فصلنامهٔ علمی - پژوهشی زبان‌پژوهی دانشگاه الزهرا(س)
سال دهم، شمارهٔ ۲۹، زمستان ۱۳۹۷

بازشناسی تمایز خیشومی-انسدادی در شرایط نامطلوب شنیداری[۱]

زهرا محمودزاده[۲]

تاریخ دریافت: ۱۳۹۵/۱۲/۲۲
تاریخ پذیرش: ۱۳۹۶/۲/۱۷

چکیده

برای بهبودِ بازشناسیِ خودکارِ تمایزهای واجی، می‌توان از سرنخ‌های درکی که شنوندگان برای بازشناسی طبیعی، آن‌ها را در موقعیت‌های نامطلوب شنیداری مانند گفتار تلفنی یا نوفهٔ محیط به‌کار می‌برند، استفاده نمود. در این پژوهش، برای یافتن سرنخ‌های درکیِ مؤثر در بازشناسیِ طبیعیِ تمایزِ خیشومی-انسدادی در جفت‌رقم‌های «دو-نه» [noh]-[do] در شرایطِ نامطلوبِ گفتارِ تلفنی، از آزمون‌های درکی استفاده شد. بررسی سیگنال آکوستیکیِ [no] نشان می‌دهد که سرنخ‌های گذر واکه‌ای و زمزمهٔ خیشومی، تحتِ تأثیر عوامل مختلفِ اختلالِ سیگنال قرار گرفته، از جنبهٔ آکوستیکی کاهش پیدا می‌کنند و سببِ ابهام در درک خیشومی [n] می‌شوند. در شرایط مطلوبِ شنیداری، دقت بازشناسی طبیعی [n] بر اساس تنها پارامتر زمزمهٔ خیشومی، درحدود ۴۰ درصد است. اما با افزودن ۱۰ میلی‌ثانیه از ابتدای گذرهای واکه‌ای به آن، دقت بازشناسی به ۹۶ درصد افزایش می‌یابد. در گفتار تلفنی، دقت بازشناسیِ طبیعی براساس زمزمهٔ

---

[۱] شناسه دیجیتال (DOI): 10.22051/jlr.2017.14294.1293

[۲] دکترای تخصصی، استادیار گروه زبان‌شناسی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایران‌داک)؛
mahmoodzadeh@irandoc.ac.ir

خیشومی ۲۹ درصد و براساس هر دو پارامتر، فقط ۴۸ درصد است. به نظر می‌رسد عدمِ قطعیت و ابهامِ واژگانی شنونده، به دلیل حذف یا کاهش اطلاعات آوایی در شرایط نامطلوب شنیداری از یک سو و گرایش درکی شنونده به سمت همخوان بی‌نشان [d] از سوی دیگر منجر به کاهش بازشناسی خیشومی [n] شده‌است. براساس یافته‌های پژوهش، پارامتر آکوستیکی زمزمهٔ خیشومی، نقش مؤثری در بازشناسیِ طبیعیِ خیشومی در شرایط مطلوب یا نامطلوبِ شنیداری ندارد و فقط با افزودن اطلاعاتِ گذرهای سازه‌ای است که بازشناسی به طور معناداری افزایش می‌یابد. بنابراین، برای بازشناسی خودکار این واژه‌ها، لازم است اطلاعات زمانی و طیفی واکه‌های مجاور و گذرهای آن‌ها به‌کار گرفته شود.

واژه‌های کلیدی: شرایط نامطلوب شنیداری، ارقام فارسی، آزمایش درکی، تمایز خیشومی- انسدادی، گذر واکه‌ای

۱. مقدمه

در زندگی روزمره، شنوندگان، گفتار را در گسترهٔ وسیعی از شرایط نامطلوب یا پائین‌تر از مطلوب بازشناسی می‌کنند. شرایط نامطلوب، هر عاملی را شامل می‌شود که سبب کاهش فهم[1] گفتار، نسبت به سطح فهم آن در شرایط مطلوبِ شنیداری شود. شرایطِ مطلوبِ شنیداری، وضعیتی است که در آن شنوندگانِ بومیِ هر زبان، با سلامتِ کاملِ شنوایی در محیطی ساکت و با تمرکز کامل به گفتار ضبط شدهٔ آزمایشگاهی گوش می‌دهند (Mattys et al., 2012, p. 953). بنابراین مطالعهٔ پارامترها و عوامل مؤثر در درک شنیداریِ انسان در شرایط گفتاری متفاوت، می‌تواند به یافتن سرنخ‌ها و اطلاعات آکوستیکی مؤثر در فرایند بازشناسی خودکارِ واژه، یاری رساند. در فرایندِ بازشناسیِ خودکارِ واژه، برنامهٔ بازشناسی آواهای گفتاری ضبط‌شده را به یکی از گزینه‌های واژگانی موجود در فهرست واژگان می‌نگارد.

بایبی (Bybee, 2001) در نظریهٔ واج‌شناسی کاربردبنیاد[2] خود، فرکانس کاربرد واژه را عامل وقوع فرایندهای آوایی مانند کاهش و حذف می‌داند؛ به این معنا که تضعیف مشخصه‌های آوایی یا حذف واحدهای واجی در واژه‌های پربسامد رایج‌تر است. رقم‌های فارسی، از صفر تا نه، از جمله پرکاربردترین واژه‌های گفتاری به شمار می‌آیند. بررسی سیگنال آکوستیکی رقم‌های فارسی

[1] intelligibility
[2] usage-based phonology

نیز نشان می‌دهد این واژه‌ها با درجات متفاوتی از کاهش اطلاعات آوایی یا حذف واحدهای واجی تولید می‌شوند. برای نمونه، پژوهش محمودزاده و بحرانی (Mahmoodzadeh & Bahrani, 2014, p. 111-134) نشان می‌دهد که سیگنال آکوستیکی «صفر» [sefr] در بسیاری از نمونه‌ها با درجه‌های مختلفی از کاهشِ مشخصه‌های آوایی، در خوشهٔ همخوانی پایانی و یا با حذف واحد واجی /r/ همراه است. از آن‌جایی‌که این تغییرات آوایی معنادار هستند، اغلب سبب ایجاد ابهام در سطح انتخاب واژگانی، بین جفت‌واژه‌های [se]-[sefr] در سیستم‌های بازشناسی خودکار می‌شود.

بازشناسی خودکارِ رقم‌ها و عددهای فارسی دامنهٔ کاربرد وسیعی در نهادهای تجاری و اداری گوناگون دارد و در نتیجه میزان دقت آن از اهمیت بسیاری برخوردار است. به دلیل اهمیتِ بازشناسیِ خودکارِ ارقام، این مقاله به بررسی عوامل ابهام و کاهش میزان دقت در بازشناسی طبیعی جفت‌های «دو- نُه» [noh]-[do] می‌پردازد. همچنین پژوهش حاضر، با طراحی آزمون‌های درکی در شرایط مطلوب و نامطلوب شنیداری سعی در یافتن پارامترهای مؤثر در تشخیص تمایز اشاره‌شده دارد. از آن‌جایی‌که اطلاعات بافتی، از نوع نحوی-معنایی کمکی به بازشناسی رقم‌ها نمی‌کند، بنابراین بازشناسی آن‌ها فقط وابسته به اطلاعات آوایی و آکوستیکی است. یافته‌های این پژوهش می‌تواند برای افزایش دقتِ بازشناسی خودکار [noh]-[do] در شرایط مطلوب یا نامطلوب گفتاری بسیار کاربرد داشته باشد.

۱. ۱. شرایط نامطلوب شنیداری و اثر آنها بر درک گفتار

همان‌طور که در بالا گفته شد، وضعیت مطلوب شنیداری هنگامی رخ می‌دهد که شنوندهٔ بومی یک زبان، به گفتار تولیدشده در آزمایشگاه گوش می‌دهد. روشن است که بازشناسی در شرایط آزمایشگاهی با دقت بسیار بالایی انجام می‌شود. در حالی‌که وضعیت نامطلوب شنیداری، ناشی از عامل یا عواملی است که سبب کاهش فهم گفتار نسبت به شرایط مطلوب می‌شوند. متیز و همکاران (Mattys et al., 2012, p. 954-959) شرایط نامطلوب را بر پایهٔ منشأ و اثر آن به دو دسته گروه‌بندی می‌کنند. منظور از منشأ، کانون یا علت ایجاد وضعیت نامطلوب شنیداری است. منشأ اختلال، می‌تواند بیرون یا درونِ شنونده باشد. از جمله عوامل درونی اختلال، محدودیت‌های شنونده مانند آگاه نبودن از دانش زبان بومی، نقص شنوایی، کاشت حلزون شنوایی و نداشتنِ تمرکز است. از جمله عوامل بیرونی اختلال می‌تواند تولید غیرعادی یک گویشور، صداها یا نوفه‌های پس‌زمینه باشد. گفتار محاوره، خود می‌تواند به عنوان کانون ایجاد اختلال در نظر گرفته شود، زیرا در مقایسه با صورت رسمی و گفتار خواندنی[1] فهم گفتار در آن کاهش پیدا می‌کند. از

[1] read speech

جمله ویژگی‌های گفتار محاوره، تغییرات آوایی مانند کاهش یا حذف آوا، حذف هجا، همگونی، هم‌تولیدی و نیز متغیر بودن سرعت گفتار است. بنابراین در گفتار محاوره، اطلاعات زمانی حذف یا کاهش می‌یابند. یکی دیگر از عوامل بیرونی ایجاد اختلال، نقص در کانال انتقال سیگنال بین گوینده و شنونده است. در این صورت افت انتقال می‌تواند به دلایل گوناگونی، مانند فاصلهٔ بین گوینده و شنونده پدیدار شود. این عامل سبب انتقال بهتر مؤلفه‌های با شدت انرژی بالاتر، نسبت به مؤلفه‌های کم‌شدت‌تر می‌شود. اما اغلب، افت انتقال ناشی از وجود سیگنال‌های رقیب در محیط – به عنوان مثال نوفه یا صداهای پس‌زمینه– است یا بر اثر اعوجاج آکوستیکی[1] و تغییر شکل سیگنال در محیط فیزیکی یا کانال انتقال روی می‌دهد. برای نمونه، فیلتر گفتار در تلفن می‌تواند سبب اعوجاج آکوستیکی و تغییر پارامترهای شنیداری سیگنال شود. در گفت‌وگوی تلفنی، معمولاً فرکانس‌های پائین‌تر از ۴۰۰ هرتز و بالاتر از ۳۴۰۰ هرتز فیلتر می‌شود. در حالی‌که محدودهٔ فرکانسی اطلاعات گفتاری انسان بین ۱۰۰ تا ۵۰۰۰ هرتز است. بنابراین در گفت‌وگوی تلفنی اطلاعات طیفی تغییر، تضعیف و یا پنهان می‌شوند. بر اساس آن‌چه گفته شد، عوامل بیرونی اختلال می‌توانند سببِ کاهشِ بازشناسیِ طبیعی و نیز خودکار گفتار شوند (همان).

شرایط نامطلوب برای بازشناسی گفتار را می‌توان براساس تأثیری که بر نظام درک گفتار دارند و مکانیسم‌های جبران‌کنندهٔ شناختی یا عصبی که برمی‌انگیزانند، نیز تقسیم‌بندی نمود. توقف بازشناسی، تداخل درکی[2]، کاهش توانایی تمرکز و کاهش ظرفیت حافظه را می‌توان از جمله تأثیراتِ شرایط نامطلوب بر درک گفتار به شمار آورد. توقف بازشناسی یا عدم نگاشتِ مشخصه‌های آکوستیکی–آوایی به بازنمایی واجی و به دنبال آن، بازنمایی واجی به واژگانی از جمله پیامدهای رایج بسیاری از شرایط نامطلوب شنیداری است. توقف بازشناسی، می‌تواند ناشی از افت اطلاعات در حوزه‌های زمان، شدت (ناشی از فاصله) و طیف (در انتقال تلفنی یا کاشت حلزون) باشد و یا نتیجهٔ انحراف[3] آکوستیکی– آوایی از ویژگی‌های آوایی هدف مانند نتیجهٔ اختلالات گفتاری، تلفظ اشتباهی، گفتار لهجه‌دار، گفتار محاوره و موارد مشابه باشد. پیامد همهٔ این موارد می‌تواند عدمِ قطعیت و ابهام واژگانی، انتخاب واژهٔ نادرست و یا در نهایت انتخاب‌نکردن واژه باشد. اینکه چگونه سیستم گفتاری با افت اطلاعات مواجه می‌شود، به ماهیت و مقدار افت بستگی دارد. شنوندگان معمولاً با استفاده از حشو زمانی سیگنال که نتیجهٔ هم‌تولیدی و حافظهٔ پژواکی[4] است، وقفه‌های کوتاه اختلال در سیگنال را به خوبی جبران می‌کنند. اطلاعات نیز

---

[1] acoustic distortion
[2] perceptual interference
[3] deviation
[4] echoic memory

در شرایطی که حذف واحد آوایی به دلیل اطلاعات غنی بافتی مجاز است، به آسانی بازسازی می‌شود، همان‌گونه‌که اغلب در گفتار محاوره رخ می‌دهد (Ernestus et al., 2002).

تأثیر ناشی از انحراف آکوستیکی-آوایی هنگام تولید صدای هدف، به بسامد کاربرد واژه، میزان ابهام‌زایی واژهٔ مورد نظر (به بیانی میزان تراکم واژگانی حول واژه) و نیز به درجهٔ حمایت بافتی واژه بستگی دارد. در شرایط مساوی، درک واحدهای زبانی که حشو، قابل‌پیش‌بینی و بدون ابهام هستند، کمتر تحت تأثیر افت سیگنال قرار می‌گیرد. در واقع، اگر بازشناسی به کمک بافت باشد، تغییر بنیادی در مکانیسم‌های درکی ایجاد نمی‌شود، بلکه اثرگذاری بافت در شرایط نامطلوبی که سینگال افت می‌کند، بیشتر می‌شود. برای نمونه، وقتی به دلیل نوفه یا ابهام در سیگنال امکان چندین تعبیر وجود دارد، تأثیر دانش معنایی واژگانی و جمله بسیار بیشتر است. تداخل درکی، اثرِ دیگر شرایط نامطلوب است. تداخل درکی، زمانی روی می‌دهد که یک سیگنال غیرهدف مانند نوفه، رقیب سیگنال گفتاری باشد. تداخل می‌تواند در سطح پایینی روی دهد، مانند زمانی‌که سیگنال رقیب، فقط بخش‌هایی از سیگنال هدف را پنهان می‌کند یا در سطح بالاتری مانند هنگامی‌که محتوای سیگنال رقیب بر تعبیر سیگنال هدف تأثیر می‌گذارد و یا توجه شنونده را از آن سلب می‌کند. به طور کلی، سیستم شنیداری پنهان‌سازی سیگنال را همانند نبودِ اطلاعات یا انحراف آکوستیکی-آوایی (توقف بازشناسی) کنترل می‌کند، به استثنای مواقعی که حضور سیگنال رقیب، سبب ایجاد چالش‌های تفکیک سیگنال از نوفه و توجه انتخابی[1] شود. فهم گفتار در شرایط نوفه به میزان هم‌پوشانی طیفی بین گفتار و نوفه و نیز مدت زمان و ماهیت تکه‌های[2] گفتار بستگی دارد (Mattys et al., 2012, p. 954-959).

در شرایط نامطلوب و پُرزحمت شنیداری، علاوه بر فعال شدن برخی فرایندهای درکی برای فهم بهینهٔ ابهام سیگنال، برخی فرایندها و مکانیسم‌های جبران‌کننده مانند فرایندهای سازگاری[3] و یادگیری درکی[4] نیز برای تنظیم نظام درکی با شرایط نامطلوب به کار گرفته می‌شوند. نظام زبانی انسان می‌تواند فعالانه با هر گونه تغییر در نمودِ آکوستیکی سیگنال، سازگار شود و نظام درکی را به نحوی تنظیم کند تا سیگنال ورودی را به بهینه‌ترین شکل ممکن پردازش کند. فرایند سازگاری می‌تواند بسیار سریع رخ دهد. برای نمونه، سازگاری با تغییر طبیعی در سرعت گفتار یا با تغییر مشخصه‌های طیفی ناشی از کانال انتقال در کمتر از یک ثانیه روی می‌دهد. این تنظیمات نسبتاً به صورت خودکار و بیشتر بی‌آنکه شنونده از پیامدهای درکی آنها آگاه باشد، روی می‌دهند

---

[1] selective attention
[2] glimpses of speech
[3] adaptation
[4] perceptual learning

(Davis et al., 2005). یادگیری درکی، یکی دیگر از مکانیسم‌های جبران‌کنندهٔ نظام درکی است. گلدستون (Goldstone, 1998, p. 586) آن را به مثابه «تغییرات نسبتاً درازمدت در نظام درکی یک موجود زنده می‌داند به نحوی که سبب افزایش توانایی آن در پاسخ به محیط شده و به وسیلهٔ محیط نیز ایجاد می‌شود». آزمایش‌های گوناگون نشان می‌دهند که فرایندهای یادگیری درکی در بیشتر موقعیت‌هایی که آواهای گفتاری خارج از محدودهٔ تجربهٔ شنونده هستند، برانگیخته می‌شوند. برای نمونه، در مواجهه با گفتارِ به‌شدت لهجه‌دار، شنوندگان باید با مجموعه‌ای از ویژگی‌های ناآشنای آوایی و نوایی سازگار شوند. پژوهش‌های پیشین نشان می‌دهند درک مؤثر گفتار در مواجهه با یک لهجهٔ ناآشنا به چند دقیقه و یا بیشتر نیازمند است (Mattys et al., 2012, p. 954-959).

۲. پارامترهای آکوستیکی تمایز [n-d]

ثمره (Samareh, 2006) در زبان فارسی، محل تولید خیشومی [n] را لثوی و [d] را دندانی توصیف می‌کند. همخوان‌های خیشومی /m/ یا /n/ با پایین آوردن نرم‌کام و جفت‌شدگی آکوستیکی حفرهٔ بینی با دستگاه گفتار و همزمان ایجاد یک انسداد کامل در حفرهٔ دهان تولید می‌شوند. انسداد دهانی خیشومی‌ها همانند انسداد همخوان‌های انسدادی ایجاد می‌شود. این در حالی است که باز شدن مسیر جریان هوا در مسیر بینی و جفت‌شدگی بینی با دستگاه گفتار، سببِ بروزِ ویژگی‌های خاص خیشومی‌ها می‌شود که در طیف زمزمهٔ خیشومی[1] نمایانده می‌شود. این وضعیتِ تولیدی، سببِ ایجاد ویژگی‌های خاصِ خیشومی‌ها از جمله تولید صفر یا پادسازه[2] در طیف، تمرکز انرژی روی فرکانس‌های پایین، تراکم زیاد سازه‌ها، افت بالای انرژی و شکل‌گیری طیفیِ ایستا می‌شود. تفاوت در اندازه و شکل حفره‌های بینی و سینوسی نیز می‌تواند به تفاوت‌های معناداری در مشخصه‌های طیفی زمزمه‌های خیشومی در گویشوران متفاوت منتهی شود (Pruthi et al., 2004).

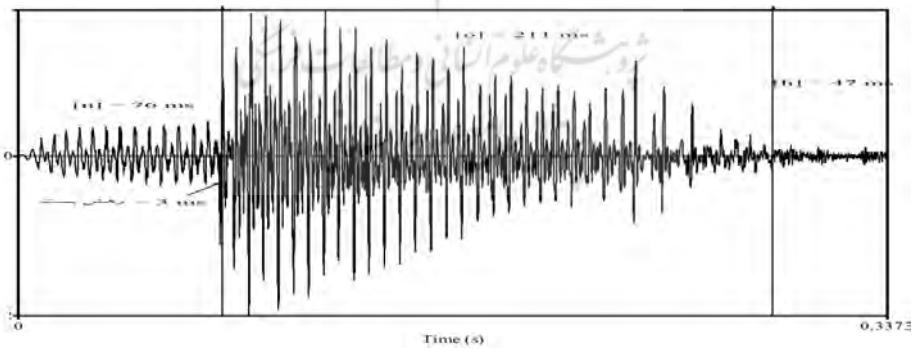
جفت‌شدگی حفره‌های دهانی و بینی در هنگامِ زمزمهٔ خیشومی سبب تولید اولین سازه در فرکانس‌های بسیار پائین می‌شود. از سویی، بست دهانی منجر به ایجاد پادسازه در طیف زمزمهٔ خیشومی می‌شود. این پادسازه‌ها معمولاً بیشتر در فرکانس‌های بالای ۱۰۰۰ هرتز تولید می‌شوند. تولید قلهٔ انرژی در فرکانس پایین به همراه پادسازه‌ها منجر به یک افت معنادار انرژی در طیف زمزمهٔ خیشومی، بی‌وقفه در بالای ناحیهٔ اولین سازه می‌شود. از آن‌جایی که عضو متحرکی در حفرهٔ

---

[1] nasal murmur
[2] anti-formant

بینی وجود ندارد، انتظار می‌رود طیف زمزمهٔ خیشومی تاحدی ثابت باشد. این وضعیت ایستا در نواحی فرکانس‌پایینِ طیف، مشاهده می‌شود، هر چند که ممکن است حرکاتی در فرکانس‌های بالاتر به دلیل تغییر جای قله‌ها و صفرها وجود داشته باشد (Fujimura, 1962).

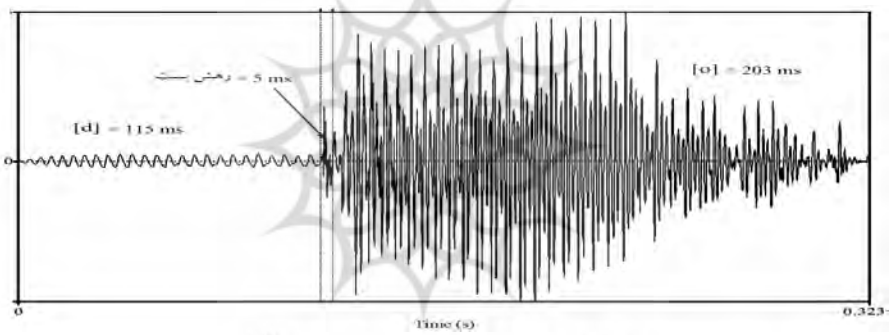
به باور پروتی و همکاران (Pruthi et al., 2004) ابهام شنیداری خیشومی‌ها با آواهای دیگر برای شنوندگان بسیار امکان‌پذیر است. زیرا ویژگی‌های طیفی خیشومی‌ها بسیار شبیه به نیم‌واکه‌ها و یا واکی است که با فرکانس پایین در طول بست همخوان‌های انسدادی تولید می‌شود. پروتی و همکاران (Pruthi et al., 2004) بیان کرده‌اند شباهت خیشومی‌ها به سایشی‌های واکدار غیرتیزی که با یک گرفتگی ضعیف تولید می‌شوند و به شکل یک همخوان رسا ظاهر می‌شوند، نیز قابل‌توجه است. بنابراین احتمال ابهام [n] با انسدادی واکدار [d] و در نتیجه افزایش ابهام -[noh] [do] بسیار زیاد است، زیرا در طول بست [d] مانند [n] تارهای صوتی با فرکانس پایین ارتعاش می‌کنند. رهش بست هر دوی این آواها در میان و انتهای واژه ضعیف و گاهی بدون رهش هستند. از آن‌جایی‌که رقم /noh/ از سه واحد زمان‌دار واجی تشکیل یافته، انتظار می‌رود که بازشناسی آن در مقایسه با /do/ چندان دشوار نباشد. اما در بسیاری از موارد، گویشوران سایشی چاکنایی /h/ را تولید نمی‌کنند و یا بسیار ضعیف تولید می‌کنند، در نتیجه تمایز شنیداری این دو واژه کاهش می‌یابد. شکل (۱)، سیگنال تلفنی یک گویشور زن را نشان می‌دهد که سایشی چاکنایی را با انرژی کم تولید کرده‌است.



شکل ۱: سیگنال [noh] در گفتار تلفنی یک گویشور زن.

همان‌طور که در سیگنال نیز قابل‌مشاهده است، نوفهٔ سایش [h] با دامنه انرژی بسیار پائین که تقریباً قابل‌تفکیک از نوفهٔ کانال انتقال نیست، تولید شده‌است. در نتیجه، بازشناسی آن در بافت انتهای واژه، حتی برای گوشِ انسان نیز به سختی صورت می‌گیرد. حذف یا کاهش این آوا در

انتهای واژهٔ [noh] سبب افزایش شباهت [no-do] می‌شود. بنابراین به نظر می‌رسد بازشناسی و درک تمایز [do]-[noh] در شرایطی که سایشی چاکنایی تولید نمی‌شود یا اطلاعات آن کاهش پیدا می‌کند، متکی به سرنخ‌های آکوستیکی انسداد خیشومی [n] و تأثیرات آکوستیکی حفرهٔ خیشوم بر واکهٔ بعدی است. کاهش اطلاعات تولیدی [noh] سبب از بین رفتن تمایز [do]-[noh] به ویژه در بازشناسی خودکار گفتار تلفنی می‌شود. گفتار تلفنی همان‌طور که گفته شد، تحت تأثیر عوامل مختلف اختلال بیرونی است. اطلاعات زمانی یا طیفی سیگنال خود به دلیل محاوره‌ای بودن گفتار ممکن است کاهش یا حذف شده باشند و از سوی دیگر به دلیل فیلتر گفتار و وجود نوفهٔ پس‌زمینه در کانال انتقال، این اطلاعات تغییر، تضعیف یا پنهان شده باشند. شکل (۲)، بست انسدادی [d] در گفتار تلفنی یک گویشور زنِ فارسی‌زبان را نشان می‌دهد. همخوان [d] در این شکل با زمان آغاز واک (وی‌اُتی[1]) منفی تولید شده‌است.



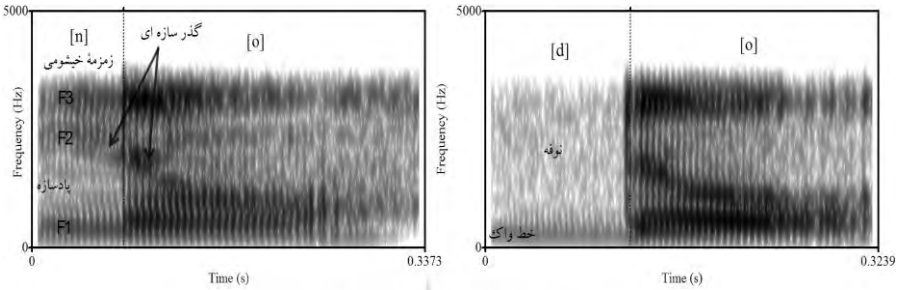
شکل ۲: سیگنال [do] در گفتار تلفنی یک گویشور زن

مقایسهٔ شکل (۱) با (۲) نشان می‌دهد که بست خیشومی انرژی بیشتری در مقایسه با بست انسدادی واک‌دار دارد و اگر بست انسدادی بدون واک تولید شود، تفاوت آن با بست خیشومی بیشتر می‌شود. شکل (۳)، طیف‌نگاشت [no] و [do] را نشان می‌دهد.

طیف‌نگاشت سمت چپ در شکل (۳)، طیف زمزمهٔ خیشومی در بست [n] را نشان می‌دهد. در طیف زمزمهٔ خیشومی سازه‌های F_1، F_2 و F_3 و نیز ناحیهٔ تشکیل پادسازه در فرکانس‌های پیرامونِ ۱۰۰۰ هرتز مشخص شده‌است. پژوهش‌های جانسون (Johnson, 2015, p. 278) نشان می‌دهد که اگر هیچ سازه‌ای نزدیک پادسازه تولید نشود، پادسازه به صورت یک نوار سفید در طیف‌نگاشت ظاهر می‌شود. در طیف‌نگاشت زمزمهٔ خیشومی [n] در بالا نیز این نوارهای سفید در مقایسه با

[1] voice onset time (VOT)

سازه‌ها به صورت نواحی کم‌رنگ‌تر تاحدی قابل‌مشاهده هستند. هر چند، حضور نوفهٔ کانال تلفن امکان تشخیص پادسازه‌ها را با مشکل مواجه می‌کند. در طیف‌نگاشت سمت راست شکل ۳، بست انسدادی [d] به همراه خط واک آن نیز مشخص شده‌است. در طول بست [d] مقداری نوفه نیز که ناشی از کانال انتقال است، دیده می‌شود.



شکل ۳: طیف‌نگاشت [noh] و [do] در گفتار تلفنی یک گویشور زن

همان‌طورکه گفته شد، پیامد آکوستیکی بازشدن حفرهٔ بینی نخست نمایانگر وقوع یک زمزمهٔ خیشومی پیش از رهش بست است و دوم، این زمزمه خیشومی طیفی دارد که برجستگی آن در فرکانس‌های پایین رخ می‌دهد. فوجیمورا (Fujimura, 1962) چهار ویژگی طیف زمزمه‌های خیشومی را به صورت زیر بر می‌شمارد: ۱. وجود سازهٔ اول با فرکانس بسیار پایین در حدود ۳۰۰ هرتز، ۲. وجود پادسازه‌ها یا صفرها در طیف، ۳. ضریب میرایی[1] نسبتاً بالای سازه‌ها و ۴. تراکم قابل‌توجه سازه‌ای. ویژگی‌های نخست و دوم نتیجهٔ عبور جریان هوا از مسیر بینی است. فرکانس پادبازخوان، بسته به محلِ گرفتگی در حفرهٔ دهانی متغیر است. ویژگی‌های سوم و چهارم ناشی از افت بالای انرژی است که به دلیل دیواره‌های نرم حفرهٔ بینی با مساحت سطح بزرگ و طول بلند روی می‌دهد. بر اساس دیدگاه وی، مشخصه‌های طیفی زمزمهٔ خیشومی در نواحی فرکانسی پایین نسبتاً پایدار است و می‌تواند معیاری معتبر برای تشخیص طبقهٔ همخوان‌های خیشومی به وجود آورد. براساس یافته‌های پروتی و همکاران (Pruthi et al., 2004)، سرنخ‌های سه ناحیهٔ متفاوت، وجود یک همخوان خیشومی را نشان می‌دهد: ۱. ناحیهٔ زمزمهٔ خیشومی، ۲. ناحیهٔ گذر بین همخوان خیشومی و واکهٔ مجاور که معمولاً همراه با یک تغییر طیفی ناگهانی است و ۳. بخش پایانی واکهٔ قبل از خیشومی و بخش آغازین واکهٔ پس از خیشومی که احتمال خیشومی‌شدگی آن‌ها به دلیل مجاورت با خیشومی امکان‌پذیر است. یافته‌های پژوهش ریکازنس (Recasens, 1982) نشان

[1] damping factor

می‌دهد که برای تمایز نحوهٔ تولید، اطلاعات زمزمهٔ خیشومی مهم‌تر هستند و سرنخ‌های حاصل از تأثیر خیشومی‌شدگی بر واکه‌های مجاور اهمیت کمتری دارند. خیشومی‌شدگی در واکه‌های مجاور با تولید صفرها و قله‌ها در تابع انتقال دستگاه گفتار و افزایش پهنایِ باندِ سازه‌های واکه به ویژه در ناحیهٔ سازهٔ اول روی می‌دهد (Pruthi et al., 2004). وینشتاین و همکاران (Weinstein et al., 1975) از پارامترهای مبتنی بر سازه برای تشخیص خیشومی‌ها در پیکره‌ای از ۱۱۱ جمله که به‌وسیلهٔ ۶ گویشور مرد و ۱ گویشور زن تولید شده بود، استفاده نمودند. آن‌ها برای بافت‌های پیش‌واکه‌ای، پس‌واکه‌ای و بین‌واکه‌ای به ترتیب نرخ شناسایی ۸۰، ۶۰ و ۸۴ درصد به دست آوردند.

در پژوهش حاضر، میزان تمایز شنیداری [do]-[noh] را براساس میزان تمایز درکی ناحیهٔ بست [n] با بست [d] و سپس اطلاعات گذرهای همخوان به واکه بررسی می‌شود. همان‌طور که طیف‌نگاشت‌ها نشان می‌دهند، رهشِ بستِ [n] با [d] تفاوت چشمگیری ندارد. اما از آن‌جا که همزمان با بستِ دهانیِ خیشومی، جریان هوا از بینی عبور می‌کند، بنابراین فشار هوای پشت بست خیشومی چندان بالا نیست و معمولاً در بیشتر نمونه‌ها رهش بست دیده نمی‌شود و یا با انرژی بسیار کم تولید شده‌است. در شکل (۳) رهش بست خیشومی ۳ میلی‌ثانیه است و در طیف‌نگاشت آن قابل مشاهده است. بست انسدادی واک‌دار می‌تواند با تولید واک همراه باشد یا نباشد و بست آن به دلیل افزایش فشار هوای پشت بست با نوفه پرانرژی‌تری نسبت به خیشومی رهش می‌کند. در هر دو همخوان خیشومی [n] و انسدادی [d] بازخوان‌های حفرهٔ جلوی بست در شکل‌گیری ویژگی‌های طیف رهش نقش دارند. به نظر می‌رسد در گفتار تلفنی علاوه بر خط واک در طی بست [d]، نوفهٔ پس‌زمینه نیز در شباهت بیشتر [n] به [d] تأثیرگذار باشد.

۳. تحلیل داده‌ها

در آزمایش نخست، برای یافتن سرنخ‌های درکی تمایزدهندهٔ خیشومی-انسدادی در شرایط نامطلوب شنیداری، محرک‌های درکی[1] از نمونه‌های تلفنی [noh] سه گویشور زن و سه گویشور مرد، موجود در پیکرهٔ گفتار تلفنی اعداد فارسی استخراج شد. رقم‌ها در این پیکره به صورت ناپیوسته و گفتار خواندنی ضبط شده‌اند. تعداد ۲۲ محرک درکی در این آزمایش از قطعهٔ بست خیشومی، از اولین پالس تا قبل از رهش انسداد خیشومی، از سیگنال استخراج شد و سپس برای بررسی میزان اهمیت زمزمهٔ خیشومی در تمایز خیشومی-انسدادی برای شش شنوندهٔ فارسی‌زبان

[1] perceptual stimuli

پخش شد. شنوندگان و نیز گویشوران شرکت‌کننده در این پژوهش از توانایی گفتاری و شنیداری سالم و طبیعی برخوردار بودند. در وظیفهٔ شناسایی[1] از شنوندگان خواسته شد که آیا سیگنال مورد نظر به صدای [n] بیشتر شباهت دارد یا به [d]. محرک‌های درکی معمولاً بیش از یک بار برای هر شنونده پخش شد و از آن‌ها خواسته شد بین گزینه‌ها یکی را انتخاب کنند. در آزمایش اول تعداد ۱۳۲ (۲۲ × ۶) نمونهٔ درکی به‌دست آمد. جدول (۱)، نشان می‌دهد شنوندگان ۹۴ مورد را به صورت [d] و ۳۸ مورد را به صورت [n] شناسایی کرده‌اند که مشتمل بر بیش از ۷۱ درصد شناسایی نادرست و در حدود ۲۹ درصد شناسایی درست است. علاوه بر آن، درصد شناسایی درست برای شنوندگان شمارهٔ ۱، ۲، ۵ و ۶ از ۲۹ درصد کمتر به دست آمد.

جدول ۱: درصد شناسایی زمزمهٔ خیشومی در گفتار تلفنی

پاسخ	ش.۱	ش.۲	ش.۳	ش.۴	ش.۵	ش.۶	جمع کل
[n]	۶	۶	۱۱	۸	۲	۵	۳۸
[d]	۱۶	۱۶	۱۱	۱۴	۲۰	۱۷	۹۴
شناسایی درست	٪۲۷	٪۲۷	٪۵۰	٪۳۶	٪۹	٪۲۳	٪۲۹

بر مبنای یافته‌های جدول (۱)، ۲۹ درصد شناسایی درست، به معنای دقت بازشناسی بسیار پائین است که دلیل آن می‌تواند فیلتر صدا و نیز نوفهٔ کانال انتقال باشد؛ شناسایی نادرست نخست، سبب افت اطلاعات طیفی و شناسایی نادرست دوم، سبب اختلال درکی زمزمهٔ خیشومی با نوفهٔ محیط می‌شود.

در آزمایش دوم، برای یافتن سرنخ‌های درکی تمایزدهندهٔ خیشومی-انسدادی در شرایط مطلوب شنیداری، ابتدا نمونه های گفتار دقیق و آزمایشگاهی [noh] در یک اتاق آرام و در سکوت کامل به وسیلهٔ یک گویشور مرد با گفتار محاورهٔ معیار تولید و ضبط شد. سپس تعداد ۹ محرک شنیداری همانند آزمایش اول از قطعهٔ زمزمهٔ خیشومی استخراج شد و در اختیار ۵ شنوندهٔ فارسی زبان قرار گرفت. در این آزمایش ۴۵ نمونه (۵ × ۹) به‌دست آمد. یافته‌های به‌دست‌آمده، نمایانگر ۲۷ مورد شناسایی [d] و یا حدود ۶۰ درصد شناسایی نادرست و ۱۸ مورد شناسایی [n] و یا حدود ۴۰ درصد شناسایی درست است. یافته‌های به دست‌آمده در جدول (۲) به نمایش درآمده‌است.

در این آزمایش نسبت به آزمایش نخست که در آن محرک‌های شنیداری از نمونه‌های تلفنی استخراج شده‌اند، میزان شناسایی درست، ۱۱ درصد بهبود یافته‌است. اما همچنان در شرایط

[1] identification task

مطلوب شنیداری نیز اطلاعات زمزمهٔ خیشومی به تنهایی در بازشناسی خیشومی-انسدادی کافی نیست. میزان ۴۰ درصد شناسایی درست یعنی در واقع سرنخ زمزمهٔ خیشومی بدونِ در نظر گرفتنِ شرایط درکی متفاوت (مطلوب یا نامطلوب) معیار مناسبی برای بازشناسی خیشومی لثوی نمی‌تواند باشد. پاسخ‌های شنوندهٔ شمارهٔ (۱) به ویژه این دیدگاه را تقویت می‌کنند؛ چراکه درصد شناسایی درست برای این شنونده برابر با صفر است.

جدول ۲: درصد شناسایی زمزمهٔ خیشومی در شرایط مطلوب شنیداری

پاسخ	ش.۱	ش.۲	ش.۳	ش.۴	ش.۵	جمع کل
[n]	۰	۶	۵	۴	۳	۱۸
[d]	۹	۳	۴	۵	۶	۲۷
شناسایی درست	٪۰	٪۶۷	٪۵۶	٪۴۴	٪۳۳	٪۴۰

جانسون (Johnson, 2015, p. 281) به این نکته اشاره می‌کند که زمزمه‌های خیشومی به دو دلیل، سرنخ‌های خوبی برای درک محل تولید نیستند. نخست، در ارتباط معمولی گفتاری که در میان سر و صدای محیط روی می‌دهد، پادسازه‌ها به‌خوبی قابل‌شنیدن نیستند. زیرا نوفهٔ محیط دره‌های پادسازه‌ها را پر می‌کند، همچنین، شنوندگان برای درک خیشومی‌ها ممکن است توجه چندانی به جزئیات شکل طیف در انسدادی‌های خیشومی نکنند. علاوه بر این، شکل طیف دربرگیرندهٔ اطلاعاتی دربارهٔ اندازه و شکل حفره‌های بینی و سینوسی افراد است و بنابراین سرنخ درکی چندان معتبری محسوب نمی‌شود، چراکه در بین افراد گوناگون و در شرایط مختلف مانند بیماری متفاوت است.

یافته‌های دو آزمایش بالا در هر دو شرایط مطلوب و نامطلوب درکی نیز نشان می‌دهد، زمزمهٔ خیشومی نمی‌تواند سرنخ درکی مؤثری در تمایز خیشومی-انسدادی باشد. در واقع شنوندگان بست خیشومی را بیشتر به صورت انسدادی می‌شنوند تا خیشومی. دلیل آن را علاوه بر پر شدن دره‌های پادسازه‌ها و کاهش اطلاعات طیفی و عوامل اختلال بیرونی که در بالا به آن‌ها اشاره شد، می‌توان در مشخصهٔ بی‌نشان [d] نیز دانست. به نظر می‌رسد شنوندگان گرایش بیشتری به شنیدن همخوان بی‌نشان نسبت به نشان‌دار در شرایطی که سیگنال نامشخص است، دارند. در واقع در شرایطی که تعیین فاصلهٔ شنیداری بین دو قطعهٔ سیگنال به دلیل کاهش اطلاعات آوایی امکان‌پذیر نیست، مکانیسم‌های درکی شنونده را به سمت همخوان بی‌نشان هدایت می‌کنند. به باور چنگ و همکاران (Chang et al., 2001, p. 89) در خطاهای درکی علاوه بر عوامل آکوستیکی- شنیداری،

نشانداری نیز نقش مهمی در ابهام همخوان‌ها بر عهده دارد. برای نمونه، انسدادی‌های لثوی نسبت به انسدادی‌های کامی بی‌نشان هستند. این واقعیت ممکن است سبب هدایت شنوندگان به این سمت شود که وقتی افراد با انسدادی‌های کامی تضعیف‌شده مواجه می‌شوند، تصور کنند که انسدادی لثوی می‌شنوند. آزمایش کینگستون و شینیا (Kingston & Shinya, 2003) در مورد درک محل تولید در خوشه‌های همخوانی نیز نشان می‌دهد که شنوندگان بیشتر همخوان لثوی بی‌نشان می‌شنوند تا همخوان‌های لبی و کامی که نشان‌دار هستند.

بر اساس یافته‌های به‌دست‌آمده، به نظر می‌رسد برای بازشناسی خودکار تمایز [do]-[noh] نمی‌توان فقط به اطلاعات زمزمهٔ خیشومی در بست، بسنده نمود و لازم است اطلاعات گذر واکه نیز در نظر گرفته شود. در آزمایش سوم، برای بررسی تأثیر گذرهای واکه‌ای بر بازشناسی تمایز [d]-[n]، محرک‌های درکی از قطعهٔ زمزمهٔ خیشومی به اضافهٔ ۱۰ میلی‌ثانیه از اول گذرهای واکهٔ [o] استخراج شدند. این ۱۰ میلی‌ثانیه معمولاً شامل دو یا سه پالسِ آغازی واکه است. بنابراین در این محرک‌ها علاوه بر اطلاعات زمزمهٔ خیشومی در بست [n]، اطلاعات ابتدای گذرهای واکهٔ [o] و نیز تأثیرات حفرهٔ خیشوم بر آن‌ها به صورت خیشومی‌شدگی نیز وجود دارد. تعداد ۵ محرک درکی از همان نمونه‌های گفتار تلفنی آزمایش اول برای شرایط نامطلوب شنیداری و ۵ محرک دیگر نیز از نمونه‌های گفتاری آزمایش دوم در شرایط مطلوب شنیداری استخراج شد. محرک‌های درکی در اختیار ۵ شنونده قرار گرفت و از آن‌ها خواسته شد که آیا سیگنال مورد نظر به [noh] بیشتر شبیه است یا به [do]. پاسخ شنوندگان و یافته‌های به‌دست‌آمده از شناسایی آن‌ها در جدول (۳) آمده‌است.

جدول ۳: درصد بازشناسی خیشومی براساس اطلاعات زمزمهٔ خیشومی و گذرهای واکه‌ای

شنیداری	پاسخ	ش.۱	ش.۲	ش.۳	ش.۴	ش.۵	جمع کل
نامطلوب	**[n]**	۳	۱	۳	۳	۲	۱۲
	[d]	۲	۴	۲	۲	۳	۱۳
	شناسایی درست	٪۶۰	٪۲۰	٪۶۰	٪۶۰	٪۴۰	٪۴۸
مطلوب	**[n]**	۵	۵	۵	۴	۵	۲۴
	[d]	۰	۰	۰	۱	۰	۱
	شناسایی درست	٪۱۰۰	٪۱۰۰	٪۱۰۰	٪۸۰	٪۱۰۰	٪۹۶

یافته‌های پژوهش نشان می‌دهد افزودن ۱۰ میلی‌ثانیه از ابتدای واکه به همراه قطعهٔ زمزمهٔ خیشومی، نقش تعیین‌کننده و مؤثری در بازشناسی تمایز [d]-[n] در شرایط مطلوب شنیداری دارد.

دقت بازشناسی نسبت به زمانی‌که فقط اطلاعات زمزمهٔ خیشومی در اختیار شنوندگان قرار گرفت، ۵۶ درصد بهبود یافت و از ۴۰ درصد به ۹۶ درصد رسید. به نظر می‌رسد گذرهای واکه سرنخ‌های مهم‌تری نسبت به اطلاعات زمزمهٔ خیشومی در درک به شمار می‌آیند. اما در گفتار تلفنی اگرچه بازشناسی به میزان ۱۹ درصد بهبود یافت و از ۲۹ درصد به ۴۸ درصد رسید، اما همچنان درصد بازشناسی بسیار پائینی محسوب می‌شود. افت بازشناسی در گفتار تلفنی را می‌توان ناشی از عوامل مختلف اختلال بیرونی دانست، این عوامل سبب افت اطلاعات در حوزه‌های زمان، شدت و طیف می‌شوند. برای نمونه، نوفهٔ کانال انتقال سبب پنهان‌شدنِ اطلاعات طیفی و تداخل درکی و فیلتر صدا، سبب کاهش اطلاعات زمانی و طیفی و نیز اعوجاجِ سیگنال می‌شود. در این میان، همان‌طور که گفته شد در این آزمایش فقط ۱۰ میلی‌ثانیه از ابتدای گذر واکه به قطعهٔ بست خیشومی افزوده شد، شنوندگان در گفتار تلفنی روزمره کل گذر و واکهٔ [o] را برای بازشناسی در اختیار دارند. در واقع، به دلیل حشو زمانی، سیگنال ناشی از هم‌تولیدی [n] با [o] و خیشومی‌شدگی جزئی ابتدای واکه، ممکن است افت اطلاعات طیفی تاحدی جبران شود و درصد بازشناسی افزایش یابد. علاوه بر آن، در گفتار تلفنی فعال شدن فرایندهای سازگاری نیز به افزایش بازشناسی سیگنال کمک می‌کند. همان‌طور که گفته شد، فرایند سازگاری می‌تواند بسیار سریع رخ دهد و نظام درکی را به گونه‌ای تنظیم کند که سیگنال ورودی را به بهینه‌ترین شکل ممکن پردازش کند. برای نمونه، سازگاری با تغییر مشخصه‌های طیفی ناشی از کانال انتقال در کمتر از یک ثانیه روی می‌دهد (Davis et al., 2005). اما با این وجود درصد خطا در بازشناسی [noh]-[do] همچنان در گفتار تلفنی و دیگر شرایط نامطلوب شنیداری، مانند افزایش فاصله بین گوینده و شنونده، بیشتر از شرایط مطلوب شنیداری است. دلیل آن می‌تواند ناشی از تداخل درکی نوفه‌های محیط با زمزمهٔ خیشومی باشد.

همان‌طور که جانسون (Johnson, 2015, p. 281) اشاره می‌کند نوفه‌های محیط دره‌های پادسازه‌ها را پر می‌کند که در این صورت میزان خیشومی‌شدگی واکهٔ مجاور کاهش می‌یابد و در نتیجه اطلاعات گذرهای واکه‌ای مجاور خیشومی به گذرهای واکه‌ای مجاور انسدادی شباهت بیشتری پیدا می‌کنند. بنابراین تمایز خیشومی-انسدادی براساس گذر واکه‌ای کاهش یافته، فاصلهٔ شنیداری خیشومی-انسدادی کاهش می‌یابد. دلیل دیگری که وی به آن اشاره می‌کند از این قرار است که شنوندگان برای درک خیشومی‌ها ممکن است توجه چندانی به جزئیات شکل طیف در انسدادی‌های خیشومی نکنند. زیرا شکل طیف حاوی اطلاعاتی دربارهٔ اندازه و شکل حفره‌های بینی و سینوسی افراد است و بنابراین سرنخ درکی چندان معتبری محسوب نمی‌شود. علاوه بر آن،

همان‌طور که گفته شد با کاهش یا حذف اطلاعات آوایی، اطلاعات معنایی و بافتی جمله کمکی به بازشناسی رقم‌ها نمی‌کند. بنابراین با کاهش فاصلهٔ شنیداری و نبودِ حمایت بافتی، نبودِ قطعیت و ابهام واژگانی افزایش یافته، منتهی به انتخاب واژهٔ نادرست یا توقف بازشناسی می‌شود.

٤. نتیجه‌گیری

یافته‌های پژوهش حاضر نشان می‌دهد که اطلاعات آکوستیکی و آوایی تمایز [do]-[noh] تحتِ تأثیر شرایطِ نامطلوبِ شنیداری در گفتار، به شدت کاهش پیدا می‌کند. بنابراین در زندگی روزمره میزان ابهام تمایز [do]-[noh] به دلیل تضعیف یا از بین رفتن اطلاعات آوایی سیگنال بالا است. سرنخ های زمزمهٔ خیشومی و گذرهای واکه‌ای تحتِ تأثیر شرایطِ نامطلوبِ شنیداری مانند فاصلهٔ گوینده با شنونده و یا نوفهٔ کانال تلفن نمی‌توانند نقشی در بازشناسی انسداد خشومی [n] داشته باشند، بنابراین شباهت درکی [n-d] افزایش می‌یابد. یکی از راه‌های رفع ابهام در سطح واژگانی، استفاده از اطلاعات بافتی است. اما از آن‌جایی‌که در تمایز اشاره‌شده، اطلاعات بافتی و معنایی جمله نیز چندان تأثیری ندارند، بنابراین شنونده دچار عدم قطعیت شده و از بازشناسی تمایز اشاره‌شده باز می‌ماند. در این شرایط شنونده برای رفع ابهام از گوینده می‌خواهد که رقم را تکرار کند.

در شرایط مطلوب شنیداری نیز سرنخ‌های بست خیشومی نقش محدودی در تمایز همخوان‌های [n-d] بر عهده دارند. به نظر می‌رسد برای فهم تمایز [d]-[n] در شرایط مطلوب شنیداری در حدود ۱۰ میلی‌ثانیه از ابتدای گذرهای واکه‌ای به همراه اطلاعات زمزمهٔ خیشومی سرنخ‌های لازم برای درک را فراهم می‌کند. فهم گفتار در شرایط نامطلوب شنیداری مستلزم اطلاعات زمانی و طیفی بیشتری از قطعهٔ واکه و گذرهای آن است. بنابراین، برای افزایش دقت بازشناسی خودکار، این رقم‌ها بهتر است اطلاعات طیفی کل واکه در نظر گرفته شود، زیرا به دلیل هم‌تولیدی خیشومی با واکه، اطلاعات همخوان خیشومی در سیگنال واکه به صورت حشو موجود است.

فهرست منابع

ثمره، یدالله (۱۳۸۵). *آواشناسی زبان فارسی*. تهران: مرکز نشر دانشگاهی.

جانسون، کیث (۱۳۹۴). *آواشناسی آکوستیک و شنیداری*. ترجمهٔ زهرا محمودزاده و رحیمه روح‌پرور. تهران: نشر چاپار.

محمودزاده، زهرا و محمد بحرانی (۱۳۹۳). «بررسی موردی کاهش آوایی در ارقام تلفنی فارسی». *زبان‌پژوهی*. دورهٔ ۶. شمارهٔ ۱۳. صص ۱۱۰-۱۳۴.

Bybee, J. L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Chang, S. S., Plauche, M. C. and Ohala, J. J. (2001). Markedness and consonant confusion asymmetries. In E. Hume and K. Johnson (Eds.), *The Role of Speech Perception in Phonology* (pp. 79-101). San Diego. Academic press.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor. K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134, 222- 241.

Ernestus M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language***.** 81, 162- 173.

Fujimura, O. (1962). Analysis of nasal consonants. *Journal of the Acoustical Society of America*, 34 (12), 1865–1875.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*. 49, 585-612.

Johnson, K. (2015). *Acoustic and auditory phonetics*. Translated by Z. Mahmoodzadeh and R. Rouhparvar. Tehran: Chapar publication [in Persian].

Kingston, J., & Shinya, T. (2003). Markedness asymmetries in place perception in consonant clusters*. In M. J. Solé, D. Recasens, and J. Romero; Barcelona.* International Congress of Phonetic Sciences (pp. 399-402). Barcelona, Spain.

Mahmoodzadeh, Z., & Bahrani, M. (2014). A case study of phonetic reduction in Persian telephonic numerals. *Quarterly Journal of Language Studies Alzahra University*, 6 (13), 110-134 [in Persian].

Mattys, S. L., M. H. Davis, A. R. Bradlow & S. K. Scott (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27 (7-8), 953-978.

Pruthi, T., & Espy-Wilson, C. Y. (2004). Acoustic parameters for automatic detection of nasal manner. *Speech Communication*, 43, 225–239.

Recasens, D. (1982). Place cues for nasal consonants with special reference to Catalan. *Journal of the Acoustical Society of America*, 73(4), 1346-1353.

Samareh, Y. (2006). *Phonology of Persian language*. Tehran: University Publication Center [in Persian].

Weinstein, C. J., McCandless, S. S., Mondshein, L. F., & Zue, V. W. (1975). A system for acoustic-phonetic analysis of continuous speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23, 54–67.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

Quarterly Journal of Research in Linguistics, Al-ZahraUniversity
VOL. 10,NO. 29, 2018-2019
http://jlr.alzahra.ac.ir

The recognition of nasal-stop distinction in adverse listening conditions[1]

Zahra Mahmoodzadeh

Received: 2017/03/13
Accepted: 2017/05/07

Abstract

The automatic recognition of Persian numerals [sefr-se] zero˚ three , [do-noh] two-nine and [haft-ha t] seven-eight is considered as a challenge for speech recognition systems. Mahmoodzadeh and Bahrani (2014) found that the acoustic reduction of telephone speech triggers lexical ambiguity for the automatic recognition of [sefr-se] pair. The numeral [sefr] is produced with the deletion of [r] at word final position and the weak labial friction of [f] is masked by the channel noise, which results in an increase of acoustic similarity between [sefr] and [se].

The automated recognition of phonological distinctions can be improved by using perceptual cues which listeners apply for the natural recognition of sounds in adverse listening conditions such as telephone speech or noisy environment. In this research for discovering efficient perceptual cues responsible for distinction of numerals [do]-[noh] two-nine in both natural and telephone speech, perception tests were used. The acoustic signal of [noh] shows a weak or practically deleted final [h], which is not audible and recognizable from background noise. Therefore, the acoustic differences of nasal-stop distinction and the co-articulatory effects of nasal on the following vowel play an important role in [do-no] recognition.

---

[1] . (DOI): 10.22051/JLR.2017.14294.1293

[2] . Assistant professor of General Linguistics, Iranian research institute for information science and technology (IRANDOC). Iran; mahmoodzadeh@irandoc.ac.ir

Results show that the acoustic landmarks, nasal murmur and nasal-vowel transitions affected by various sources of signal disruption, undergo phonetic reduction, which leads to perceptual similarity of [do-noh] pair and ultimately listeners` lexical ambiguity. In optimal listening condition, natural recognition of [n] based on nasal murmur is about 40% which increases to 96% after the addition of 10ms of the beginning of vowel [o]. However in telephone speech, natural recognition of [n] based on murmur is about 29% and after the addition of 10ms of vowel [o] transitions rises to just about 48%. According to the outcomes, nasal murmur is not an influential perceptual cue for the recognition of [n] in both optimal and adverse listening conditions; however recognition rises significantly after adding only 10 ms of the beginning of following vowel. It is likely that listeners lexical confusion and uncertainty due to lack and reduction of phonetic information reinforced by listeners perceptual bias toward unmarked consonant, give rise to [d] responses and failure of [n] identification in adverse listening conditions. The automatic recognition of these words should be done by capturing further temporal and spectral information from neighboring vowel transitions.

Keywords: adverse listening conditions, Persian numerals, perception test, nasal-stop distinction, vowel transition.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی