و در صورتی که با رهش، سایش و دمش تولید شوند، دیرش هر یک از این پارامترها، کوتاه است و شدت صوت‌شناختی کمی دارد. به طور کلی همخوانهای ضعیفتر انفجار رهش کمتر، دامنه رهش پایینتر، دمش و سایش کوتاهتر و ضعیفتر دارند.

بر اساس آنچه که گفته شد، پارامترهای زمانی و طیفی دو همخوان /f, r/ در پایان کلمه /sefr/ و نیز کیفیت و ساختار سازه‌ای واکه /e/ مورد تجزیه و تحلیل اکوستیکی قرار می‌گیرد تا با تعیین میزان و شدت کاهش، سرنخهای اکوستیکی لازم برای بازشناسی تمایز /se/-/sefr/ در بازشناسی گفتار بدست آید.

¹² . casual

¹³ . Kirchner, R. M.

¹⁴ . linguopalatal

¹⁵ . Lavoie, L., M.

۲-۲-۱. مشخصات واجی

همخوان /r/

ثمره (۶۷، ۱۳۸۵) همخوان /r/ را یک همخوان لرزشی^{۱۶} یا غلتان^{۱۷} در زبان فارسی می‌نامد. وی (۶۸، همان) واجگونه‌های این همخوان را به سه صورت زنشی^{۱۸} [r]، سایشی^{۱۹} [I] و واک گونه نشان می‌دهد:

"تولید این همخوان ممکن است سایشی باشد. بدین ترتیب که نوک زبان به جای تماس با لثه بالا در فاصله اندکی از آن قرار گرفته، مجرای تنگی ایجاد می‌کند. هنگام عبور هوا از این مجرای تنگ، سایش به وجود می‌آید و بدین طریق گونه دیگری از این همخوان که با علامت [I] مشخص می‌شود، به وجود می‌آید."

ثمره (۶۹، همان) محل وقوع گونه سایشی [I] را در پایان واژه، بعد از واکه‌های پیشین و نیز بعد از همخوان می‌داند و معتقد است که [I] معمولاً واکرفته می‌شود و به ویژه پس از همخوانهای بیواک. مثالهایی که وی ارائه می‌کند به صورت زیر است:

[macI], [sadI], [ʔasaI], [p^hiI], [ʔasI]

در این پژوهش همخوان /r/ در پایان کلمه و پس از همخوان سایشی بیواک /f/ در کلمه "صفر" /sefr/ تولید شده است. بر اساس دیدگاه ثمره (همان)، گونه تولید شده در این کلمه از نوع سایشی و واکرفته [I] است زیرا جایگاه تولید این همخوان در پایان کلمه، پس از واکه پیشین /e/ و نیز پس از همخوان بیواک /f/ است. لازم به ذکر است در این گزارش برای اشاره به صورت واجی، علامت همخوان لرزش /r/ به تنهایی یا در رقم /sefr/ به کار رفته است اما صورتهای آوایی با علامت گونه سایشی آن [I] یا [sefrI] در متن آوانویسی شده‌اند. آوانویسی در شکلها یا نمودارها، به صورت [r] یا [sefr] است.

¹⁶ . trill

¹⁷ . roll

¹⁸ . flap

¹⁹ . fricative

همخوان /f/

همخوان /f/ یک آوای سایشی است که از نزدیک شدن لب پایین با دندانهای بالا تولید می‌شود. هنگام تولید این آوا تارهای صوتی در وضعیت ارتعاش قرار ندارد و /f/ یک همخوان بیواک است. ثمره (۱۳۸۵، ۵۸) معتقد است میزان سایش این آوا در آغاز کلمه و به ویژه در موضع تکیه بیشتر از پایان کلمه و یا بین دو واکه بدون تکیه است و سایش آن قبل از همخوانهای سایشی کمتر از جاهای دیگر به گوش می‌رسد. وی (همان) اظهار می‌کند که زبان در تولید این آوا نقشی ندارد و از این رو در مرحله آمادگی برای آوای بعدی قرار می‌گیرد. وی (همان) معتقد است که کشش این همخوان در جایگاه عضو اول خوشه دو همخوانی به ویژه آن که عضو دوم خوشه یک همخوان انفجاری باشد، بیشتر از جاهای دیگر به گوش می‌رسد. وی (۱۳۸۵، ۵۹)، تولید این آوا را در جایگاه قبل از همخوانهای دیگر به جز انفجاری‌ها، ملایم و غیر کشیده توصیف می‌کند.

از آنجا که در کلمه [sefi] همخوان سایشی [f] قبل از سایشی [I] تولید شده است، انتظار می‌رود بر طبق گفته ثمره با سایش کمتر و ملایمتری تولید شود.

واکه /e/

ثمره (۱۳۸۵، ۸۹) تولید این آوا را به صورت زیر توصیف می‌کند:

”اندامهای سازنده این آوا بخش پیشین زبان و سخت کام هستند. فاصله جلوی زبان و سخت کام، هنگام تولید این آوا دو برابر فاصله ای است که برای تولید /i/ لازم است. فاصله دندانهای بالا و پایین حدود یک سانتیمتر است. نوک زبان در پشت دندانهای پایین قرار می‌گیرد. تارهای صوتی در وضعیت تولید واکه قرار دارند و لبها به شکل نیمه گسترده در می‌آیند. در این حال فاصله دو لب حدود یک سانتی متر است.”

واکه /e/ یک واکه کوتاه محسوب می‌شود اما ثمره (همان) معتقد است در بعضی جایگاه ها کشش آن از /i/ هم بیشتر و در حدود ۱۷۰ میلی ثانیه می‌شود. وی (همان) محل وقوع /e/ بسیار کشیده را قبل از خوشه‌های همخوانی پایانی مانند [dʒe:sm] و [ʃe:bh] می‌داند. از سویی ثمره (همان) معتقد است که تولید /e/ با کشش طبیعی در جایگاه میان و پایان واژه مانند کلمه [dʒelve] است.

از آنجا که جایگاه [e] در کلمه /sefr/ نیز قبل از خوشه همخوانی [fi] است، انتظار می‌رود بر طبق دیدگاه ثمره (همان) به صورت بسیار کشیده تولید شود. اما تولید آن در کلمه [se] انتظار می‌رود، به صورت طبیعی و در نتیجه کوتاه‌تر از تولید آن در [sefi] باشد.

۳-۲. روش‌شناسی

۳-۲-۱. داده‌ها

داده‌ها عبارتند از سیگنالهای تلفنی ارقام "صفر" /sefr/ و "سه" /se/ که توسط افراد مختلف، زن و مرد تولید و توسط موسسه پردازش هوشمند علائم ضبط شده‌اند و از یک رقم منفرد همانند "۳" تا هفت رقم متوالی همانند "۰۱۲۰۹۸۰" را شامل می‌شوند. گویشوران رقمهای منفرد را به صورت یک کلمه مجزا تولید کرده‌اند اما در توالی‌های رقمی (بیش از یک رقم) وضعیت یکسان نیست بدین صورت که برخی ارقام را با فاصله و مکث تولید کرده‌اند و برخی با سرعت تولید گفتار بالاتر، بدون مکث و متوالی تولید شده‌اند. می‌توان براساس تعاریف موجود در روشهای بازشناسی گفتار، داده‌ها را شامل هر دو نوع گفتار مجزا^{۲۰} و متصل^{۲۱} دانست. کلمات مجزا، گفتاری است که در آن کلمات مورد آزمایش در درون جمله قرار ندارد بدین نحو که گویشوران لیستی از کلمات را می‌خوانند و ضبط می‌شود و بنابراین هر کلمه به طور مجزا از دیگری تولید می‌شود. کلمات مجزا متفاوت از گفتار پیوسته^{۲۲} است که در آن کلمات مورد آزمایش در درون جملات قرار دارند. تولید کلمات در گفتار مجزا با دقت بیشتری نسبت به گفتار پیوسته صورت می‌گیرد و بنابراین خصوصیات آوایی واجها بیشتر حفظ می‌شود و کمتر تحت تأثیر فرایندهای آوایی قرار می‌گیرند. ویبل و لی^{۲۳} (۱۹۹۰، ۱۴۵) در Readings in speech recognition، گفتار متصل را نوعی گفتار خواندنی تعریف می‌کنند که در آن گویشور لیستی از کلمات مجزا را با دقت مناسب اما بدون مکث و توقف تولید می‌کند.

ارقام "صفر" /sefr/ و "سه" /se/ مورد آزمایش از هر سه جایگاه آغاز، میان و پایان رشته ارقام انتخاب شده‌اند. به عنوان مثال در نمونه "۰۳۰"، رقم "سه" در میان رشته ارقام و رقم "صفر" هم در آغاز و هم در میان این زنجیره تولید شده است. افراد معمولاً در آغاز گفتار نسبت به میان یا پایان آن، کلمات را با تأکید بیشتری تولید می‌کنند. از سویی ارقام تولیدشده در میان یا پایان گفتار ممکن است با دقت کمتر یا سرعت گفتار بیشتر تولید شوند و بنابراین بیشتر دستخوش فرایندهای آوایی قرار گیرند. نکته مهم

²⁰. isolated word form

²¹. connected speech

²². continuous speech

²³. Weibel, Alex and Kai-Fu Lee

این است که فرایند کاهش آوایی حتی در نمونه‌هایی که گویشور فقط رقم [sefr] را به صورت یک کلمه منفرد و مجزا تولید کرده است، نیز دیده می‌شود. در نمونه‌هایی متوالی ارقام، هنگامی که سرعت گفتار بیشتر است، طبیعتاً کاهش بیشتر روی داده است. در این نمونه‌ها، گاهی آن دقت لازم در تولید که در تعریف گفتار متصل بدان اشاره شد، نیز دیده نمی‌شود.

از آنجا که در این آزمایش نمونه‌های کاهش‌یافته‌ای مورد بررسی قرار گرفته‌اند که در آنها همخوان /r/ به طور کامل حذف شده است، این نمونه‌ها به صورت [seF] آوانویسی شده‌اند. علامت [F] بزرگ، برای آوانویسی همخوان /f/ در نمونه‌های کاهش‌یافته در این پژوهش به کار رفته است. تعداد نمونه‌های مورد بررسی (۱۸۱ نمونه) برحسب جنسیت در جدول ۱-۲ ارائه شده است. این نمونه‌ها از گفتار تلفنی ضبط‌شده ۱۳ زن و ۱۴ مرد انتخاب شده‌اند.

جدول ۱-۲: نمونه‌های مورد مطالعه و تعداد آنها

		Token						Total
		[ɹ]	[f]	[F]	[sefr]	[se]	[seF]	
		Count	Count	Count	Count	Count	Count	
sex	male	16	16	16	16	15	16	95
	female	16	16	12	16	14	12	86
Total		32	32	28	32	29	28	181

نمونه‌های /sefr/ مورد بررسی به دو دسته تقسیم می‌شوند نمونه‌هایی که در آنها /sefr/ به طور کامل تولید شده است و نمونه‌هایی که در آنها /sefr/ تحت تأثیر فرایندهای آوایی به طور ناقص و کاهش‌یافته تولید شده است. تشخیص دسته دوم براساس سیگنال آنها و نیز گوش دادن به نمونه مورد نظر صورت گرفته است. سیگنال نمونه‌های کاهش‌یافته تفاوت فاحشی دارد با نمونه‌هایی که به طور کامل تولید شده‌اند. در نمونه‌های کاهش‌یافته، منحنی انرژی دامنه^{۲۴} نوفه سایش همخوان [ɹ] دیده نمی‌شود و نوفه سایش /f/ با انرژی و دیرش کم مشاهده می‌شود، گاهی نیز تنها رهش بسیار کوتاه و جزئی سایش /f/ دیده می‌شود که نشان می‌دهد در دستگاه گفتار بست سایش تنها برای چند لحظه صورت گرفته است و یا تنها حرکت لب پایین به سمت دندانها صورت گرفته اما مرحله انجام بست صورت نگرفته است. در بسیاری از نمونه‌های کاهش‌یافته تولید بست /f/ شنیده می‌شود و منحنی انرژی دامنه تفاوت چندانی بین [se]-[sefr] نشان نمی‌دهد. میترر و ارنستس^{۲۵} (۲۰۰۶) نیز پارامتر

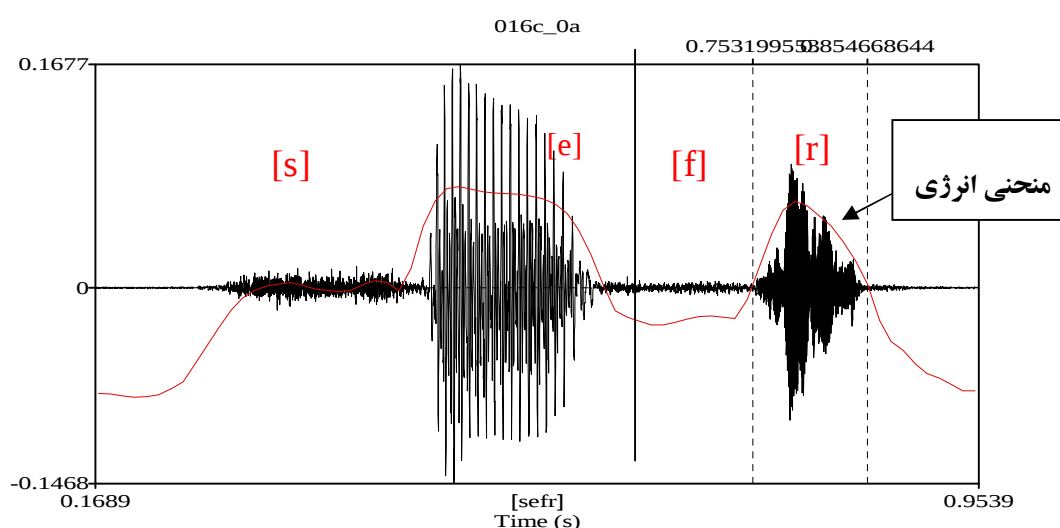
²⁴ . intensity contour

²⁵ . Mitterer, H., and M. Ernestus.

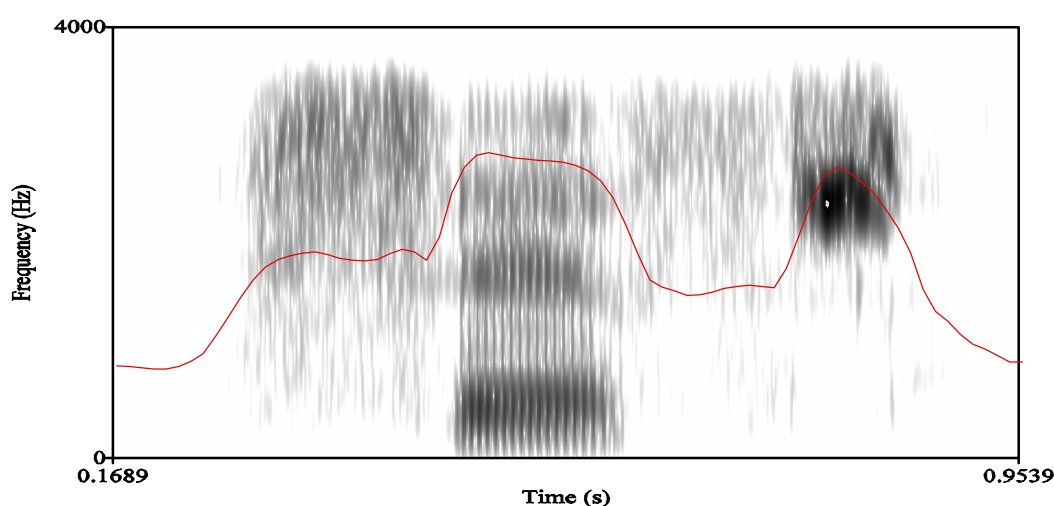
بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۱۴

دامنه انرژی را به عنوان معیاری برای حذف یا عدم حذف /t/ در خوشه‌های پایانی به کار بردند. در این پژوهش نیز، تولید [ɪ] بر اساس حضور منحنی شدت انرژی سنجیده می‌شود که در طول سایش [ɪ] ابتدا افزایش و سپس کاهش می‌یابد.

شکلهای ۱-۲ و ۲-۲ سیگنال و طیف‌نگاشت رقم [sefɪ] را که توسط یک گویشور مرد تولید شده است، نشان می‌دهند. سیگنال [ɪ]، نوفه سایشی بیواکی است که دامنه انرژی آن نیز کم است. این سیگنال متفاوت از سیگنال همخوان لرزشی /r/ است که در آن ارتعاش تارهای صوتی ملاحظه می‌شود.



شکل ۱-۲: سیگنال و منحنی انرژی [sefɪ]

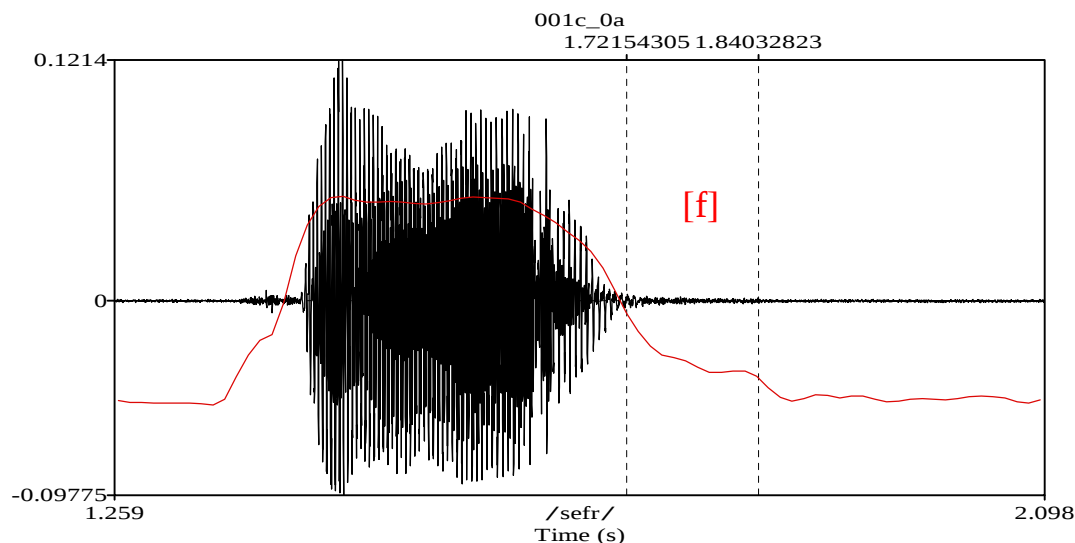


شکل ۲-۲: طیف‌نگاشت و منحنی انرژی [sefɪ]

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۱۵

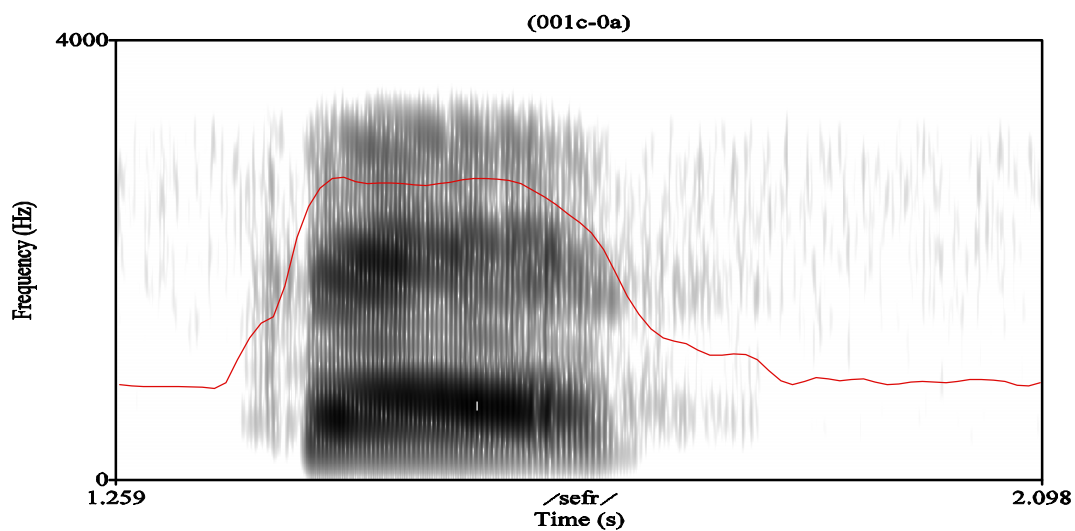
همانطور که مشخص است همخوان سایشی [f] با دامنه انرژی بسیار کمی نسبت به [I] تولید شده است و تغییرات انرژی نوفه سایش همخوان [I] همانند یک همخوان سایشی با شیبی معین افزایش و سپس کاهش یافته است. در نمونه‌هایی که تولید همخوان سایشی [I] صورت گرفته، بازشناسی رقم [sefI] با استفاده از دیرش نوفه سایش [f] و [I]، تغییرات شدت و دامنه انرژی همخوان لثوی [I] و نیز لبی [f] و نیز گذرهای واکه [e] امکان‌پذیر است.

سیگنال و طیف‌نگاشت ۳-۲ و ۴-۲، نمونه‌ای از تولید کاهش‌یافته واژه [sefI] را که توسط یک گویشور زن تولید شده، نشان می‌دهد. مقایسه سیگنال و طیف‌نگاشت ۳-۲ و ۴-۲ با شکل‌های فوق نشان می‌دهد که منحنی شدت انرژی، تولید [I]، را نشان نمی‌دهد. علاوه بر آن از قدرت و شدت نوفه سایش [f] هم کاسته شده است.



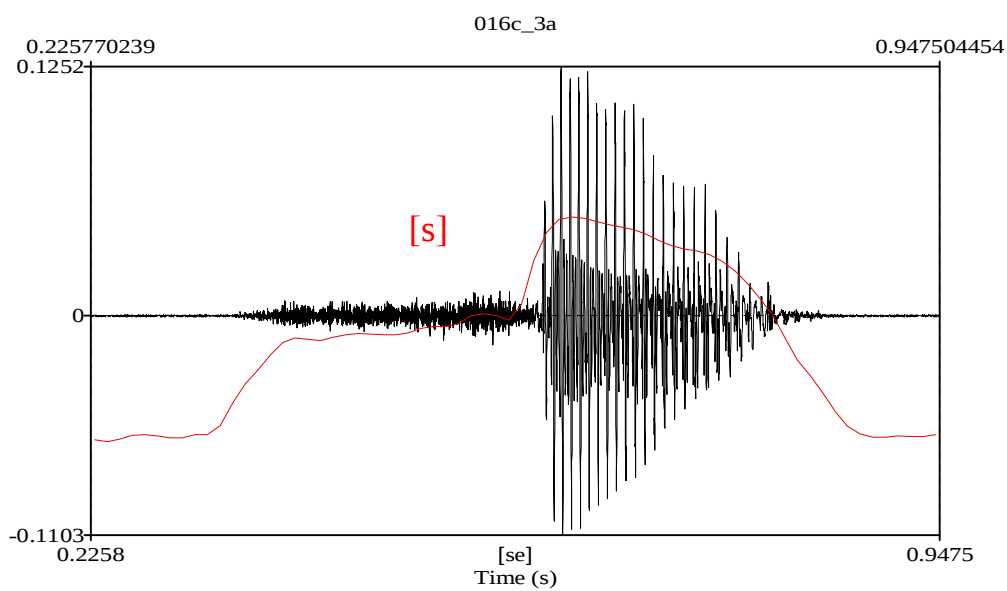
شکل ۳-۲: سیگنال و منحنی انرژی در نمونه‌های کاهش‌یافته [sefI]

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۱۶

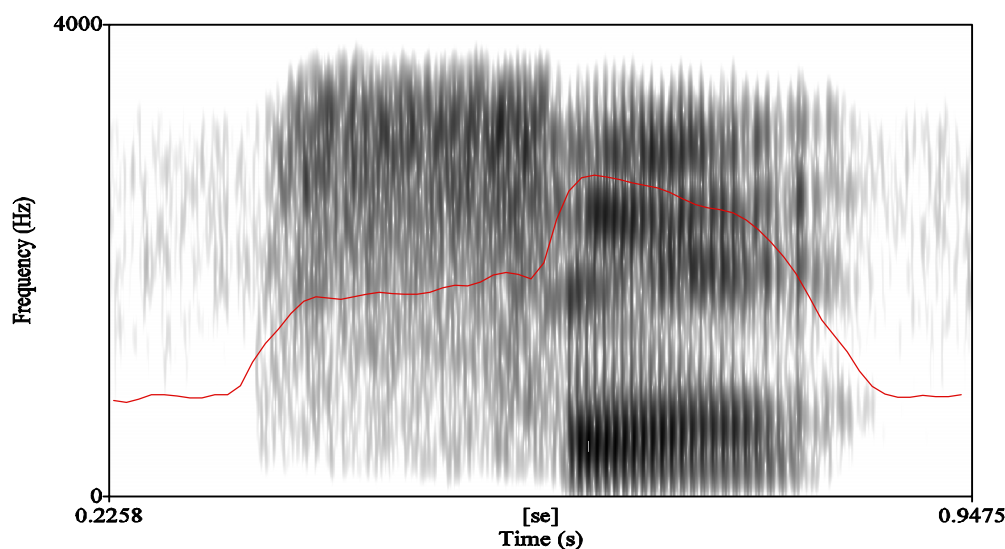


شکل ۲-۴: طیف نگاشت و منحنی انرژی در نمونه‌های کاهش یافته [seF]

سیگنال و طیف نگاشت رقم [se] نیز در شکل‌های زیر (۲-۵ و ۲-۶) نشان داده شده است. این نمونه توسط یک گویشور مرد تولید شده است.



شکل ۲-۵: سیگنال و منحنی انرژی [se]



شکل ۲-۶: طیف‌نگاشت و منحنی انرژی [se]

کاهش واژه [sefi] سبب می‌شود تا مشخصات اکوستیکی آن به واژه [se] نزدیک شود. مقایسه منحنی شدت انرژی، تفاوت چندانی بین [sef] کاهش یافته و [se] نشان نمی‌دهد.

لازم به ذکر است که در رشته‌هایی از ارقام که رقم [sefi] به طور متوالی تولید شده، کاهش بیشتر روی داده است. زیرا تکرار از جمله فاکتورهای بسیار مهم در کاهش و حذف محسوب می‌شود. در واقع بین بسامد تکرار یک تکواژ و طول آن رابطه عکس وجود دارد، به نحوی که هرچه بسامد تولید یک واحد آوایی بیشتر باشد، تضعیف آن بیشتر است (Ussishkin And Wedel, 2009, 274).

۲-۳-۲. نرخ و سبک گفتار

نرخ گفتار عبارت است از تعداد هجاها در واحد زمان (syll/sec). در اندازه‌گیری نرخ گفتار، مکث، درنگ و تپه‌های گفتاری نیز محاسبه می‌شود. گویشوران زنجیره ارقام را با نرخ گفتار متفاوت تولید کرده‌اند. گوش دادن به نمونه‌ها و مشاهده سیگنال‌ها نشان می‌دهد برخی از گویشوران ارقام را با فاصله و مکث کوتاهی تولید کرده‌اند، بدین معنا که هر رقم را به صورت یک کلمه مجزا در نظر گرفته و بنابراین پس از تولید آن، کلمه بعدی را با توقف و مکث کوتاهی تولید نموده‌اند اما برخی دیگر رشته ارقام را با نرخ گفتار بالاتر تولید کرده‌اند به نحوی که پس از پایان زنجیره متوالی ارقام، مکث و تنفس نموده‌اند. این تفاوت در تولید سبب ایجاد تفاوت‌های قابل توجهی در سیگنال اکوستیکی حاصل می‌شود. بنابراین نمونه‌های مورد آزمایش نیز از هر سه نرخ گفتار، معمولی، کند و تند انتخاب

شده‌اند. در سرعت گفتار معمولی و نیز کند معمولاً خصوصیات آوایی و اکوستیکی کاملتر و بیشتر در سیگنال گفتار انتقال می‌یابد زیرا افراد معمولاً ارقام را با دقت و تأکید بیشتری تولید می‌کنند و بنابراین بازشناسی آواها با احتمال بیشتری صورت می‌گیرد اما با افزایش سرعت گفتار، احتمال وقوع فرایندهای آوایی نظیر هم‌تولیدی، همگونی و کاهش نیز افزایش می‌یابد زیرا دیرش زمانی هر آوا کاهش یافته و بسیاری از سرنخهای اکوستیکی حذف می‌شوند. در نتیجه بازشناسی در گفتار تند با اختلال بیشتری مواجه می‌شود.

سبک گفتاری افراد نیز، فارسی غیررسمی محاوره‌ای است. ثمره (۱۹۷۷) در مقدمه کتاب خود تحت عنوان "A course in colloquial Persian" فارسی محاوره‌ای یا سبک غیررسمی فارسی معیار را لهجه معیار جدیدی برمی‌شمارد که افراد تحصیلکرده بدان صحبت می‌کنند و دارای اعتبار اجتماعی است به نحوی که افراد زیادی در کشور تمایل دارند به آن صحبت کنند و آن را به گویشهای محلی خود ترجیح می‌دهند. این سبک علاوه بر تفاوتهایی که به لحاظ واژگان و اصطلاحات با سبک رسمی دارد، حاوی فرایندهای آوایی، نظیر حذف، همگونی، کاهش و سایر تغییرات آوایی است.

کرچنر (۱۹۹۸: ۲۱۴) معتقد است بین نرخ گفتار و سبک آن رابطه وجود دارد. بدین معنا که اگر یک همخوان در یک سبک یا نرخ گفتار کاهش یابد، در گفتار سریعتر یا سبکهای محاوره‌ای تر گفتار نیز کاهش پیدا می‌کند. وی (همان: ۲۱۸-۲۱۷) تضعیف همخوانها را در گفتار سریع به دو دلیل می‌داند: اولاً اینکه در گفتار سریع، اندامهای گفتاری سریعتر به گرفتگی مورد نظر می‌رسند که این سبب افزایش سرعت ایجاد بستار گرفتگی می‌شود و یا اینکه خود گرفتگی کوتاه می‌شود که سبب کاهش زمانی گرفتگی می‌شود و بنابراین انرژی و کوشش^{۲۶} بیشتری مورد نیاز است. وی بدین ترتیب تضعیف در سبکهای محاوره‌ای و سریع گفتار را بر اساس اصل کوشش توجیه می‌کند.

۲-۳-۳. پارامترهای اکوستیکی

نوفه سایش و گذرهای سازه‌ای واکه حاوی سرنخهای ادراکی سایشی‌ها هستند (Hedrick and Ohde, 1993). در همخوانهای سایشی لبی نوفه سایش با انرژی کم بر روی همه فرکانسها پراکنده است. اندازه‌گیری و محاسبه پارامترها در نرم‌افزار Praat صورت گرفت.

²⁶ . effort

۲-۳-۱. پارامترهای نوفه سایش

پارامترهای نوفه سایش عبارتند از: دیرش سایش^{۲۷}، مرکز ثقل طیف^{۲۸}، بسامد قله طیف^{۲۹} و شدت انرژی قله طیف^{۳۰} سایش.

دیرش سایش: دیرش سایش عبارت است از مدت زمانی که نوفه سایش هر یک از همخوانهای [f] و [I] طول می کشد و واحد آن میلی ثانیه (ms) است. اندازه گیری دیرش سایش [I] بر اساس سیگنال و طیف نگاشت صورت گرفت. در نمونه هایی که [sefi] کاهش نیافته است، افزایش میزان نوفه سایش بر روی سیگنال و نیز مشاهده نواحی نوفه ای با شدت انرژی بیشتر، از بسامدهای ۲۰۰۰ هرتز به بعد بر روی طیف نگاشت، ناحیه سایش [I] را نشان می دهد. این ناحیه در مقایسه با ناحیه سایش [f] که انرژی نوفه ای کمتر و پراکنده بر روی همه بسامدها دارد، شدت انرژی بیشتری داشته و بنابراین منحنی شدت انرژی، افزایش انرژی طیفی در این همخوان را نشان می دهد. اندازه گیری دیرش سایش [f] نیز بر اساس سیگنال و طیف نگاشت صورت گرفت. همانطور که گفته شد، شدت نوفه این همخوان در مقایسه با [I] بسیار کم است و منحنی انرژی تنها افت انرژی را نشان می دهد و هیچ قله طیفی در نوفه دیده نمی شود. گاهی به دلیل افزایش نوفه حین ضبط داده ها، تشخیص محل پایان دیرش سایش [f] و آغاز [I] دیده نمی شود. در این موارد، زمان شروع افزایش منحنی انرژی به عنوان نقطه شروع سایش [I] در نظر گرفته شده است. سیگنال و طیف نگاشت ۲-۷ و ۲-۸ روش اندازه گیری دیرش سایش را نشان می دهد.

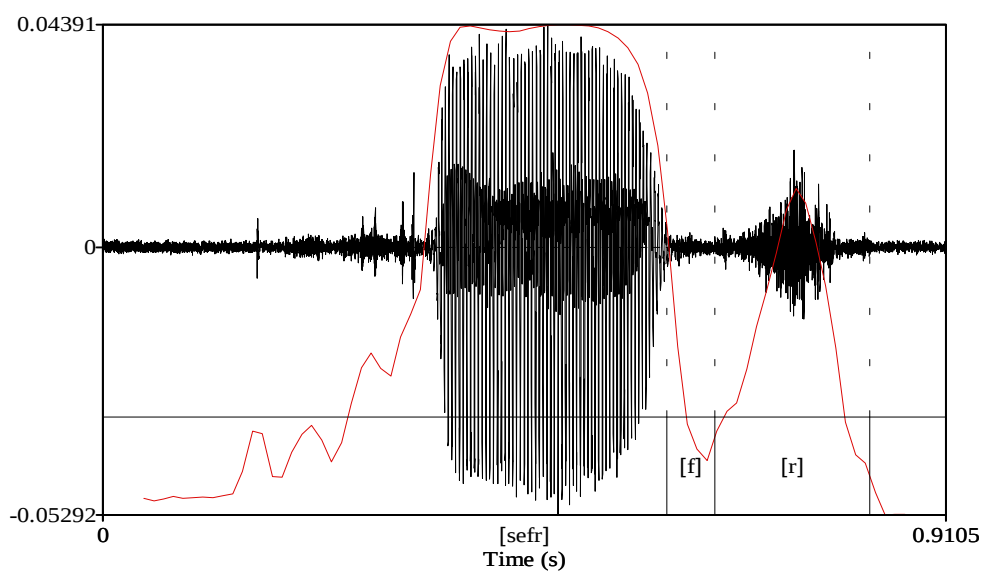
^{۲۷}. frication duration

^{۲۸}.center of gravity (COG)

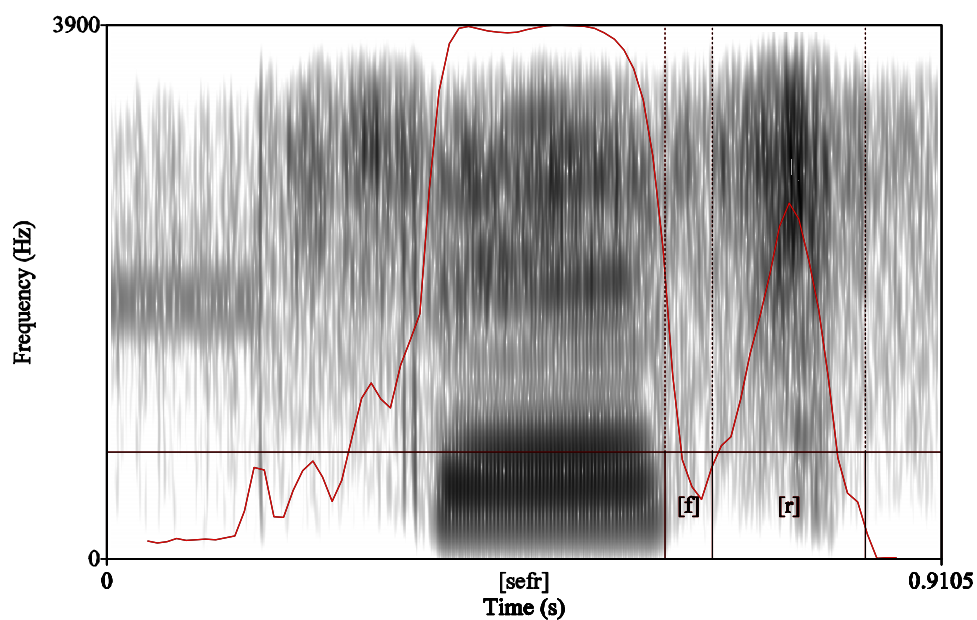
^{۲۹}. spectral peak frequency

^{۳۰}. spectral peak intensity

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۲۰



شکل ۲-۷: ناحیه سایش [f] و [r] بر روی سیگنال



شکل ۲-۸: ناحیه سایش [f] و [r] بر روی طیف نگاشت

مرکز ثقل طیف: گشتار^{۳۱} مرکز ثقل طیف، عبارت است از میانگین بسامدی طیف. در واقع به بسامدی اطلاق می شود که طیف انرژی را به دو بخش تقسیم می کند به نحوی که در این بسامد،

^{۳۱}. moment

میزان انرژی فرکانسهای بالا با میزان انرژی فرکانسهای پایین برابر می‌شود و بنابراین به آن مرکز ثقل طیف می‌گویند. براساس آنچه که گفته شد، مرکز ثقل طیف پارامتری است که از طیف سایش استخراج می‌شود و واحد آن هرتز (Hz) است.

پارامتر COG با ساینز حفره جلویی دهان ارتباط دارد، بدین معنا که مقدار آن در همخوانهای پیشین یا قدیمی بیشتر از همخوانهای پسین است (Gordon et al. 2002). آواهایی که انرژی فرکانسهای بالای آنها زیاد است، بسامد COG آنها نیز بر روی فرکانسهای بالا قرار دارد. ون سان و پولز^{۳۲} (۱۹۹۷) کاهش اکوستیکی COG به همراه دیرش همخوان را از جمله نشانه‌های جهانی کاهش آوایی برمی‌شمارند. بدین معنا که کاهش مقدار بسامدی COG نشان می‌دهد که انرژی فرکانسهای بالا کاهش یافته است و بنابراین اطلاعات تولیدی کمتری در سیگنال وجود دارد. از سویی نتایج مطالعات ناکامورا^{۳۳} و دیگران (۲۰۰۷) نیز نشان می‌دهد که کاهش طیفی یکی از دلایل عمده کاهش دقت بازشناسی گفتار پیوسته است.

محاسبه این پارامتر یکی از برنامه‌های نرم‌افزار Praat است و بنابراین می‌توان مقدار این پارامتر را پس از محاسبه طیف سایش، در این نرم‌افزار بدست آورد. مقدار COG براساس کل دیرش سایش محاسبه شد. این پارامتر روشن می‌کند که طیف نوفه بیشتر تحت تسلط کدام نواحی بسامدی است. در کل نمونه‌ها، به منظور حذف سیگنال واک و نیز حذف انرژی نوفه‌ای جریان برق، بسامدهای زیر ۱۰۰۰ هرتز فیلتر شدند. در این نمونه‌ها، گاهی مقداری ارتعاش واک همراه با نوفه سایش [f] مشاهده می‌شود. گاهی نیز هنگام تولید [I] در برخی نمونه‌ها، ارتعاش بسیار ضعیف تارهای صوتی دیده می‌شود. لازم به ذکر است پیش تنظیمات Praat برای محاسبه مقادیر COG بدون تغییر مورد استفاده قرار گرفت.

بسامد قله طیف: بسامد قله طیف که دارای بیشترین شدت انرژی طیف است. واحد بسامد هرتز (Hz) است. همخوان سایشی [f] یک همخوان لبی-دندانی است که از نزدیک شدن لب پایین با دندانه‌های بالا تولید می‌شود. از آنجا که در مقابل گرفتگی لبی-دندانی تقریباً هیچ فیلتری در دستگاه گفتار نیست، بنابراین طیف این همخوان هیچ قله طیفی ندارد. طیف سایش همخوان [f] یکدست است یعنی انرژی بر روی همه فرکانسها دیده می‌شود. بنابراین تنها طیف [I] مورد بررسی قرار گرفت.

³² . Van Son R.J.J.H., and C. W. Pols

³³ . Nakamura, M.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۲۲

صادقی (۱۳۸۴، ۲۵۷-۲۶۳) طیف‌نگاشت رشته آوایی [fe] در کلمه "فعل" را بررسی می‌کند. وی (همان) تمرکز انرژی سایش [f] را بر روی نواحی بسامدی ۱۸۰۰ تا ۳۷۰۰ هرتز در این رشته، بدست می‌آورد.

شدت انرژی قله طیف: این پارامتر بیانگر شدت انرژی قله طیف است و واحد آن dB است. از آنجا که در منحنی انرژی سایش [f] همواره سیر نزولی دیده می‌شود و به عبارتی هیچ قله انرژی مشاهده نمی‌شود، بنابراین فقط شدت انرژی قله طیفی [I] اندازه‌گیری شد.

۲-۳-۲. پارامترهای واکه

پارامترهای واکه عبارتند از دیرش واکه، شیب گذر سازه‌های F1، F2 و F3 و بسامد مکانی هر یک از سازه‌ها

دیرش واکه: عبارت است از مدت زمان تولید واکه [e] و واحد آن میلی‌ثانیه (ms) است. دیرش واکه از اولین پالس حنجره تا آخرین پالس محاسبه می‌شود.

گذرهای سازه‌ای: سازه‌ها وضعیت دستگاه گفتار حین تولید یک آوا را نشان می‌دهند و به شکل دستگاه گفتار بستگی دارند. بنابراین سیگنال حنجره تأثیری در مقادیر سازه‌ای یک آوا ندارد. سازه‌ها در واقع فیلترهای بسامدی هستند که حاصل وضعیت دستگاه گفتار هستند و بنابراین بر حسب آواها متفاوتند. این فیلترها که در حفره دهان، حلق و بینی شکل می‌گیرند برخی بسامدها را تضعیف و برخی را تقویت نموده سبب بوجود آمدن باندهای بسامدی انرژی بر روی طیف‌نگاشت می‌شوند. در مطالعات آواشناسی، معمولاً بسامدهای اول تا سوم F1، F2 و F3 مورد مطالعه قرار می‌گیرند (Reetz and Jongman, 2009, 177)

سازه‌های اول، دوم و سوم سرنخهای اکوستیکی محسوب می‌شوند که کیفیت واکه را تعیین می‌کنند. سازه‌های بالاتر چهار و پنج به نظر می‌رسد بیشتر اطلاعاتی در مورد هویت گویشور تا واکه فراهم می‌کنند (همان، ۱۸۴)

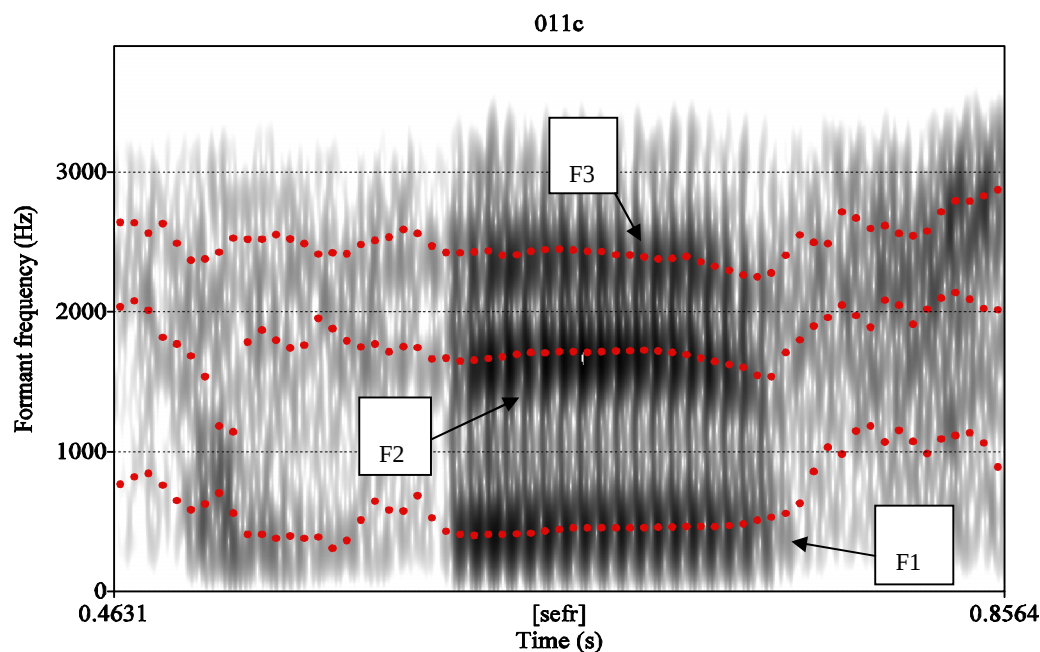
گذرهای سازه‌ای از واکه به همخوان یا از همخوان به واکه بر حسب مشخصه‌های سازه‌ای واکه و نیز محل و نحوه تولید همخوان تغییر می‌کنند. به نظر می‌رسد گوش انسان، اطلاعات گذرهای سازه‌ای را برای تعیین محل و نحوه تولید همخوان به کار می‌برد حتی وقتی که سایر سرنخهای اسپکتروگرافی همخوان از بین رفته و یا مبهم باشند.

نتایج آزمایشهایی که بر روی درک گفتار صورت گرفته است، نشان می‌دهد که جهت یک گذر سازه‌ای می‌تواند یک سرنخ اکوستیکی درک محل یا نحوه تولید باشد. جهت سازه دوم و سوم به مشخصه محل تولید، و جهت اولین سازه به نحوه تولید (انسدادی، سایشی، خیشومی و غیره) بستگی دارد (Delattre, Liberman and Cooper, 1962). تفاوت $F_3 - F_2$ با محل تولید واکه تطابق دارد.

گذرهای همخوانهای لبی و کامی بر حسب واکه بعدی تغییر می‌کند. به طور کلی همخوانهای لبی معمولاً F_2 و F_3 پایتتری نسبت به واکه مجاور خود دارند. زیرا گرفتگی لبی معمولاً سبب افت همه گذرهای واکه‌ای می‌شود (Flemming, 2002, 22). در همخوانهای لثوی، F_2 ، بسامد بالاتری دارد.

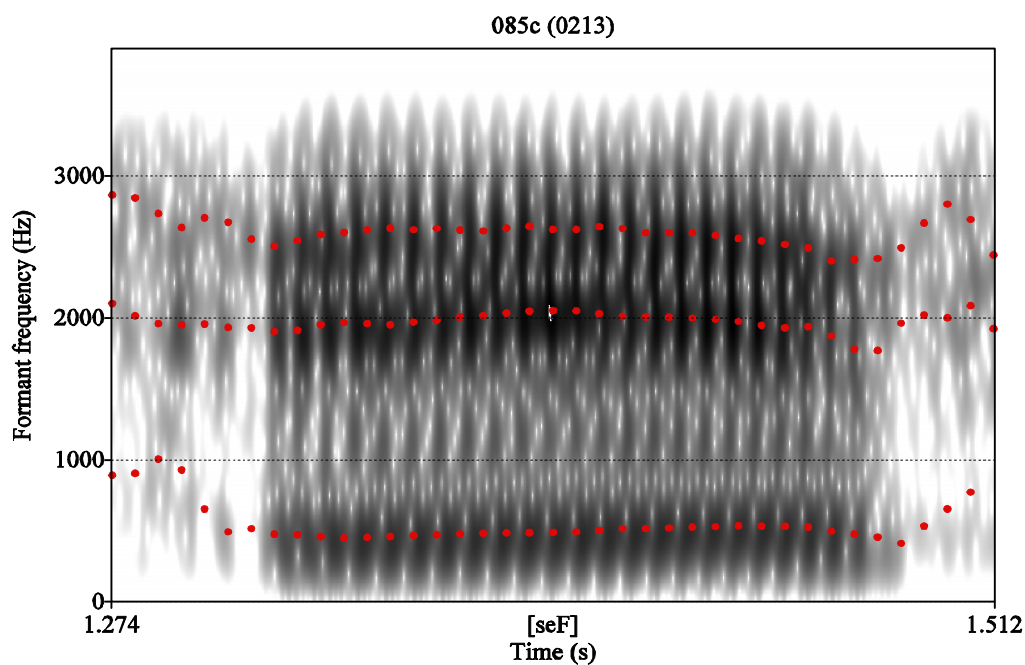
گذرهای سازه اول F_1 تمایزات محل تولید در همخوانها را نشان نمی‌دهند. زیرا گرفتگی یک همخوان سبب افت F_1 فارغ از مشخصه محل تولید می‌شود. سایر ویژگیهای گذر F_1 مانند نرخ تغییر آن ممکن است با مشخصه محل تولید ارتباط داشته باشد (Flemming, 2002, 22).

در شکل‌های ۲-۹ تا ۲-۱۱ هر یک از سه سازه مشخص شده‌اند. اندازه‌گیری سازه‌ها در میان و پایان واکه پس از استخراج الگوی سازه‌ای در نرم‌افزار Praat صورت گرفت. الگوی گذر سازه‌ها در هر یک از شکل‌های زیر متفاوت است. در شکل اول (۲-۹)، بسامد گذر سازه اول در انتهای واکه افزایش پیدا کرده است.

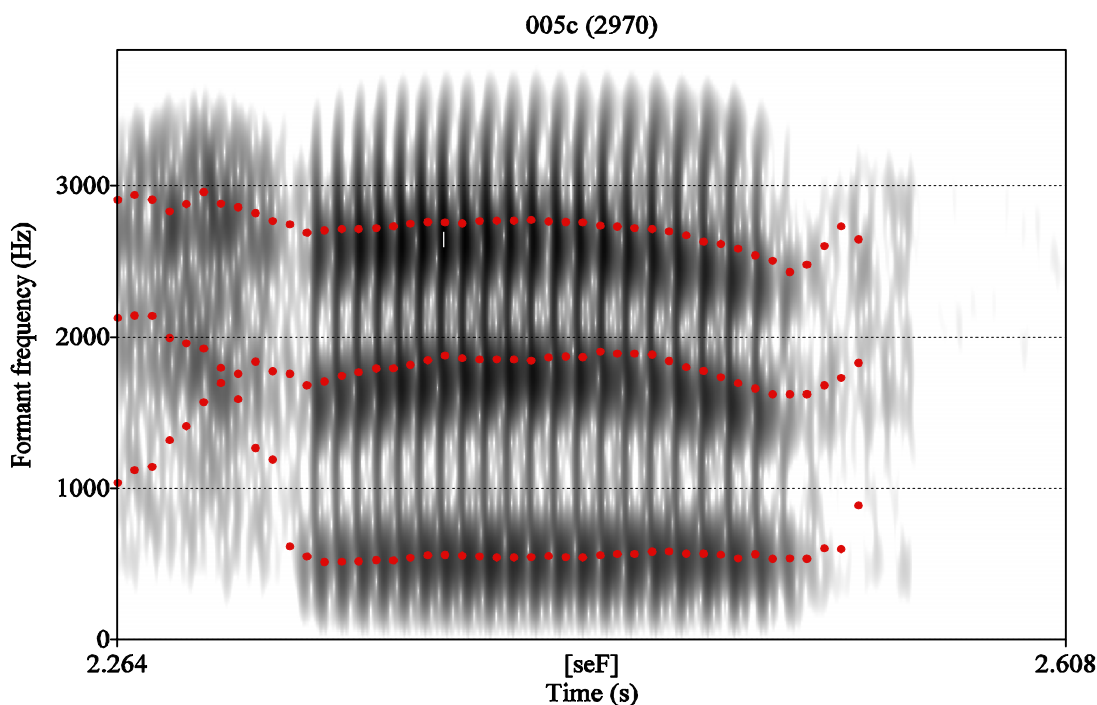


شکل ۲-۹: بسامدهای سازه‌ای، افزایش F1 و کاهش دو سازه F2 و F3 در انتهای واکه

در شکل (۲-۱۰) هر سه سازه در انتهای واکه افت پیدا کرده‌اند و در شکل سوم (۲-۱۱) الگوی گذر سازه اول یکدست است و در واقع بسامد سازه اول از میان به انتهای واکه تغییر نکرده است. این دو شکل (۲-۱۰ و ۲-۱۱) از نمونه‌هایی استخراج شده‌اند که در آنها، رقم /sefr/ کاهش پیدا کرده است.



شکل ۲-۱۰: کاهش هر سه سازه F1, F2 و F3 در انتهای واکه



شکل ۱-۲: الگوی گذر F1 از میان به انتهای واکه تغییر بسامدی نیافته، اما F2 و F3 در انتهای واکه کاهش یافته‌اند.

مطالعات کی و کلت^{۳۴} (۱۹۸۸) نشان داد که همخوانهای [f] و [θ] به لحاظ ادراکی بر اساس گذرهای F2 و F3 متمایز می‌شوند.

صادقی (۱۳۸۴، ۲۸۶ و ۵۱۶) وضعیت گذر سازه‌های واکه [e] را در رشته آوایی [ef] در کلمه "عارف" نشان می‌دهد. وی (همان، ۵۱۶) منحنی تغییرات بسامدهای بازخوانی اول تا چهارم را به عنوان تابعی از طول حفره حلق (۶ تا ۱۱ سانتی متر) به هنگام تولید یک واکه پیشکامی ترسیم نموده‌است. با استفاده از این شکل می‌توان میزان سازه اول از ۵۰۰ تا ۸۰۰ هرتز (طول کمتر به بیشتر)، سازه دوم بین ۱۵۰۰ تا ۱۸۰۰ و سازه سوم ۲۵۰۰ تا ۲۹۰۰ هرتز تخمین زد.

انصارین (۲۰۰۴) مقادیر سازه‌ای واکه‌های فارسی را با انجام آزمایش اکوستیکی بر روی ۱۲ گویشور فارسی زبان و ارائه فضای واکه‌ای تهیه نموده است. وی (همان) میانگین مقادیر سازه اول و دوم واکه /e/ را به صورت $F1 = 644_{\text{Hz}}$ و $F2 = 2115_{\text{Hz}}$ در کل داده‌ها بدست آورده است.

³⁴ . Key K. and D. H. Klatt

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۲۶

بسامد مکانی: بسامد انتهای گذر وا که به همخوان یا انتهای گذر همخوان به وا که است که در بافتهای وا که ای متفاوت، مقدار آن تا حدودی ثابت است زیرا اطلاعات محل و نحوه تولید همخوان را انتقال می دهد. بنابراین هر سازه برای هر همخوان یک فرکانس مکانی دارد که شنونده را در تشخیص همخوان کمک می کند.

۴-۲. تجزیه و تحلیل آماری داده ها

تجزیه و تحلیل آماری داده ها با استفاده از نرم افزار SPSS 18 صورت گرفت. روشهای آماری به کار رفته عبارتند از آزمونهای تحلیل واریانس یک طرفه^{۳۵} و دوطرفه (مانوا^{۳۶}) و نیز آزمونهای تعقیبی LSD (post Hoc) و Tamhane. آزمون تعقیبی LSD زمانی به کار می رود که تست همگنی واریانسها (homogeneity test) معنادار نباشد و Tamhane هنگامی که این تست معنادار باشد.

۴-۲-۱. دیرش کلمه

مقادیر دیرش زمانی هر یک از نمونه های آزمایشی در جدول زیر ارائه شده است.

جدول ۲-۲: آمار توصیفی پارامتر دیرش زمانی (ms) آواها و کلمات

Token	Mean	N	Std. Deviation	Minimum	Maximum	Range
[ɹ]	110.09	32	41.270	42	242	200
[f]	79.91	32	37.497	26	169	143
[F]	86.39	28	37.247	28	155	127
[sefɹ]	496.44	32	99.023	293	711	418
[se]	348.34	29	102.021	200	623	423
[seF]	366.18	28	90.298	235	639	404
Total	247.18	181	180.300	26	711	685

³⁵. Univariate analysis of variance

³⁶. Multivariate analysis of variance

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۲۷

مقادیر دیرش زمانی هر یک از نمونه‌های آزمایشی [sef.I]، [seF] و [se] در جدول ۲-۲ ارائه شده است. تحلیل واریانس یک‌طرفه نشان می‌دهد که تفاوت معناداری بین مقادیر دیرش زمانی این نمونه‌ها وجود دارد.

جدول ۲-۳: نتایج آزمون واریانس یک‌طرفه، مقایسه پارامتر میانگین دیرش کلمات

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	402397.826 ^a	2	201198.913	21.216	.000
Intercept	1.446E7	1	1.446E7	1524.282	.000
Token	402397.826	2	201198.913	21.216	.000
Error	815556.534	86	9483.216		
Total	1.598E7	89			
Corrected Total	1217954.360	88			

تست Post Hoc (LSD) در جدول زیر نشان می‌دهد:

۱. دیرش زمانی [seF] تفاوت معناداری با [se] ندارد.

۲. دیرش زمانی [sef.I] به طور معناداری بیشتر از [se] و [seF] است.

جدول ۲-۴: نتایج آزمون LSD، مقایسه میانگین دیرش زمانی کلمات با یکدیگر

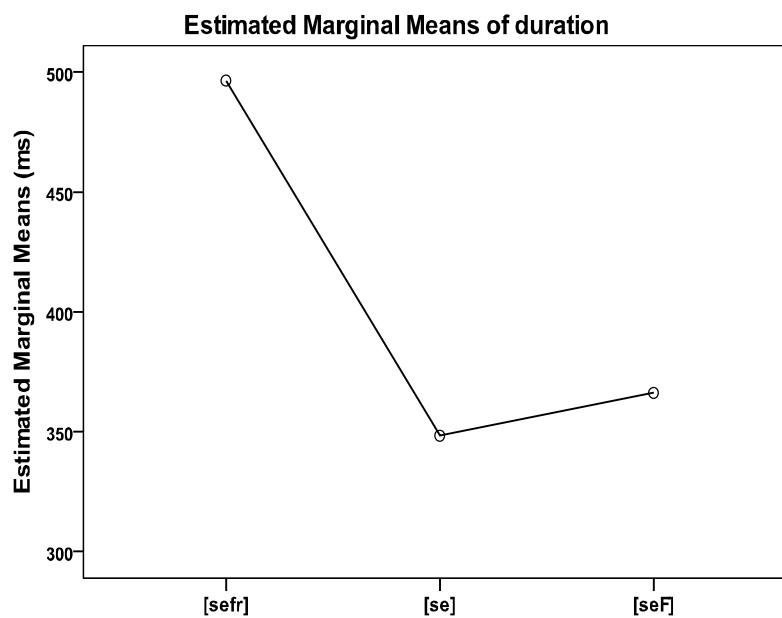
Token	Token	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
[sef.I]	[se]	148.09*	24.967	.000	98.46	197.73
	[seF]	130.26*	25.200	.000	80.16	180.35
[se]	[sef.I]	-148.09*	24.967	.000	-197.73	-98.46
	[seF]	-17.83	25.801	.491	-69.12	33.46
[seF]	[sef.I]	-130.26*	25.200	.000	-180.35	-80.16
	[se]	17.83	25.801	.491	-33.46	69.12

*. The mean difference is significant at the .05 level.

میانگین دیرش زمانی [seF] تنها ۱۸ میلی ثانیه بیشتر از [se] است. این تفاوت اندک ناشی از حذف همخوان [ɪ] است. دیرش واکه در [se] نیز بسیار بیشتر از [seF] است که در بخشهای بعدی مورد بحث قرار خواهد گرفت. کاهش دیرش زمانی واژه /sefi/ در این نمونه‌ها، به میزان دیرش زمانی واژه /se/، سبب کاهش ادراکی تفاوت شنیداری این دو واژه می‌شود. در بازشناسی گفتار نیز این افت دیرش زمانی یکی از عواملی است که بازشناسی این دو واژه از یکدیگر را، با مشکل مواجه می‌کند.

از سویی، میانگین دیرش زمانی [seF] به میزان ۱۳۰ میلی ثانیه کمتر از [sefi] است. این تفاوت دیرش زمانی ناشی از حذف همخوان [ɪ] است که میانگین دیرش زمانی آن ۱۱۰ میلی ثانیه است. پیوستار دیرش زمانی نمونه‌های مذکور به صورت زیر است:

$$[sefi] > [seF] > [se]$$



نمودار ۱-۲: تفاوت دیرش زمانی نمونه‌ها

۲-۴-۲. دیرش سایش [f]

مقادیر بدست آمده دیرش سایش [f] در نمونه‌های [sefi] و [seF] در جدول ۲-۵ ارائه شده است.

جدول ۲-۵: آمار توصیفی پارامتر دیرش سایش [f] در نمونه‌های کاهش یافته و نیافته

Token	Mean	N	Std. Deviation	Minimum	Maximum	Range
[f]	79.91	32	37.497	26	169	143
[F]	86.39	28	37.247	28	155	127
Total	82.93	60	37.206	26	169	143

دیرش سایش [f] در نمونه‌هایی که تضعیف یافته‌اند، تقریباً ۷ میلی ثانیه بیشتر از نمونه‌هایی است که در آن [sefi] کاملتر 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 این آوا را در جایگاه قبل از همخوانهای دیگر به جز انفجاری‌ها، ملایم و غیرکشیده توصیف می‌کند. نتایج این پژوهش نیز نشان می‌دهد که دیرش سایش [f] در جایگاه قبل از یک همخوان سایشی نظیر [I] کمتر از حتی در پایان کلمه، [seF] است. به طور کلی سایشهای لبی کم‌انرژی‌تر و طبیعتاً کوتاه‌تر از سایر سایشی‌ها هستند زیرا به دلیل کوتاه شدن طول دستگاه گفتار در جلوی محل گرفتگی، هیچگونه فیلتری بر روی آنها صورت نمی‌گیرد.

تولید [I] مستقل از [f] است، ثمره (۱۳۸۵، ۵۸) اظهار می‌کند که زبان در تولید [f] نقشی ندارد و از این رو در مرحله آمادگی برای آوای بعدی قرار می‌گیرد. اما می‌توان افزایش دیرش سایش را در نمونه‌های [seF] چنین توجیه کرد که در [sefi] گرفتگی لبی-دندانی با سرعت بیشتری از وضعیت سایش خود خارج شده است. زیرا دستگاه گفتار به سمت تولید همخوان [I] که متوالیاً در پی آن است، رفته اما در [seF]، دستگاه گفتار در وضعیت تولید سایش اندکی بیشتر باقی می‌ماند زیرا این آوا در پایان گفتار قرار دارد و پس از آن آوای دیگری برای تولید نیست.

مقدار کمینه سایش در هر دو وضعیت تقریباً یکسان (۲۶ و ۲۸ میلی ثانیه) است. اما مقدار بیشینه سایش [f] در نمونه‌های کاهش نیافته بیشتر است.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۳۰

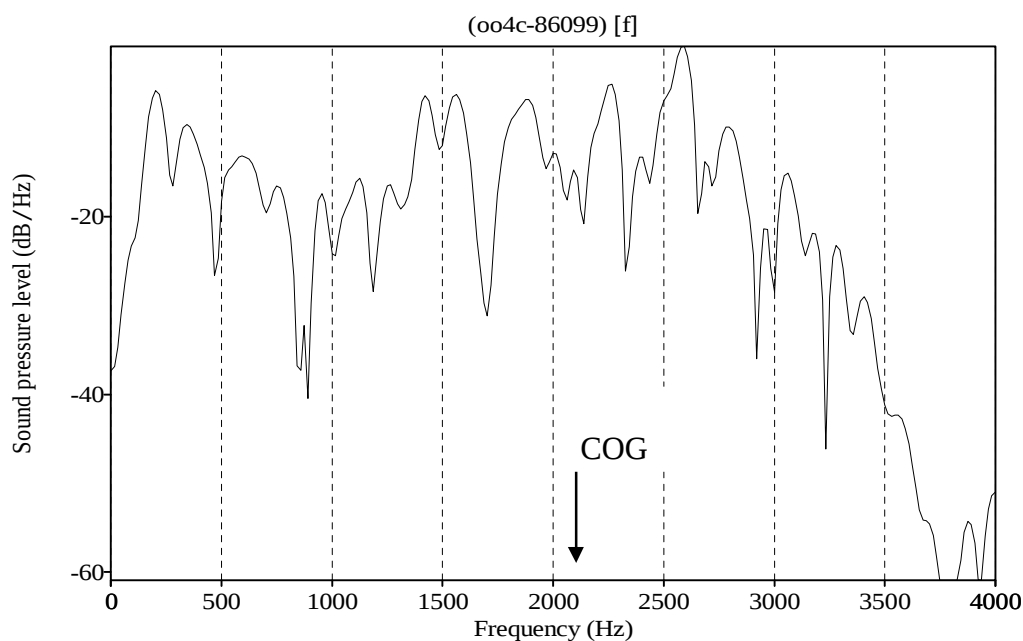
۳-۴-۲. مرکز ثقل طیف COG

میانگین مقادیر COG سایش [f] در نمونه‌های [sefi] و [seF] در جدول زیر آمده است.

جدول ۲-۶: آمار توصیفی پارامتر مرکز ثقل طیف سایش (COG) در نمونه‌های کاهش یافته و نیافته

Token	Mean	N	Std. Deviation	Minimum	Maximum	Range
[f]	2115.0006	32	269.09352	1679.00	2676.80	997.80
[F]	2040.4998	28	184.87599	1688.05	2372.08	684.03
Total	2080.2335	60	234.71848	1679.00	2676.80	997.80

نمودار ۲-۲ طیف سایش [f] و بسامد COG را بر روی آن نشان می‌دهد.



نمودار ۲-۲: طیف سایش [f] و بسامد COG

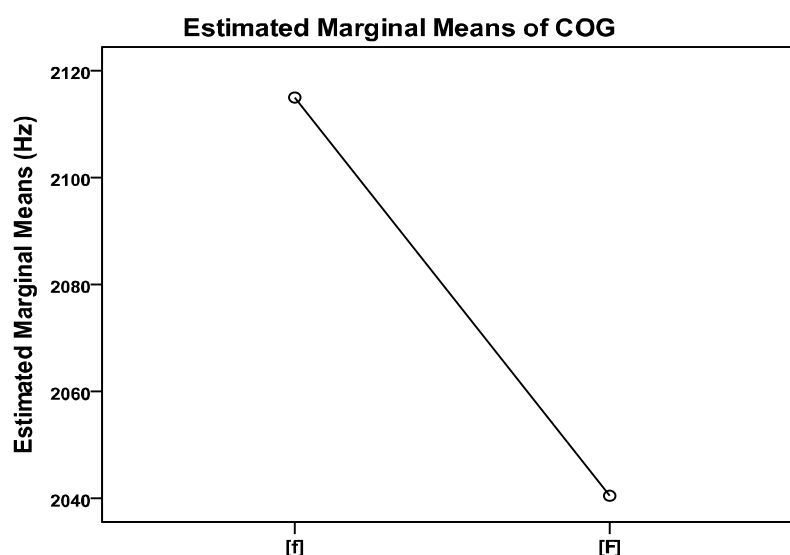
میانگین مقادیر COG نشان می‌دهد که مقدار این پارامتر در نمونه‌های تضعیف‌شده، کاهش یافته است. به عبارتی میانگین بسامدی همخوان سایشی کاهش یافته است. آزمون واریانس یک‌طرفه نشان می‌دهد که این تفاوت معنادار نیست.

جدول ۲-۷: نتایج آزمون واریانس یک طرفه، مقایسه پارامتر میانگین مرکز ثقل طیف

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	82885.531 ^a	1	82885.531	1.518	.223
Intercept	2.579E8	1	2.579E8	4721.748	.000
Token	82885.531	1	82885.531	1.518	.223
Error	3167587.557	58	54613.579		
Total	2.629E8	60			
Corrected Total	3250473.088	59			

a. R Squared = .025 (Adjusted R Squared = .009)

از آنجا که دو پارامتر دیرش سایش و COG تفاوت معناداری بین [sef.I] و [seF] نشان نمی‌دهند، چنین می‌توان نتیجه گرفت که کاهش /sefr/ تنها سبب حذف همخوان /r/ شده، اما تأثیر چندانی بر پارامترهای نوفه سایش /f/ نداشته است. بنابراین، این دو پارامتر به عنوان سرخهای اکوستیکی عمل می‌کنند که سبب درک صورتهای کاهش یافته /sefr/ می‌شوند و تمایز شنیداری و ادراکی آنها را از واژه /se/ سبب می‌شوند. در سیستمهای بازشناسی گفتار نیز می‌توان با به کارگیری پارامترهای نوفه سایش، تمایزات واجی /se/ - /seF/ را بدست آورد.



نمودار ۲-۳: تفاوت میانگین COG سایش در نمونه‌های کاهش یافته با نیافته

۴-۴-۲. دیرش واکه [e]

مقدار دیرش واکه [e] در هر یک از نمونه‌ها در جدول زیر آمده است.

جدول ۲-۸: آمار توصیفی پارامتر دیرش واکه [e]

Token	Mean	N	Std. Deviation	Minimum	Maximum	Range
[sefɪ]	187.19	32	54.036	100	330	230
[se]	230.93	29	92.896	109	466	357
[seF]	181.04	28	44.289	124	314	190
Total	199.51	89	69.752	100	466	366

برخلاف دیدگاه ثمره (۱۳۸۵) میزان دیرش [e] در پایان کلمه [se] بیشتر از مقدار آن در جایگاه ماقبل خوشه در کلمه [sefɪ] است. دلیل آن را می‌توان خارج شدن سریعتر اندامهای گفتاری از وضعیت واکه برای تولید همخوان بعدی در [sefɪ] دانست. آزمون تحلیل واریانس یک‌طرفه نشان می‌دهد که دیرش واکه در این نمونه‌ها با یکدیگر تفاوت معنادار دارد.

جدول ۲-۹: نتایج آزمون واریانس یک‌طرفه، مقایسه پارامتر میانگین دیرش واکه [e] و نیز تأثیر جایگاه (place) بر آن

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Token	45321.560	2	22660.780	5.578	.005
place	31285.045	2	15642.522	3.850	.025
Token * place	20844.824	4	5211.206	1.283	.284
Error	325013.981	80	4062.675		
Total	3970576.000	89			
Corrected Total	428154.247	88			

تست Post Hoc (LSD) در جدول زیر نشان می‌دهد:

۱. دیرش واکه در واژه [se] به طور معناداری بیشتر از [sefɪ] و [se] است.

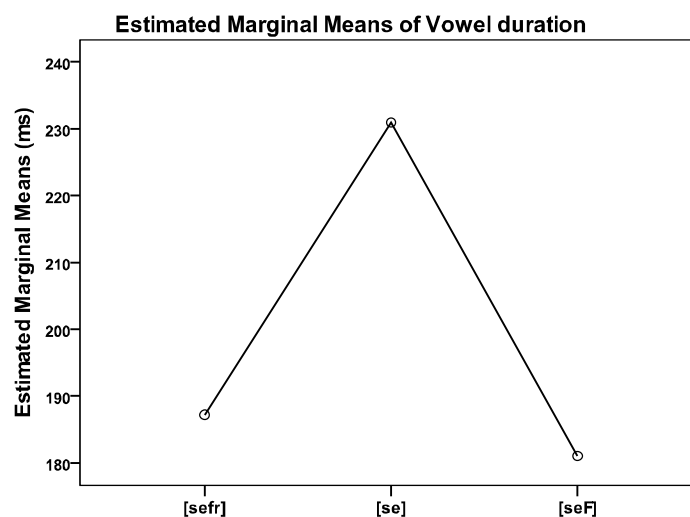
۲. دیرش واکه [e] تفاوت معناداری بین [sefɪ] و [seF] ندارد.

جدول ۲-۱۰: نتایج آزمون LSD، مقایسه میانگین دیرش واکه در نمونه ها با یکدیگر

Token	Token	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
[sefɪ]	[se]	-43.74*	16.342	.009	-76.26	-11.22
	[seF]	6.15	16.494	.710	-26.67	38.98
[se]	[sefɪ]	43.74*	16.342	.009	11.22	76.26
	[seF]	49.90*	16.888	.004	16.29	83.50
[seF]	[sefɪ]	-6.15	16.494	.710	-38.98	26.67
	[se]	-49.90*	16.888	.004	-83.50	-16.29

*, The mean difference is significant at the .05 level.

میانگین دیرش واکه در واژه [se] بیشتر از دو نمونه دیگر است. بنابراین این پارامتر، سرنخ اکوستیکی است که تمایز ادراکی /sefɪ-/se/ را حتی در صورتهای کاهش یافته سبب می شود و بنابراین می تواند به عنوان یک سرنخ اکوستیکی در بازشناسی گفتار به کار رود.



نمودار ۲-۴: میانگین دیرش واکه در نمونه ها

مقدار مینیمم دیرش [e] در [seF] بیشتر از [sefɪ] و [se] است. نتایج بدست آمده با دیدگاه ثمره در مورد میزان دیرش این واکه در جایگاه قبل از خوشه های همخوانی پایانی و نیز پایان واژه مغایرت دارد. وی (۱۳۸۵، ۸۹) معتقد است میزان دیرش این واکه در جایگاه قبل از خوشه های همخوانی پایان

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۳۴

نظیر [fi] در کلمه [sefi] بیشتر از تولید این آوا در پایان کلمه [se] است. اما نتایج این پژوهش نشان داد که میزان دیرش آن در پایان واژه بسیار بیشتر، بیش از ۴۰ میلی ثانیه، از تولید آن در جایگاه قبل از خوشه‌های پایانی است.

۲-۴-۵. تأثیر جایگاه رقم در رشته ارقام بر دیرش واک

میانگین مقادیر دیرش واک بر حسب جایگاه رقم در رشته ارقام در جدول ۲-۱۱ آمده است.

جدول ۲-۱۱: آمار توصیفی پارامتر دیرش واکه [e] بر حسب جایگاه تولید رقم در رشته ارقام

Token	place	Mean	N	Std. Deviation	Minimum	Maximum	Range
[sefr]	initial	211.57	7	47.116	139	255	116
	middle	195.21	19	52.645	136	330	194
	final	133.33	6	30.078	100	177	77
	Total	187.19	32	54.036	100	330	230
[se]	initial	262.88	8	82.532	189	402	213
	middle	247.83	12	105.871	144	466	322
	final	180.00	9	66.817	109	316	207
	Total	230.93	29	92.896	109	466	357
[seF]	initial	166.75	8	37.144	124	236	112
	middle	187.93	15	52.862	125	314	189
	final	183.20	5	21.288	165	217	52
	Total	181.04	28	44.289	124	314	190
Total	initial	213.83	23	69.958	124	402	278
	middle	206.57	46	73.090	125	466	341
	final	166.80	20	52.156	100	316	216
	Total	199.51	89	69.752	100	466	366

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی £ ۳۵

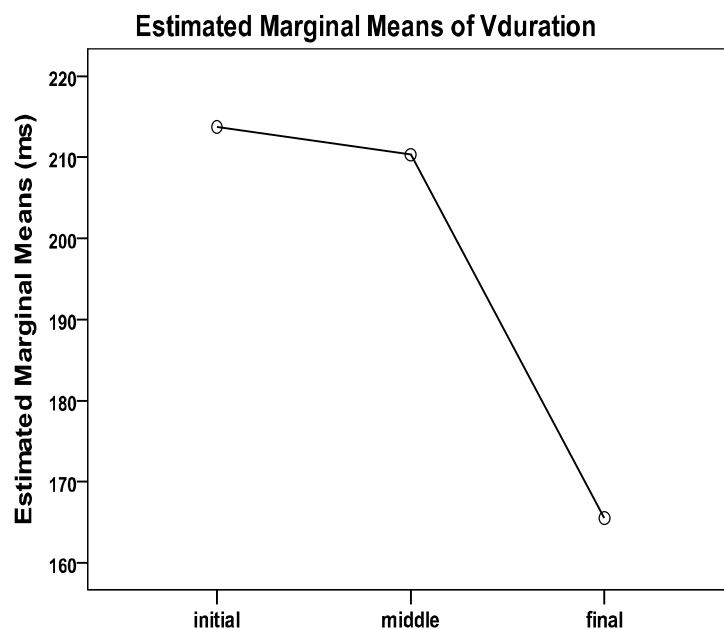
جایگاه تأثیر معناداری بر دیرش واک دارد (جدول ۲-۹). میانگین دیرش واکه هنگامی که رقم تولید شده در پایان رشته ارقام تولید شود، به طور معناداری کمتر از آغاز و میان رشته ارقام است (جدول ۲-۱۲).

جدول ۲-۱۲: نتایج آزمون LSD، مقایسه میانگین دیرش واکه [e] بر حسب جایگاه رقم در رشته ارقام

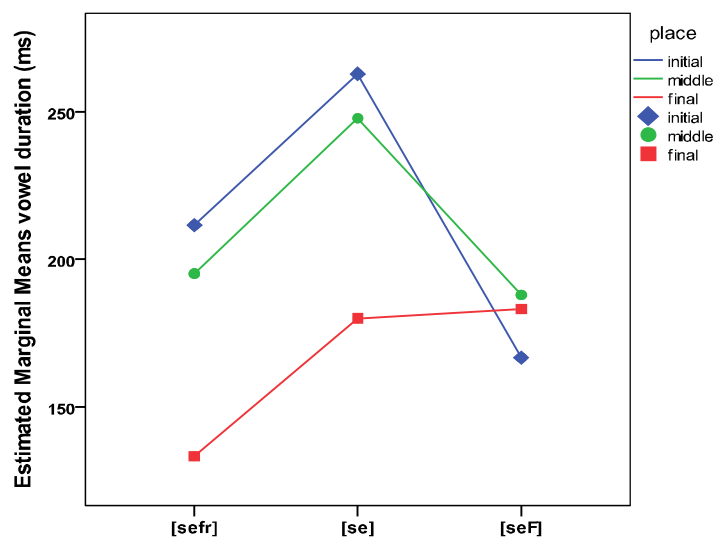
(I) place (J) place	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
				Lower Bound	Upper Bound
initial middle final	7.26	16.278	.657	-25.13	39.65
	47.03*	19.488	.018	8.24	85.81
middle initial final	-7.26	16.278	.657	-39.65	25.13
	39.77*	17.072	.022	5.79	73.74
Final initial middle	-47.03*	19.488	.018	-85.81	-8.24
	-39.77*	17.072	.022	-73.74	-5.79

*, The mean difference is significant at the .05 level.

نمودارهای ۲-۵ و ۲-۶ تأثیر جایگاه رقم بر میزان دیرش واکه را نشان می دهد.



نمودار ۲-۵: دیرش واکه بر حسب جایگاه رقم در رشته ارقام



نمودار ۲-۶: تأثیر جایگاه رقم بر میزان دیرش واکه آن در هر نمونه

۶-۴-۲. گذرهای سازه‌ای و بسامد مکانی (locus frequency)

مقادیر سازه‌ای F1، F2 و F3 در وضعیت ایستان در میان واکه و پایان واکه [e] در هر یک از سه نمونه در جدول ۲-۱۳ آمده است (حرف m برای اشاره به بسامد سازه در میان واکه (middle) و حرف f برای اشاره به بسامد سازه در پایان واکه (final) به کار رفته است. لذا F1m یعنی مقدار بسامدی سازه اول در میان واکه و F1f یعنی مقدار بسامدی سازه اول در انتهای واکه).

جدول ۲-۱۳: آمار توصیفی پارامترهای سازه‌ای واکه

Token		F1m	F1f	F2m	F2f	F3m	F3f
[sefr]	Mean	542.47	494.44	1915.22	1684.66	2669.75	2601.22
	N	32	32	32	32	32	32
	Std. Deviation	85.256	84.898	169.313	126.384	165.968	196.554
[se]	Mean	547.76	519.07	2026.00	1991.14	2660.72	2683.45
	N	29	29	29	29	29	29
	Std. Deviation	72.038	113.831	172.939	193.761	138.597	161.966
[seF]	Mean	557.36	530.39	1962.54	1633.21	2663.21	2513.00
	N	28	28	28	28	28	28
	Std. Deviation	71.526	101.539	162.458	156.217	118.164	144.905
Total	Mean	548.88	513.78	1966.20	1768.34	2664.75	2600.26
	N	89	89	89	89	89	89
	Std. Deviation	76.286	100.291	172.743	223.102	141.826	181.941

میانگین مقادیر گذر هر سازه در هر سه نمونه آوایی، به جز مقدار F3 واکه در نمونه [se]، به تدریج از بخش میانی واکه به انتهای آن کاهش یافته است. مقدار F3 واکه در نمونه [se] به میزان ۲۳ هرتز افزایش فرکانس داشته است. کاهش مقادیر سازه F2 در انتهای واکه [e] در نمونه‌های [sefi] و [seF] بیش از نمونه [se] است. میزان F2 و F3 واکه در نمونه [seF] نیز بیش از [sefi] کاهش پیدا کرده است.

جدول ۲-۱۴: تفاوت میانگین بسامدی سازه‌ها از میان به انتهای واکه در هر یک از نمونه‌ها

Token	F1m – F1f	F2m – F2f	F3m – F3f
[sefi]	48	231	68
[se]	28	35	+23
[seF]	27	329	150

مقادیر گذر هر سه سازه واکه [e] از بخش ایستان واکه به انتهای آن در کلمه [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 به تدریج در انتهای آن در کلمات [sefi] و [se] کاهش یافته است. کاهش مقادیر سازه‌های F2 و F3 در انتهای واکه [e] در نمونه‌های [sefi] و [seF] به دلیل تولید گرفتگی سایش [f] رخ داده است. در واقع تغییر وضعیت دستگاه گفتار از واکه به همخوان سبب بروز تغییر در الگوی فرکانسی سازه‌های دوم و سوم شده است. شیب گذر این سازه‌ها حاوی سرنخهای اکوستیکی لازم برای درک همخوانهای سایشی است و مقادیر سازه‌های F1، F2 و F3 در انتهای واکه فرکانسهای مکانی همخوانی محسوب می‌شوند.

لازم به ذکر است مقادیر F1 و F3 بدست آمده با مقادیر صادقی (۱۳۸۴، ۵۱۶) تطابق دارد. وی مقدار تغییرات سازه اولی برحسب طول دستگاه گفتار را ۵۰۰-۸۰۰ هرتز و سازه سوم را ۲۵۰۰-۲۹۰۰ تعیین کرده است.

در این پژوهش میانگین مقدار سازه اول ۵۴۸ هرتز، سازه دوم ۱۹۶۶ هرتز و سازه سوم ۲۶۶۴ هرتز برای واکه [e] بدست آمد. اما میانگین مقدار سازه دوم با مقادیر صادقی (همان) ۱۵۰۰-۱۸۰۰ تطابق ندارد. مقدار مینیمم بدست آمده در داده‌ها ۱۶۰۷ هرتز و مقدار ماکزیمم ۲۳۹۰ هرتز است. نتایج آزمون واریانس دوطرفه برای هر یک از مقادیر سازه‌ای در میان و پایان واکه در جدول ۲-۱۵ ارائه شده است.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۳۸

جدول ۲-۱۵: نتایج آزمون واریانس دوطرفه، مقایسه میانگین بسامدی سازه‌ها در نمونه‌ها و تأثیر جنسیت (sex) بر آن

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Token	F1m	7168.216	2	3584.108	1.042	.357
	F2m	198039.330	2	99019.665	4.630	.012
	F3m	941.625	2	470.812	.029	.972
	F1f	25366.014	2	12683.007	1.347	.266
	F2f	2220596.192	2	1110298.096	51.569	.000
	F3f	357335.886	2	178667.943	10.323	.000
sex	F1m	220639.647	1	220639.647	64.117	.000
	F2m	663105.010	1	663105.010	31.005	.000
	F3m	341513.932	1	341513.932	20.758	.000
	F1f	61969.817	1	61969.817	6.581	.012
	F2f	193706.426	1	193706.426	8.997	.004
	F3f	1000269.233	1	1000269.233	57.794	.000
Token * sex	F1m	403.980	2	201.990	.059	.943
	F2m	4950.579	2	2475.290	.116	.891
	F3m	44591.099	2	22295.550	1.355	.264
	F1f	24813.146	2	12406.573	1.318	.273
	F2f	220582.242	2	110291.121	5.123	.008
	F3f	38457.129	2	19228.564	1.111	.334

بر طبق جدول فوق، تنها میانگین مقادیر فرکانس مکانی F2 در میان و پایان واژه [e] و نیز F3 در پایان واژه تفاوت معناداری در هر یک از نمونه‌ها نشان می‌دهد.

۲-۴-۶-۱. گذر و بسامد مکانی F2

نتایج تست Post Hoc LSD در جدول ۲-۱۶ نشان می‌دهد کدام یک از میانگین مقادیر F2m و F2f با یکدیگر تفاوت معنادار دارد.

جدول ۲-۱۶: نتایج آزمون LSD، مقایسه میانگین مقادیر سازه F2 در نمونه‌ها

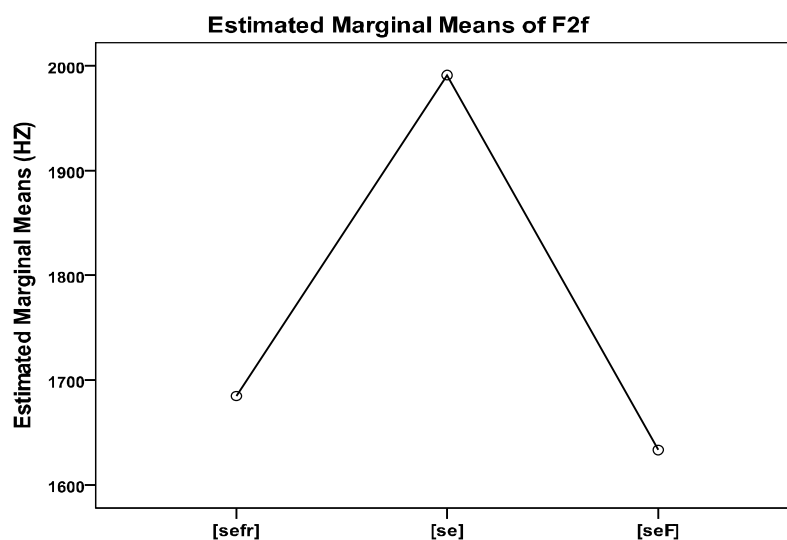
Dependent Variable	(I) Token	(J) Token	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
F2m	[sef.ɪ]	[se]	-110.78*	43.174	.012	-196.61	-24.95
		[seF]	-47.32	43.576	.281	-133.94	39.31
	[se]	[sef.ɪ]	110.78*	43.174	.012	24.95	196.61
		[seF]	63.46	44.616	.159	-25.23	152.16
	[seF]	[sef.ɪ]	47.32	43.576	.281	-39.31	133.94
		[se]	-63.46	44.616	.159	-152.16	25.23
F2f	[sef.ɪ]	[se]	-306.48*	41.056	.000	-388.10	-224.87
		[seF]	51.44	41.438	.218	-30.94	133.82
	[se]	[sef.ɪ]	306.48*	41.056	.000	224.87	388.10
		[seF]	357.92*	42.427	.000	273.58	442.27
	[seF]	[sef.ɪ]	-51.44	41.438	.218	-133.82	30.94
		[se]	-357.92*	42.427	.000	-442.27	-273.58

*, The mean difference is significant at the .05 level.

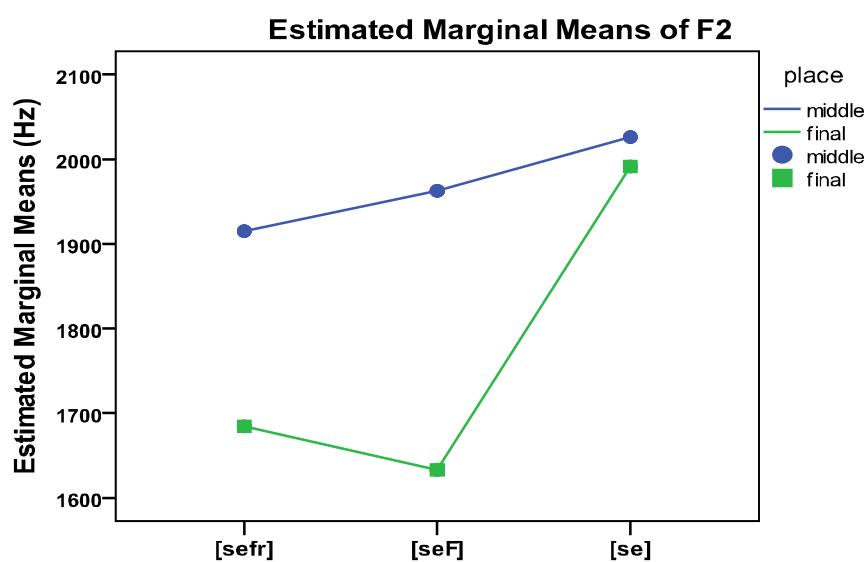
نتایج تست Post Hoc LSD نشان می‌دهد که گذر F2f در پایان واکه و پیش از گرفتگی [f] در هر یک از نمونه‌های [sef.ɪ] و [seF] به طور معناداری کمتر از مقدار فرکانسی همین گذر در پایان واکه در کلمه [se] دارد. همانطور که گفته شد، دلیل این افت فرکانسی به میزان ۲۰۰-۳۰۰ هرتز تغییر وضعیت دهان از واکه به همخوان [f] است. بدین معنا که منبع انرژی واکه قطع می‌شود و در نتیجه ساختار سازه‌ای از بین می‌رود زیرا فیلتر واکه [e] برای تولید آوای بعدی از وضعیت خود خارج می‌شود و پس از آن منبع انرژی دیگری که سایش لبی-دندانی آغار می‌شود. در [se] میانگین مقادیر نشان می‌دهد که تغییر چندانی در وضعیت واکه روی نداده است. این نتایج قابل توجه است زیرا نشانگر رفتار گویشور در مواجهه با رشته‌ای از اعداد است که قرار است به طور منفرد تولید شوند و نه به صورت ترکیبی و به شکل عبارات، مثلاً به صورت "سه، صفر، چهار" و نه "سیصد و چهار" (۳۰۴). از آنجا که گویشور قصد (intention) داشته، هر رقم را مستقل از رقم دیگر تولید کند، هر چند رقم یا رقمهای بعدی را بلافاصله تولید کرده (به صورت connected)، اما به نظر می‌رسد انفصال و استقلال آوایی و اکوستیکی واجها تاحدی حفظ شده است و بازتاب آن در سیگنال گفتار در مشخصه‌های اکوستیکی تظاهر نموده است، به نحوی که مقادیر سازه‌ای واکه در [se] در رشته‌های متوالی ارقام،

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۴۰

تحت تأثیر واج آغازین کلمه بعدی قرار نگرفته است. نمودارهای ۷-۲ و ۸-۲ تفاوت میانگین بسامد سازه دوم را در نمونه‌ها در انتهای واکه و نیز از میان به پایان واکه نشان می‌دهد.



نمودار ۷-۲: تفاوت بسامد گذر F2 در پایان واکه در هر یک از نمونه‌ها



نمودار ۸-۲: تفاوت بسامد گذر F2 از میان به پایان واکه

۲-۴-۶-۲. گذر و بسامد مکانی F3

جدول ۱۷-۲: نتایج آزمون Tamhane، مقایسه میانگین مقادیر F3 در نمونه‌ها

(I) Token	(J) Token	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
[sefɪ]	[se]	-82.23	45.955	.218	-195.19	30.73
	[seF]	88.22	44.240	.145	-20.64	197.08
[se]	[sefɪ]	82.23	45.955	.218	-30.73	195.19
	[seF]	170.45*	40.676	.000	70.26	270.63
[seF]	[sefɪ]	-88.22	44.240	.145	-197.08	20.64
	[se]	-170.45*	40.676	.000	-270.63	-70.26

*, The mean difference is significant at the .05 level.

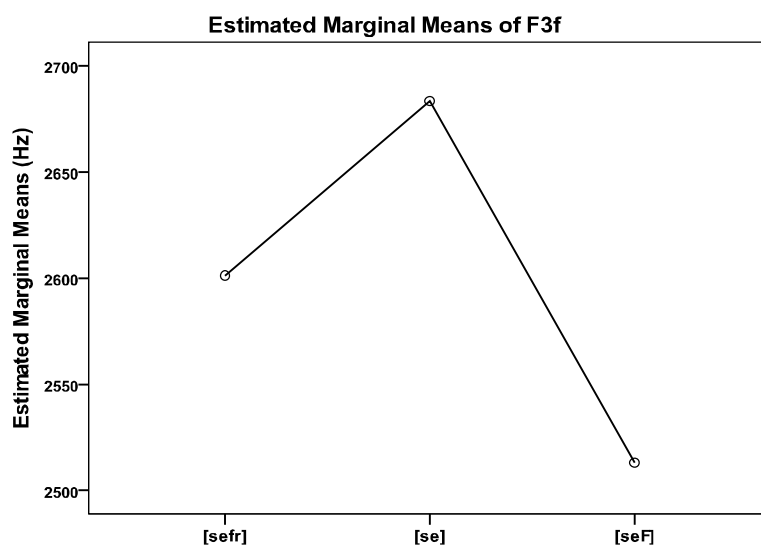
گذر سازه سوم نیز به میزان قابل توجهی در نمونه‌های [sefɪ] و [seF] کاهش پیدا کرده است (نمودار ۲-۹ و ۲-۱۰). نتایج تست Tamhane (جدول ۱۷-۲) نشان می‌دهد که:

۱. میانگین بسامد سازه سوم در [seF] به طور معناداری کمتر از [se] است.

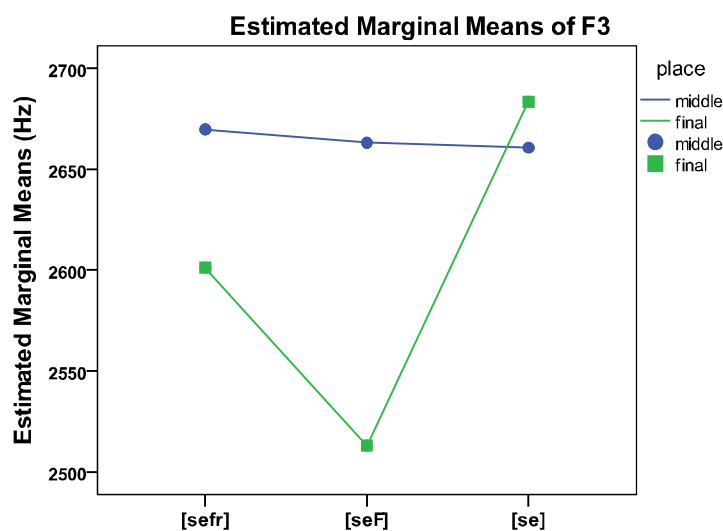
۲. میانگین بسامد سازه سوم در [sefɪ] تفاوت معناداری با [se] و [seF] ندارد

کاهش بسامد سازه سوم در پایان واکه در [sefɪ] و [seF] به ترتیب ۶۸ و ۱۵۰ هرتز به دلیل خارج شدن دهان از وضعیت واکه به هخوان است. اما میانگین آن در پایان واکه در [seF] کاهش بیشتری نسبت به [sefɪ] نشان می‌دهد زیرا هنگامی که لبها برای ایجاد گرفتگی لبی-دندانی به دندانه‌های بالا نزدیک می‌شوند، به نظر می‌رسد همزمان تیغه زبان هم به سمت لثه برای تولید سایش [ɪ] حرکت می‌کند و از آنجا که گذر F3 در همخوان لثوی افزایش پیدا می‌کند، به نظر می‌رسد سبب شده تا بسامد این سازه در [sefɪ] کاهش کمتری نسبت به [seF] پیدا کند. از سویی دیرش سایش در [sefɪ] کوتاهتر از [seF] است و بنابراین تأثیر کمتری هم بر F2 و هم بر F3 واکه دارد.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۴۲



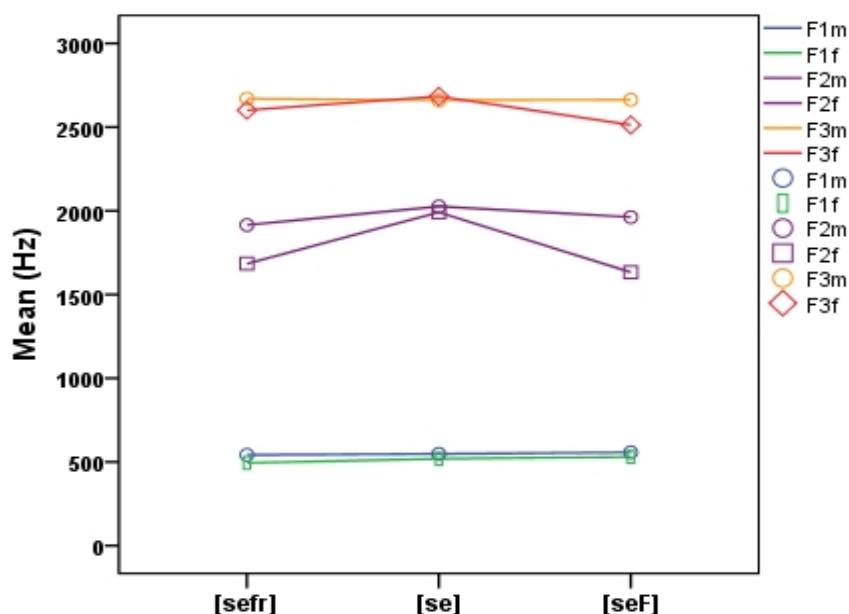
نمودار ۲-۹: تفاوت بسامد گذر F3 در پایان واژه در هر یک از نمونه‌ها



نمودار ۲-۱۰: تفاوت بسامد گذر F3 از میان به پایان واژه در هر یک از نمونه‌ها

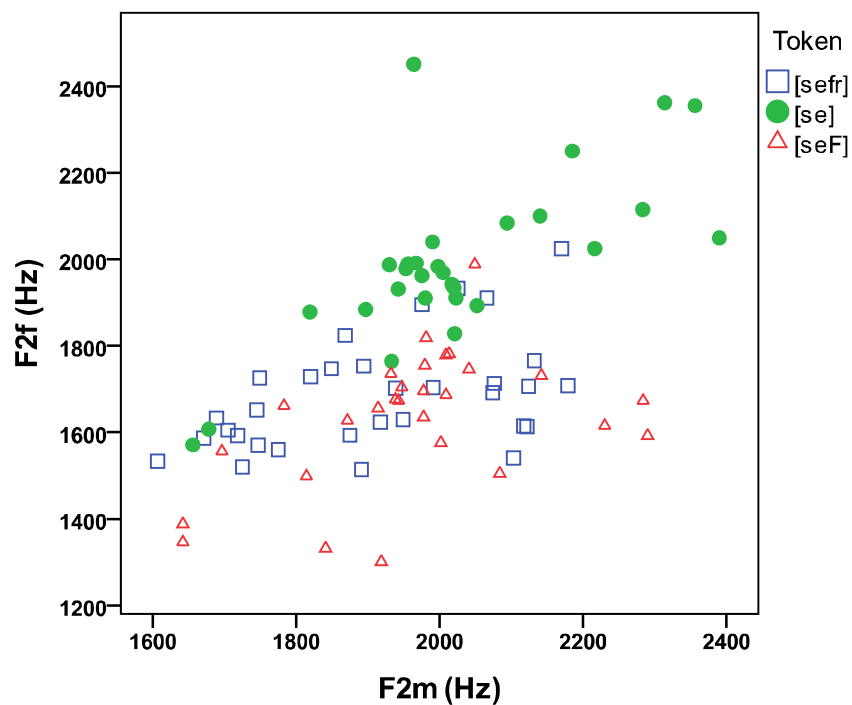
جهت سازه دوم و سوم به مشخصه محل تولید ارتباط دارد و جهت اولین سازه به نحوه تولید (انسدادی، سایشی، خیشومی و غیره) بستگی دارد (دیلتر ۱۹۶۲).

تغییر مقادیر سازه‌های F1، F2 و F3 واکه [e] در هر یک از نمونه‌ها در نمودارهای زیر (۲-۱۱ تا ۲-۱۳) نشان داده شده است. بررسی داده‌ها نشان می‌دهد که مقادیر F2 در نمونه‌های [sefɪ] و [seF] همواره از میان واکه به انتهای آن کاهش یافته است. اما در نمونه [se] این یکدستی مشاهده نمی‌شود به نحوی که تقریباً در یک سوم داده‌ها مقدار بسامدی F2 از میان واکه به انتهای آن افزایش پیدا کرده است. به طور کلی سایر مقادیر سازه‌های F1 و نیز F3 الگوی گذر یکسانی را در هر سه نمونه نشان نمی‌دهند و در برخی از نمونه‌ها افزایش گذر این دو سازه از میان به انتهای واکه و در برخی افت این دو گذر سازه‌ای مشاهده می‌شود. تغییرات F1، F2 و F3 از گذر میان به انتهای واکه در نمودار ۲-۱۱ نشان داده شده است. در نمودارهای ۲-۱۲ و ۲-۱۳، پراکندگی مقادیر سازه‌های F2 و F3 برحسب میان به انتهای واکه، در داده‌ها نشان داده شده است. نمودار ۲-۱۲ نشان می‌دهد به طور کلی بسامد سازه دوم در [se] در پایان واکه بیشتر از دو نمونه دیگر است اما در مورد F3، نمودار ۲-۱۳ نشان می‌دهد که داده‌ها با یکدیگر همپوشانی دارند.

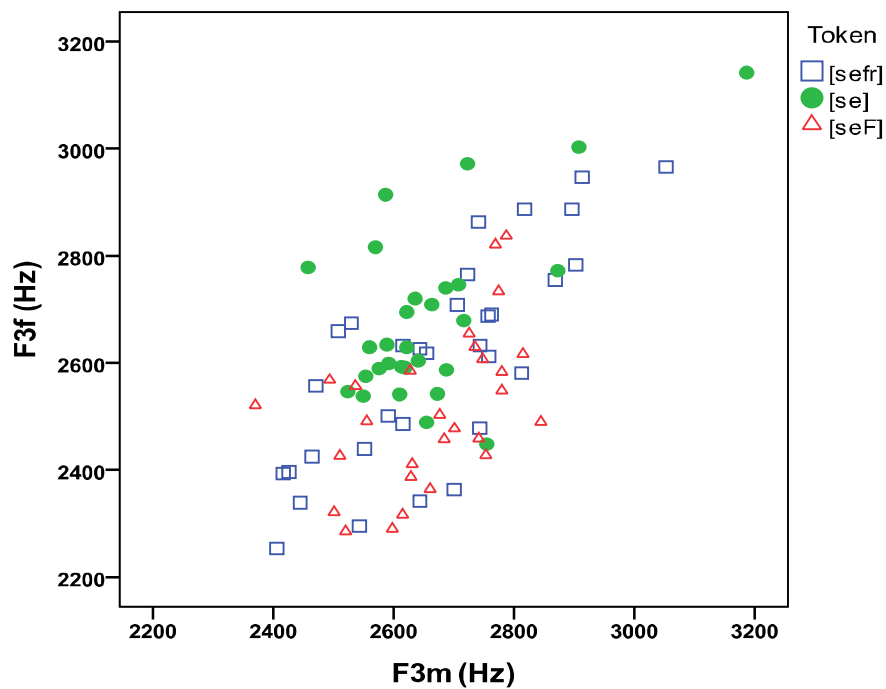


نمودار ۲-۱۱: تغییر بسامد سازه‌ها از میان به پایان واکه

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۴۴



نمودار ۲-۱۲: پراکندگی داده‌ها برحسب تغییر F2 از میان به پایان واکه



نمودار ۲-۱۳: پراکندگی داده‌ها برحسب تغییر F3 از میان به پایان واکه

۲-۴-۳. تأثیر جایگاه کلمه [sefɪ] در رشته ارقام بر مقادیر سازه‌ای واکه

در نمونه [seF] و [sefɪ]، مقادیر سازه واکه [e] تحت تأثیر واج مجاور [f] در همه جایگاه‌ها چه در آغاز، میان یا پایان رشته ارقام قرار می‌گیرد، و بنابراین ویژگیهای این واکه تغییر می‌کند و این خود عامل تمایز [sefɪ] یا [seF] از [se] است.

جدول ۲-۱۸: آمار توصیفی مقادیر سازه‌ای واکه برحسب جایگاه رقم (place) در رشته ارقام

Token	Place		F1m	F1f	F2m	F2f	F3m	F3f
[sefɪ]	initial	Mean	578.86	504.43	1938.71	1714.86	2764.71	2713.29
		N	7	7	7	7	7	7
		Std. Deviation	98.570	103.802	178.106	196.890	203.871	250.584
	middle	Mean	535.74	483.53	1948.26	1666.89	2676.63	2607.84
		N	19	19	19	19	19	19
		Std. Deviation	76.419	78.983	166.864	94.066	145.939	170.752
	final	Mean	521.33	517.33	1783.17	1705.67	2537.17	2449.50
		N	6	6	6	6	6	6
		Std. Deviation	99.095	89.877	116.551	132.868	101.649	117.650
[se]	initial	Mean	523.25	453.13	2108.88	2059.38	2689.13	2767.13
		N	8	8	8	8	8	8
		Std. Deviation	54.426	86.221	159.805	207.970	138.829	180.525
	middle	Mean	543.92	504.33	1967.42	1919.75	2606.75	2621.08
		N	12	12	12	12	12	12
		Std. Deviation	48.256	105.921	180.586	207.142	69.934	102.225
	final	Mean	574.67	597.33	2030.44	2025.67	2707.44	2692.22
		N	9	9	9	9	9	9
		Std. Deviation	104.556	108.877	159.542	144.623	189.766	189.760
[seF]	initial	Mean	535.25	490.13	1983.13	1704.00	2709.13	2536.88
		N	8	8	8	8	8	8
		Std. Deviation	60.311	92.102	177.105	174.444	68.863	146.713
	middle	Mean	573.47	546.40	1964.87	1589.60	2641.73	2497.93
		N	15	15	15	15	15	15
		Std. Deviation	73.474	109.752	178.662	159.990	136.716	163.513
	final	Mean	544.40	546.80	1922.60	1650.80	2654.20	2520.00
		N	5	5	5	5	5	5
		Std. Deviation	84.334	89.720	93.147	65.903	120.346	91.842

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۴۶

در نمونه‌های [se] نیز نتایج (جدول ۱۸-۲، ۱۹-۲ و نمودارهای ۱۴-۲ تا ۱۸-۲) نشان می‌دهد که مقادیر سازه‌ای واکه در [se] تحت تأثیر واجهای بعدی در جایگاه‌های آغاز، میان یا پایان رشته متوالی ارقام قرار نگرفته است.

تحلیل واریانس دوطرفه (جدول ۱۹-۲) نشان می‌دهد که مقادیر سازه‌ای [se] تفاوت معناداری به جزء در بسامد F1f برحسب جایگاه آغاز، میان یا در پایان رشته ارقام با یکدیگر ندارد. بدین معنا که الگوی بسامدی واکه [e] تحت تأثیر واجهای بعدی حین تولید رشته ارقام قرار نگرفته است. همانطور که نتایج نشان داد (جدول ۱۵-۲)، تغییر بسامد سازه اول F1 تأثیری در تمایز نمونه‌ها ندارد.

جدول ۱۹-۲: نتایج آزمون تحلیل واریانس دوطرفه، تأثیر جایگاه بر پارامترهای سازه‌ای واکه در [se]

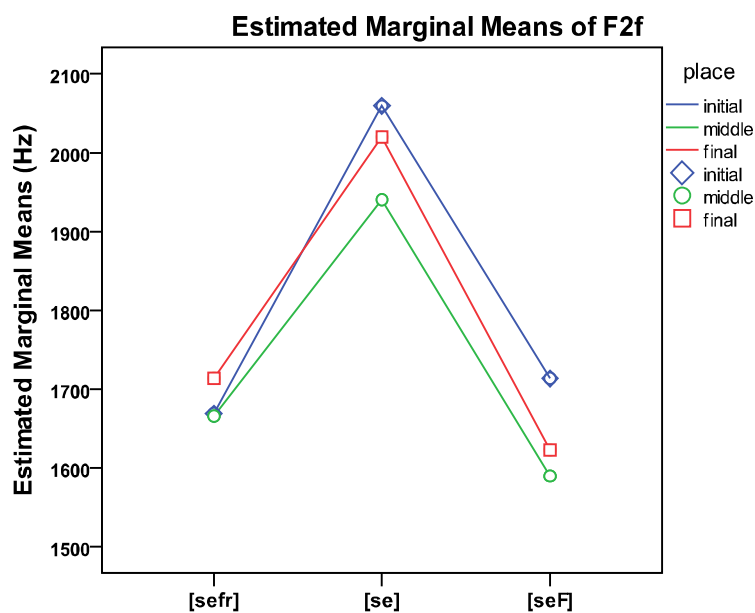
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
place	F1m	11498.894	2	5749.447	1.117	.342
	F2m	96307.986	2	48153.993	1.689	.204
	F3m	61056.446	2	30528.223	1.665	.209
	F1f	92522.320	2	46261.160	4.450	.022
	F2f	109135.323	2	54567.662	1.506	.241
	F3f	103379.825	2	51689.913	2.129	.139
Error	F1m	133806.417	26	5146.401		
	F2m	741116.014	26	28504.462		
	F3m	476801.347	26	18338.513		
	F1f	270285.542	26	10395.598		
	F2f	942074.125	26	36233.620		
	F3f	631145.347	26	24274.821		
Total	F1m	8846451.000	29			
	F2m	1.199E8	29			
	F3m	2.058E8	29			
	F1f	8176353.000	29			
	F2f	1.160E8	29			
	F3f	2.096E8	29			
Corrected Total	F1m	145305.310	28			
	F2m	837424.000	28			
	F3m	537857.793	28			
	F1f	362807.862	28			
	F2f	1051209.448	28			
	F3f	734525.172	28			

نمودارهای ۲-۱۴ تا ۲-۱۸ تأثیر جایگاه را بر مقادیر سازه‌ای نشان می‌دهند. مقادیر سازه‌ دوم و سوم واکه در نمونه‌های [sef] و [sefi] در هر سه جایگاه کاهش پیدا می‌کند و این دو پارامتر تمایز بین نمونه [se] با [sefi] و [sef] را در هر سه حالت، چه این ارقام در آغاز، میان یا پایان رشته ارقام تولید شوند، نشان می‌دهد. این نتایج قابل توجه است زیرا نشانگر رفتار گویشور در مواجهه با رشته‌ای از ارقام است که قرار است به طور منفرد تولید شوند و نه به صورت ترکیبی و به شکل عبارات، مثلاً به صورت "سه، صفر، چهار" و نه "سیصد و چهار" (۳۰۴). از آنجا که گویشور قصد (intention) داشته، هر رقم را مستقل از رقم دیگر تولید کند، هر چند رقم یا رقمهای بعدی را بلافاصله تولید کرده (به صورت connected)، اما به نظر می‌رسد انفصال و استقلال آوایی و اکوستیکی واجها تاحدی حفظ شده است و بازتاب آن در سیگنال گفتار در مشخصه‌های اکوستیکی تظاهر نموده است، به نحوی که مقادیر سازه‌ای واکه در [se] در رشته‌های متوالی ارقام، تحت تأثیر واج آغازین کلمه بعدی قرار نگرفته است.

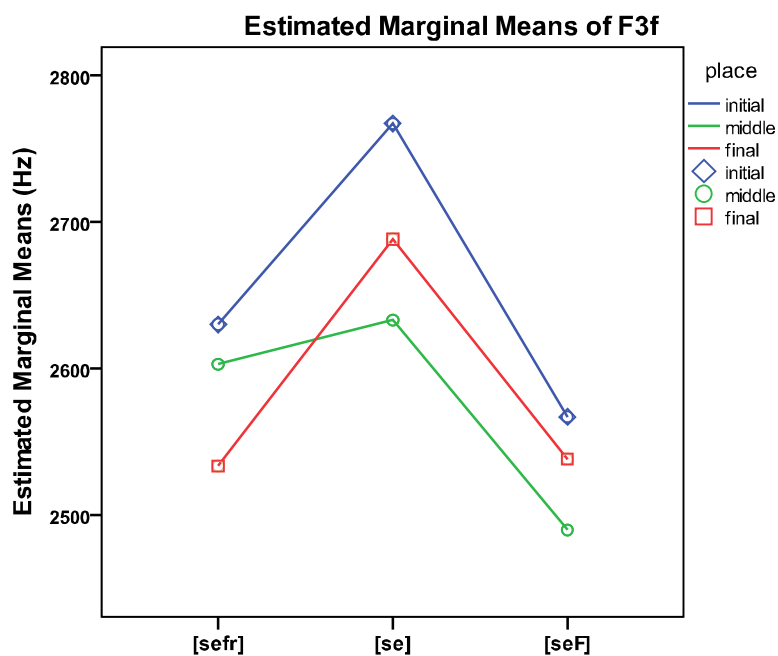
براین اساس به نظر می‌رسد در این نوع گفتار، عامل فرایندهای آوایی مثل کاهش کمتر ناشی از برهم کنش واجها و یا هم‌تولیدی^{۳۷} است زیرا اولاً همان‌طور که گفته شد، کاهش در کلمات منفرد نیز دیده شد. ثانیاً گفتار به صورت جمله و یا عبارات نیست و فقط کلمات مجزایی است که متوالیاً تولید می‌شوند، و نهایتاً اینکه نتایج این پژوهش نیز نشان می‌دهد که مقاومت در برابر هم‌تولیدی واحدهای آوایی در مرز کلمات مجزا در گفتار متصل قابل توجه است.

³⁷ . coarticulation

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۴۸

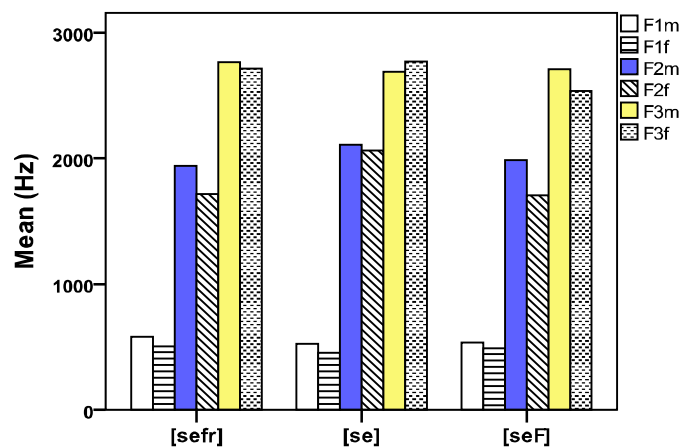


نمودار ۲-۱۴: بسامد F2f برحسب جایگاه نمونه در رشته ارقام

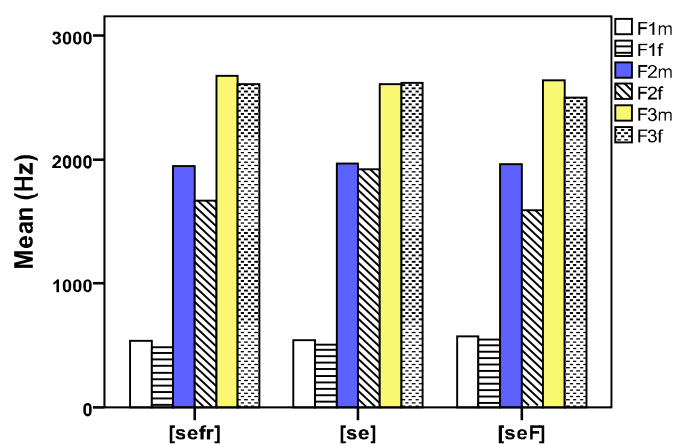


نمودار ۲-۱۵: بسامد F3f برحسب جایگاه نمونه در رشته ارقام

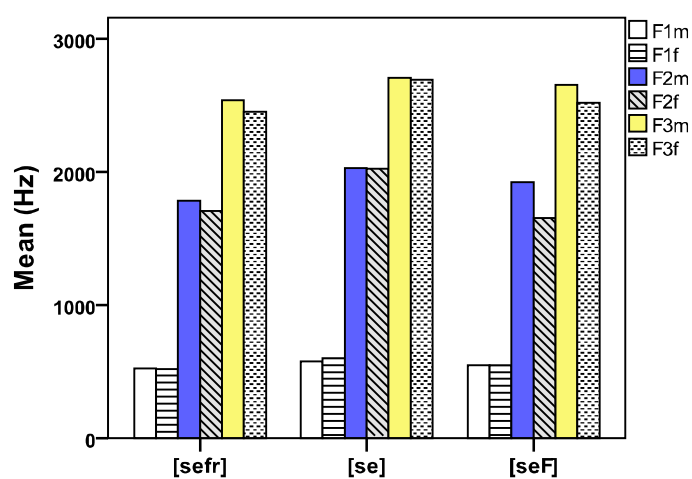
بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۴۹



نمودار ۲-۱۶: تغییر بسامد هر سه سازه در جایگاه آغازین رشته ارقام برحسب نمونه



نمودار ۲-۱۷: تغییرات هر سه سازه در جایگاه میانی رشته ارقام برحسب نمونه



نمودار ۲-۱۸: تغییرات هر سه سازه در جایگاه پایانی رشته ارقام برحسب نمونه

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۵۰ £

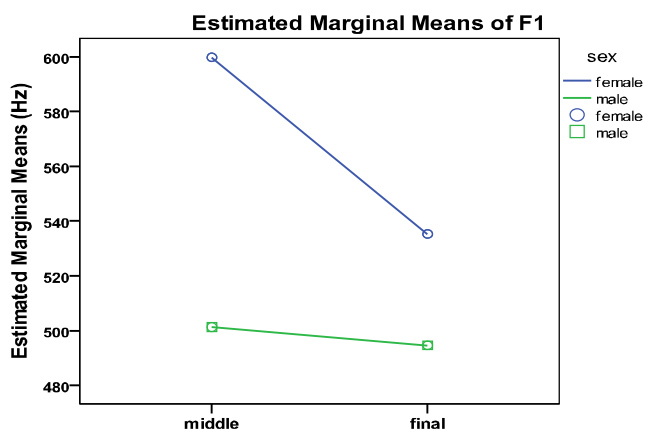
۴-۶-۴-۲. تأثیر جنسیت بر پارامترهای سازه‌ای

همانطور که نتایج آزمون آماری (جدول ۲-۱۵) نشان می‌دهد مقادیر پارامترهای سازه‌ای $F1$ ، $F2$ و $F3$ برحسب جنسیت تفاوت معناداری با یکدیگر دارد. همانطور که انتظار می‌رود، مقادیر بسامدی سازه‌ها در زنان بالاتر از مردان است (جدول ۲-۲۰).

جدول ۲-۲۰: آمار توصیفی مقادیر سازه‌ای برحسب جنسیت (sex)

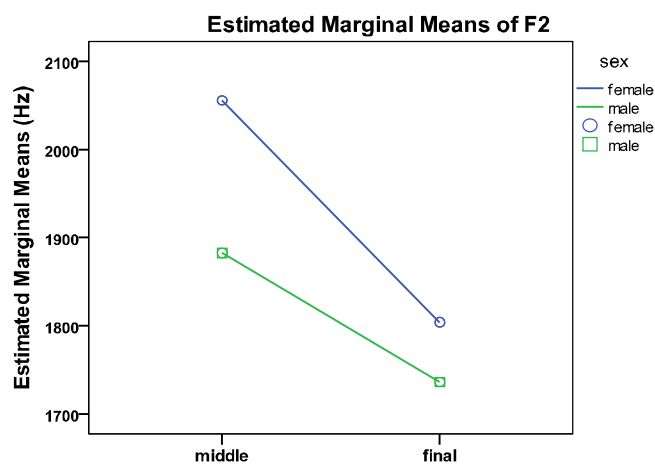
sex		F1m	F1f	F2m	F2f	F3m	F3f
male	Mean	502.02	490.40	1885.51	1720.49	2604.74	2496.21
	N	47	47	47	47	47	47
	Std. Deviation	47.736	68.864	140.248	169.126	94.926	134.220
female	Mean	601.31	539.93	2056.50	1821.88	2731.90	2716.69
	N	42	42	42	42	42	42
	Std. Deviation	67.832	122.234	161.749	263.020	155.935	156.654
Total	Mean	548.88	513.78	1966.20	1768.34	2664.75	2600.26
	N	89	89	89	89	89	89
	Std. Deviation	76.286	100.291	172.743	223.102	141.826	181.941

نمودارهای ۲-۱۹ تا ۲-۲۱ تأثیر جنسیت را بر هر یک از پارامترهای سازه‌ای و نیز از میان به پایان واکه نشان می‌دهد.

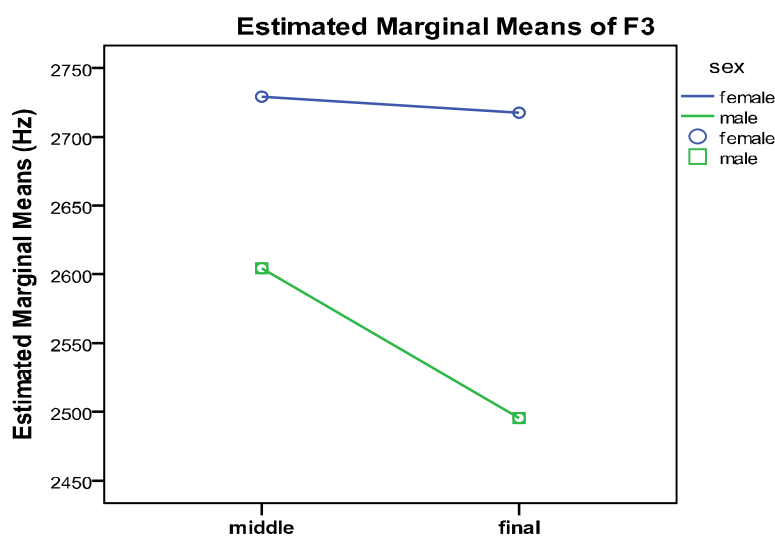


نمودار ۲-۱۹: تأثیر جنسیت بر $F1$ از میان به پایان واکه

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۵۱ £



نمودار ۲-۲۰: تأثیر جنسیت بر F2 از میان به پایان واکه



نمودار ۲-۲۱: تأثیر جنسیت بر F3 از میان به پایان واکه

۵-۲. همخوان سایشی [ɹ]

در جدول ۲-۲۱ دیرش سایش، COG و نیز قله فرکانسی و شدت انرژی قله فرکانسی همخوان لرزشی سایشی [ɹ] ارائه شده است.

جدول ۲-۲۱: آمار توصیفی پارامترهای همخوان [ɹ]

Token		duration	COG	Peak Frequency	Intensity Peak
[ɹ]	Mean	110.09	2263.1383	2326.72	46.5783
	N	32	32	32	32
	Std. Deviation	41.270	342.40658	485.167	6.51189
[f]	Mean	79.91	2115.0006		
	N	32	32		
	Std. Deviation	37.497	269.09352		
[F]	Mean	86.39	2040.4998		
	N	28	28		
	Std. Deviation	37.247	184.87599		

مقادیر دیرش سایش و COG این همخوان با سایشی لبی [f] و صورت کاهش یافته آن [F] تفاوت معناداری دارد (جدول ۲-۲۲).

جدول ۲-۲۲: نتایج آزمون واریانس دوطرفه، مقایسه میانگین دیرش و COG همخوان [ɹ] با [f] در نمونه‌های کاهش یافته و نیافته

Source	variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	duration	16023.569 ^a	2	8011.784	5.327	.007
	COG	781058.664 ^b	2	390529.332	5.110	.008
Intercept	duration	777819.753	1	777819.753	517.213	.000
	COG	4.195E8	1	4.195E8	5488.559	.000
Token	duration	16023.569	2	8011.784	5.327	.007
	COG	781058.664	2	390529.332	5.110	.008
Error	duration	133844.116	89	1503.866		
	COG	6802097.729	89	76428.064		
Total	duration	935009.000	92			
	COG	4.304E8	92			
Corrected Total	duration	149867.685	91			
	COG	7583156.394	91			

نتایج تستهای Post Hoc در جدول زیر (۲-۲۳) نشان می‌دهد که:

۱- دیرش سایش [ɪ] به طور معناداری بیشتر از [f] و [F] است.

۲- بسامد COG در [ɪ] تفاوت معناداری با [f] ندارد ولی به طور معناداری بیشتر از [F] است.

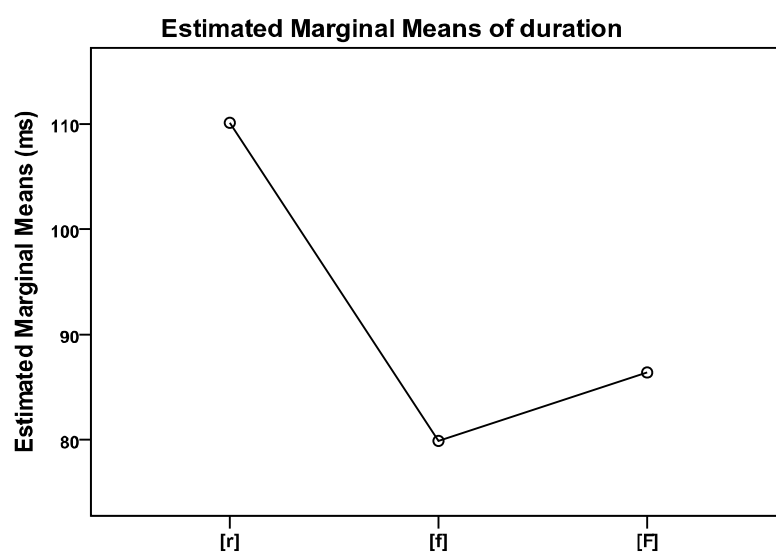
جدول ۲-۲۳: نتایج آزمون LSD و Tamhane، مقایسه میانگین پارامترهای دیرش و COG در نمونه‌ها

Dependent Variable	(I) Token	(J) Token	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Duration LSD	[ɪ]	[f]	30.19*	9.695	.002	10.92	49.45
		[F]	23.70*	10.035	.020	3.76	43.64
	[f]	[ɪ]	-30.19*	9.695	.002	-49.45	-10.92
		[F]	-6.49	10.035	.520	-26.43	13.45
	[F]	[ɪ]	-23.70*	10.035	.020	-43.64	-3.76
		[f]	6.49	10.035	.520	-13.45	26.43
COG Tamhane	[ɪ]	[f]	148.1377	76.98490	.167	-41.0723	337.3476
		[F]	222.6385*	69.88923	.008	49.8489	395.4281
	[f]	[ɪ]	-148.1377	76.98490	.167	-337.3476	41.0723
		[F]	74.5008	59.02150	.511	-70.8416	219.8432
	[F]	[ɪ]	-222.6385*	69.88923	.008	-395.4281	-49.8489
		[f]	-74.5008	59.02150	.511	-219.8432	70.8416

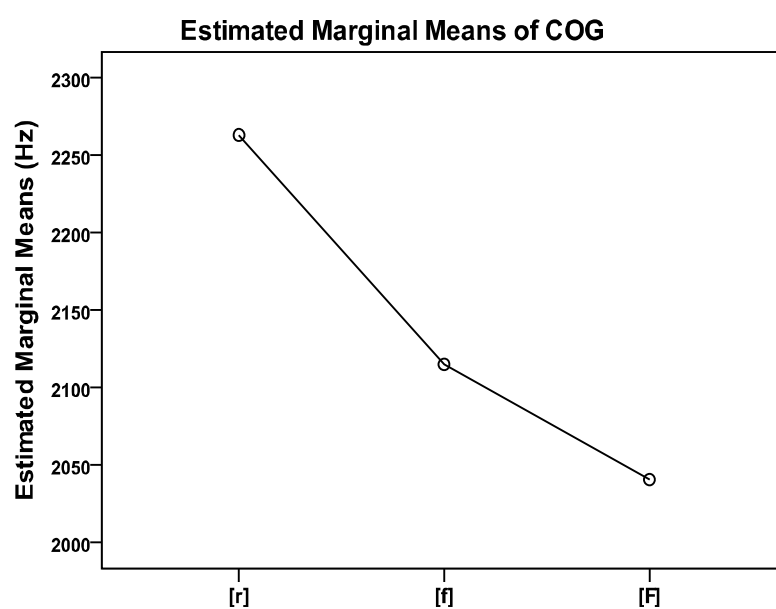
*. The mean difference is significant at the .05 level.

این نتایج ویژگیهای اکوستیکی گونه سایشی [ɪ] را نشان می‌دهد. تفاوت میزان دیرش سایش و COG واجگونه سایش [ɪ] با همخوان سایشی لبی-دندانی [f] و نیز صورت کاهش یافته آن [F] در نمودارهای زیر (۲-۲۲ و ۲-۲۳) نشان داده شده است.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۵۴



نمودار ۲-۲۲: تفاوت دیرش سایش [ɾ] با [f] در نمونه های کاهش یافته و نیافته



نمودار ۲-۲۳: تفاوت COG سایش [ɾ] با [f] در نمونه های کاهش یافته و نیافته

۲-۶. بحث و تحلیل نتایج

نتایج آزمایش نشان می‌دهد که سرنخهای اکوستیکی لازم برای بازشناسی صورتهای کاهش یافته /sefr/ در دو همخوان [f] و واکه [e] قرار دارند. پارامترهای نوفه سایش شامل دیرش سایش و مرکز ثقل طیف و نیز پارامترهای واکه شامل دیرش واکه، شیب گذر F2 و مکان فرکانسی F2 همگی سرنخهای اکوستیکی لازم برای ایجاد تمایز شنیداری [se] - [seF] و بنابراین ادراک صورتهای کاهش یافته /sefr/ را فراهم می‌آورند. چو و مک کوین^{۳۸} (در انتشار) با انجام آزمایشهای ادراکی بر روی خوشه‌های کاهش یافته و صورتهای آوایی حذف شده، نتیجه گرفتند که شنوندگان قادر به درک نتایج فرایندهای گفتار پیوسته که در آنها کاهش به لحاظ اکوستیکی به طور کامل روی می‌دهد، با استفاده از جزئیات آوایی هستند. آنها دریافتند که بازیابی آواهای حذف شده‌ای که نه به همراه سرنخهای اکوستیکی و نه اطلاعات بافتی هستند، نشان می‌دهد که دانش واجی خاص هر زبان که مبتنی بر جزئیات مفصل آوایی است، توسط شنونده برای درک این صورتهای به کار گرفته می‌شود.

عمده‌ترین تفاوت آوایی [se] - [sefi]، پس از حذف [ɪ]، همخوان سایشی [f] است. اما از آنجا که شدت انرژی سایش این همخوان لبی-دندانی بسیار کم است و در واقع در طی تولید این همخوان، منحنی انرژی همواره سیر نزولی دارد، بنابراین بازشناسی سایش از سکوت، حتی برای گوش به سختی صورت می‌گیرد. زیرا در صورتهای کاهش یافته [seF]، منحنی شدت انرژی شباهت بسیاری به منحنی [se] دارد یعنی پس از تولید [e] همان سیر نزولی انرژی مشاهده می‌شود و همین امر سبب می‌شود سیستم شنیداری گوش انسان و به ویژه سیستمهای بازشناسی گفتار این دو صورت آوایی را

³⁸ . Cho, T. and McQueen, J.

به صورت یکسانی درک و بازشناسی کنند. کاواساکی^{۳۹} (۱۹۹۲) دو فاکتور اکوستیکی را عامل درک یا عدم درک توالی‌های آوایی برمی‌شمارد: ۱) شدت^{۴۰} تغییر اکوستیکی^{۴۱} در درون یک توالی^{۴۲} (۲) میزان و درجه تفاوت اکوستیکی^{۴۳} بین توالی‌ها. عامل اول براساس این فرضیه است که "شدت تغییر اکوستیکی در یک توالی آوایی مستقیماً به برجستگی ادراکی^{۴۴} آن ارتباط دارد. بدین معنا که تغییر اکوستیکی به لحاظ ادراکی برجسته‌تر از وضعیت سکون است و بنابراین توالی‌هایی که تغییر طیفی کمی دارند، اولویت ادراکی ندارند. عامل دوم منوط به این استلزام است که تمایز تقابلهای باید به حداکثر خود برسد. بنابراین تقابلهایی که به لحاظ شنیداری متمایز نیستند، نیز به لحاظ ادراکی اولویت ندارند. براساس عامل اکوستیکی اول مطرح‌شده از سوی کاواساکی (۱۹۹۲)، درک [f] در این بافت آوایی ([sefi]) به دشواری صورت می‌گیرد زیرا شدت تغییر طیفی در درون این واحد آوایی بسیار کم است و در نتیجه برجستگی ادراکی آن با وضعیت سکون (یعنی وضعیت Ø که هیچ آوایی تولید نشده است) برابر است و بنابراین به لحاظ شنیداری درک آن در صورت‌های کاهش‌یافته از [se] دشوار است.

آزمایشهای ادراکی فیجو^{۴۵} و سایرین (۱۹۹۸، ۲۰۲۵-۲۰۲۶) نیز نشان می‌دهد که درک /f/ وابسته با بافت است به طوری که درک این همخوان سایشی در بافت واکه‌های پسین، بیشتر (در حدود ۸۵ درصد) از بافت واکه‌های پیشین (حدوداً ۶۰ درصد) است. آنها دلیل این امر را ایجاد قله طیفی حاصل از جفت‌شدن^{۴۶} اکوستیکی حفره جلو و عقب گرفتگی در نتیجه گرد شدن لبها و هم‌تولیدی با واکه پسین ذکر می‌نمایند.

از آنجا که در صورتهایی که گویشوران [ɪ] را تولید کرده‌اند، مهمترین عامل تمایز واجی - sefr/se، انرژی نوفه سایش همخوان [ɪ] است، بنابراین مهمترین پارامتری که سرنخهای اکوستیکی لازم برای تمایز شنیداری se/sefr - را فراهم می‌کند، شدت انرژی قله فرکانسی نوفه سایش [ɪ] است و پارامترهای نوفه سایش [f] شامل دیرش و طیف آن اطلاعات چندانی را برای بازشناسی این آواها به ویژه در سیستمهای بازشناسی گفتار فراهم نمی‌کنند. انرژی نوفه سایش [ɪ] بر روی فرکانسهای حدود

³⁹ . Kawasaki, H.

⁴⁰ . magnitude

⁴¹ . acoustic modulation

⁴² . sequence

⁴³ . acoustic difference

⁴⁴ . perceptual saliency

^{۴۵} . Feijoo, S.

^{۴۶} . acoustic coupling

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۵۷

۱۵۰۰ تا ۳۰۰۰ هرتز متمرکز است. قله فرکانسی انرژی سایش این همخوان در حدود ۲۳۰۰ هرتز با میانگین شدت انرژی تقریبی 47dB است.

سرنخهای اکوستیکی دیگری که تمایز شنیداری /se/ - sefr/ را سبب می‌شوند، پارامترهای واکه شامل دیرش، شیب و پایانه گذرهای واکه‌ای F2 و F3 هستند که حاوی اطلاعات محل تولید همخوان هستند. مقادیر F2 و F3 به طور متوسط به ترتیب ۲۰۰ و ۱۰۰ هرتز در پایانه واکه [e] و قبل از همخوان سایشی [f] افت می‌کند. به نظر می‌رسد مهمترین سرنخ اکوستیکی که می‌تواند تمایز شنیداری /se/ - sefr/ را انتقال دهد، این دو پارامتر هستند. با عملیاتی کردن این دو پارامتر در سیستم بازشناسی ارقام انتظار می‌رود، بازشناسی صورت کاهش یافته /sefr/ به میزان زیادی بهبود یابد. بنابراین استفاده از این پارامتر، به همراه پارامترهای نوفه سایش راه حل دیگری است که می‌تواند برای بازشناسی تمایز /se/ - sefr/ در یک سیستم بازشناسی گفتار به کار رود.

فصل سوم

بازشناسی ارقام

۱-۳. مقدمه

تجزیه و تحلیل اکوستیکی پارامترها نشان داد که پارامترهای واکه، سرنخهای لازم برای بازشناسی [sefi] را در صورتهای کاهش یافته به همراه خود انتقال می دهند. این پارامترها شامل دیرش واکه، مقدار بسامدی F2 در پایانه واکه و تفاوت میزان گذر F2 از میان به انتهای واکه یا شیب گذر F2 هستند. بنابراین براساس آنچه در مرحله قبل بدست آمد، این سه پارامتر برای افزایش دقت بازشناسی ارقام تلفنی به کار گرفته می شوند.

۲-۳. روش شناسی

۱-۲-۳. نمونه ها

در این مرحله سه پارامتر دیرش واکه و مقادیر F2 در ۸۱۱ نمونه آوایی (۴ زن و ۹ مرد) مورد اندازه گیری قرار گرفت (پیوست ۱). اندازه گیری این پارامترها به صورت دستی و به همان روش توضیح داده شده در فصل قبل انجام گرفت.

۲-۲-۳. شناسایی ارقام

۱-۲-۲-۳. شناسایی ارقام "سه" و "صفر" بر اساس پارامترهای واکه

پس از تعیین پارامتر دیرش واکه [e] و همچنین پارامترهای سازه F2 مربوط به این واکه (F2 و F2f transition)، برای سیگنالهای "سه" و "صفر" داده آزمون، می توان بر اساس هر یک از این پارامترها، در مورد "سه" یا "صفر" بودن سیگنال مورد آزمون تصمیم گیری کرد. تصمیم گیری بر اساس میانگین و انحراف معیار هر یک از این پارامترها برای نمونه های [se]، [sefi] و [seF] که در بررسی های پیشین به دست آمده (جدول های ۲-۸ و ۲-۱۳) انجام می شود. تصمیم گیری بدین صورت است که ابتدا هر

یک از ارقام [se]، [sefi] و [seF] به عنوان یک دسته یا کلاس مجزا در نظر گرفته می‌شوند سپس به مسئله تصمیم‌گیری مانند یک مسئله دسته‌بندی^{۴۷} برخورد می‌شود؛ به این صورت که ما باید یک سیگنال ورودی را پس از استخراج پارامترهای آن به یکی از دسته‌های [se]، [sefi] و [seF] منتسب کنیم. البته چون در نهایت هدف ما انتخاب یکی از ارقام "سه" یا "صفر" است، در صورت تشخیص یکی از کلاس‌های [sefi] و [seF] می‌توان سیگنال ورودی را "صفر" و درغیراین صورت "سه" بازشناسی کرد. به منظور دسته‌بندی سیگنال ورودی بر اساس رویه طبقه‌بندی بیزی^{۴۸} عمل می‌کنیم؛ یعنی به ازای یک سیگنال ورودی S باید کلاس C را طوری پیدا کنیم که $P(C|S)$ بیشینه شود. بیشینه شدن $P(C|S)$ طبق رابطه بیز معادل بیشینه شدن رابطه زیر است:

$$C = \arg \max_i P(C_i | S) = \arg \max_i \frac{P(S | C_i) P(C_i)}{P(S)} = \arg \max_i P(S | C_i) P(C_i)$$

در رابطه بالا $P(C_i)$ احتمال پیشین کلاس C_i و $P(S|C_i)$ درجه شباهت^{۴۹} سیگنال S با کلاس C_i می‌باشد. از آنجا که احتمال پیشین هر یک از دسته‌های [se]، [sefi] و [seF] را می‌توان مساوی فرض کرد، پس مسئله دسته‌بندی به مسئله بیشینه کردن درجه شباهت^{۵۰} محدود می‌شود.

برای محاسبه درجه شباهت فرض می‌کنیم توزیع داده‌ها در هر یک از دسته‌ها از چگالی گوسی تبعیت می‌کنند؛ بنابراین به ازای هر یک از دسته‌ها با استفاده از میانگین و انحراف معیار پارامترهای مربوط به هر دسته یک توزیع گوسی تخمین می‌زنیم. بنابراین داریم:

$$P(S | C_i) = \frac{1}{s_i \sqrt{2p}} e^{-\frac{(x_s - u_i)^2}{2s_i^2}}$$

که x_s پارامتر استخراج شده از سیگنال S و μ_i و σ_i به ترتیب میانگین و انحراف معیار این پارامتر در دسته i هستند. از آنجا که سه پارامتر دیرش واکه [e] فرکانس انتهای سازه F2 (F2f) و شیب گذر سازه (F2 transition) برای تصمیم‌گیری مدنظر است، می‌توان سه طبقه‌بندی بیزی داشت که هر یک شامل سه توزیع گوسی (به تعداد دسته‌ها) می‌شود. با استفاده از توزیع‌های گوسی مربوط به هر پارامتر می‌توان تعیین کرد که یک پارامتر ورودی (مثلاً پارامتر دیرش واکه مربوط به یک سیگنال

⁴⁷ classification

⁴⁸ Bayesian classification

⁴⁹ likelihood

⁵⁰ Maximum Likelihood

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۶۱

ناشناخته) به کدام یک از سه دسته [se]، [sefi] و [sef] متعلق است. این کار از طریق بیشینه کردن درجه شباهت انجام می شود یعنی پارامتر ورودی مربوط به کلاسی است که $P(S|Ci)$ را بیشینه کند.

با استفاده از روش دسته بندی فوق، بازشناسی سیگنال های مربوط به ۸۱۱ نمونه "سه" و "صفر" با استفاده از سه پارامتر مذکور صورت گرفت. جدول زیر دقت بازشناسی را به ازای هر یک از پارامترها نشان می دهد.

جدول ۳-۱: دقت بازشناسی سیگنال های "سه" و "صفر" بر اساس پارامترهای مختلف

پارامتر	duration	F2f	F2 transition
دقت (%)	۷۱,۰۸	۷۸,۸۶	۸۱,۹۵

لازم به ذکر است که دقت شناسایی ارقام "سه" و "صفر" با استفاده همزمان از این سه پارامتر نیز اندازه گیری گردید. روش کار بدین صورت است که سه پارامتر مذکور پس از استخراج از داده های آموزشی به عنوان یک بردار سه بعدی در نظر گرفته می شوند. سپس با استفاده از میانگین و کوواریانس این بردارها یک توزیع گوسی سه بعدی برای هر کدام از کلاس ها تخمین زده می شود و رویه دسته بندی بر اساس معیار بیشترین شباهت و با استفاده از این توزیع های گوسی انجام می گیرد. دقت شناسایی در این حالت ۸۸٪ می باشد.

۳-۲-۲-۲. شناسایی ارقام "سه" و "صفر" با استفاده از سیستم بازشناسی گفتار مبتنی بر مدل مخفی مارکوف

بازشناسی نمونه های آزمون با استفاده از یک سیستم بازشناسی گفتار مبتنی بر مدل مخفی مارکوف^{۵۱} (HMM) نیز صورت گرفت. سیستم بازشناسی مذکور با استفاده از یک موتور بازشناسی کلمات متصل، قادر به شناسایی رشته های متصل ارقام (از صفر تا نه) می باشد (فاضل دهکردی ۱۳۸۳). این سیستم، مدل سازی آکوستیک را در سطح کلمه با استفاده از مدل مخفی مارکوف انجام می دهد. تعداد حالت های مدل مخفی مارکوف ۱۶ حالت و تعداد توزیع های گوسی در هر حالت سه عدد می باشد. به منظور آموزش مدل های آکوستیک نیز از دادگان اعداد متصل FCDigits (فاضل

⁵¹ Hidden Markov Model

دهکردی و دیگران (۱۳۸۴) استفاده می‌شود. تعداد ضرایب بازنمایی استخراج شده از هر قاب سیگنال گفتار ۳۹ ضریب می‌باشد که شامل ۱۳ ضریب مل-کپستروم و مشتقات اول و دوم آنها است.

در مرحله بازشناسی نیز این سیستم با استفاده از الگوریتم ویتربی و یک گرامر ساده، بازشناسی رشته‌های ارقام را انجام می‌دهد. با استفاده از سیستم مذکور، بازشناسی ۶۱۴ رشته از ارقام که همگی شامل تعدادی رقم "سه" یا "صفر" بودند، انجام شد. تعداد کل نمونه‌های "سه" و "صفر" در این رشته‌ها ۸۱۱ نمونه است. دقت بازشناسی سیستم برای کل ارقام ۹۵٫۸۰٪ و برای ارقام "سه" و "صفر" ۹۳٫۳۳٪ است. با توجه به اینکه دقت بازشناسی "سه" و "صفر" کمتر از میانگین دقت برای کل ارقام می‌باشد، می‌توان انتظار داشت که با افزایش دادن دقت شناسایی ارقام "سه" و "صفر" با استفاده از دستاوردهای این تحقیق، دقت کل بازشناسی افزایش یابد.

۳-۲-۲-۳. ترکیب نتایج بازشناسی حاصل از دو روش فوق

برای افزایش دقت سیستم بازشناسی می‌توان علاوه بر بازشناسی بر اساس مدل مخفی مارکوف و ضرایب مل-کپستروم، بازشناسی را بر اساس پارامترهای استخراج شده از واژه نیز انجام داد. بدین منظور می‌توان نتایج بازشناسی حاصل از دو روش را به طریقه‌های مختلفی با هم ترکیب کرد. روش ساده‌ای که در این پژوهش به کار برده شده است، روش رأی‌گیری^{۵۲} می‌باشد. روش رأی‌گیری بدین صورت است که ابتدا شناسایی سیگنال ورودی با استفاده از سیستم بازشناسی گفتار و همچنین با استفاده از سه پارامتر مذکور (به روشی که در بالا توضیح داده شد)، انجام می‌گیرد؛ سپس بین نتایج مختلف بازشناسی رأی‌گیری می‌شود و هر کدام از ارقام "سه" و "صفر" که توسط پارامترهای بیشتری شناسایی شده باشد، به عنوان خروجی نهایی سیستم در نظر گرفته می‌شود. اگر تعداد آراء مساوی شد، نتیجه حاصل از سیستم بازشناسی به عنوان خروجی نهایی انتخاب می‌شود.

نتایج بازشناسی حاصل از این روش در جدول زیر آمده است. در این جدول پارامترهای مختلف دوبه‌دو انتخاب شده‌اند و در کنار سیستم بازشناسی به کار رفته‌اند. همان‌طور که مشاهده می‌شود حداکثر دقت زمانی حاصل می‌شود که از هر سه پارامتر در بازشناسی کمک گرفته شود. نتایج همچنین نشان می‌دهد که با استفاده از پارامترهای استخراج شده از واژه [e] دقت بازشناسی ارقام "سه" و "صفر" اندکی افزایش می‌یابد.

⁵² voting

جدول ۲-۳: نتایج نهایی بازشناسی ارقام "سه" و "صفر" با استفاده از روش رأی گیری بین نتایج حاصل از پارامترها

دقت (%)	HMM-based Speech	F2 transition	F2f	Duration
۹۳,۳۳	P			
۹۰,۷۳	P	P	P	
۹۱,۱۰	P		P	P
۹۱,۳۵	P	P		P
۹۳,۵۰	P	P	P	P

یکی دیگر از راه‌های ترکیب نتایج حاصل از دو روش فوق، استفاده از معیار اطمینان^{۵۳} می‌باشد. در سیستم بازشناسی فعلی از هیچ معیار اطمینانی استفاده نمی‌شود ولی در صورتی که معیار اطمینان در دسترس باشد می‌توان هر جا که معیار اطمینان سیستم بازشناسی پایین باشد از نتایج حاصل از پارامترها بهره گرفت و همین‌طور اگر برای روش مبتنی بر پارامترها نیز معیار اطمینانی تعریف شود می‌توان هر جا که این معیار اطمینان پایین است نتیجه حاصل از سیستم بازشناسی را مورد توجه قرار داد. همان‌طور که گفتیم دقت حاصل از سیستم مبتنی بر پارامتر ۸۸٪ و دقت سیستم بازشناسی مبتنی بر HMM حدود ۹۳٪ است؛ بنابراین هر کدام از این سیستم‌ها به تنهایی دقت بالایی دارند. در صورت تعریف شدن یک معیار اطمینان مناسب برای این سیستم‌ها می‌توان با بهره‌گیری از نتایج هر دو سیستم دقت بازشناسی را افزایش داد.

یکی از روش‌هایی که می‌توان برای سیستم مبتنی بر پارامتر، معیار اطمینان تعریف کرد استفاده از درجه شباهت نرمال شده‌ای است که با استفاده از توزیع گوسی به دست می‌آید. همان‌طور که گفتیم یک توزیع گوسی سه‌بعدی برای داده‌های آموزشی تخمین زده شده است. با نرمال‌سازی درجه شباهت حاصل از این توزیع گوسی (تقسیم درجه شباهت‌ها بر حاصل جمع آنها بر روی کلاس‌های مختلف)، عددی بین صفر و یک به دست می‌آید که می‌توان از آن به عنوان معیار اطمینان استفاده کرد:

$$P(S) = \sum_i P(S | C_i)P(C_i) = P(C) \sum_i P(S | C_i)$$

⁵³ confidence measure

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۶۴

$$Confidence = \frac{P(S | C_i)P(C_i)}{P(S)} = \frac{P(S | C_i)}{\sum_i P(S | C_i)}$$

در روابط بالا از این واقعیت استفاده شده است که احتمال پیشین همه کلاس‌ها با هم مساوی هستند بنابراین به جای همه آنها مقدار $P(C)$ قرار داده شده است.

حال روش زیر را برای ترکیب نتایج حاصل از دو سیستم به کار می‌بریم: ابتدا نتیجه هر دو سیستم را با هم مقایسه می‌کنیم (نتیجه حاصل از سیستم مبتنی بر پارامتر در اینجا با استفاده از درجه شباهت توزیع گوسی سه‌بعدی بر روی کلاس‌های مختلف به دست می‌آید). اگر هر دو سیستم جواب یکسانی داشته باشند ("سه" یا "صفر")، همان جواب به عنوان نتیجه نهایی در نظر گرفته می‌شود؛ ولی اگر جواب دو سیستم متفاوت باشد به معیار اطمینان سیستم مبتنی بر پارامتر نگاه می‌کنیم. اگر این معیار اطمینان از یک حد آستانه بالاتر باشد، نتیجه حاصل از سیستم مبتنی بر پارامتر را به عنوان نتیجه نهایی انتخاب می‌کنیم و در غیر این صورت نتیجه سیستم مبتنی بر HMM را قبول می‌کنیم.

بر اساس روش فوق آزمایش‌هایی با مقادیر مختلف حد آستانه برای معیار اطمینان انجام گردید. جدول زیر نتایج نهایی بازشناسی را به ازای مقادیر مختلف حد آستانه نشان می‌دهد.

جدول ۳-۳: نتایج نهایی بازشناسی ارقام "سه" و "صفر" با استفاده از مقادیر مختلف معیار اطمینان

آستانه	دقت نهایی
۰,۵	۹۲,۳۴
۰,۶	۹۴,۱۹
۰,۷	۹۴,۳۱
۰,۸	۹۳,۸۲
۰,۹	۹۳,۸۲
۱,۰	۹۳,۳۳

مشاهده می‌شود که حداکثر دقت زمانی حاصل می‌شود که حد آستانه برابر با ۰,۷ انتخاب شود. این دقت حدود ۱٪ بیشتر از دقت سیستم بازشناسی مبتنی بر HMM (۹۳,۳۳٪) است. بنابراین می‌توان گفت روش ترکیبی مبتنی بر معیار اطمینان خطای بازشناسی ارقام "سه" و "صفر" را به طور نسبی حدود ۱۵٪ کاهش داده است.

داده های بدست آمده در این مرحله مورد تجزیه و تحلیل آماری نیز قرار گرفت تا روشن شود که آیا مقادیر بدست آمده در این مرحله با حجم نمونه بیشتر، نتایج بدست آمده در مرحله قبل با حجم نمونه کمتر را تأیید می کند.

۳-۳. تجزیه و تحلیل آماری پارامترها

۳-۳-۱. دیرش واکه

آمار توصیفی مقادیر دیرش واکه در جدول ۳-۳ آمده است.

جدول ۳-۳: میانگین مقادیر دیرش واکه برحسب جایگاه

token	place	Mean	N	Std. Deviation	Minimum	Maximum	Range
[sefr]	initial	193.66	150	50.744	111	317	206
	middle	203.29	305	54.197	106	381	275
	final	155.14	113	35.832	80	234	154
	Total	191.17	568	53.334	80	381	301
[se]	initial	264.76	78	92.053	132	564	432
	middle	283.51	103	80.744	127	553	426
	final	179.63	62	55.851	77	291	214
	Total	250.99	243	89.702	77	564	487
Total	initial	217.98	228	75.546	111	564	453
	middle	223.54	408	71.031	106	553	447
	final	163.82	175	45.365	77	291	214
	Total	209.09	811	71.743	77	564	487

همانطور که داده ها نشان می دهد تفاوت معناداری در دیرش واکه بین نمونه های [se]-[sefr] دیده می شود و دیرش واکه در [se] بیشتر از [sefr] است (جدول ۳-۳). علاوه بر آن دیرش واکه هنگامی که رقم در پایان رشته ارقام تولید شده است، در هر دو نمونه کمتر از زمانی است که ارقام در میان یا آغاز رشته ارقام تولید شده اند (جدول ۳-۳) و این تفاوت معنادار است (جدول ۳-۴ و ۳-۵). این نتایج، یافته های فصل قبل را نیز تأیید می کند.

جدول ۳-۴: نتایج آزمون تحلیل واریانس دو طرفه، تأثیر جایگاه و جنسیت بر دیرش واکه

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
token	595815.794	1	595815.794	187.033	.000
place	542811.324	2	271405.662	85.197	.000
sex	336474.082	1	336474.082	105.623	.000
token * place	83622.345	2	41811.172	13.125	.000
token * sex	54766.504	1	54766.504	17.192	.000
place * sex	3269.728	2	1634.864	.513	.599
token * place * sex	3422.687	2	1711.344	.537	.585
Error	2545311.309	799	3185.621		
Total	3.963E7	811			
Corrected Total	4169064.064	810			

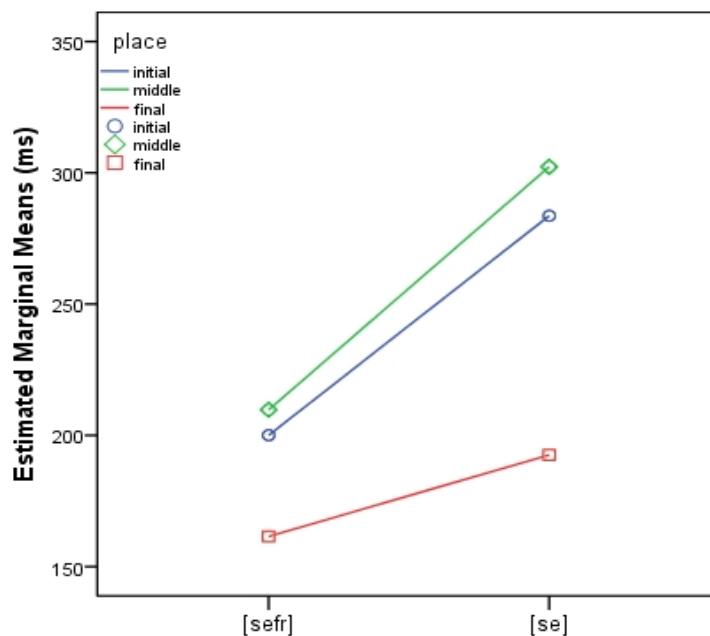
جدول ۳-۵: نتایج آزمون تعقیبی LSD، تفاوت دیرش واکه برحسب جایگاه

(I) place	(J) place	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
initial	middle	-5.56	4.667	.234	-14.72	3.60
	final	54.17*	5.672	.000	43.03	65.30
middle	initial	5.56	4.667	.234	-3.60	14.72
	final	59.73*	5.100	.000	49.72	69.74
final	initial	-54.17*	5.672	.000	-65.30	-43.03
	middle	-59.73*	5.100	.000	-69.74	-49.72

*. The mean difference is significant at the .05 level.

نمودار ۳-۱ نشان می‌دهد که میانگین دیرش واکه در پایان رشته ارقام کمتر از میان یا پایان کلمه است.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۶۷



نمودار ۳-۱: تفاوت دیرش واکه در نمونه‌ها برحسب جایگاه

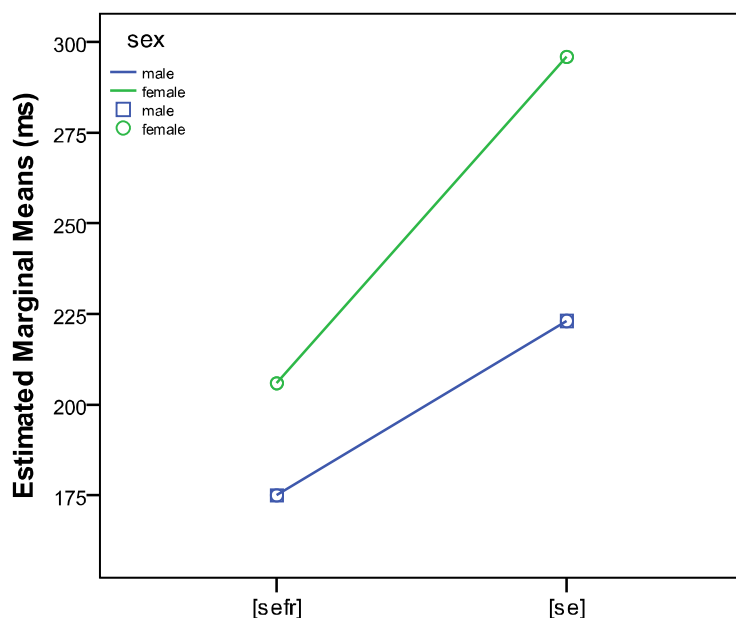
جدول ۳-۴ نشان می‌دهد که جنسیت نیز تأثیر معناداری بر دیرش واکه دارد. جدول ۳-۶ میانگین مقادیر دیرش واکه برحسب جنسیت را نشان می‌دهد.

جدول ۳-۶: میانگین دیرش واکه برحسب جنسیت در هر یک از نمونه‌ها

token	sex	Mean	N	Std. Deviation	Minimum	Maximum	Range
[sefr]	male	181.13	398	47.959	80	317	237
	female	214.67	170	57.858	114	381	267
	Total	191.17	568	53.334	80	381	301
[se]	male	232.84	179	83.719	77	553	476
	female	301.73	64	86.913	130	564	434
	Total	250.99	243	89.702	77	564	487
Total	male	197.17	577	65.761	77	553	476
	female	238.48	234	77.351	114	564	450
	Total	209.09	811	71.743	77	564	487

میانگین دیرش واکه در زنان به طور معناداری بیشتر از مردان است (جدول ۳-۶).

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۶۸



نمودار ۲-۳: تفاوت میانگین دیرش واکه بر حسب جنسیت در هریک از نمونه ها

۲-۳-۳. بسامد پایانه F2

جدول ۷-۳ نشان می دهد که در ۵۶۴ مورد از نمونه های [sefr] مقدار F2m-F2f مثبت بوده است یعنی بسامد این سازه از میان به انتهای واکه کاهش یافته و تنها در ۴ مورد بسامد از میان به انتهای واکه افزایش یافته است. در مورد نمونه [se] در ۱۲۵ نمونه بسامد سازه از میان به انتهای واکه افزایش یافته و در ۱۱۶ نمونه کاهش یافته است. جدول ۸-۳ نشان می دهد که هنگامی که [se] در میان رشته ارقام تولید شده است، بسامد سازه کاهش یافته است، اما این مقدار به لحاظ آماری معنادار نیست.

جدول ۷-۳: بسامد افزایش و کاهش F2 از میان به انتهای واکه

		F2m- F2f	
		+	-
		Count	Count
token	[sefr]	564	4
	[se]	116	125

آمار توصیفی میانگین بسامد سازه دوم در هر یک از سه جایگاه تولید در رشته ارقام در جدول ۸-۳ ارائه شده است.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۶۹

جدول ۳-۸: آمار توصیفی مقادیر F2 از میان به انتهای واکه برحسب جایگاه

F2f				F2m				N	place	token
Max	Min	St. de	Mean	Max	Min	St. de	Mean			
2135	1225	156.815	1700.07	2411	1556	168.55	1924.49	150	آغاز	[sefr]
2067	1269	146.865	1693.26	2327	1483	162.095	1940.74	305	میان	
2137	1082	199.197	1743.89	2359	1462	186.970	1904.44	113	پایان	
2137	1082	162.008	1705.13	2411	1462	169.288	1929.23	568	کل	
2450	1483	160.989	1965.87	2344	1574	147.668	1947.41	78	آغاز	[se]
2302	1407	170.896	1920.97	2288	1517	157.607	1942.57	103	میان	
2406	1592	154.242	1966.02	2200	1511	158.183	1925.76	62	پایان	
2450	1407	164.465	1946.88	2344	1511	154.232	1939.84	243	کل	

همانطور که داده‌ها نیز نشان می‌دهد میانگین بسامد سازه دوم از میان به انتهای واکه [e] در نمونه‌های [sefr] تقریباً بیش از ۲۰۰ هرتز کاهش یافته است اما چنین کاهشی در مورد نمونه‌های [se] مشاهده نمی‌شود. نتایج داده‌ها، یافته‌های فصل قبل را تأیید می‌کند.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی £ ۷۰

جدول ۳-۹ میانگین تغییرات گذر F2 را برحسب جنسیت نشان می دهد. میانگین آن در زنان بیشتر از مردان است.

جدول ۳-۹: میانگین مقادیر گذر F2 برحسب جنسیت

token	sex		F2m	F2f
[sefr]	male	Mean	1865.38	1664.97
		N	398	398
		Std. Deviation	144.338	151.960
	female	Mean	2078.71	1799.14
		N	170	170
		Std. Deviation	123.013	145.349
[se]	male	Mean	1900.99	1916.03
		N	179	179
		Std. Deviation	145.428	146.316
	female	Mean	2048.48	2033.16
		N	64	64
		Std. Deviation	123.853	182.066
Total	male	Mean	1876.43	1742.86
		N	577	577
		Std. Deviation	145.488	189.847
	female	Mean	2070.44	1863.14
		N	234	234
		Std. Deviation	123.716	187.657

جدول ۳-۱۰ نشان می دهد که میانگین بسامد F2 در پایانه واکه تفاوت معناداری بین نمونه های [sefr]-[se] دارد. میانگین بسامد F2 در زنان به طور معناداری بیشتر از مردان است.

جدول ۳-۱۰: نتایج آزمون تحلیل واریانس دوطرفه

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
token	F2m	978.378	1	978.378	.051	.822
	F2f	7947468.848	1	7947468.848	344.135	.000
sex	F2m	4397573.168	1	4397573.168	228.087	.000
	F2f	2132842.866	1	2132842.866	92.355	.000
token * sex	F2m	146403.177	1	146403.177	7.593	.006
	F2f	9798.269	1	9798.269	.424	.515

۳-۳-۳. گذر F2

جدول ۱۱-۳ نشان می دهد که میانگین کاهش بسامد F2 واکه از میان به انتها در هر یک از نمونه ها در سه جایگاه مختلف به چه میزان بوده است. این مقادیر نشان می دهد که کاهش سازه دوم قبل از [f] در میان و آغاز رشته ارقام بیش از ۲۰۰ هرتز است. در پایان رشته ارقام بیش از ۱۵۰ هرتز است. کاهش بسامد سازه دوم در نمونه [sefr] به طور معناداری بیشتر از [se] است. در نمونه [se] تنها در میان رشته ارقام کاهش بسامد F2 به میزان ۱۰۰ هرتز است.

جدول ۱۱-۳: میانگین کاهش سازه دوم بر حسب جایگاه

token	place	Mean	N	Std. Deviation	Minimum	Maximum
[sefr]	initial	224.42	150	139.053	-14	658
	middle	251.73	301	133.087	-164	823
	final	160.55	113	110.517	-59	522
	Total	226.20	564	134.906	-164	823
[se]	initial	86.15	39	76.301	1	263
	middle	102.30	60	88.870	8	410
	final	50.41	17	72.610	-93	203
	Total	89.27	116	83.811	-93	410

جدول ۱۲-۳ نشان می دهد که میانگین افزایش سازه دوم از میان به انتهای واکه در هر یک از نمونه ها به چه میزان است. این مقادیر منفی است زیرا مقدار $F2m - F2f$ منفی است یعنی سازه دوم از میان به انتها افزایش پیدا کرده است. افزایش سازه دوم در نمونه [sefr] تنها در ۴ مورد روی داده که قابل توجه نیست. اما در ۱۲۵ مورد از نمونه های [se] سازه دوم افزایش یافته که نشانگر تفاوت تولید این واکه در نمونه های [sefr]-[se] است.

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۷۲ £

جدول ۳-۱۲: میانگین افزایش سازه دوم بر حسب جایگاه

token	place	Mean	N	Std. Deviation	Minimum	Maximum
[sefɪ]	middle	-72.00	4	83.789	-192	-11
	Total	-72.00	4	83.789	-192	-11
[se]	initial	-126.32	38	145.565	-675	131
	middle	-91.00	43	69.234	-250	-4
	final	-76.20	44	71.776	-348	-1
	Total	-96.53	125	100.806	-675	131

۴-۳. نتیجه گیری

برای افزایش دقت سیستم بازشناسی می توان علاوه بر بازشناسی بر اساس مدل مخفی مارکوف و ضرایب مل-کپستروم، بازشناسی را بر اساس پارامترهای استخراج شده از واکه نیز انجام داد. بدین منظور می توان نتایج بازشناسی حاصل از دو روش را به طریقه های مختلفی با هم ترکیب کرد. حداکثر دقت زمانی حاصل می شود که از هر سه پارامتر در بازشناسی کمک گرفته شود. نتایج نشان می دهد که با استفاده از پارامترهای استخراج شده از واکه [e] دقت بازشناسی ارقام "سه" و "صفر" اندکی افزایش می یابد. در حال حاضر دقت سیستم بازشناسی در حدود ۹۳/۳۳ است که با تلفیق پارامترهای واکه به روشی که ذکر شد، دقت بازشناسی به ۹۳/۵۰ می رسد. یکی دیگر از راه های ترکیب نتایج حاصل از دو روش فوق، استفاده از معیار اطمینان می باشد. در سیستم بازشناسی فعلی از هیچ معیار اطمینانی استفاده نمی شود ولی در صورتی که معیار اطمینان در دسترس باشد می توان هر جا که معیار اطمینان سیستم بازشناسی پایین باشد از نتایج حاصل از پارامترها بهره گرفت و همین طور اگر برای روش مبتنی بر پارامترها نیز معیار اطمینانی تعریف شود می توان هر جا که این معیار اطمینان پایین است نتیجه حاصل از سیستم بازشناسی را مورد توجه قرار داد. یکی از روش هایی که می توان برای سیستم مبتنی بر پارامتر، معیار اطمینان تعریف کرد استفاده از درجه شباهت نرمال شده ای است که با استفاده از توزیع گوسی به دست می آید. مشاهده می شود که حداکثر دقت زمانی حاصل می شود که حد آستانه برابر با ۰,۷ انتخاب شود. این دقت حدود ۱٪ بیشتر از دقت سیستم بازشناسی مبتنی بر HMM (۹۳,۳۳٪) است و دقت بازشناسی را به ۹۴/۳۱٪ می رساند. بنابراین می توان گفت روش ترکیبی مبتنی بر معیار اطمینان خطای بازشناسی ارقام "سه" و "صفر" را به طور نسبی حدود ۱۵٪ کاهش داده است.

به نظر می‌رسد با روشهای متفاوتی می‌توان مقادیر پارامترهای اکوستیکی را با نتایج سیستم تلفیق نمود و نتایج مؤثری در بهبود بازشناسی بدست آورد. این مسئله خود زمینه پژوهشی جدیدی است که نیازمند بررسی و مطالعه بیشتر است.

مسئله دیگری که قابل ذکر است این است که به دلیل آماده نبودن نمونه‌هایی که در محیطهای واقعی ضبط شده‌اند، امکان بررسی دقت بازشناسی با استفاده از پارامترهای اکوستیکی در این مدت زمانی صورت نگرفت. نمونه‌هایی که در محیط واقعی گفتاری تحت تأثیر فاکتورهایی مانند سن، جنسیت، تحصیلات، شرایط جسمانی (سلامت، بیماری)، شرایط روحی (خندان، غمگین، گریه) تولید می‌شوند به میزان قابل توجهی بر دقت بازشناسی سیستم تأثیر می‌گذارند. بنابراین بررسی میزان بهبود دقت بازشناسی با استفاده از پارامترهای اکوستیکی در این نمونه‌ها، به دلیل در اختیار نداشتن این قبیل داده‌ها و زمان پروژه صورت نگرفت و در واقع تأثیر استفاده از پارامترهای اکوستیکی در دقت بازشناسی این نمونه‌ها بسیار حائز اهمیت و مهم است.

فهرست منابع

- الماس گنج، فرشاد، سید محمدرضا هاشمی گلپایگانی، کارو لوکس، سیدعلی سید صالحی و محمود بی جن خان. ۱۳۸۹. افزایش دقت بازشناسی گفتار واجهای زبان فارسی با استفاده از خصوصیات نوایی گفتار در سطح واجها. حسین صامتی و محمود بی جن خان، زبان و فارسی و رایانه: برگزیده مقالات: کنفرانس بین المللی سالانه انجمن کامپیوتر ایران، کنفرانس مهندسی برق ایران و.... انتشارات سمت. جلد دوم. ۹۰۳-۹۱۰
- ثمره، یدالله. ۱۳۸۵. آواشناسی زبان فارسی: آواها و ساخت آوایی هجا. مرکز نشر دانشگاهی
- شیخ سنگ تجن، شهین. ۱۳۸۹. بررسی صوت شناختی کاهش واکه های زبان فارسی در مکالمات تلفنی. دانشگاه تهران: پایان نامه دکتري
- صادقی، وحید. ۱۳۸۴. بازشناسی واجی کلمات فارسی: رویکردی مبتنی بر نظریه بهینگی. دانشگاه تهران: پایان نامه دکتري.
- فاضل دهکردی، امین. ۱۳۸۳. بازشناسی اعداد پیوسته از طریق خط تلفن. دانشگاه صنعتی شریف، دانشکده مهندسی کامپیوتر: پایان نامه کارشناسی ارشد.
- فاضل دهکردی، ا.، صامتی، ح.، بحرانی، م. ۱۳۸۴. طراحی و تهیه دادگان تلفنی اعداد متصل زبان فارسی و استفاده از آن در یک سیستم بازشناسی خودکار گفتار مبتنی بر کلمه. زنجان: سیزدهمین کنفرانس مهندسی برق ایران.

- Ansarin, Ali Akbar. 2004. An acoustic analysis of Modern Persian Vowels. SPECOM: 9th conference: Speech and Computer. St. Petersburg, Russia.
- Cho, Taehong & James McQueen (in press). Perceptual recovery from consonant-cluster simplification in Korean using language-specific phonological knowledge. *Journal of Psycholinguistic Research*. http://tcho.hanyang.ac.kr/papers/Cho_McQueen_in_press_JPR.pdf
- Delattre, P. C., A. M. L. Liberman, and F. S. Cooper. 1962. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica* (16): 104–122.
- Feijoo, S., S. Fernandez, and 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*: 2025-2026.
- Flemming, Edward. 2002. Auditory representations in phonology. Edited by Laurence Horn; outstanding dissertations in Linguistics. Newyork, Routledge.
- Gordon, M., P. Barthmaier, and K. Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International phonetic association* 32 (2): 141-174
- Hedrick, Mark S., and Ralph N. Ohde. 1993. Effect of relative amplitude of frication on perception of place of articulation. *Journal of Acoustical Society of America* 94 (4): 2005-2026.
- Kawasaki, Haruko. 1992. An acoustical basis for universal phonotactic constraints. *Language and Speech* 35 (73-86)
- Key, Kelsey, and Dennis H. Klatt. 1988. Analysis of formant transitions for English fricatives spoken by five male and five female talkers. *Journal of Acoustical Society of America* 83 (1).
- Kirchner, R. M. 1998. An effort based approach to consonant lenition. UCLA: PhD dissertation.

- Lavoie, L. M. 2001. Consonant strength: phonological patterns and phonetic manifestations. Garland publishing.
- Mitterer, Holger., and Mirjam Ernestus. 2006. Listeners recover /t/s that speakers reduce: evidence from /t/-lenition in Dutch. *Journal of phonetics* 34: 73- 103.
- Nakamura, Masanobu, k. Iwano, and s. furui. 2007. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language* 22: 171-184.
- Reetz, Henning., and Allerd Jongman. 2009. *Phonetics: transcription, production, acoustics, and perception*. UK, Wiley-Blackwell.
- Ussishkin, Adam., and Andrew Wedel. 2009. Lexical access, effective contrast, and patterns in the lexicon. Edited by Boersma, Paul., and Silke Hamann. *Phonology in Perception*. Mouton de Gruyter.
- van Son, R. J. J. H., and Louis C. W. Pols. 1997. The correlation between consonant identification and the amount of acoustic consonant reduction. <http://www.fon.hum.uva.nl/IFA-publications/Eurospeech97/A0301/A0301.html>
- Weibel, Alex., and Kai-Fu Lee. 1990. *Readings in speech recognition*. Morgan Kaufmann

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۷۸

پیوست ۱: داده ها: مقادیر اکوستیکی پارامترهای واژه [e] و نمونه ای از بازشناسی

	بازشناسی	بازشناسی				بازشناسی		
F2m- F2f	F2 transition	F2f	F2f	F2m	V duration	دیرش واکه	رقم	نمونه
146	0	3	2024	2170	245	3	0	0a
82	3	3	1874	1956	260	3	0	0b
92	0	3	1933	2025	226	3	0	01a
141	0	0	1753	1894	209	3	0	07a
217	0	3	1982	2199	230	3	0	40
169	0	3	1896	2065	224	3	0	0482a
418	0	0	1706	2124	261	3	0	602
231	0	3	1910	2141	312	3	0	0875a
181	0	3	1910	2091	233	3	0	1051
346	0	0	1823	2169	259	3	0	3043
95	0	3	2048	2143	212	0	0	4740
61	3	3	1988	2049	236	3	0	05781a
171	0	3	1916	2087	245	3	0	06959a
251	0	3	1857	2108	253	3	0	9008
268	0	0	1788	2056	245	3	0	
162	0	3	1919	2081	223	3	0	9056
182	0	3	1849	2031	230	3	0	11009
145	0	0	1812	1957	208	0	0	
179	0	3	2024	2203	289	3	0	60110
93	0	3	2034	2127	212	0	0	
335	0	3	1856	2191	221	3	0	61806
62	3	3	2065	2127	241	3	0	73803
281	0	3	1907	2188	276	3	0	79033
137	0	3	2021	2158	263	3	0	95080
85	3	3	2137	2222	229	3	0	
128	0	3	1997	2125	242	3	0	0187529a
249	0	0	1803	2052	259	3	0	2700023
305	0	0	1757	2062	270	3	0	
271	0	0	1814	2085	253	3	0	
126	0	3	1954	2080	290	3	0	3070222
98	0	3	1854	1952	260	3	0	
372	0	0	1829	2201	304	3	0	3316045
58	3	0	1773	1831	309	3	0	5207932
245	0	3	1904	2149	297	3	0	7154049
242	0	3	1957	2199	301	3	0	7307401
110	0	3	2022	2132	242	3	0	
295	0	3	1843	2138	247	3	0	8672005
363	0	0	1763	2126	223	3	0	
132	0	3	2015	2147	207	0	0	9832770
-89	3	3	2208	2119	256	3	3	83

-67	3	3	2213	2146	271	3	3	3a
-73	3	3	2195	2122	281	3	3	3b
-65	3	3	2088	2023	232	3	3	243
-65	3	3	2250	2185	402	3	3	361
59	3	3	2077	2136	344	3	3	377
-133	3	3	2247	2114	242	3	3	493
32	3	3	2152	2184	330	3	3	3043
-17	3	3	2082	2065	252	3	3	
-75	3	3	2219	2144	332	3	3	7034
267	0	3	2016	2283	358	3	3	21382
27	3	3	2112	2139	308	3	3	67639
311	0	3	1896	2207	308	3	3	73803
169	0	3	2052	2221	318	3	3	79033
25	3	3	2175	2200	246	3	3	
-100	3	3	2212	2112	295	3	3	2341484
210	0	3	1929	2139	180	0	0	0a
67	3	3	1986	2053	195	0	0	0b
467	0	0	1679	2146	252	3	0	02a
591	0	0	1619	2210	221	3	0	03a
274	0	3	1882	2156	273	3	0	006a
361	0	0	1766	2127	210	0	0	
563	0	0	1540	2103	252	3	0	09a
621	0	0	1519	2140	254	3	0	019a
69	3	3	2028	2097	163	0	0	70
199	0	3	1883	2082	177	0	0	110
467	0	0	1622	2089	258	3	0	0249a
281	0	0	1823	2104	220	3	0	707
207	0	0	1774	1981	228	3	0	1008
76	3	3	1934	2010	210	0	0	
265	0	0	1752	2017	235	3	0	1040
10	3	3	1962	1972	138	0	0	
247	0	0	1829	2076	290	3	0	2012
282	0	3	1846	2128	232	3	0	2810
25	3	3	2044	2069	127	0	0	
129	0	3	1935	2064	243	3	0	3017
122	0	3	1978	2100	176	0	0	7550
210	0	3	1840	2050	281	3	0	20626
241	0	3	1896	2137	248	3	0	35023
27	3	3	2020	2047	177	0	0	43220
345	0	0	1800	2145	284	3	0	0050161a
374	0	0	1740	2114	212	3	0	
334	0	0	1650	1984	198	0	0	
447	0	0	1739	2186	246	3	0	80607
236	0	3	1911	2147	231	3	0	

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۸۰ £

295	0	3	1863	2158	240	3	0	84700
36	3	3	2073	2109	156	0	0	
407	0	0	1629	2036	245	3	0	2816080
146	0	3	2029	2175	157	0	0	
217	0	3	1833	2050	196	3	0	5887908
300	0	0	1773	2073	235	3	0	6086644
317	0	3	1858	2175	223	3	0	6180340
62	3	3	2105	2167	132	0	0	
277	0	3	1907	2184	230	3	0	6650036
149	0	3	2001	2150	214	3	0	
473	0	0	1638	2111	253	3	0	8073639
337	0	0	1683	2020	250	3	0	9006215
335	0	0	1706	2041	233	3	0	
-55	3	3	2047	1992	210	3	3	03a
43	3	3	2107	2150	279	3	3	3a
17	3	3	2089	2106	257	3	3	3b
-65	3	3	2110	2045	217	3	3	93
-33	3	3	1989	1956	227	3	3	413
-60	3	3	2088	2028	339	3	3	3017
-255	3	3	2280	2025	388	3	3	3533
79	3	3	2010	2089	374	3	3	
26	3	3	2114	2140	190	0	3	
-19	3	3	2178	2159	332	3	3	35023
-7	3	3	2149	2142	213	3	3	
-32	3	3	2036	2004	369	3	3	2343784
37	3	3	2032	2069	323	3	3	
-119	3	3	2302	2183	325	3	3	6650036
24	3	3	2146	2170	337	3	3	8073639
31	3	3	2007	2038	318	3	3	
9	3	0	1724	1733	148	0	0	0a
384	0	0	1560	1944	154	0	0	0b
238	0	0	1536	1774	165	0	0	001a
87	3	0	1677	1764	162	0	0	
63	3	0	1632	1695	138	0	0	10
86	3	0	1664	1750	164	0	0	059a
275	0	0	1457	1732	180	0	0	60
141	0	0	1625	1766	177	0	0	0178a
150	0	0	1657	1807	227	3	0	300
87	3	0	1624	1711	160	0	0	
260	0	0	1563	1823	175	0	0	0873a
212	0	0	1602	1814	170	0	0	0971a
18	3	0	1655	1673	179	0	0	01034a
241	0	0	1576	1817	177	0	0	
236	0	0	1601	1837	198	0	0	2066

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۸۱ £

165	0	0	1605	1770	207	0	0	6050
101	0	0	1604	1705	163	0	0	
200	0	0	1574	1774	209	0	0	8040
141	0	0	1529	1670	186	0	0	
221	0	0	1564	1785	170	0	0	9030
60	3	0	1707	1767	170	0	0	
96	0	0	1710	1806	189	0	0	09275a
151	0	0	1610	1761	191	0	0	40560
144	0	0	1528	1672	141	0	0	
168	0	0	1611	1779	189	0	0	80819
52	3	0	1716	1768	137	0	0	86920
96	0	0	1661	1757	197	0	0	87088
316	0	0	1547	1863	183	0	0	90658
151	0	0	1636	1787	178	0	0	0297395a
137	0	0	1660	1797	169	0	0	0355210a
308	0	0	1417	1725	136	0	0	
27	3	0	1801	1828	143	0	0	0601569a
227	0	0	1599	1826	162	0	0	
309	0	0	1555	1864	177	0	0	2504667
95	0	0	1718	1813	170	0	0	5089376
166	0	0	1629	1795	182	0	0	5328309
90	0	0	1662	1752	192	0	0	5863107
187	0	0	1663	1850	181	0	0	6207614
94	0	0	1640	1734	180	0	0	9427000
133	0	0	1587	1720	163	0	0	
74	3	0	1588	1662	136	0	0	
-110	3	3	1889	1779	253	3	3	3a
-146	3	3	1885	1739	220	3	3	3b
-116	3	3	1949	1833	290	3	3	31
-69	3	3	1792	1723	202	0	3	63
-129	3	3	1966	1837	264	3	3	300
65	3	0	1735	1800	275	3	3	357
-145	3	3	1943	1798	254	3	3	367
-60	3	3	1832	1772	194	0	3	713
-12	3	0	1824	1812	291	3	3	0873a
-78	3	3	1907	1829	359	3	3	01034a
18	3	0	1781	1799	262	3	3	9030a
-36	3	3	1858	1822	224	3	3	0297395a
-66	3	0	1792	1726	221	3	3	532309
100	0	0	1715	1815	239	3	3	
-39	3	3	1848	1809	295	3	3	9113884
202	0	0	1786	1988	172	0	0	0a
325	0	0	1650	1975	189	0	0	0b
165	0	0	1699	1864	260	3	0	007a

402	0	0	1522	1924	233	3	0	
298	0	0	1626	1924	171	0	0	10
279	0	0	1665	1944	182	0	0	70
177	0	0	1834	2011	193	0	0	80
266	0	0	1613	1879	279	3	0	096a
231	0	0	1683	1914	228	3	0	100
479	0	0	1548	2027	155	0	0	
323	0	0	1601	1924	287	3	0	0471a
382	0	0	1592	1974	245	3	0	601
400	0	0	1634	2034	275	3	0	602
233	0	0	1634	1867	290	3	0	0830a
182	0	0	1766	1948	151	0	0	
318	0	0	1561	1879	282	3	0	0896a
446	0	0	1460	1906	239	3	0	2506
224	0	0	1708	1932	280	3	0	4004
249	0	0	1668	1917	245	3	0	
232	0	0	1705	1937	300	3	0	5208
404	0	0	1514	1918	280	3	0	05339a
273	0	0	1632	1905	312	3	0	05505a
150	0	0	1797	1947	268	3	0	
423	0	0	1543	1966	175	0	0	15490
340	0	0	1564	1904	317	3	0	0026609a
289	0	0	1640	1929	262	3	0	
336	0	0	1600	1936	240	3	0	
352	0	0	1543	1895	245	3	0	28907
366	0	0	1666	2032	289	3	0	68042
250	0	0	1663	1913	177	0	0	47880
170	0	0	1605	1775	247	3	0	97002
244	0	0	1567	1811	219	3	0	
137	0	3	1869	2006	199	0	0	98290
354	0	0	1598	1952	261	3	0	0124407a
341	0	0	1693	2034	293	3	0	
436	0	0	1501	1937	301	3	0	0956062a
209	0	0	1681	1890	273	3	0	
327	0	0	1568	1895	284	3	0	2018793
488	0	0	1494	1982	236	3	0	2102326
359	0	0	1594	1953	304	3	0	6130158
294	0	0	1616	1910	280	3	0	7091000
262	0	0	1607	1869	296	3	0	
385	0	0	1587	1972	267	3	0	
137	0	0	1746	1883	174	0	0	
475	0	0	1549	2024	291	3	0	8407619
352	0	0	1619	1971	299	3	0	8804663
54	3	3	1963	2017	254	3	3	3b

76	3	3	1849	1925	383	3	3	342
-12	3	3	1965	1953	385	3	3	373
6	3	3	1944	1950	166	0	3	
18	3	3	1949	1967	407	3	3	381
-31	3	3	1958	1927	206	0	3	693
25	3	3	1913	1938	369	3	3	0830a
-13	3	3	1974	1961	393	3	3	4362
-19	3	3	2021	2002	360	3	3	05339a
-10	3	3	1924	1914	365	3	3	
-187	3	3	2214	2027	381	3	3	7731
11	3	3	1915	1926	360	3	3	35451
-3	3	3	1936	1933	379	3	3	38675
-222	3	3	2296	2074	335	3	3	97136
-157	3	3	1982	1825	384	3	3	6130158
-26	3	3	2027	2001	361	3	3	2102326
-54	3	3	2036	1982	162	0	3	8804663
386	0	0	1623	2009	241	3	0	0a
274	0	0	1657	1931	218	3	0	0b
59	3	3	1863	1922	203	0	0	06a
299	0	0	1732	2031	230	3	0	007a
202	0	0	1665	1867	235	3	0	
261	0	0	1767	2028	265	3	0	026a
151	0	0	1780	1931	202	0	0	30
178	0	0	1757	1935	224	3	0	033a
219	0	0	1718	1937	194	0	0	106
176	0	0	1771	1947	230	3	0	0246a
154	0	0	1684	1838	316	3	0	409
315	0	0	1701	2016	303	3	0	00520a
198	0	0	1785	1983	268	3	0	
194	0	3	1847	2041	213	0	0	
229	0	0	1644	1873	219	3	0	720
255	0	0	1733	1988	296	3	0	0913a
37	3	0	1666	1703	214	3	0	1201
307	0	0	1747	2054	287	3	0	01598a
326	0	0	1628	1954	317	3	0	3800
190	0	0	1676	1866	216	3	0	
318	0	0	1656	1974	269	3	0	9001
312	0	0	1667	1979	284	3	0	
126	0	0	1775	1901	191	0	0	35640
386	0	0	1629	2015	343	3	0	41808
350	0	0	1661	2011	344	3	0	45042
295	0	0	1646	1941	262	3	0	99076
193	0	0	1659	1852	254	3	0	0414008a
148	0	0	1649	1797	266	3	0	

308	0	0	1621	1929	257	3	0	
379	0	0	1627	2006	243	3	0	0580750a
322	0	0	1675	1997	279	3	0	
288	0	0	1649	1937	198	0	0	
195	0	0	1678	1873	204	0	0	1019309
239	0	0	1690	1929	247	3	0	
256	0	0	1726	1982	274	3	0	2375206
314	0	0	1640	1954	381	3	0	4461084
230	0	0	1734	1964	301	3	0	6503008
288	0	0	1677	1965	222	3	0	
138	0	0	1743	1881	282	3	0	
250	0	0	1683	1933	314	3	0	7025788
359	0	0	1628	1987	302	3	0	7032535
219	0	0	1703	1922	220	3	0	8777100
121	0	0	1731	1852	174	0	0	
49	3	3	1851	1900	216	3	3	3a
228	0	0	1751	1979	251	3	3	3b
231	0	0	1714	1945	359	3	3	033a
99	0	0	1813	1912	219	3	3	
-86	3	3	2071	1985	406	3	3	34
132	0	3	1877	2009	363	3	3	739
263	0	0	1794	2057	423	3	3	356
189	0	0	1766	1955	436	3	3	30
-59	3	3	1878	1819	166	0	3	1863
194	0	0	1763	1957	278	3	3	813
295	0	0	1706	2001	332	3	3	832
127	0	3	1845	1972	266	3	3	0913a
-79	3	3	2011	1932	477	3	3	3800
65	3	3	2039	2104	564	3	3	31958
208	0	3	1827	2035	426	3	3	35640
15	3	3	1939	1954	376	3	3	49363
-8	3	3	1835	1827	238	3	3	
-55	3	3	1871	1816	348	3	3	1019309
150	0	3	1853	2003	402	3	3	2375206
74	3	0	1812	1886	432	3	3	7032535
214	0	0	1737	1951	370	3	3	
53	3	3	2023	2076	448	3	3	7317023
203	0	0	1756	1959	234	3	3	
15	3	0	1563	1578	215	3	0	000a
57	3	0	1508	1565	198	0	0	
-1	3	0	1609	1608	82	0	0	
8	3	0	1643	1651	138	0	0	0a
-14	3	0	1570	1556	132	0	0	0b
62	3	0	1546	1608	187	0	0	08a

14	3	0	1633	1647	93	0	0	10
-59	3	0	1778	1719	120	0	0	20
174	0	0	1520	1694	230	3	0	044a
112	0	0	1621	1733	109	0	0	60
78	3	0	1649	1727	231	3	0	205
-25	3	0	1696	1671	107	0	0	240
25	3	0	1645	1670	205	0	0	401
37	3	0	1590	1627	169	0	0	0477a
114	0	0	1571	1685	230	3	0	0740a
65	3	0	1713	1778	190	0	0	
465	0	0	1313	1778	190	0		804
42	3	0	1584	1626	212	0		02569a
-5	3	0	1648	1643	196	0		3064
85	3	0	1568	1653	204	0		03519a
2	3	0	1706	1708	194	0		05063a
121	0	0	1718	1839	181	0	0	
51	3	0	1696	1747	204	0	0	7001
101	0	0	1634	1735	200	0	0	
143	0	0	1709	1852	199	0	0	8088
76	3	0	1640	1716	201	0	0	09011a
127	0	0	1517	1644	172	0	0	
217	0	0	1600	1817	166	0	0	9805
173	0	0	1572	1745	193	0	0	27085
21	3	0	1629	1650	126	0	0	53430
67	3	0	1853	1920	214	3	0	69068
12	3	0	1615	1627	238	3	0	0222095a
34	3	0	1533	1567	222	3	0	
85	3	0	1561	1646	226	3	0	2920071
69	3	0	1677	1746	218	3	0	
0	3	0	1741	1741	207	0	0	4576035
291	0	0	1525	1816	215	3	0	5755074
132	0	0	1628	1760	164	0	0	6010063
6	3	0	1698	1704	166	0	0	
63	3	0	1575	1638	159	0	0	
57	3	0	1674	1731	205	0	0	7230526
59	3	0	1547	1606	173	0	0	9080009
54	3	0	1530	1584	179	0	0	
-93	3	0	1641	1548	195	0	0	
11	3	0	1581	1592	174	0	0	
214	0	0	1561	1775	214	3	0	9965030
26	3	0	1717	1743	124	0	0	
62	3	0	1754	1816	314	3	3	1347283
-134	3	0	1821	1687	111	0	3	
-212	3	3	1931	1719	249	3	3	3064

-68	3	0	1744	1676	322	3	3	3368
-236	3	3	1919	1683	286	3	3	
-98	3	0	1816	1718	133	0	3	3a
-131	3	3	1848	1717	134	0	3	3b
-205	3	3	1929	1724	265	3	3	32
-83	3	0	1737	1654	302	3	3	03519a
-135	3	0	1696	1561	121	0	3	05063a
-168	3	3	2048	1880	291	3	3	53430
51	3	0	1658	1709	322	3	3	
54	3	3	1839	1893	237	3	3	55386
41	3	0	1762	1803	321	3	3	1397283
-182	3	3	1861	1679	102	0	3	
-16	3	3	1849	1833	253	3	3	3215873
16	3	0	1777	1793	113	0	3	
5	3	0	1785	1790	268	3	3	3894484
-165	3	3	1837	1672	241	3	3	4576035
-130	3	0	1768	1638	107	0	3	6010063
73	3	0	1668	1741	278	3	3	7230526
-35	3	0	1756	1721	291	3	3	9965030
260	0	0	1722	1982	221	3	0	0a
219	0	0	1728	1947	228	3	0	0b
214	0	0	1732	1946	206	0	0	08a
318	0	0	1670	1988	230	3	0	021a
254	0	0	1685	1939	207	0	0	40
252	0	0	1688	1940	237	3	0	0044a
335	0	0	1606	1941	226	3	0	
367	0	0	1584	1951	233	3	0	056a
232	0	0	1735	1967	234	3	0	60
198	0	0	1762	1960	216	3	0	550
300	0	0	1657	1957	213	0	0	709
223	0	0	1681	1904	205	0	0	0734a
361	0	0	1611	1972	215	3	0	0860a
159	0	0	1751	1910	191	0	0	
219	0	0	1695	1914	230	3	0	02163a
239	0	0	1735	1974	268	3	0	3065
237	0	0	1749	1986	205	0	0	04131a
365	0	0	1649	2014	250	3	0	4406
194	0	0	1774	1968	214	3	0	6780
252	0	0	1731	1983	225	3	0	07027a
385	0	0	1630	2015	246	3	0	
252	0	0	1713	1965	227	3	0	07242a
261	0	0	1711	1972	254	3	0	8016
377	0	0	1582	1959	239	3	0	8092
283	0	0	1732	2015	237	3	0	10845

357	0	0	1667	2024	254	3	0	76089
316	0	0	1714	2030	268	3	0	2205397
302	0	0	1691	1993	215	3	0	3819770
285	0	0	1700	1985	255	3	0	4610020
258	0	0	1718	1976	252	3	0	
229	0	0	1707	1936	210	0	0	
365	0	0	1595	1960	266	3	0	6483305
266	0	0	1702	1968	253	3	0	6907098
294	0	0	1680	1974	259	3	0	
318	0	0	1660	1978	280	3	0	8230600
332	0	0	1673	2005	241	3	0	
220	0	0	1695	1915	218	3	0	
391	0	0	1570	1961	260	3	0	8820371
383	0	0	1622	2005	251	3	0	9503037
307	0	0	1695	2002	250	3	0	
192	0	0	1734	1926	238	3	0	9900126
236	0	0	1714	1950	247	3	0	
81	3	3	1931	2012	295	3	3	1347
31	3	3	2001	2032	306	3	3	3819770
70	3	3	1911	1981	316	3	3	32
-6	3	3	1942	1936	231	3	3	93
93	0	3	1946	2039	372	3	3	6483305
82	3	3	1915	1997	379	3	3	
1	3	3	1934	1935	229	3	3	3a
5	3	3	1923	1928	235	3	3	3b
4	3	3	1970	1974	214	3	3	173
1	3	3	1986	1987	323	3	3	393
14	3	3	1957	1971	233	3	3	
8	3	3	1944	1952	303	3	3	0734a
28	3	3	1970	1998	278	3	3	3065
-11	3	3	1972	1961	240	3	3	02163a
-4	3	3	2012	2008	327	3	3	04131a
15	3	3	2010	2025	366	3	3	5327
50	3	3	1920	1970	298	3	3	23685
-20	3	3	2012	1992	243	3	3	54183
18	3	3	1944	1962	262	3	3	92243
60	3	3	1957	2017	330	3	3	2205397
-51	3	3	2007	1956	370	3	3	8230600
46	3	3	1919	1965	333	3	3	8820371
34	3	3	1964	1998	340	3	3	9503037
62	3	3	1886	1948	307	3	3	
139	0	3	1847	1986	132	0	0	0a
119	0	3	1852	1971	138	0	0	0b
202	0	0	1693	1895	133	0	0	20

162	0	3	1831	1993	177	0	0	00022a
278	0	0	1716	1994	160	0	0	
319	0	0	1619	1938	194	0	0	
232	0	0	1704	1936	118	0	0	40
131	0	0	1778	1909	130	0	0	60
244	0	0	1737	1981	119	0	0	70
125	0	3	1861	1986	161	0	0	0087a
340	0	0	1640	1980	156	0	0	
107	0	0	1822	1929	130	0	0	90
193	0	3	1847	2040	173	0	0	99
256	0	0	1753	2009	154	0	0	349
258	0	0	1739	1997	141	0	0	506
149	0	0	1828	1977	131	0	0	510
302	0	0	1662	1964	170	0	0	704
115	0	3	1881	1996	152	0	0	806
49	3	3	1907	1956	165	0	0	02436a
52	3	3	1901	1953	115	0	0	4310
326	0	0	1619	1945	167	0	0	7075
237	0	0	1700	1937	126	0	0	7907
166	0	0	1713	1879	132	0	0	8409
212	0	0	1797	2009	158	0	0	10017
226	0	0	1759	1985	161	0	0	
226	0	0	1746	1972	169	0	0	20542
87	3	3	1888	1975	144	0	0	60155
96	0	3	1884	1980	153	0	0	0066774a
165	0	0	1792	1957	190	0	0	
109	0	3	1897	2006	146	0	0	67103
139	0	0	1771	1910	131	0	0	82509
260	0	0	1774	2034	159	0	0	0088944a
347	0	0	1638	1985	164	0	0	
127	0	0	1676	1803	122	0	0	91620
223	0	0	1743	1966	129	0	0	97801
213	0	0	1776	1989	166	0	0	0797335a
254	0	0	1736	1990	174	0	0	3708402
261	0	0	1655	1916	141	0	0	
262	0	0	1726	1988	166	0	0	4030125
179	0	0	1776	1955	195	0	0	
346	0	0	1624	1970	158	0	0	9630050
201	0	0	1726	1927	165	0	0	
81	3	3	1865	1946	92	0	0	
-93	3	3	2036	1943	121	0	3	283
-135	3	3	2130	1995	140	0	3	04613a
9	3	3	2029	2038	187	0	3	0797335a
50	3	3	1948	1998	144	0	3	

193	0	3	1828	2021	206	0	3	3399561
11	3	3	1931	1942	257	3	3	
159	0	3	1893	2052	189	0	3	32
-24	3	3	1991	1967	109	0	3	0213a
70	3	3	1910	1980	144	0	3	0349a
112	0	3	1911	2023	209	0	3	3708402
20	3	3	2016	2036	143	0	3	3a
0	3	3	1979	1979	132	0	3	3b
-20	3	3	2003	1983	151	0	3	02436a
167	0	3	1880	2047	197	0	3	3589
52	3	3	1904	1956	128	0	3	5873
-250	3	3	2193	1943	205	0	3	1198234
108	0	3	1880	1988	212	3	3	4030125
8	3	3	1952	1960	195	0	3	8653968
-38	3	3	1954	1916	195	0	3	9630050
47	3	3	1903	1950	133	0	0	0a
97	0	3	1884	1981	127	0	0	0b
15	3	3	1917	1932	100	0	0	30
73	3	0	1812	1885	145	0	0	033a
73	3	0	1749	1822	163	0	0	0075a
823	0	0	1442	2265	163	0	0	
253	0	0	1620	1873	150	0	0	083a
108	0	0	1820	1928	111	0	0	90
-164	3	3	1998	1834	155	0	0	106
158	0	0	1700	1858	143	0	0	309
108	0	0	1745	1853	98	0	0	620
212	0	0	1636	1848	157	0	0	0705a
144	0	3	1853	1997	123	0	0	
286	0	0	1631	1917	147	0	0	02933a
177	0	0	1686	1863	162	0	0	03082a
344	0	0	1681	2025	145	0	0	
64	3	0	1738	1802	152	0	0	3504
208	0	0	1725	1933	166	0	0	4105
152	0	0	1668	1820	156	0	0	9400
211	0	0	1677	1888	126	0	0	
412	0	3	1855	2267	166	0	0	2701
102	0	3	1876	1978	173	0	0	20404
306	0	0	1785	2091	142	0	0	
225	0	0	1622	1847	168	0	0	40102
230	0	0	1572	1802	134	0	0	
229	0	3	2058	2287	158	0	0	42045
382	0	0	1667	2049	148	0	0	90509
41	3	0	1775	1816	142	0	0	
211	0	0	1643	1854	166	0	0	0166607a

بهبود سیستم بازشناسی ارقام تلفنی با استفاده از پارامترهای اکوستیکی آواهای زبان فارسی ۹۰ £

141	0	0	1819	1960	154	0	0	
162	0	0	1722	1884	144	0	0	0327645a
181	0	0	1682	1863	145	0	0	0397100a
297	0	0	1740	2037	153	0	0	
107	0	0	1821	1928	110	0	0	
148	0	0	1629	1777	146	0	0	0806255a
58	3	0	1763	1821	133	0	0	
87	3	0	1752	1839	168	0	0	2006957
216	0	0	1622	1838	173	0	0	
4	3	3	1910	1914	80	0	0	6185890
163	0	0	1685	1848	144	0	0	7026784
169	0	0	1625	1794	135	0	0	7240012
187	0	3	1845	2032	149	0	0	
189	0	0	1647	1836	131	0	0	8098968
133	0	0	1670	1803	156	0	0	8808483
176	0	0	1689	1865	135	0	3	309
53	3	3	1885	1938	187	0	3	34187
-132	3	3	2079	1947	225	3	3	8638
129	0	3	1828	1957	191	0	3	3504
223	0	3	1858	2081	194	0	3	0327645a
160	0	3	1894	2054	193	0	3	0397100a
-39	3	3	2011	1972	94	0	3	8808483
-66	3	3	2037	1971	132	0	3	3a
-45	3	3	1967	1922	138	0	3	3b
79	3	3	1868	1947	214	3	3	30
179	0	3	1811	1990	195	0	3	033a
-103	3	3	2081	1978	109	0	3	
228	0	0	1736	1964	196	0	3	37
-87	3	3	2159	2072	98	0	3	083a
-22	3	3	2047	2025	95	0	3	73
-158	3	3	2082	1924	176	0	3	234
-18	3	3	2047	2029	109	0	3	593
410	0	3	1878	2288	215	3	3	02933a
-28	3	3	2060	2032	77	0	3	
130	0	3	1857	1987	264	3	3	03082a
248	0	0	1699	1947	219	3	3	13749
-4	3	3	1958	1954	94	0	3	28263
0	3	3	2023	2023	104	0	3	78143
318	0	0	1665	1983	156	0	0	00a
153	0	3	2067	2220	118	0	0	
55	3	3	2135	2190	141	0	0	0a
222	0	3	2041	2263	138	0	0	0b
495	0	0	1673	2168	147	0	0	014a
419	0	0	1728	2147	129	0	0	0060a

317	0	3	1842	2159	134	0	0	
331	0	3	1893	2224	146	0	0	
323	0	3	2036	2359	159	0	0	60
533	0	0	1739	2272	128	0	0	074a
423	0	0	1716	2139	193	0	0	120
555	0	0	1518	2073	134	0	0	0239a
620	0	0	1667	2287	148	0	0	504
592	0	0	1810	2402	134	0	0	0612a
166	0	3	2017	2183	152	0	0	630
363	0	3	1931	2294	158	0	0	830
625	0	0	1578	2203	171	0	0	3100
168	0	3	1891	2059	151	0	0	
658	0	0	1721	2379	141	0	0	0003508a
365	0	0	1691	2056	130	0	0	
461	0	0	1665	2126	159	0	0	
515	0	0	1617	2132	149	0	0	
523	0	0	1653	2176	114	0	0	9051
476	0	0	1656	2132	155	0	0	9709
573	0	0	1682	2255	130	0	0	10315
440	0	0	1779	2219	137	0	0	10671
570	0	0	1638	2208	163	0	0	29007
320	0	0	1566	1886	160	0	0	
314	0	0	1704	2018	151	0	0	47240
607	0	0	1678	2285	145	0	0	52034
676	0	0	1592	2268	131	0	0	70422
295	0	3	1920	2215	141	0	0	88020
114	0	3	2107	2221	145	0	0	
628	0	0	1783	2411	136	0	0	0101892a
556	0	0	1640	2196	145	0	0	
112	0	3	1881	1993	131	0	0	0305544a
476	0	0	1712	2188	120	0	0	
202	0	0	1778	1980	137	0	0	0400818a
348	0	0	1658	2006	137	0	0	
349	0	0	1702	2051	131	0	0	
481	0	0	1606	2087	142	0	0	2790938
522	0	0	1805	2327	138	0	0	3040747
347	0	0	1718	2065	130	0	0	
47	3	3	2012	2059	141	0	0	5914260
266	0	3	1957	2223	140	0	0	7346440
543	0	0	1653	2196	155	0	0	7987708
-181	3	3	2188	2007	211	0	3	8633
-217	3	3	2406	2189	130	0	3	
131	0	3	2213	2344	184	0	3	3a
-120	3	3	2332	2212	212	3	3	3b

-348	3	3	2090	1742	205	0	3	13
-349	3	3	2244	1895	189	0	3	36
224	0	3	1893	2117	226	3	3	0239a
-675	3	3	2450	1775	185	0	3	365
196	0	3	1896	2092	266	3	3	630
229	0	0	1472	1701	136	0	0	0a
203	0	0	1446	1649	134	0	0	0b
100	0	0	1472	1572	142	0	0	20
241	0	0	1377	1618	122	0	0	40
328	0	0	1341	1669	146	0	0	099a
279	0	0	1277	1556	146	0	0	230
303	0	0	1362	1665	133	0	0	00388a
314	0	0	1349	1663	142	0	0	
377	0	0	1322	1699	123	0	0	0370a
195	0	0	1394	1589	134	0	0	
235	0	0	1466	1701	114	0	0	604
207	0	0	1400	1607	122	0	0	801
188	0	0	1339	1527	128	0	0	870
239	0	0	1516	1755	131	0	0	2210
357	0	0	1326	1683	144	0	0	0002493a
351	0	0	1359	1710	150	0	0	
344	0	0	1393	1737	135	0	0	
522	0	0	1082	1604	131	0	0	4340
146	0	0	1444	1590	120	0	0	04720a
92	3	0	1514	1606	131	0	0	
300	0	0	1326	1626	111	0	0	05002a
289	0	0	1369	1658	128	0	0	
502	0	0	1322	1824	126	0	0	
246	0	0	1309	1555	113	0	0	7050
238	0	0	1452	1690	123	0	0	
288	0	0	1373	1661	136	0	0	7730
329	0	0	1360	1689	131	0	0	9033
228	0	0	1492	1720	138	0	0	20600
221	0	0	1438	1659	129	0	0	
121	0	0	1470	1591	121	0	0	
321	0	0	1281	1602	152	0	0	56071
109	0	0	1374	1483	128	0	0	57407
229	0	0	1328	1557	106	0	0	60523
119	0	0	1343	1462	123	0	0	86940
388	0	0	1225	1613	114	0	0	0576800a
369	0	0	1323	1692	117	0	0	
108	0	0	1488	1596	131	0	0	
252	0	0	1379	1631	124	0	0	0906820a
317	0	0	1353	1670	152	0	0	

177	0	0	1443	1620	132	0	0	
353	0	0	1340	1693	111	0	0	0915445a
280	0	0	1359	1639	115	0	0	0975536a
262	0	0	1301	1563	134	0	0	1034164
169	0	0	1423	1592	133	0	0	4270674
297	0	0	1383	1680	136	0	0	9004954
413	0	0	1269	1682	150	0	0	
329	0	0	1372	1701	117	0	0	9014798
-21	3	0	1745	1724	171	0	3	53286
-442	3	3	2109	1667	159	0	3	3141
-121	3	0	1786	1665	158	0	3	0002493a
-109	3	0	1686	1577	156	0	3	1034164
-237	3	3	1916	1679	179	0	3	3a
-44	3	0	1739	1695	168	0	3	3b
91	3	0	1483	1574	151	0	3	39
-81	3	0	1592	1511	165	0	3	63
-432	3	3	2090	1658	159	0	3	313
-86	3	0	1664	1578	165	0	3	
189	0	0	1494	1683	130	0	3	03070a
110	0	0	1407	1517	158	0	3	1132
-67	3	0	1685	1618	182	0	3	4340
-36	3	0	1654	1618	204	0	3	9033
-98	3	0	1614	1516	168	0	3	
-166	3	3	1922	1756	162	0	3	0975536a
141	0	0	1451	1592	127	0	3	1221735
71	3	0	1607	1678	162	0	3	00338a
85	3	0	1571	1656	148	0	3	
222	0	3	1831	2053	172	0	0	00a
267	0	0	1713	1980	152	0	0	
243	0	0	1773	2016	187	0	0	0a
183	0	3	1840	2023	174	0	0	0b
152	0	3	1922	2074	166	0	0	04a
127	0	3	1842	1969	195	0	0	0006a
-11	3	3	1834	1823	173	0	0	
175	0	0	1775	1950	169	0	0	
71	3	3	1942	2013	161	0	0	06a
220	0	0	1754	1974	193	0	0	030a
282	0	0	1771	2053	179	0	0	
230	0	0	1789	2019	181	0	0	085a
173	0	3	1835	2008	152	0	0	280
4	3	3	1887	1891	190	0	0	405
236	0	0	1724	1960	161	0	0	0718a
157	0	3	1901	2058	191	0	0	90
238	0	0	1805	2043	164	0	0	909

241	0	0	1744	1985	214	3	0	4066
329	0	0	1687	2016	169	0	0	05818a
254	0	0	1794	2048	166	0	0	7803
269	0	0	1665	1934	175	0	0	8008
108	0	0	1818	1926	153	0	0	
270	0	0	1762	2032	155	0	0	9303
106	0	3	1866	1972	167	0	0	09766a
173	0	3	1860	2033	173	0	0	37710
270	0	0	1752	2022	175	0	0	44024
316	0	0	1716	2032	181	0	0	49705
234	0	0	1707	1941	160	0	0	0103572a
297	0	0	1745	2042	172	0	0	
206	0	0	1768	1974	171	0	0	1095500
267	0	0	1783	2050	184	0	0	
154	0	3	1874	2028	144	0	0	
257	0	0	1761	2018	197	0	0	3010904
226	0	3	1838	2064	174	0	0	
414	0	0	1642	2056	167	0	0	
268	0	0	1813	2081	170	0	0	3560623
77	3	3	1940	2017	174	0	0	5960137
-67	3	3	1969	1902	193	0	0	6083443
421	0	0	1652	2073	177	0	0	6907070
213	0	3	1836	2049	154	0	0	
367	0	0	1681	2048	176	0	0	
155	0	3	1872	2027	187	0	0	7820845
157	0	3	1913	2070	189	0	0	9165500
166	0	3	1831	1997	172	0	0	
-18	3	3	2011	1993	192	0	3	263
44	3	3	2076	2120	241	3	3	6083443
108	0	3	1929	2037	151	0	3	030a
7	3	3	2045	2052	227	3	3	3b
21	3	3	2064	2085	210	0	3	3a
12	3	3	1969	1981	170	0	3	23
-217	3	3	2106	1889	204	0	3	93
9	3	3	2023	2032	246	3	3	7394
27	3	3	2028	2055	208	0	3	7803
29	3	3	1944	1973	256	3	3	9303
77	3	3	1944	2021	215	3	3	
31	3	3	2007	2038	259	3	3	37710
22	3	3	2081	2103	232	3	3	94399
41	3	3	1852	1893	178	0	3	103572
131	0	3	1892	2023	207	0	3	3010904
187	0	0	1732	1919	186	0	3	3560623
-18	3	3	2049	2031	201	0	3	

85	3	3	2041	2126	243	3	3	5960137
251	0	0	1703	1954	220	3	0	8603035
137	0	0	1747	1884	217	3	0	
81	3	3	1869	1950	216	3	0	8529094
127	0	0	1764	1891	151	0	0	0a
106	0	0	1773	1879	210	0	0	00a
34	3	0	1800	1834	128	0	0	
79	3	0	1816	1895	200	0	0	02a
109	0	0	1769	1878	126	0	0	570
136	0	0	1754	1890	213	0	0	078a
185	0	0	1731	1916	212	0	0	5605
204	0	0	1767	1971	246	3	0	3300931
176	0	0	1797	1973	245	3	0	
-192	3	3	1992	1800	139	0	0	25104
187	0	0	1722	1909	210	0	0	01680a
13	3	3	1872	1885	124	0	0	
42	3	0	1826	1868	132	0	0	920
438	0	0	1509	1947	247	3	0	9506047
165	0	0	1801	1966	222	3	0	
212	0	0	1729	1941	217	3	0	8022
206	0	0	1733	1939	217	3	0	049a
234	0	0	1681	1915	159	0	0	0b
255	0	0	1665	1920	265	3	0	200
79	3	3	1902	1981	144	0	0	
131	0	0	1742	1873	176	0	0	08a
330	0	0	1697	2027	206	0	0	69501
144	0	0	1820	1964	218	3	0	81046
215	0	0	1762	1977	213	0	0	0302800a
285	0	0	1714	1999	175	0	0	
314	0	0	1645	1959	192	0	0	
28	3	0	1818	1846	129	0	0	
173	0	0	1736	1909	217	3	0	020a
53	3	0	1823	1876	127	0	0	
128	0	0	1758	1886	215	3	0	078a
22	3	3	1896	1918	134	0	0	270
229	0	0	1743	1972	197	0	0	0795a
147	0	0	1747	1894	212	0	0	0965a
201	0	0	1687	1888	211	0	0	9064
150	0	0	1770	1920	249	3	0	29083
94	0	3	1901	1995	240	3	0	37055
-18	3	3	1921	1903	140	0	0	82107
92	3	0	1795	1887	265	3	0	1036184
38	3	3	1898	1936	145	0	0	6632820
201	0	0	1716	1917	243	3	0	2359005

98	0	0	1826	1924	224	3	0	
-137	3	3	2092	1955	320	3	3	3300931
-14	3	3	2012	1998	279	3	3	
-59	3	3	2067	2008	296	3	3	
58	3	3	1945	2003	175	0	3	3a
-26	3	3	1949	1923	182	0	3	3b
-179	3	3	2033	1854	169	0	3	13
-78	3	3	2063	1985	332	3	3	336
-169	3	3	2155	1986	337	3	3	
-84	3	3	2016	1932	178	0	3	423
-20	3	3	2057	2037	288	3	3	1737
-30	3	3	2025	1995	314	3	3	3873
-1	3	3	1968	1967	153	0	3	
-50	3	3	2017	1967	150	0	3	29083
-51	3	3	2084	2033	309	3	3	37055
-37	3	3	1972	1935	288	3	3	302800
-162	3	3	2138	1976	350	3	3	1306184
-37	3	3	1949	1912	553	3	3	2359005
-67	3	3	2065	1998	330	3	3	3998414
-122	3	3	2130	2008	286	3	3	6632820
-123	3	3	2092	1969	305	3	3	8603035
164	0	3	1904	2068	263	3	3	
-53	3	3	1988	1935	322	3	3	9434991