پیش‌گویی قابلیت فهم همخوان‌ها در افراد دارای شنوایی عادی با استفاده از مدل‌های میکروسکوپی دارای معیار فاصله متفاوت در بازشناساگر خودکار گفتار

مسعود گراوانچی‌زاده[1]، علی فلاح[2] و میرعلی اعتراف اسکویی[3]

[1و2] دانشکدهٔ مهندسی برق و کامپیوتر- دانشگاه تبریز- تبریز- ایران

[3] دانشکدهٔ توانبخشی- دانشگاه علوم پزشکی تبریز- تبریز- ایران

چکیده

در این مطالعه، نرخ تشخیص همخوان‌های موجود در ساختار هجایی «واکه- همخوان- واکه»، در آزمون‌های شنوایی و دو مدل میکروسکوپی ادراک گفتار مورد بررسی قرار می‌گیرد. چنین ساختار هجایی در زبان فارسی و ترکی آذری وجود ندارد؛ با وجود این، نتایج آزمون‌های شنوایی نشان می‌دهد که شنوندهٔ آذری یا فارسی‌زبان در شرایط بدون نوفه، قادر به تشخیص صحیح همخوان‌ها هستند. برای این پژوهش که در آن هدف، تشخیص صحیح آواها و نه کلمات بامعنی است، استفاده از این دادگان صوتی فاقد معنی مناسب است، چون با استفاده از این دادگان، دانش زبانی شنوندگان در پیش‌بینی کلمات نادیده گرفته می‌شود. نتایج آزمون‌های شنوایی با نتایج دو مدل میکروسکوپی که بر پایهٔ دستگاه شنوایی انسان است، مقایسه می‌شود. تفاوت دو مدل در مرحلهٔ نهایی استخراج ویژگی به‌منظور استفاده در شناساگر خودکار گفتار DTW است. در مدل میکروسکوپی اول، در مرحلهٔ پایانی استخراج ویژگی، از فیلتر ۸ هرتز و در مدل دوم، از فیلتربانک مدولاسیون استفاده می‌شود. در ادامه، نرخ تشخیص صحیح آواها در مقادیر مختلف سیگنال به نوفه با استفاده از معیارهای فاصلهٔ اقلیدسی و لگاریتمی با یکدیگر مقایسه می‌شود. در این تحقیق، نرخ تشخیص همخوان‌ها برای شنوندهٔ آذری‌زبان مورد بررسی قرار گرفته است. در کنار جنبهٔ تجربی این مطالعه، نوآوری این مقاله در بررسی دو معیار فاصلهٔ مختلف برای مدل هلوب و نیز مقایسهٔ مستقیم دو مدل میکروسکوپی در پیش‌بینی میانگین نرخ تشخیص و نیز نرخ تشخیص تک‌تک همخوان‌ها است.

واژگان کلیدی: قابلیت فهم، ادراک گفتار، مدل میکروسکوپی، بردار ویژگی، نرخ تشخیص آوا، معیار فاصله، شناساگر خودکار گفتار.

۱- مقدّمه

آزمون‌های شنوایی به‌عنوان دقیق‌ترین معیار ارزیابی در الگوریتم‌های بهبود کیفیت و جداسازی سیگنال‌های گفتار مطرح هستند. در کنار این آزمون‌ها، روش‌های ارزیابی بدون استفاده از شنونده انسانی وجود دارند که متخصصان شنوایی‌سنجی و مهندسان را از انجام آزمون‌های وقت‌گیر شنوایی بی‌نیاز می‌کند. دو روش ساده برای ارزیابی میزان بهبود کیفیت گفتار و جداسازی وجود دارند که روش اوّل، مقایسهٔ نسبت سیگنال به نوفهٔ[1] ورودی با نسبت سیگنال به نوفهٔ خروجی و روش دوم، مقایسه ماسک باینری طراحی‌شده با ماسک باینری ایده‌آل است (Wang *et al.*, 2005).

مطالعاتی که بر روی معیارهای ارزیابی کیفیت گفتار صورت گرفته، نشان می‌دهد که بهبود در مقدار سیگنال به نوفه به‌حتم به معنی بهبود کیفیت گفتار نیست (Loizou, P., 2007)؛ از این رو مدل‌هایی مطرح شده‌اند که برای پیش‌گویی قابلیت ادراک گفتار مورد استفاده قرار گرفته‌اند. مطالعهٔ میزان تأثیر نوفه بر قابلیت فهم گفتار منجر به ابداع معیارهای استانداردشده‌ای مانند شاخص بیان[2] و

[1] Signal-to-Noise Ratio (SNR)

[2] Articulation Index (AI)

شاخص قابلیت فهم[1] و شاخص انتقال گفتار[2] شده است. در شاخص بیان و شاخص قابلیت فهم ، مقدار سیگنال به نوفه در باندهای فرکانسی مختلف محاسبه می‌شود و این مقادیر سیگنال به نوفه پس از وزن‌دهی براساس اهمیت باندهای فرکانسی، با هم جمع می‌شوند. این مقدار به‌دست آمده، با استفاده از تبدیل غیرخطی وابسته به داده‌های گفتار در آزمون‌های شنوایی، به شاخص قابلیت فهم نگاشت می‌شود. در شاخص انتقال گفتار ، تابع تبدیل مدولاسیون[3] برای بررسی میزان مدلاسیون‌های تخریب‌شده ناشی از اکوستیک اتاق برای سیگنال گفتار، مورد استفاده قرار می‌گیرد. از شاخص انتقال گفتار برای پیش‌بینی قابلیت فهم گفتار در محیط‌های نوفه‌ای و پژواک‌دار استفاده شده است. به‌تازگی، روشی جدیدی در مرجع (Taghia et al 2014) از نقطه‌نظر تئوری اطلاعات برای پیش‌بینی قابلیت فهم ارائه شده است. در این روش، اطلاعات متقابل[4] بین پوش زمانی سیگنال تمیز و سیگنال آلوده در باندهای فرکانسی مختلف تخمین زده می‌شود. تمامی این مدل‌های نامبرده شده، مدل‌های ماکروسکوپی نامیده می‌شوند. علّت این نام‌گذاری، پیش‌بینی نرخ متوسط تشخیص جملات و عدم بررسی نرخ‌های تشخیص واحدهای کوچک‌تر گفتار مانند آوا، سیلاب و کلمه است. در کنار مدل‌های ماکروسکوپی، مدل‌های میکروسکوپی پیشنهاد شده‌اند که در آن‌ها سعی شده است از پردازشی مشابه دستگاه شنوایی انسان استفاده شود. همچنین در این مدل‌ها، برخلاف مدل‌های ماکروسکوپی، نرخ تشخیص برای واحدهای کوچک‌تر گفتار نیز بررسی می‌شود. ضعف مدل‌های میکروسکوپی در پیچیدگی و زمان لازم برای پیاده‌سازی است (Bronkhorst *et al.*, 2000). دو مدل میکروسکوپی پیش‌بینی ادراک گفتار بر مبنای مدل اولیه پردازش شنوایی گوش انسان، پیشنهاد شده‌اند. این مدل اولیه پیش‌تر توسط *داو* (Dau *et al.*, 1996) ارائه شده است. الگوی استخراج‌شده از مراحل پیش‌پردازش به‌عنوان بردار ویژگی در شناساگر خودکار گفتار[5] DTW[6] (Sakoe *et al.*, 1978)، استفاده می‌شود. همچنین، از این بردار ویژگی در دستگاه بازشناسی گفتار با استفاده از مدل مخفی مارکوف[7] نیز استفاده شده است (Tchortz *et al.*,1999). *هلوب* (Holube *et al.*, 1996)، با استفاده از مدل اولیه پردازش شنوایی انسان و شناساگر گفتار DTW، اولین مدل میکروسکوپی را برای پیش‌بینی قابلیت فهم آواها، ارائه کرده است. *یورگنز* (Jürgen *et al.*, 2009)، با حفظ ساختار کلی مدل هلوب و با ایجاد تغییراتی در بردار ویژگی، مدل میکروسکوپی جدیدی ارائه نموده است. موضوع جدیدی که در مدل یورگنز مطرح است، استفاده از معیار فاصله متفاوت در DTW است. بر حسب این معیار، نرخ تشخیص تغییر می‌کند و درنتیجه یکی از مدل‌ها برای برازش داده‌های آزمون شنوایی مناسب‌تر خواهد بود. تاکنون مقایسه‌ای در مورد نرخ تشخیص این دو مدل و نیز نوع معیار فاصله مناسب در آنها، صورت نگرفته است. در این مقاله، نرخ تشخیص آواها برای افراد دارای شنوایی عادی مورد بررسی قرار می‌گیرد. در ادامه، نرخ تشخیص آواها در دو مدل میکروسکوپی مورد مطالعه قرار خواهد گرفت و مدل و معیار فاصله بهینه تعیین می‌شود. همچنین، نرخ تشخیص تک‌تک آواها در دو مدل مقایسه خواهد شد.

بخش‌بندی ادامهٔ مقاله به این ترتیب است که در بخش دوم، ساختار مدل‌های شنوایی، بازشناساگر DTW، دادگان مورد استفاده و شرایط انجام تست شنوایی با جزئیات شرح داده می‌شود. در ادامه و در بخش سوم، نتایج تست شنوایی و شبیه‌سازی با رویکرد مقایسه مدل‌ها وتأثیر معیار فاصلهٔ انتخابی بر میزان پیش‌بینی قابلیت فهم، مورد تحلیل قرار می‌گیرد. در پایان و در بخش چهارم نیز بحث و نتیجه‌گیری‌های لازم از نتایج بالا لحاظ شده است.

۲- ساختار مدل‌ها

در شکل (۲-۱)، مدل ادراک گفتار پیشنهادشده توسط هلوب نشان داده شده است (Holube *et al.*, 1996). در قسمت بالای این مدل، مرحلهٔ آموزش و در قسمت پایین، مرحلهٔ آزمون داده‌های صوتی آمده است. سیگنال‌های گفتار آلوده به نوفه، به‌عنوان ورودی‌های آموزش و آزمون، به مدل اعمال می‌شوند.

در اولین مرحلهٔ پیش‌پردازش، سیگنال از فیلتربانک گاماتون عبور داده می‌شود که این فیلتربانک مدلی برای فیلترینگ سیگنال در حلزون گوش انسان است. فیلتربانک گاماتون با استفاده از روش هومن (Hohmann, 2002) پیاده‌سازی شده است. در فیلتربانک گاماتون پاسخ ضربه فیلتر به‌صورت زیر است:

[1] Speech Intelligibility Index (SII)
[2] Speech Transmission Index (STI)
[3] Modulation Transfer Function (MTF)
[4] Mutual Information
[5] Automatic Speech Recognizer (ASR)
[6] Dynamic Time Warping (DTW)
[7] Hidden Markov Model (HMM)

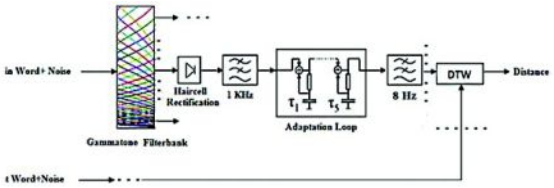
$$g_\gamma[n] = n^{\gamma-1}\tilde{a}^n \quad n \geq 0, \qquad \tilde{a} = \lambda.\exp(j\beta), \tag{۱}$$

که در آن γ مرتبهٔ فیلتر، β فرکانس نوسان و λ پارامتر میرایی یا پهنای باند است. این پاسخ ضربه، با سری قراردادن چهار فیلتر مرتبهٔ اول میان‌گذر با ضرایب مختلط پیاده‌سازی می‌شود. پاسخ فرکانسی یک فیلتر گاماتون از رابطهٔ زیر به‌دست می‌آید:

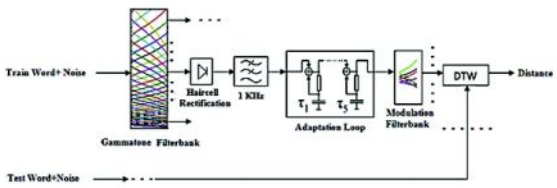
$$H(z) = \frac{1}{\left(1-\tilde{a}z^{-1}\right)^4}. \tag{۲}$$

برای استخراج بردارهای ویژگی، سیگنال ابتدا از ۲۷ فیلتر گاماتون عبور داده می‌شود. این فیلترها، طبق مقیاس پهنای باند مستطیلی معادل[1] و یک فیلتر به‌ازای هر پهنای باند مستطیلی معادل در محور فرکانس، توزیع شده‌اند. این فیلتربانک، بازهٔ فرکانسی ۲۳۴ هرتز تا ۸ کیلوهرتز را پوشش می‌دهد. در ادامه، خروجی هر فیلتر، از یکسوساز نیم‌موج عبور داده می‌شود. پس از عبور از یکسوساز نیم‌موج، سیگنال توسط فیلتر پایین‌گذر مرتبهٔ اوّل با فرکانس قطع یک کیلوهرتز، فیلتر می‌شود. یکسوساز نیم‌موج و فیلتر یک کیلوهرتز، برای مدل‌سازی سلول‌هایی مویی[2] گوش استفاده شده‌اند. در مرحله بعدی، خروجی مدل سلول‌های مویی از پنج حلقه انطباق[3] متوالی با ثابت‌های زمانی $\tau_1 = 5\,\text{ms}$ $\tau_2 = 50\,\text{ms},$ $\tau_3 = 129\,\text{ms},$ $\tau_4 = 253\,\text{ms}$ و $\tau_5 = 500\,\text{ms}$ که برای مدل‌سازی پردازش‌های غیر خطی گوش استفاده شده است، مطابق با (Dau *et al.*, 1996)، عبور داده می‌شود. حلقه‌های انطباق، سیگنال‌های زمانی ایستان را به‌طور تقریبی به‌صورت لگاریتمی فشرده می‌سازند. در مورد سیگنال‌های غیرایستان، مانند سیگنال گفتار، شروع[4] و پایان[5]های سیگنال در خروجی حلقه‌ها، به‌طور برجسته‌تر تقویت می‌شود. هر مسیر بازگشتی در این حلقه‌ها شامل یک مقسم و یک فیلتر پایین‌گذر با ثابت زمانی τ_i است. سیگنال ورودی به حلقه بر خروجی فیلتر پایین‌گذر تقسیم می‌شود. برای سیگنال با مقدار ثابت C، سیگنال خروجی برای هر حلقه انطباق به‌صورت $O = \sqrt{C}$ است. بنابراین خروجی پنج حلقه برابر $O = C^{1/32}$ خواهد بود که به این صورت عملیات فشرده‌سازی لگاریتمی سیگنال ورودی تخمین زده می‌شود. با تغییر ثابت‌های زمانی با استفاده از ظرفیت خازن‌ها، می‌توان پاسخ سامانه به نوسانات سیگنال ورودی را تغییر داد. اگر نوسانات سیگنال ورودی در مقایسه با ثابت‌های زمانی حلقه‌های انطباق سریع‌تر باشد، سیگنال به‌طور تقریبی بدون تغییر از حلقه‌ها عبور می‌کند و در غیر این صورت، سیگنال فشرده می‌شود؛ حتی هنگامی که سیگنال تحریک در ورودی حلقه‌ها قطع می‌شود، به‌دلیل پردازش غیرخطی تا مدت زمانی در خروجی، تحریک وجود دارد. بنابراین، فشرده‌سازی سیگنال و انطباق زمانی[6] به‌طور هم‌زمان انجام می‌گیرد. در این پردازش غیرخطی، اطلاعات مربوط به پوش سیگنال حفظ می‌شود. در مدل هلوب، مرحله قبل از DTW ، عبور سیگنال از فیلتر پایین‌گذر با ثابت زمانی بیست میلی ثانیه (که به‌طور تقریبی معادل فیلتر هشت هرتز با شیب شش دسیبل در هر اکتاو می‌باشد) است.

شماتیک مدل دوم که مربوط به کار یورگنز است، در شکل (۲-۲) نشان داده شده است (Jürgen *et al.*, 2009). تفاوت آن با مدل هلوب، در مرحلهٔ پیش پردازش قبل از DTW است.



(شکل ۲-۱): شمای مدل پیش‌بینی قابلیت فهم هلوب. پیش‌پرازش شامل عبور از فیلتربانک گاماتون، یکسوسازی نیم‌موج و حلقه‌های انطباق است که برای مدل‌سازی برخی از آزمون‌های شنوایی مورد استفاده قرار گرفته است. برای آزمایش‌های مربوط به ادراک گفتار از شناساگر DTW استفاده می‌شود که برای سنجش فاصله بین کلمه الگو با کلمه آزمون مورد استفاده قرار می‌گیرد (Holube *et al.*, 1996).



(شکل ۲-۲): شمای مدل پیش‌بینی ادراک گفتار ارائه‌شده توسط یورگنز. مراحل پیش‌پردازش، به جز مرحلهٔ آخر که از فیلتربانک مدولاسیون به‌جای فیلتر ۸ هرتز استفاده شده است، مشابه مدل هلوب است. IR در این مدل ماتریس دوبعدی است: بعد اول برای فیلتربانک گاماتون و بعد دوم برای فیلتربانک مدولاسیون تعریف می‌شود (Jürgen, et al., 2009).

[1] Equivalent Rectangular Bandwidth (ERB)
[2] Haircells
[3] Adaptation Loop
[4] Onset
[5] Offset
[6] Adaptation

در مدل یورگنز، از فیلتربانک مدولاسیون[1] مطابق کار داو (Dau *et al.*, 1997) استفاده شده است. این فیلتربانک، شامل چهار کانال مدولاسیون به‌ازای هر کانال فرکانسی گاماتون است که یک فیلتر پایین‌گذر با فرکانس قطع ۲/۵ هرتز و سه فیلتر میان‌گذر با فرکانس‌های مرکزی ۵، ۱۰ و ۱۶/۷ هرتز را در بر می‌گیرد. پهنای باند برای فیلتر پایین‌گذر و فیلترهای با فرکانس‌های مرکزی ۵ و ۱۰ هرتز، برابر ۵ هرتز و برای فیلتر ۱۶/۷ برابر ۸/۳ هرتز است.

فیلترهای مدولاسیون با استفاده از رابطهٔ بازگشتی زیر پیاده‌سازی می‌شوند:

$$y[n] = e^{-\pi\beta\Delta}\ e^{j2\pi f_0\Delta}\ y[n-1] + (1-e^{-\pi\beta\Delta})\ x[n], \qquad (۳)$$

که در آن β پهنای باند فیلتر، Δ پریود نمونه‌برداری و f_0 فرکانس مرکزی فیلتر است.

تابع تبدیل رابطه (۳) برابر است با:

$$H(z) = \frac{1-e^{-\pi\beta\Delta}}{1-e^{-\pi\beta\Delta}.e^{j2\pi f_0\Delta}z^{-1}}, \qquad (۴)$$

که تابع تبدیل فیلتر میان‌گذر مرتبهٔ اول است.

پس از کاهش نرخ نمونه‌برداری به مقدار صد هرتز، نمایش داخلی سیگنال[2] به‌دست می‌آید. بنابراین، نمایش‌های داخلی سیگنال گفتار برای مدل‌های هلوب و یورگنز، به‌ترتیب، به‌صورت بردار یک‌بعدی و ماتریس دوبعدی در هر ده میلی‌ثانیه خواهد بود. واحد عناصر این بردار و ماتریس، واحد مدل[3] نامیده می‌شود. یک واحد مدل معادل با یک دسیبل شدت صوت است (Dau *et al.*, 1997).

۲-۱- بازشناساگر خودکار DTW

در شکل‌های (۲-۱) و (۲-۲)، مسیر پردازشی بالا برای سیگنال مرحلهٔ یادگیری و مسیر پایین برای سیگنال آزمون در نظر گرفته شده است. الگوهای به‌دست‌آمده از این دو مسیر، توسط شناساگر خودکار گفتار DTW (Sakoe *et al.*, 1978) مقایسه می‌شوند. در مدل‌های میکروسکوپی، از DTW برای حالتی که تفاوت سیگنال الگو و آزمون تنها در نوفه افزوده‌شده به آنهاست، به‌منظور پیش‌بینی قابلیت فهم گفتار استفاده شده است. یک تبدیل زمانی[4] خاص برای انطباق بهینهٔ نمایش‌های داخلی سیگنال الگو و سیگنال آزمون، با محاسبهٔ ماتریس فاصلهٔ D به‌دست می‌آید. هر عنصر D(i, j) در این ماتریس، از محاسبهٔ فاصله بین ماتریس ویژگی مربوط به نمایش داخلی الگو در شاخص زمانی i و ماتریس ویژگی مربوط به نمایش داخلی سیگنال آزمون در شاخص زمانی j به‌دست می‌آید. سپس، یک مسیر به‌هم پیوسته از روی این ماتریس فاصله به‌دست می‌آید. این مسیر به‌گونه‌ای انتخاب می‌شود که مقدار نهایی فاصله بین دو الگوی آموزش و آزمون در نقطهٔ انتهایی مسیر حداقل شود. با مقایسهٔ تمامی سیگنال‌های آزمون با سیگنال الگو، سیگنال دارای کمترین فاصله، به‌عنوان الگوی تشخیص داده‌شده تعیین می‌شود.

۲-۲- الگوی فاصله

ساده‌ترین رویکرد در DTW استفاده از فاصلهٔ اقلیدسی بین بردارهای ویژگی الگو ($\mathbf{IR}_{template}$) و آزمون ($\mathbf{IR}_{test}$) است:

$$D_{Euclidean} = \sqrt{\sum_{f}\sum_{f_{\mathrm{mod}}}\left(\mathbf{IR}_{templ}(i,f,f_{\mathrm{mod}}) - \mathbf{IR}_{test}(i,f,f_{\mathrm{mod}})\right)^2}, \qquad (۵)$$

که در آن، f بیان‌گر کانال فرکانسی مربوط به فیلتربانک گاماتون و f_{mod} مربوط به باند فرکانسی مدولاسیون در مدل یورگنز است. عمل جمع‌بندی در رابطهٔ (۵) در مدل هلوب، تنها برای خروجی تمامی فیلترهای گاماتون صورت می‌گیرد؛ و شاخصی برای مدولاسیون در محاسبه فاصله وجود ندارد. در حالتی که اختلاف عناصر ماتریس الگو و ماتریس آزمون دارای توزیع گوسی باشد، فاصلهٔ اقلیدسی بهینه خواهد بود؛ اما هنگامی که این توزیع گوسی نباشد، فاصله اقلیدسی دیگر بهترین معیار، و به‌عبارتی ساده‌تر، کم‌ترین مقدار فاصله نخواهد بود. در مدل یورگنز، برای ماتریس‌های ویژگی نشان داده شده است، هنگامی که تفاوت سیگنال الگو و سیگنال آزمون تنها در نوفه افزوده‌شده به آنهاست، فاصله بین این ماتریس‌های ویژگی از توزیع کوشی تبعیت می‌نماید. برای توزیع کوشی، معیار فاصله لگاریتمی بهینه است (Jürgens *et al.*, 2009). فاصلهٔ لگاریتمی از رابطه زیر به‌دست می‌آید:

$$D_{Log} = \sum_{f}\sum_{f_{\mathrm{mod}}}\log\left(1+\frac{1}{2}\left(\mathbf{IR}_{templ}(i,f,f_{\mathrm{mod}}) - \mathbf{IR}_{test}(i,f,f_{\mathrm{mod}})\right)^2\right). \qquad (۶)$$

با استفاده از دو معیار فاصله و دو مدل میکروسکوپی می‌توان چهار حالت را برای بررسی نتایج مربوط به نرخ‌های تشخیص برای پیش‌بینی قابلیت فهم گفتار در نظر گرفت.

[1] Modulation Filter Bank (MFB)
[2] Internal Representation (IR)
[3] Model Unit (MU)
[4] Time-Transformation

برای پیاده‌سازی فیلتربانگ گاماتون، از کدهای موجود در پایگاه اینترنتی دانشگاه Oldenburg استفاده شده که در آن روش هومن (Hohmann, 2002) پیاده‌سازی شده است؛ همچنین، مدل سلول‌های مویی و حلقه‌های انطباق، با استفاده از فایل‌های MEX موجود در تارنما همین دانشگاه شبیه‌سازی شده است. کدهای مربوط به فیلتر هشت هرتز، فیلتربانک مدولاسیون و DTW توسط نویسندگان مقاله در محیط C و در قالب MEX تهیه شده‌اند.

۲-۳- داده‌های گفتاری و افراد شرکت‌کننده در آزمون شنوایی

دادگان گفتاری، دارای ساختار «واکه-همخوان-واکه[1]» هستند که از پایگاه دادگان OLLO (Wesker *et al.*, 2005) موجود در تارنمای دانشگاه الدنبورگ[2] انتخاب شده‌اند. ساختار هجایی «واکه- همخوان- واکه» که توسط گویندگان آلمانی ادا شده است، به‌طور معمول در زبان‌های فارسی و ترکی آذری وجود ندارد؛ اما از آنجایی که هدف از استفاده از این دادگان، تنها تشخیص همخوان است، نتایج آزمون‌های شنوایی که در ابتدا برای سیگنال تمیز بدون نوفه انجام گرفت، نشان داد که شنوندگان آذری‌زبان در تشخیص همخوان‌ها در این ساختار هجایی مشکلی ندارند و استفاده از این دادگان در آزمون‌های شنوایی برای سیگنال‌های نوفه‌ای، بلامانع است.

دادگان OLLO از نوع «واکه- همخوان- واکه» دارای واکه‌های یکسان در ابتدا و انتها هستند. برای مقایسه مستقیم نتایج مدل‌ها با آزمون‌های شنوایی، از سیگنال‌های صوتی مشابهی در مدل و آزمون استفاده می‌شود. فایل‌های صوتی انتخاب‌شده، مربوط به گویندهٔ مرد با شناسه S10M_NO و دارای سرعت بیان طبیعی آواها است. همخوان‌های مورد استفاده یکی از سیزده آوای /د/، /ت/، /گ/، /ک/، /ف/، /س/، /ب/، /پ/، /و/، /م/، /ن/، /ش/ و /ل/ است که در بین یکی از پنج واکه /آ/، /اِ/، /ای/، /اُ/ و /او/ قرار می‌گیرد. درنتیجه ۶۵ فایل صوتی متفاوت برای آزمون شنوایی و آزمون مدل‌ها خواهیم داشت. از همخوان /ts/ پایگاه داده OLLO که دارای صدایی مابین صدای دو همخوان /ت/ و /س/ می‌باشد، به‌علت ناآشنایی شنوندگان با آن استفاده نشده است.

ده فرد دارای شنوایی طبیعی (نه مرد و یک زن) که بین ۲۲ تا ۲۹ سال دارند، در آزمون‌های شنوایی شرکت کرده‌اند.

۲-۴- شرایط آزمون شنوایی و آزمون مدل

دادگان آزمون به دسته‌هایی تقسیم‌بندی می‌شوند که تنها در آوای وسط با یکدیگر متفاوت باشند. برای آزمون‌های شنوایی، فرد آزمون‌شونده پس از شنیدن فایل صوتی با استفاده از موش‌واره، یکی از گزینه‌ها را از روی صفحه نمایش انتخاب می‌کند. این انتخاب از بین یکی از سیزده گزینه موجود صورت می‌گیرد. از این رو شرایط آزمون به صورت ارزیابی بسته[3] است. شنونده امکان بازپخش مجدد یک داده صوتی را برای دفعات دلخواه دارد. از آنجایی که دانش زبانی افراد در آزمون شنوایی با توجه به بی‌معنابودن مواد گفتار نادیده گرفته می‌شود، امکان مقایسهٔ دقیق‌تر مدل‌ها با آزمون شنوایی وجود دارد. همچنین، علاوه‌بر مقایسه نرخ تشخیص آواها، می‌توان ماتریس ابهام[4] را نیز در مورد آنها مورد بررسی قرار داد.

مقدار جذر میانگین مربعات[5] سیگنال گفتار با حذف بازه‌های سکوت در ابتدا و انتهای فایل‌های صوتی در شصت دسیبل تنظیم می‌شود. نوفهٔ ایستان که دارای طیف زمان-طولانی مشابه سیگنال گفتار[6] است (Dreschler *et al.*, 2001)، در شدت خاصی که برمبنای نسبت سیگنال به نوفهٔ مورد نظر تنظیم شده است، چهارصد میلی‌ثانیه قبل از شروع پخش سیگنال گفتار به داده‌های صوتی افزوده می‌شود. زمان چهارصد میلی‌ثانیه به‌منظور مقداردهی اولیه به خروجی حلقه‌های انطباق است. برای سیگنال آلوده به نوفه، در ابتدا و انتها از شیب هن[7] دارای بازهٔ زمانی صد میلی‌ثانیه استفاده می‌شود. بعد از محاسبه نمایش داخلی برای سیگنال تولیدشده، بردارهای ویژگی مربوط به چهارصد میلی‌ثانیه پخش نوفه پیش از شروع گفتار، حذف می‌شوند. این عمل به‌منظور درنظرگرفتن اطلاعات لازم از زمان پخش آواها صورت می‌گیرد.

برای ارزیابی مدل، نمونه‌های از نوفه در نسبت سیگنال به نوفهٔ مورد نظر به هر کدام از دادگانی که تفاوت آنها تنها در هجای وسط است، افزوده می‌شود و نمایش

[1] Vowel-Consonant-Vowel (VCV)
[2] Oldenburg

[3] Closed Test
[4] Confusion Matrix
[5] Root-Mean-Square
[6] ICRA noise
[7] Hann

داخلی سیگنال نوفه‌ای به‌عنوان الگویی که می‌بایست تشخیص داده شود، ذخیره می‌شود. نوفه‌ای متفاوت از نوفهٔ اضافه‌شده به الگو، این‌بار به هر کدام از سیگنال‌های گفتار آزمون افزوده می‌شود که یکی از این سیگنال‌های آزمون همان سیگنال استفاده شده در الگوی ذخیره‌شده است. با محاسبهٔ فاصله هر سیگنال آزمون با الگوهای ذخیره‌شده و محاسبه الگوی دارای کم‌ترین فاصله در DTW، نرخ تشخیص صحیح آواها در سیگنال به نوفه‌های مختلف محاسبه می‌شود. برای افزایش دقت نتایج شبیه‌سازی در مدل‌ها، برای هر کدام از دادگان الگو، ده‌بار شبیه‌سازی تکرار می‌شود. درنتیجه با داشتن ۶۵ فایل صوتی، درمجموع ۶۵۰ بار مقایسه الگو با سیگنال‌های آزمون صورت می‌گیرد. حداکثر مدت شبیه‌سازی مربوط به مدل یورگنز با استفاده از فاصلهٔ لگاریتمی است که این زمان در حدود پنج ساعت است.

۲-۵- آزمون‌های شنوایی

نرخ تشخیص سیزده هجا با استفاده از دستگاه شنوایی‌سنجی AC40 که قابلیت پخش سیگنال گفتار در شدت دسیبل دلخواه را دارد، در آزمایشگاه گفتار دانشکدهٔ توانبخشی دانشگاه تبریز که مجهز به یک اتاقک عایق در برابر صداست، انجام گرفته است. از نسبت سیگنال به نوفه‌های صفر، ۵-، ۱۰-، ۱۵- و ۲۰- دسیبل در آزمون‌های شنوایی و آزمون مدل‌ها استفاده شده است. ترتیب پخش هجاها به‌صورت تصادفی است. شنونده می‌تواند یکی از سیزده گزینهٔ موجود را انتخاب کند. قبل از انجام آزمون شنوایی، تمامی افراد با داده‌های صوتی آشنا شده‌اند.

برای توصیف میانگین قابلیت فهم برای تمامی هجاها، تابع مدل یا تابع روان‌سنجی[1] که با رابطه زیر توصیف می‌شود:

$$P(L, L_{50}, s_{50}) = \frac{1-g}{1+\exp\left(4.s_{50}.(L_{50}-L)\right)} + g, \qquad (۷)$$

به میانگین نرخ تشخیص شنوندگان در تمامی سیگنال به نوفه‌ها، برازش می‌شود که در آن s_{50} شیب تابع روانسنجی، L_{50} آستانه ادراک گفتار[2]، L مقدار سیگنال به نوفه و g احتمال تشخیص صحیح یک آوا به‌صورت تصادفی است که برابر ۰/۰۷ است. عملیات برازش با استفاده از cftool نرم‌افزار متلب انجام می‌شود.

[1] Psychometric
[2] Speech Reception Threshold (SRT)

۳- شبیه‌سازی و نتایج آزمون‌های شنوایی

۳-۱- میانگین نرخ تشخیص

شکل (۳-۱) میانگین نرخ تشخیص آواها به درصد برحسب مقدار سیگنال به نوفه برای تمامی آواها را نشان می‌دهد. میله‌های خطا[3] در شکل (۳-۱)، انحراف معیار برای ده شنونده را نشان می‌دهد. تابع روان‌سنجی برازش‌شده برای تشخیص صحیح همخوان‌ها، دارای شیب ۷ %/dB و آستانه ادراک گفتار برابر ۹/۹۲- دسیبل است. نتایج آزمون‌های شنوایی با نتایج به‌دست‌آمده از اندازه‌گیری‌های یورگنز مطابقت خوبی دارد. مقدار میانگین قدر مطلق خطا بین اندازه‌گیری‌ها و مقاله بالا کمتر از ۵٪ است. بیشترین مقدار خطا در نسبت سیگنال به نوفهٔ ۲۰- دسیبل و در حدود ۱۰٪ است که بیشتر ناشی از شرایط آزمایشگاهی و افراد استفاده‌شده در آزمون شنوایی است.

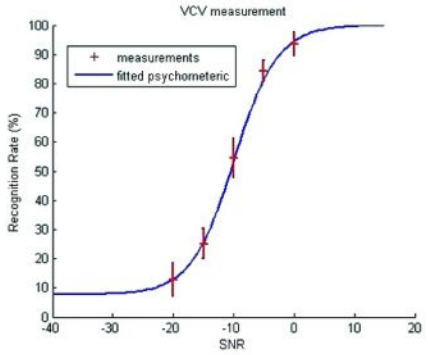
در شکل (۳-۲)، نتایج مربوط به شبیه‌سازی‌های دو مدل که در هرکدام از مدل‌ها نیز از دو معیار فاصله استفاده شده، رسم شده است. با توجه به شکل (۳-۲)، استفاده از معیار فاصله متفاوت در مدل هلوب، تأثیر چندانی در تغییر نرخ تشخیص ندارد؛ اما در مدل یورگنز، معیار فاصله، تأثیر قابل ملاحظه‌ای در نرخ تشخیص دارد. استفاده از فاصلهٔ لگاریتمی، نرخ تشخیص را به‌ویژه در سیگنال به نوفه‌های میانی نسبت به فاصله اقلیدسی بهبود می‌بخشد. البته این بهبود به معنی مناسب‌بودن مدل نیست؛ چون ممکن است مدل دارای نرخ تشخیص پایین‌تر، نتایج آزمون شنوایی را بهتر پیش‌بینی کند.

برای مقایسهٔ مدل‌ها با نتایج آزمون شنوایی، پارامترهای تابع روان‌سنجی برازش‌شده برای هر کدام از منحنی‌های شکل (۳-۲) با پارامترهای تابع روان‌سنجی مربوط به آزمون شنوایی، مقایسه شده است. پارامترهای حاصل از برازش تابع روان‌سنجی در آزمون‌های شنوایی و مدل‌ها با دو معیار فاصلهٔ مختلف در جدول (۳-۱) نشان داده شده است. کمترین تفاوت در آستانهٔ ادراک گفتار در مدل‌ها نسبت به آستانهٔ ادراک گفتار در آزمون شنوایی، مربوط به مدل یورگنز با استفاده از فاصلهٔ لگاریتمی است. استفاده از فاصلهٔ اقلیدسی در مدل یورگنز، نتیجهٔ بدتری در پیش‌بینی مقدار آستانهٔ ادراک گفتار در مقایسه با مدل هلوب دارد.

[3] Error Bars

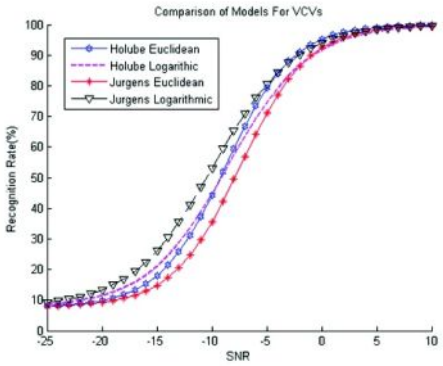
مدل هلوب با استفاده از فاصلهٔ اقلیدسی دومین پیش‌بینی نزدیک به آستانهٔ ادراک گفتار اندازه‌گیری شده را دارد. همچنین، در پیش‌بینی شیب تابع روان‌سنجی، مدل هلوب با فاصلهٔ لگاریتمی و مدل یورگنز با فاصلهٔ لگاریتمی، نزدیک‌ترین پیش‌بینی را نسبت به آزمون‌های شنوایی دارد. مربع ضرایب همبستگی پیرسون [1] (r^2) در ستون آخر جدول (۳-۱) آمده است. ضریب همبستگی پیرسون (r) بین دو جمعیت آماری از میانگین‌گیری حاصل‌ضرب مشاهدات هنجارسازی‌شده دو جمعیت به‌دست می‌آید. مشاهده هنجارسازی‌شده برای هر جمعیت از اختلاف مقدار مشاهده از میانگین جمعیت تقسیم بر انحراف معیار جمعیت به‌دست می‌آید. مقدار «یک» برای مربع این ضریب نشان‌دهندهٔ همبستگی کامل و مقدار «صفر» نشان‌دهنده عدم همبستگی بین داده‌های دو جمعیت است. بیشترین همبستگی با مقادیر آزمون شنوایی در مدل یورگنز با فاصلهٔ لگاریتمی مشاهده می‌شود.

استفاده از معیار فاصلهٔ متفاوت، در مدل هلوب تفاوت چندانی را در نتایج نشان نمی‌دهد (اختلاف کمتر از ۰/۲ دسیبل است)؛ اما در مدل یورگنز، تفاوت مقادیر آستانه ادراک گفتار پیش‌بینی‌شده با استفاده از دو فاصلهٔ مختلف، در حدود ۲/۵ دسیبل است.



(شکل۳-۱): نتایج آزمون شنوایی برای اندازه‌گیری متوسط مقدار قابلیت فهم گفتار بر حسب مقدار سیگنال به نوفه. در آزمون شنوایی از نوفهٔ دارای طیف مشابه سیگنال گفتار (ICRA noise) استفاده شده است. میله‌های خطای مربوط به انحراف معیار نرخ تشخیص بین ده شنونده و نیز تابع روان‌سنجی برازش‌شده بر اندازه‌گیری‌ها در شکل نشان داده شده است.

[1] Pearson



(شکل۳-۲): نتایج پیش‌بینی نرخ تشخیص در مدل‌ها برای پیش‌بینی مقدار قابلیت فهم گفتار با استفاده از تابع روان‌سنجی برازش‌شده بر شبیه‌سازی‌ها.

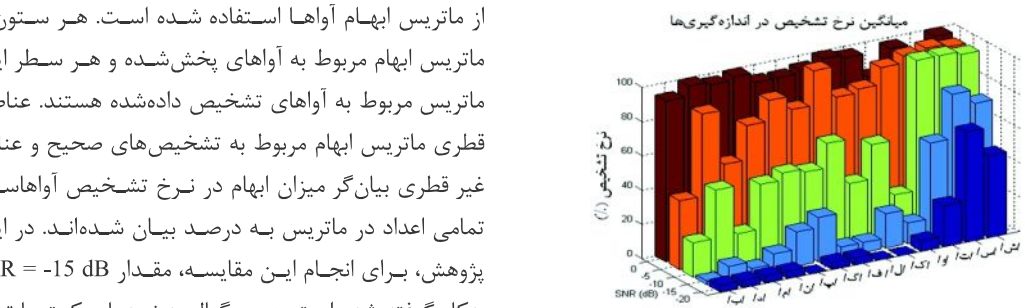
(جدول۳-۱): فهرست پارامترهای تخمین‌زده‌شده تابع روان‌سنجی برازش‌شده بر نرخ تشخیص آواها در اندازه‌گیری‌ها و دو مدل میکروسکوپی با دو معیار فاصله.

	SRT (dB)	اختلاف با SRT آزمون شنوایی	شیب (%/dB)	پیرسون(r^2)
اندازه‌گیری‌ها	-۹/۹۲۸۲	صفر	۷	۱
مدل هلوب با فاصله اقلیدسی	-۸/۷۴۰۷	۱/۱۹	۸/۳	۰/۹۷۸۵
مدل هلوب با فاصله لگاریتمی	-۸/۵۶۴۱	۱/۳۶	۶/۸۴	۰/۹۸۵۵
مدل یورگنز با فاصله اقلیدسی	-۷/۴۰۹۶	۲/۵۱	۸/۱۵	۰/۹۵۹۰
مدل یورگنز با فاصله لگاریتمی	-۹/۸۸۹۳	۰/۰۳۸۹	۶/۷۷	۰/۹۹۵۳

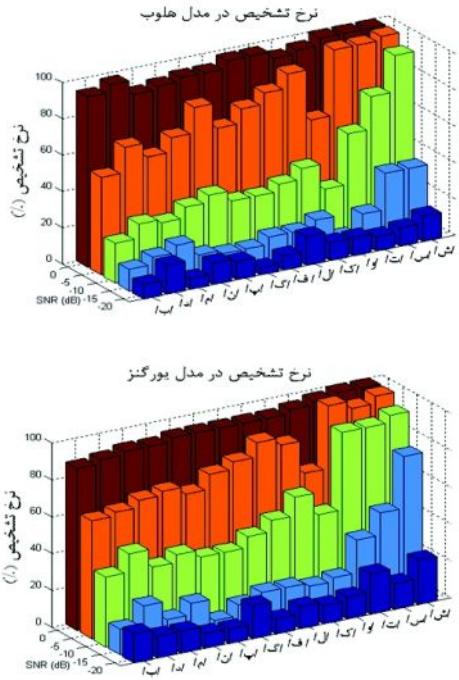
۳-۲- نرخ تشخیص هر کدام از آواها

نمودار میله‌ای شکل (۳-۳)، نتایج مربوط به نرخ تشخیص هر کدام از همخوان‌ها در آزمون‌های شنوایی را نشان می‌دهد. حداکثر نرخ تشخیص مربوط به آواهای /ت/، /س/ و /ش/ است. نرخ تشخیص برای آواهای /م/، /و/، /ب/ و /گ/ به‌طور متوسط در تمامی سیگنال به نوفه‌ها، پایین‌تر از بقیه است. در شکل (۳-۴)، نتایج مربوط به نرخ تشخیص دو مدل رسم شده است. نمودار میله‌ای اول در شکل (۳-۴)، مربوط به مدل هلوب با استفاده از فاصله اقلیدسی و نمودار دوم در این شکل، نتایج مدل یورگنز با استفاده از فاصلهٔ لگاریتمی است. این نمودارها نشان می‌دهند که نرخ تشخیص دو مدل برای سه آوای /ت/، /س/ و /ش/، مشابه آزمون‌های شنوایی به‌طور مشخص بیشتر از بقیه آواهاست.

تفاوت اصلی بین نتایج آزمون شنوایی و نتایج مربوط به دو مدل در نرخ تشخیص آواها برای سیگنال به نوفه‌های پایین (به‌ویژه در SNR = -20 dB) است. در سیگنال به نوفه‌های پایین، نرخ تشخیص برای سه آوای /ت/، /س/ و /ش/ در آزمون‌های شنوایی، خیلی بالاتر از سایر آواها، اما در مدل‌ها این اختلاف کمتر است. نرخ تشخیص مدل‌ها در SNR = -5 dB به‌طور متوسط پایین‌تر از شنونده انسانی است. این امر به‌ویژه در مورد آواهای /ن/، /و/ و /ب/ برجسته است. در SNR = -10 dB برای آواهای /م/ و/ب/، مدل‌ها نرخ تشخیص بالاتری از شنوندهٔ انسانی دارند؛ همچنین، نرخ تشخیص آوای /و/ در آزمون شنوایی برای SNR = 0 مقداری در حدود ۸۰٪ را نشان می‌دهد؛ درحالی‌که نتیجهٔ پیش‌بینی مدل‌ها بیش از ۹۰٪ است. به‌طور کلی، نتایج مدل در مقایسه با نتایج آزمون‌های شنوایی یک‌نواختی بیش‌تری در نرخ تشخیص برای سیگنال به نوفه‌های مختلف دارند. مدل یورگنز در مقایسه با مدل هلوب نرخ تشخیص بالاتری به‌ویژه در سیگنال به نوفه‌های کم دارد. این امر به‌ویژه در سیگنال به نوفه‌های ۱۰- و ۱۵- دسیبل مشهود است. به‌عنوان نمونه در SNR = -15 dB برای آوای /ش/ درحالی‌که نرخ تشخیص برای مدل هلوب برابر ۳۵٪ است، مدل یورگنز نرخ تشخیص ۷۵٪ را نشان می‌دهد. در سیگنال به نوفه‌های بالاتر از ۵- و صفر دسیبل، تفاوت دو مدل با توجه به نزدیک‌شدن نرخ تشخیص به مقدار سقف ۱۰۰٪، کاهش می‌یابد. ازاین‌رو می‌توان بیشتربودن شیب تابع روان‌سنجی برای مدل هلوب در مقایسه با مدل یورگنز را از روی نمودارهای میله‌ای نرخ تشخیص پیش‌بینی کرد.



(شکل۳-۳): نتایج نرخ تشخیص برای هر کدام از همخوان‌ها در آزمون‌های شنوایی. نرخ تشخیص برای سه آوای /ت/، /س/ و /ش/ در تمامی سیگنال به نوفه‌ها بیشتر از بقیه آواهاست. به‌ویژه در سیگنال به نوفه‌های کم اختلاف بسیار زیادی بین نرخ تشخیص این سه آوا و بقیه وجود دارد. در مقابل همخوان‌های /م/، /و/، /ب/ و /گ/ نرخ تشخیص پایین‌تری را نسبت به بقیه آواها دارند.



(شکل۳-۴): مقایسهٔ نرخ تشخیص مدل هلوب با استفاده از فاصلهٔ اقلیدسی (شکل بالا) و مدل یورگنز با استفاده از فاصلهٔ لگاریتمی (شکل پایین) برحسب مقدار سیگنال به نوفه. مشابه نتایج آزمون شنوایی، نرخ تشخیص در هر دو مدل برای همخوان‌های /ت/، /س/ و /ش/ بیشتر از بقیه است. مدل یورگنز در سیگنال به نوفه‌های پایین‌تر نرخ تشخیص بالاتری نسبت به مدل هلوب دارد.

۳-۳- ماتریس ابهام آواها

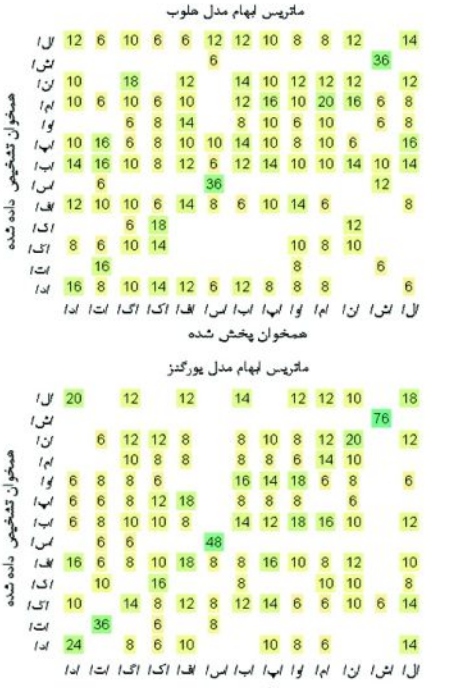
برای مقایسهٔ دقیق‌تر نرخ تشخیص مدل‌ها با آزمون شنوایی از ماتریس ابهام آواها استفاده شده است. هر ستون از ماتریس ابهام مربوط به آواهای پخش‌شده و هر سطر این ماتریس مربوط به آواهای تشخیص داده‌شده هستند. عناصر قطری ماتریس ابهام مربوط به تشخیص‌های صحیح و عناصر غیر قطری بیان‌گر میزان ابهام در نرخ تشخیص آواهاست. تمامی اعداد در ماتریس به درصد بیان شده‌اند. در این پژوهش، برای انجام این مقایسه، مقدار SNR = -15 dB به‌کار گرفته شده است. در سیگنال به نویزهای کمتر با توجه به پایین‌بودن نرخ تشخیص برای تمامی آواها، نتایج مقایسه، چندان مطلوب نخواهد بود. در سیگنال به نوفه‌های بالاتر نیز به‌علت نزدیک‌شدن نرخ تشخیص به نرخ ۱۰۰٪، باز هم نمی‌توان مقایسه را انجام داد؛ لذا، استفاده از نسبت سیگنال به نوفهٔ متوسط ۱۵- مناسب است. ماتریس ابهام آزمون

شنوایی برای SNR = -15dB در شکل (۳-۵) رسم شده است. در این ماتریس، از نرخ‌های تشخیص پایین‌تر از ۶٪ صرف نظر شده است. برای آواهای /ت/، /س/ و /ش/ نرخ تشخیص صحیح، به‌ترتیب، برابر ۵۷٪، ۸۳٪ و ۷۴٪ و خیلی بالاتر از نرخ تشخیص بقیه آواهاست.

شکل (۳-۶) ماتریس‌های ابهام را، به‌ترتیب، برای مدل هلوب با استفاده از فاصلهٔ اقلیدسی و مدل یورگنز با استفاده از فاصلهٔ لگاریتمی نشان می‌دهد. در مقایسه با ماتریس ابهام آزمون شنوایی، پراکندگی نرخ‌های تشخیص در هر دو مدل بسیار بیشتر است. در مدل هلوب نرخ تشخیص بالاتری برای دو همخوان /س/ و /ش/ دیده می‌شود؛ اما در مدل یورگنز هر سه همخوان /س/، /ش/ و /ت/ منطبق بر ماتریس ابهام آزمون شنوایی، دارای نرخ تشخیص بالاتری هستند. ماتریس ابهام مدل یورگنز در مقایسه با مدل هلوب مقادیر بالاتری از متوسط نرخ تشخیص را بر روی قطر اصلی نشان می‌دهد که بیان‌گر قابلیت تشخیص بالاتر این مدل در سیگنال به نویزهای پایین است.

برای مقایسهٔ ماتریس‌های ابهام آزمون شنوایی و دو مدل، از معیار ضرایب همبستگی پیرسون طبق جدول (۳-۲) استفاده شده است. هر کدام از ضرایب همبستگی مربوط به مقایسهٔ هر کدام از ستون‌های ماتریس شکل (۳-۵) با ستون متناظر در ماتریس‌های شکل (۳-۶) می‌باشد. همبستگی بالا برای همخوان‌های /ت/، /س/ و /ش/ در مدل یورگنز با توجه به نرخ تشخیص صحیح در آزمون شنوایی و در مدل‌ها قابل مشاهده است. همچنین، آوای /ک/ در مدل هلوب همبستگی خوبی با آزمون‌های شنوایی دارد. کم‌ترین عدم تطابق در مدل هلوب مربوط به آواهای/گ/، /پ/، /ن/ و /ل/ و در مدل یورگنز مربوط به /د/، /ب/ و /ن/ است. برای مجموع اصوات انفجاری[1] (/د/،/ت/، /گ/ و /ک/) و اصوات انفجاری-انسدادی (/ب/ و /پ/) میانگین مربع ضرایب همبستگی پیرسون در مدل یورگنز برابر ۰/۱۹ و در مدل هلوب برابر ۰/۱۵ است. برای اصوات سایشی[2] (/س/، /ش/، /ف/ و /و/) میانگین این ضرایب دارای مقادیر قابل توجه و به‌ترتیب، برابر ۰/۵۲ و ۰/۴۷ است. برای مجموع تمامی اصوات غیر انفجاری شامل اصوات سایشی، اصوات خیشومی[3] (/م/ و /ن/) و صوت روان[4] /ل/ این ضریب، به‌ترتیب، برای مدل‌های یورگنز و هلوب برابر ۰/۲۷ و ۰/۲۴ است. با توجه به میانگین مربع ضرایب پیرسون، مدل یورگنز در مقایسه با مدل هلوب دارای همبستگی بیشتری با نتایج تست شنوایی است.

(شکل۳-۵): ماتریس ابهام مربوط به آواها در آزمون شنوایی در SNR = -15dB. عناصر قطر اصلی مربوط به نرخ تشخیص صحیح هر آوا می‌باشند. بیشترین نرخ تشخیص مربوط به آواهای /ت/، /س/ و /ش/ است. نرخ‌های تشخیص صحیح کمتر از ۶٪ در ماتریس ابهام گنجانده نشده‌اند.



(شکل ۳-۶): ماتریس ابهام مربوط به مدل هلوب (شکل بالا) و ماتریس ابهام مدل یورگنز (شکل پایین). بیشترین نرخ تشخیص در مدل هلوب مربوط به آواهای /س/ و /ش/ است. در مدل یورگنز، علاوه‌بر این دو آوا، همخوان‌های /ت/ و /د/ نیز دارای نرخ تشخیص صحیح بالاتری هستند.

[1] Plosives
[2] Fricatives
[3] Nasals
[4] Liquid

(جدول۳-۲): مربع ضرایب همبستگی پیرسون بین ستون‌های ماتریس ابهام آزمون شنوایی با ماتریس‌های ابهام مدل هلوب و یورگنز در SNR = -15dB.

همخوان	مدل هلوب با فاصله اقلیدسی	مدل یورگنز با فاصله لگاریتمی
/د/	۰/۰۲۲	۰/۰۰۷
/ت/	۰/۱۵	۰/۸۷
/گ/	۰/۰۰۶	۰/۱۰۴
/ک/	۰/۵۰	۰/۰۵۶
/ف/	۰/۰۵۴	۰/۰۵۳
/س/	۰/۸۶	۰/۹۶
/ب/	۰/۲۲	۰/۰
/پ/	۰/۰۰۹	۰/۱۳
/و/	۰/۰۲۳	۰/۱۵
/م/	۰/۰۳	۰/۰۷۶
/ن/	۰/۰۰۲	۰/۰۰۰۱
/ش/	۰/۹۵	۰/۹۲
/ل/	۰/۰۰۵	۰/۰۶۹

۴- بحث و نتیجه‌گیری

در این پژوهش، نرخ تشخیص همخوان‌ها برای افراد دارای شنوایی طبیعی با استفاده از آزمون شنوایی در سیگنال به نوفه‌های مختلف، اندازه‌گیری شد. در ادامه، نتایج دو مدل میکروسکوپی در پیش‌بینی نرخ تشخیص صحیح آواها مورد بررسی قرار گرفت. با مقایسه توابع روان‌سنجی برازش‌شده بر آزمون شنوایی با توابع روان‌سنجی برازش‌شده بر مدل‌ها، نشان داده شد که هر دو مدل قابلیت پیش‌بینی صحیح مقادیر آستانه ادراک گفتار در آزمون‌های شنوایی را دارند. در مورد متوسط نرخ تشخیص برای تمامی آواها، دقیق‌ترین پیش‌بینی مربوط به مدل یورگنز با فاصله لگاریتمی است. در این مدل با تغییر معیار فاصله لگاریتمی به اقلیدسی، اختلاف آستانه ادراک گفتار مدل با آستانه ادراک گفتار آزمون شنوایی از مقدار ناچیز ۰/۰۳۸۹ دسیبل به مقدار ۲/۵۱ دسیبل افزایش می‌یابد. از این رو انتخاب معیار فاصله در مدل یورگنز اهمیت زیادی دارد. همچنین، با تغییر معیار فاصله در مدل هلوب، اختلاف زیادی در مقادیر آستانه ادراک گفتار به‌وجود می‌آید. تفاوت آستانه ادراک گفتار مدل هلوب با استفاده از معیار اقلیدسی با آستانه ادراک گفتار آزمون شنوایی برابر ۱/۱۹ دسیبل است. با تغییر معیار فاصله اقلیدسی به فاصله لگاریتمی در این مدل، این اختلاف به ۱/۳۶ دسیبل می‌رسد. نکتهٔ قابل توجه در مورد مدل هلوب، بدترشدن پیش‌بینی مدل با تغییر فاصله اقلیدسی به فاصله لگاریتمی است، درحالی‌که در مدل یورگنز عکس این روند مشاهده می‌شود.

یکی از مزیت‌های مدل‌های میکروسکوپی در مقایسه با مدل‌های ماکروسکوپی امکان ارزیابی نرخ تشخیص تک‌تک آواها با استفاده از نمودار میله‌ای نرخ تشخیص صحیح و نیز ماتریس ابهام است. نرخ تشخیص آواهای /ت/، /س/ و /ش/ در آزمون‌های شنوایی دارای مقادیر بیشتری نسبت به بقیه آواها به‌ویژه در سیگنال به نوفه‌های پایین هستند. مدل‌های میکروسکوپی شبیه‌سازی‌شده در این پژوهش نیز نتایج مشابهی را در تشخیص صحیح این سه همخوان نسبت به بقیه نشان می‌دهند. بیشتربودن نرخ تشخیص صحیح مدل یورگنز در مقایسه با مدل هلوب در سیگنال به نوفه‌های کم منجر به افزایش شیب تابع روان‌سنجی در مدل هلوب می‌شود. مدل یورگنز علاوه‌بر تخمین دقیق‌تر مقدار آستانه ادراک گفتار آزمون شنوایی، شیب تابع روانسنجی را نیز با اختلاف کمتری در مقایسه با مدل هلوب تخمین می‌زند.

پژوهش‌هایی که در ادامه کار پیشنهاد می‌شود، مطالعهٔ پیش‌بینی قابلیت فهم در حضور پژواک و گویندگان دیگر است. پژواک می‌تواند قابلیت فهم سیگنال گفتار را تا حد زیادی کاهش دهد. ارائه مدل میکروسکوپی که بتواند علاوه‌بر پیش‌بینی قابلیت فهم در شرایط ایده‌ال، برای شرایط پژواک‌دار نیز تخمین مناسبی از آستانه ادراک گفتار داشته باشد، می‌تواند قدم بعدی در تکمیل مدل‌های میکروسکوپی ادراک گفتار باشد.

تقدیر و تشکر

نویسندگان این مقاله از هم‌فکری خانم نگین صالحی در دانشکده توانبخشی دانشگاه علوم پزشکی تبریز در استفاده از آزمایشگاه صوت و انجام آزمون‌های شنوایی کمال تشکر را دارند.

۵- مراجع

Bronkhorst, A. W., 2000. The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple Talker Conditions, Acta Acoustical United With Acustica, vol. 86, pp. 117–128.

Dau, T., Püschel, D. and Kohlrausch, A., 1996. A quantitative model of the "effective" signals proce-

Wesker, T., Meyer, B., Wagener, K., Anemüller, J., Mertins, A., and Kollmeier, B., 2005. Oldenburg logatom speech corpus (OLLO) for speech recognition experiments with humans and machines, in Proceedings of Interspeech, pp. 1273-1276, Lisbon, Portugal.

مسعود گراوانچی‌زاده مدرک کارشناسی را در رشتهٔ مهندسی الکترونیک در سال ۱۳۶۵ از دانشگاه تبریز اخذ کرده است. سپس، مدارک کارشناسی ارشد و دکتری در رشتهٔ پردازش سیگنال را به‌ترتیب، در سال‌های ۱۳۷۴ و ۱۳۸۰ از دانشگاه Ruhr-University Bochum کسب کرده است. ایشان هم‌اکنون عضو هیأت علمی دانشکدهٔ مهندسی برق و کامپیوتر دانشگاه تبریز با مرتبه علمی دانشیاری هستند. زمینه‌های تحقیقاتی مورد علاقهٔ ایشان بهبود کیفیت گفتار، مکان‌یابی و جداسازی سیگنال گفتار، پردازش سیگنال‌های تصادفی و پردازش سیگنال دوگوشی است .

نشانی رایانامهٔ ایشان عبارت است از:

geravanchizadeh@tabrizu.ac.ir

علی فلاح مدارک کارشناسی و کارشناسی ارشد در رشتهٔ برق گرایش مخابرات را به‌ترتیب در سال‌های ۱۳۸۵ و ۱۳۸۸ از دانشگاه‌های صنعتی امیرکبیر و دانشگاه شاهد اخذ کرده است. وی هم اکنون دانشجوی دکترای رشتهٔ مهندسی برق گرایش مخابرات سیستم در دانشگاه تبریز است. زمینه‌های پژوهشی مورد علاقهٔ ایشان پردازش سیگنال آرایه‌ای، پردازش سیگنال گفتار، پردازش سیگنال دوگوشی و مدل‌های ادراک شنوایی است.

نشانی رایانامهٔ ایشان عبارت است از:

ali.fallah@tabrizu.ac.ir

ssing in the auditory system: I. Model structure, J. Acoust. Soc. Am, vol. 99, pp. 3615–3622.

Dau, T. and Kohlrausch, A., 1997. Modeling auditory processing of amplitude modulation I. Detection and masking with narrowband-carriers, J. Acoust. Soc. Am, vol. 102, pp. 2893–2905.

Dreschler, W. A., Verschuure, H., Ludvigsen, C. and Westermann, S., 2000. ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment, Audiology, vol. 40, pp. 148–157.

Hohmann, V., 2002. Frequency analysis and synthesis using a gammatone filterbank, Acta Acoustical United With Acustica, vol. 88, pp. 433–442.

Holube, I. and Kollmeier, B., 1996. Speech Intelligibility Prediction in Hearing-Impaired Listeners Based on a Psychoacoustically Motivated Perception Model, J. Acoust. Soc. Am, vol. 100, pp. 1703–1716.

Jürgens, T. and Brand, T., 2009. Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model, J. Acoust. Soc. Am, vol. 126, pp. 2635–2648.

Loizou, P., 2007. Speech Enhancement: Theory and Practice, CRC Press, Boca Raton: FL.

Press, W., Teukolsky, S., Vetterling, W. T. and Flannery, B. P., 1992. Numerical Recipes in C, Cambridge University, Press.

Sakoe, H. and Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recog-nition, IEEE Trans. Acoust, Speech, Signal Pro-cess, vol 26, pp. 43–49.

Tchorz, J. and Kollmeier, B., 1999. A model of an auditory perception as front end for automatic speech recognition, J. Acoust. Soc. Am, vol. 106, pp. 2040-2050.

Taghia, J. and Martin, R. , 2014. Objective intelligibility measures based on mutual information for speech subjected to speech enhancement processing, IEEE Trans. Audio, Speech, Signal Process, vol 22, pp. 6–16.

Wang, D. L. and Brown, G., 2005. Computational Auditory Scene Analysis: Principals, Algorithms and Applications, NewYork: IEEE Press, Wiley-Interscience.

میر علی اعتراف اسکوئی مدارک کارشناسی و کارشناسی ارشد را در رشتهٔ فیزیوتراپی، به‌ترتیب، در سال‌های ۱۳۷۰ و ۱۳۷۳ از دانشگاه علوم پزشکی ایران و تهران اخذ کرده است. سپس، با اعطای بورسیهٔ تحصیلی خارج از کشور در سال ۱۳۸۰ عازم کشور کانادا شده و در سال ۱۳۸۵ دکترای تخصصی و Post Doctorate Fellowship خود را در زمینه بیومکانیک عضله از University of Calgary کشور کانادا اخذ کرد. ایشان هم‌اکنون دانشیار دانشکدهٔ توانبخشی دانشگاه علوم پزشکی تبریز بوده و زمینه‌های مورد علاقهٔ ایشان بیومکانیک عضله، بیماری‌های عضلانی اسکلتی، ارگونومی و پردازش سیگنال‌های الکتریکی عضله است.

نشانی رایانامهٔ ایشان عبارت است از:

eterafoskouei@tbzmed.ac.ir