DOI: 10.22041/ijbme.2013.13121

Iranian Journal of Biomedical Engineering 6 (2013) 257- 265

Study of VTLN Method to Recognize Common Speech Disorders in Speech-Therapy of Persian Children

Sh. Azizi [1], F. Towhidkhah [2]*, F. Almasganj [3]

[1]M. Sc, Bioelectric Department, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran, azizishahla@yahoo.com

[2]Associate Professor, Bioelectric Department, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran.

[3]Associate Professor, Bioelectric Department, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran, almas@aut.ac.ir

Abstract

In present work, recognition of isolated word has been studied. The purpose of this research is to increase the performance of children's speech recognizer using Vocal Tract Length Normalization. This recognition system has been created to design a speech therapy software. Recognition of correct and wrong pronunciation and help children to improve it using some feedbacks are the goals of this software. In test phase, some speech data that are related to correct and incorrect pronunciation of 47 words have been utilized. Four Baseline models have been Trained, one for children, one combined model (females and children) and two for Adults (by exploiting one Persian database). Children's model was trained and tested with data that have been collected from 38 children (5 to 8 years old). These experiments were implemented in HTK toolkit. Poor performance was improved using VTLN. Improvement of adult's model was more than children's model.

Key words: Children speech recognition, Vocal Tract Length Normalization, Speaker adaptation, Children speech therapy software, Hidden Markov Models.

*Corresponding author

Address: *Faculty of Biomedical Engineering, Amirkabir University of Technology (Tehran Polytechnic), P.O.Box: 15875-3413, I.R. Iran., Postal Code: 15914, Tehran, I.R. Iran*
Tel: +982164542372
Fax: +982166468186
E-mail: *towhidkhah@aut.ac.ir*

DOI: 10.22041/ijbme.2013.13121

بررسی اثر استفاده از روش تطبیق هنجارسازی طول مسیر صوتی به منظور تشخیص اختلالات گفتاری رایج و گفتاردرمانی کودکان فارسی زبان

شهلا عزیزی[١]، فرزاد توحیدخواه[٢*]، فرشاد الماس‌گنج[٣]

[١]دانش‌آموختهٔ کارشناسی ارشد، گروه بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، تهران
Azizishahla@yahoo.com
[٢]دانشیار، گروه بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، تهران.
[٣]دانشیار، گروه بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، تهران Almas@aut.ac.ir

چکیده

در این مقاله، یک سیستم بازشناسی کلمات جداگانه بررسی شده است. هدف این تحقیق، افزایش کارایی سیستم بازشناسی گفتار کودکان با استفاده از روش هنجارسازی طول مسیر صوتی است. این سیستم بازشناسی، برای استفاده در طراحی نرم‌افزار گفتاردرمانی ایجاد شده است به طوری که این نرم‌افزار با استفاده از سیستم بازشناسی، درست یا نادرست بودن تلفظ کودک را تشخیص می‌دهد و تلاش می‌کند تا با استفاده از بازخوردها گفتار کودک را بهبود بخشد. دادگان گفتاری -که در فاز بازشناسی این سیستم استفاده شده است- مربوط به ٤٧ کلمه و اختلالات تولیدی رایج در آنها است. در این مطالعه، ٤ مدل پایه شامل مدل کودکان، مدل ترکیبی کودکان و زنان و دو مدل بزرگسالان (با استفاده از داده‌های فارس‌دات) آموزش داده شده است. داده‌هایی که برای آموزش و آزمون مدل کودکان استفاده شده، مربوط به ٣٨ کودک در بازه سنی ٥ تا ٨ است. همه مراحل آموزش و آزمون سیستم بازشناسی با استفاده از ابزار *HTK* انجام شده است. نتایج این پژوهش نشان می‌دهد که کارایی کم سیستم بازشناسی با استفاده از روش تطبیق هنجارسازی طول مسیر صوتی افزایش می‌یابد و بهبود مدل بزرگسالان چشمگیرتر از مدل کودکان است.

کلیدواژگان: بازشناسی گفتار کودکان، هنجارسازی طول مسیر صوتی، تطبیق گوینده، نرم‌افزار گفتاردرمانی کودکان، مدل‌های مارکوف پنهان.

*عهده‌دار مکاتبات
نشانی: تهران، خیابن حافظ، روبروی خیابان سمیه، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی، **کدپستی:** ١٥٩١٤
تلفن: ٦٤٥٤٢٣٧٢-٠٢١ ، **دورنگار:** ٦٦٤٦٨١٨٦-٠٢١ ، **پیام نگار:** Towhidkhah@aut.ac.ir

۱- مقدمه

سیستم‌های بازشناسی گفتار کودکان بمنظور ایجاد سیستم‌های مختلفی مانند آموزش خودکار گفتار، آموزش زبان، نرم‌افزارهای سرگرم‌کننده، بازی‌ها، نرم‌افزارهای آموزشی کودکان و غیره مورد توجه قرار گرفته‌اند [۱]. هدف این پژوهش، طراحی یک نرم‌افزار گفتاردرمانی کودکان است که وجود اختلالات گفتاری را در کودکان فارسی زبان شناسایی کند. در این مقاله، اختلال تولیدی مطالعه شده است. در این نوع اختلال، کودک هیچ نوع مشکل آناتومیک و فیزیولوژیک ندارد و تنها بر اثر مجموعه‌ای از عوامل محیطی مثل تقلید از الگوی نامناسب، آموزش نادرست والدین و اطرافیان دچار مشکل شده است. در این نرم‌افزار آزمون و آموزش ٤٧ کلمه انجام می‌شود و برای هر کدام از این کلمات مجموعه‌ای از همسان‌ها تعریف شده است. اگر اختلال کودک در بازه اختلالات تعریف شده برای آن کلمه نباشد، یعنی مشکل کودک حاد است و باید به گفتاردرمانگر مراجعه کند. در همسان‌های این کلمات اختلالاتی چون حذف واج، جایگزینی واج با واج دیگر و جابجایی دو واج کناری در نظر گرفته شده است.

برای آموزش و آزمون سیستم بازشناسی گفتار کودکان به نمونه‌های گفتاری کودکان نیاز است؛ ولی جمع‌آوری نمونه از کودکان بسیار سخت است، زیرا اغلب کودکان، نفس‌آلود، ناواضح و آرام صحبت می‌کنند، خجالت می‌کشند و یا توانایی و سواد خواندن متن را ندارند. به همین دلیل محققان به دنبال راه‌هایی هستند تا از گفتار بزرگسالان برای بازشناسی گفتار کودکان استفاده کنند؛ ولی بیشتر ویژگی‌های گفتاری کودکان با ویژگی‌های گفتاری بزرگسالان متفاوت است. برای نمونه، کودکان دارای طول مسیر صوتی کوتاه‌تری در مقایسه با بزرگسالان هستند که این منجر به افزایش فرکانس گام و فرمنت‌ها در کودکان و در نتیجه زیر بودن صدای آنها می‌شود [۲]. افزون بر این، پارامترهای گفتاری کودکان به سن آنها بستگی دارد و این به دلیل دگرگونی‌های فیزیولوژیک و آناتومیکی است که به هنگام رشد ایجاد می‌شود. از مهم‌ترین مشخصه‌های گفتاری وابسته به سن کودک می‌توان به: مشخصه‌های زمانی و فرکانسی و تلفظ نادرست و نفس‌آلود اشاره کرد [۲،۳،٤،٥]. به طور کلی می‌توان گفت بازشناسی گفتار کودکان بسیار سخت‌تر از بازشناسی گفتار بزرگسالان است [٦،٥].

همان گونه که گفته شد، گفتار کودکان و بزرگسالان تفاوت‌های زیادی با یکدیگر دارند. همچنین نتایج پژوهش‌های پیشین نشان می‌دهد که به دلیل این تفاوت‌ها، سیستم‌هایی که با گفتار بزرگسالان آموزش دیده‌اند در مقایسه با سیستم‌هایی که با گفتار کودکان تعلیم دیده‌اند، برای بازشناسی گفتار کودکان کارایی کمتری دارند [۸،۷،٤،۱]. از این رو، محققان از برخی روش‌های تطبیق استفاده می‌کنند تا تفاوت‌های موجود در گفتار کودکان و بزرگسالان را کاهش و در پی آن، کارایی سیستم بازشناسی گفتار کودکان را افزایش دهند. برخی روش‌های تطبیق عبارتند از: روش هنجارسازی طول مسیر صوتی[1]، تطبیق گوینده (رگرسیون خطی با بیشینه درست نمایی[2] و رگرسیون خطی با بیشینه درست‌نمایی محدود[3])، مدل‌سازی زبان و مدل‌سازی تغییرات تلفظ [۹].

در این مطالعه برای بهبود کارایی سیستم بازشناسی گفتار کودکان، از روش تطبیق هنجارسازی طول مسیر صوتی استفاده شده است. این روش تطبیق، در مرحله استخراج ویژگی از آموزش مدل اعمال می‌شود. در این روش به صورت خطی، دوخطی و غیرخطی، مقیاس فرکانس با استفاده از یک ضریب پیچش تغییر داده می‌شود [٥]. نتایج پژوهش‌های پیشین نشان می‌دهد که روش تطبیق هنجارسازی طول مسیر صوتی روشی مناسب بمنظور بهبود مدل گفتار بزرگسالان برای استفاده در بازشناسی گفتار کودکان است [۹،۷،٤،۲]. به سخن دیگر با استفاده از این روش می‌توان مشخصه‌های گفتاری بزرگسالان را

[1] Vocal Tract Length Normalization [2] Maximum Likelihood Linear Regression [3] Constrained Maximum Likelihood Linear Regression

به مشخصه‌های گفتاری کودکان نزدیک کرد؛ آنگاه مدلی که با استفاده از این بردار ویژگی‌ها آموزش داده می‌شود، کارایی بیشتری در مقایسه با مدل بزرگسالان برای بازشناسی گفتار کودکان خواهد داشت. اگر روش هنجارسازی طول مسیر صوتی در هر دو مرحله آموزش و آزمون اعمال شود، سیستم بازشناسی کارایی بهتری خواهد داشت [۷،۲].

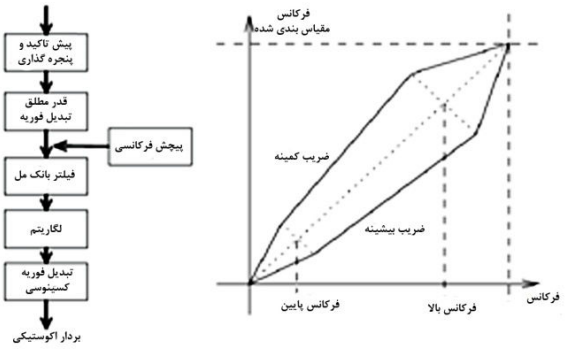
این مقاله دارای چهار قسمت اصلی است: در قسمت اول مقدمه‌ای بر این مطالعه آورده شده است. در قسمت دوم در مورد روش تطبیق و پایگاه داده استفاده شده توضیحاتی داده شده است. در قسمت سوم، الگوریتم‌های استفاده شده و نتایج به دست آمده و در قسمت چهارم جمع‌بندی این پژوهش آورده شده است.

۲- روش

در این قسمت توضیحاتی برای روش تطبیق و پایگاه داده مورد استفاده، آورده شده است.

۲-۱- روش تطبیق هنجارسازی طول مسیر صوتی

روش هنجارسازی طول مسیر صوتی بر این اصل عمل می‌کند که طول مسیر صوتی در افراد مختلف، متفاوت است [۱۰] و این تفاوت‌ها را می‌توان با تغییر خطی مقیاس محور فرکانس کاهش داد. افرادی که طول محدوده صوتی آنها کوتاه‌تر است، صدای زیرتری دارند و برعکس. این تفاوت، باعث عدم تطابق گوینده‌های آموزش با آزمون سیستم بازشناسی می‌شود و بنابراین کارایی آن کاهش می‌یابد. در طول مرحله استخراج ویژگی می‌توان این تفاوت‌ها را کاهش داد. به این فرایند، کاهش محدوده صوتی گفته می‌شود. اساس این روش، تخمین عامل α یا ضریب پیچش است به طوری که مقیاس فرکانس طیف متناسب با این ضریب تغییر می‌کند [۸]. تغییر مقیاس فرکانس می‌تواند به صورت تکه‌ای خطی، دوخطی و غیرخطی باشد که اغلب از مدل تکه‌ای خطی استفاده می‌شود[۷]. در شکل (۱) روندنمای روش تطبیق (سمت چپ) و همچنین تابع پیچش فرکانسی (سمت راست) نشان داده شده است.



شکل(۱)- تابع پیچش فرکانسی(الف) و روندنمای روش تطبیق (ب).

انتخاب ضریب پیچش مناسب در اعمال روش تطبیق بسیار مهم است. برای محاسبه ضریب پیچش روش‌های متعددی مانند تابع خطی وجود دارد که گاه مبتنی بر بیشینه درست‌نمایی و گاه بر پایه پارامترهای خاص صدای گوینده مثل فرمنت‌ها است؛ زیرا فرکانس فرمنت‌ها ارتباط معکوس با طول مسیر صوتی دارد[۱۱] .

فرض می‌شود که $O_\alpha(t)$ بردار ویژگی سیگنال ورودی با ضریب پیچش w ،(α) برچسب آوایی ورودی و λ پارامترهای مدل پایه است. آنگاه مراحل تخمین ضریب پیچش به صورت زیر است:

۱- برای هر گوینده، α برابر با ۱ قرار داده می‌شود.

۲- برای هر گوینده:

- مقدار رابطه (۱) محاسبه می‌شود:

$$S^*_{t,r} = \arg\max P\left(O^r_\alpha, S_t \middle| \lambda, w^r\right) \quad (۱)$$

- طبق $S^*_{t,r}$ ثابت و محدوده α که برابر با $[1,h]$ است، رابطه (۲) محاسبه می‌شود:

$$\alpha^*_r = \arg\max P(O^r_\alpha(t) \middle| S^*_{t,r}, \lambda, w^r) \quad (۲)$$

پارامترهای مدل طبق α تغییر می‌کنند.

۳- $\lambda^* = argmax\ P(O_\alpha(t)|\lambda, w^r)$

٤- مقدار λ برابر با λ^* ومقدار α برابر با α^* قرار داده می‌شود.

٥- اگر

مقدار مشخص خطا $< |\alpha^* - \alpha|$

آنگاه محاسبات از مرحله ۲ تکرار می‌شود.

دو تابع پیچش معمول عبارتند از: تابع تکه‌ای خطی و تابع غیر خطی [۱۲،٦].

تابع تکه‌ای خطی: این معادله دارای دو پارامتر α و ω_0 است. ω_0 مقداری ثابت است که به صورت تجربی به دست می‌آید. اگر f برابر با f^α و ω_0 همان فرکانس نایکوئیست[٤] باشد، آنگاه تابع به صورت تکه‌ای خطی است که در بخش(الف) شکل(۱) نشان داده شده است. معادله تابع تکه‌ای خطی به صورت رابطه (۳) است:

$$\psi_\alpha(w, w_0) = \begin{cases} \alpha\omega & \omega < \omega_0 \\ \alpha\omega_0 + \frac{\pi - \alpha\omega_0}{\pi - \omega_0}(\omega - \omega_0) = b\omega + c & \omega > \omega_0 \end{cases} \quad (۳)$$

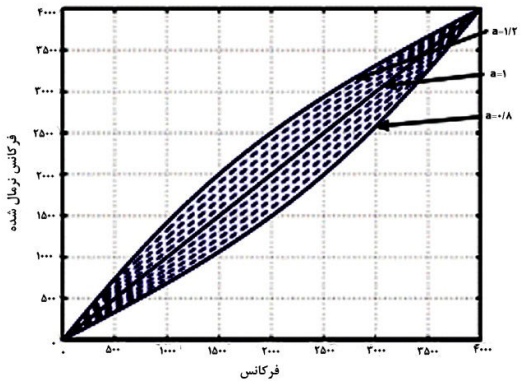
تابع غیرخطی: این تابع تنها یک پارامتر دارد و از تابع پیشین ساده‌تر است که در شکل ۲ دیده می‌شود. معادله این تابع به صورت رابطه (٤) است:

$$\psi_\alpha = \omega + 2.\tan^{-1}\left\{\frac{(1-\alpha)\sin(\omega)}{1-(1-\alpha)\cos(\omega)}\right\} \quad (٤)$$

در این مطالعه، از تابع تکه‌ای خطی برای تخمین α استفاده شد. برای اعمال این روش، در آغاز ضریب پیچش بهینه با استفاده از الگوریتم جستجو به دست آمد. بدین گونه که برای این ضریب، بازه‌ای بین ۰/۸ تا ۱/۲ در نظر گرفته شد و در این بازه، با گام ۰/۰۱، برای ضرایب مختلف، بردار ویژگی محاسبه شد. سپس با استفاده از این بردار ویژگی و یک مدل پایه، بازشناسی انجام شد و ضریب پیچشی که بردار ویژگی ناشی از آن دارای بیشینه درست‌نمایی بود، به عنوان ضریب پیچش بهینه انتخاب شد. لازم است ذکر شود که در این روش، باید برای هر گوینده یک ضریب پیچش بهینه انتخاب شود. بعد از تخمین ضریب پیچش بهینه، این ضریب در مرحله استخراج ویژگی اعمال می‌شود و سپس با استفاده از ویژگی‌های جدید، مدلی بهنجار، آموزش داده می‌شود که در آزمایش‌های بازشناسی از آن استفاده شد.

۲-۲- پایگاه داده

برای انجام این مطالعه از دو مجموعه داده استفاده شد. اولین مجموعه داده، پایگاه داده فارس‌دات است. این دادگان شامل ٦۰۸ داده گفتاری از ۳۰٤ گوینده فارسی زبان (۹۹ زن و ۲۰۵ مرد) است به طوری که هر گوینده در دو مرحله گفتارش ضبط شده است (در هر مرحله گوینده ۱۰ جمله را ادا کرده است).



شکل (۲)- پیچش فرکانسی

در جمع‌آوری این مجموعه داده از همه بازه‌های سنی، تمام لهجه‌های زبان فارسی و ۳۸٦ جمله استفاده شده است. این پایگاه داده در اتاق اکوستیک آزمایشگاه زبان‌شناسی دانشگاه تهران جمع‌آوری شده است [٤]. در آموزش مدل بزرگسالان از داده مربوط به ۲۵۰ گوینده (شامل ۵۰۰۰ جمله) و در آموزش مدل زنان از داده گفتاری مربوط به ۹۹ گوینده زن استفاده شده است.

پایگاه داده دیگر مربوط به ۳۸ کودک در بازه سنی ۵ تا ۸ است. در جمع‌آوری داده از ۲۵ کودک خواسته شده است تا متنی را که شامل ۵۰ جمله است (از جملات دادگان

[4] Nyquist

فارس‌دات)؛ بخوانند. همچنین از ۱۳ کودک دیگر خواسته شده است تا فهرستی از ۲۰۰ کلمه را بخوانند تا صدای آنها ضبط شود. این کلمات، شامل ٤٧ کلمه خاص و همسان‌های آنها است. برای هر کلمه تعدادی همسان در نظر گرفته شده است، به طوری که همسان‌ها، همان اختلالات تولیدی رایج در گفتار کودکان هستند. در واقع کودکی که مشکل آناتومیک ندارد، باید این کلمات را درست یا به صورت یکی از همسان‌ها ادا کند وگرنه مشکل کودک جدی است و باید تحت درمان قرار بگیرد. از ۱۲۵۰ جمله در دسترس، برای آموزش مدل کودکان و از ۲۵۰۰ کلمه برای مرحله آزمون استفاده شده است.

در این مطالعه نمونه‌های مزبور با فرکانس نمونه‌برداری ١٦ کیلوهرتز و به صورت ١٦ بیتی برای هر نمونه، از حالت آنالوگ به دیجیتال تبدیل، و برای پردازش‌های بعدی در دیسک سخت رایانه ذخیره شدند. لازم است ذکر شود که جمع‌آوری نمونه از کودکان بسیار سخت است؛ زیرا اغلب کودکان نفس‌آلود، ناواضح و آرام صحبت می‌کنند، خجالت می‌کشند و یا توانایی خواندن متن را ندارند و گاهی اوقات همکاری نمی‌کنند. در جمع‌آوری داده، اگر کودک توانایی خواندن و نوشتن نداشت، بعد از بازخوانی آن از سوی یک نفر، آن جمله یا واژه تکرار می‌شد. همچنین با وجود تلاش برای انتخاب یک محیط بدون نویز، همچنین صداهای ناخواسته‌ای وجود داشت که موجب افت کیفیت نمونه‌های ضبط شده، می‌شد. شایان ذکر است تمام این کودکان سالم بودند و هیچ گونه مشکل گفتاری نداشتند.

۳- نتایج بازشناسی

تمام مراحل آموزش، بازشناسی و تطبیق با استفاده از ابزار *HTK*[5] انجام شد. به طور کلی طراحی سیستم بازشناسی دارای سه مرحله اصلی است: ۱) استخراج ویژگی، ۲) آموزش مدل و ۳) بازشناسی. در ادامه، هر کدام از این قسمت‌ها توضیح داده می‌شود.

۳-۱- استخراج ویژگی

نخستین گام در بازشناسی گفتار، استخراج ویژگی است. در این مرحله برای هر فریم گفتاری یک بردار ویژگی استخراج می‌شود. بردار ویژگی استخراج شده در این مطالعه، یک بردار ۳۹ بعدی است که شامل ۱۲ ضریب کپستروم، یک ضریب انرژی، ۱۳ ضریب دلتا و ۱۳ ضریب دلتا- دلتا کپستروم است. برای فریم‌بندی، طول هر فریم ۲۵ میلی‌ثانیه و هم‌پوشانی فریم‌ها ۱۵ میلی‌ثانیه در نظر گرفته می‌شود. نوع پنجره انتخاب شده همینگ[6]، ضریب پیش تأکید ۰/۹۷۵ و تعداد فیلترهای استفاده شده ۲۲ عدد است.

۳-۲- آموزش سیستم

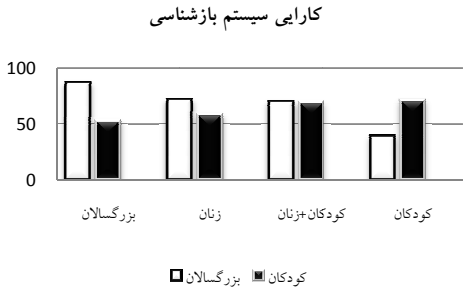
مدل‌های مارکوف پنهان[7]، ابزاری کارآمد برای مدل‌سازی آماری سیگنال متغیر با زمان و ناایستای گفتار است. در این پژوهش برای مدل هم‌خوان‌ها ۷ حالت و برای مدل واکه‌ها ۵ حالت با ۸ مخلوط گوسی در نظر گرفته شده است که با ساختار چپ به راست به هم متصل هستند. مدل‌های مارکوف آموزش داده شده، مدل‌های مبتنی بر مدل سه‌واجی هستند.

۳-۳- بازشناسی

در این مطالعه، ٤ مدل پایه آموزش داده شده است: مدل بزرگسالان که با استفاده از داده مربوط به زنان و مردان آموزش داده شده است، مدل زنان که با استفاده از ۱۹۸ داده گفتاری از ۹۹ گوینده زن تعلیم دیده است، مدل ترکیبی زنان و کودکان که با استفاده از ۲۲۳ داده گفتاری زنان و کودکان آموزش دیده است و مدل کودکان که با ۱۲۵۰ جمله از کودکان تعلیم دیده است. در مرحلهٔ اول این مطالعه، بازشناسی گفتار با مدل‌های مختلف و با استفاده از داده‌های آزمون کودکان و بزرگسالان انجام شد. شکل (۳) نمودار مربوط به نتایج بازشناسی را نشان می‌دهد. نتایج حاصل از این آزمایش، نتایج پژوهش‌های پیشین را تأیید می‌کند. همان گونه که انتظار می‌رفت مدل بزرگسالان در بازشناسی گفتار کودکان و مدل کودکان برای بازشناسی گفتار بزرگسالان دارای کارایی کمی است. بنابراین چنانچه بازشناسی گفتار

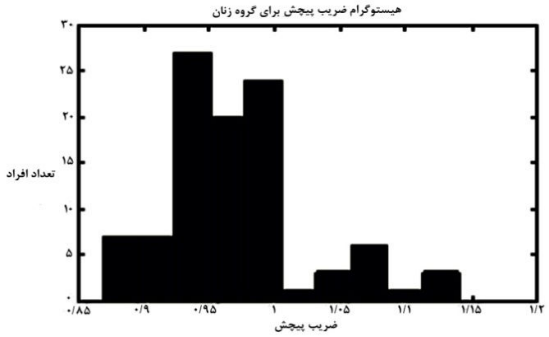
[5]Hidden Markov Model Toolkit [6]Hamming Window [7]Hidden Markov Models

کودکان با مدلی انجام شود که مشابهت بیشتری به مدل کودکان داشته باشد، کارایی آن بهتر خواهد بود. نمودار نشان می‌دهد که مدل کودکان دارای بیشترین کارایی برای بازشناسی گفتار کودکان و مدل بزرگسالان دارای بیشترین کارایی در بازشناسی گفتار بزرگسالان است. از پارامترهای مهمی که در رسیدن به این نتایج نقش دارد، تفاوت در طول مسیر صوتی افراد است. در این مقاله تلاش شده است تا با استفاده از روش تطبیق مذکور، کارایی مدل بزرگسالان برای بازشناسی گفتار کودکان بهبود داده شود. در اولین گام ضریب پیچش برای دو گروه مردان و زنان براورد شد. نمایش آماری مربوط به ضریب پیچش دو گروه در شکل (٤) و (٥) نشان داده شده است. همان طور که مشاهده می‌شود، ضریب پیچش برای گروه زنان کمتر از یک و برای گروه مردان بزرگتر از یک به دست آمده که این به دلیل طول مسیر صوتی کوتاه در زنان و طول مسیر صوتی بلند در مردان است.



شکل (۳)- کارایی سیستم‌های مختلف برای بازشناسی گفتار کودکان و بزرگسالان

نتایج بازشناسی گفتار با مدل‌های مختلف در جدول (۱) آورده شده است. نتایج نشان می‌دهد که کارایی سیستم‌های بازشناسی گفتار بعد از اعمال روش تطبیق هنجارسازی طول مسیر صوتی بهبود یافته است. در تمام آزمایش‌های مربوط به این جدول، از داده‌های گفتاری کودکان برای بازشناسی استفاده شده است.



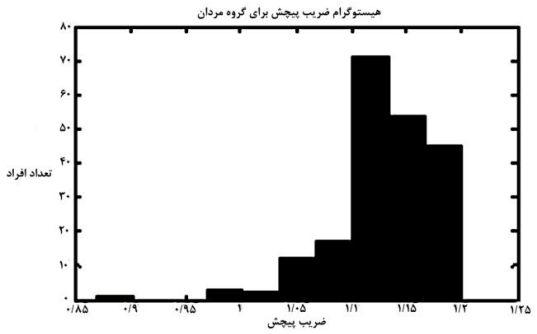
شکل (٤)- نمایش آماری ضریب پیچش گروه زنان

همچنین نتایج نشان می‌دهد که کارایی مدل بهنجار کودکان در مقایسه با مدل پایه –هنگامی که روش تطبیق مورد نظر در یکی از دو مرحله آموزش و آزمون گفتار کودکان اعمال شده باشد– کاهش یافته است. اما وقتی تطبیق در هر دو مرحله آموزش و آزمون اعمال شده، مدل بهنجار در مقایسه با مدل پایه، بهبود ناچیزی (به دلیل تفاوت کم در طول مسیر صوتی) داشته است. برای گروه بزرگسالان نیز، زمانی که تطبیق در هر دو مرحله آموزش و آزمون و یا یکی از این دو مرحله اعمال شود، کارایی مدل بهنجار در مقایسه با مدل پایه برای بازشناسی گفتار کودکان بهبود چشمگیری یافته است. این بهبود چشمگیر به دلیل کاهش تفاوت موجود در گفتار بزرگسالان و کودکان با استفاده از این روش تطبیق است. با توجه به این نتایج می‌توان گفت برای تمام مدل‌ها، بعد از اعمال این روش تطبیق در هر دو مرحله آموزش مدل و آزمون گفتار کودکان، کارایی سیستم بازشناسی بهبود یافته است. اما زمانی که تطبیق تنها در یکی از دو مرحله آزمون یا آموزش اعمال شده باشد؛ در صورت تفاوت زیاد در طول مسیر صوتی موجب بهبود و در صورت تفاوت کم در طول مسیر صوتی، باعث افت کارایی شده است. به دلیل تفاوت‌های موجود در مسیرهای صوتی کودکان و بزرگسالان، سیستم بازشناسی گفتار کودکان با استفاده از مدل بزرگسالان دارای کارایی کمی است؛ اما همین تفاوت باعث بهبود بیشتر کارایی مدل بزرگسالان بعد از اعمال روش تطبیق طول مسیر

صوتی برای بازشناسی گفتار کودکان می‌شود. زمانی که تطبیق فقط در مرحله آزمون اعمال شده باشد، سیستم بازشناسی دارای ضعیف‌ترین کارایی است.

جدول(١)- نتایج بازشناسی گفتار کودکان با مدل‌های مختلف بر حسب درصد کارایی مدل

مدل	اعمال روش تطبیق		کارایی مدل بر حسب درصد
	آموزش	آزمون	
بزرگسالان	خیر	خیر	٥٢/٧٣
	خیر	بلی	٦٠/٢٨
	بلی	خیر	٦٣/١٩
	بلی	بلی	٦٤/٥٥
زنان و کودکان	خیر	خیر	٦٩/٤٢
	خیر	بلی	٦٨/٠٨
	بلی	خیر	٦٨/٤٨
	بلی	بلی	٦٩/٦١
کودکان	خیر	خیر	٧١/٥٦
	خیر	بلی	٦٩/٣٦
	بلی	خیر	٧٢/٢٢
	بلی	بلی	٧١/٨٩



شکل (٥)- نمایش آماری ضریب پیچش گروه مردان

این نتایج نشان می‌دهد که با وجود استفاده از روش تطبیق، اگر چه مدل بزرگسالان بهبود قابل ملاحظه‌ای یافته است (١٢ درصد)؛ کارایی آن همچنان پایین‌تر از کارایی مدل کودکان برای بازشناسی گفتار کودکان است. چنین نتیجه‌ای به این دلیل است که تفاوت موجود در گفتار کودکان و بزرگسالان دلایل دیگری به غیر از تفاوت در مسیر صوتی دارد. به عنوان مثال، گفتار کودکان با تنوع طیفی بیشتر و مدت زمان قطعه‌بندی طولانی‌تری در مقایسه با گفتار بزرگسالان مشخص می‌شود.

جدول(٢)- نتایج بازشناسی گفتار کودکان با استفاده از مدل کودکان قبل و بعد از اعمال روش تطبیق برای چند گوینده

شماره گوینده	کارایی بر حسب درصد	
	مدل پایه	مدل بهنجار
اول	٧٦/٩٢	٧٧/٤٤
چهارم	٧٢/٩٢	٧٦/٠٤
پنجم	٧٣/٣٦	٧٦/٣٤
هفتم	٧٢/٦٣	٧٤/٢١
میانگین	٧٣/٩٥	٧٦

جدول(٢)، نتایج بازشناسی گفتار کودکان با استفاده از مدل پایه و بهنجار کودکان را، برای چند گوینده کودک نشان می‌دهد. برای این کودکان، کارایی سیستم بازشناسی گفتار بعد از اعمال روش تطبیق هنجارسازی طول مسیر صوتی افزایش یافته است. همان گونه که مشاهده می‌شود، کارایی سیستم برای این چهار گوینده به طور متوسط ٣ درصد افزایش یافته است.

از آنجایی که هدف ایجاد این سیستم بازشناسی گفتار کودکان، طراحی یک نرم‌افزار گفتاردرمانی است و ورودی سیستم کلمات (٤٧ کلمه و تلفظ‌های درست و نادرست رایج در آن‌ها شامل اختلالات تولیدی) هستند، کارایی سیستم بسیار مهم است. اگر در نرم‌افزار گفتاردرمانی مورد نظر، هدف تنها تشخیص درست یا نادرست بودن تلفظ کودک باشد و نیازی به تشخیص این نباشد که از بین همسان‌های یک کلمه کدام تلفظ شده است؛ کارایی سیستم بازشناسی به حد قابل قبول ٩٠ درصد می‌رسد. ولی اگر لازم باشد نرم‌افزار نوع اختلال را نیز تشخیص دهد، کارایی آن به ٧٢ درصد افت می‌کند.

٤- نتیجه‌گیری

امروزه نرم‌افزارهای گفتاردرمانی می‌توانند کمک شایانی به کودک و گفتاردرمان‌گر از دیدگاه صرفه‌جویی وقت و هزینه کنند. در این مقاله ایجاد سیستم بازشناسی گفتار کودکان بمنظور استفاده در طراحی نرم‌افزار گفتاردرمانی بررسی شده است. نتایج نشان می‌دهد که تشخیص کلمات متفاوت، برای رایانه بسیار آسان‌تر از تشخیص تلفظ‌های متفاوت کلمه است. بنابراین اگر در این مطالعه هدف تشخیص کلمه‌ای از بین کلمات معنی‌دار بود، پیچیدگی کار و آمار خطا بسیار کمتر می‌شد. دقت این نرم‌افزار، تنها در تشخیص تلفظ‌های درست و نادرست، ۹۰ درصد است ولی هنگامی که موضوع تشخیص نوع اشتباه نیز مطرح باشد، دقت سیستم به ۷۲ درصد می‌رسد.

ایجاد سیستم بازشناسی گفتار کودکان با کارایی مناسب، اغلب کار دشواری است؛ زیرا جمع‌آوری نمونه از کودکان بسیار سخت است. بنابراین محققان تلاش می‌کنند از مدل گفتاری بزرگسالان برای بازشناسی گفتار کودکان استفاده کنند. اگر چه به دلیل تفاوت‌هایی که بین گفتار بزرگسالان و کودکان وجود دارد، کارایی این سیستم‌ها بسیار کم است. یکی از این تفاوت‌ها اختلاف در مسیر صوتی بزرگسالان و کودکان است. افراد با طول مسیر صوتی کوتاه‌تر، دارای صدایی با فرکانس گام بیشتر (کودکان) و افراد با طول مسیر صوتی بلندتر، دارای فرکانس گام کوچک‌تری (بزرگسالان) هستند. تحقیقات نشان می‌دهد که می‌توان با استفاده از روش‌هایی مثل تطبیق هنجارسازی طول مسیر صوتی، این تفاوت‌ها را کاهش داد و کارایی سیستم را بهبود بخشید.

اساس روش تطبیق، تغییر مقیاس محور فرکانس است. نتایج این مطالعه نشان می‌دهد چنانچه تطبیق در هر دو مرحله آموزش و آزمون به سیستم اعمال شود، کارایی سیستم همواره بهبود خواهد داشت. یکی از آثار ناخواسته این روش آن است که مقیاس قسمت‌های غیرگفتاری سیگنال نیز تغییر می‌کند. گاهی برای انتخاب ضریب پیچش، قسمت‌هایی از سیگنال که گفتاری نیستند، حذف می‌شوند. شایان ذکر است که تابع تکه‌ای خطی در برابر این موضوع مقاوم‌تر است.

به طور کلی اعمال روش تطبیق هنجارسازی طول مسیر صوتی- زمانی که مدل با گفتار بزرگسالان آموزش می‌بینند و بازشناسی با گفتار کودکان انجام می‌شود- بیشترین بهبود را دارد. در واقع با استفاده از این روش می‌توان مدل بزرگسالان را برای استفاده در بازشناسی گفتار کودکان بهبود بخشید. از سوی دیگر، همان گونه که مشاهده شد با وجود بهبود زیاد در کارایی مدل بزرگسالان در بازشناسی گفتار کودکان، هنوز کارایی این مدل ضعیف‌تر از کارایی مدل کودکان است. از این رو نمی‌توان از این سیستم در طراحی نرم‌افزار گفتاردرمانی بهره برد. بنابراین برای بهبود بیشتر سیستم می‌توان داده‌های آموزشی را به سمت داده‌هایی با نویز کمتر بهبود بخشید و یا از روش‌های تطبیق دیگر مثل رگرسیون خطی با بیشینه درست‌نمایی استفاده کرد.

در این مقاله به انواع مختلف خطای تولیدی شامل جایگزینی، حذفی و جابجایی توجه و برای به دست آوردن سرعت و قابلیت بیشتر از ابزار بازشناسی گفتار *HTK* استفاده شده است.

سپاسگزاری

بدین وسیله از همکاری مدیریت و پرسنل محترم دبستان پسرانه شهید مهدی‌زاده ۱ (منطقه ۵ آموزش و پرورش تهران بزرگ) در جمع‌آوری داده‌ها سپاسگزاری می‌شود.

۵- مراجع

[1] Potamianos A., Robust Recognition of Children's Speech; IEEE transactions on speech and audio processing, 2003; 11(6).
[2] Giuliani D., Gerosa M., investigating recognition of children's speech; ITC-irst, Center ob Scientific and Technological Research, Trento, Italy, 2003.
[3] Potamianos A., Narayanan S., Acoustics of children's speech:Developmental changes of temporal and spectral parameters; Journal of Acoust. Soc. Amer, 1999; 105: 1455–1468.

[4] Tadayon Tabrizi Gh., HMM-Based Recognition and Adaptation of Persian Children's Speech; Department of Computer, Science and Research Branch, Islamic Azad University, Tehran, Iran, Contemporary Engineering Sciences, 2011; 4(5): 221 – 228.

[٥] تدین تبریزی ق.، ستایشی س.، ارائه روشی مبتنی بر نرمالسازی اکوستیکی و خوشه بندی برای بهبود بازشناسی گفتار کودکان فارسی زبان؛ مجله فنی مهندسی دانشگاه آزاد اسلامی مشهد، دوره سوم، شماره اول، زمستان ٨٨.

[٦] باباعلی ب.، صامتی ح.، ویسی ه.، بکارگیری نرمالسازی اثر طول مسیر صوتی گوینده‌ها در سیستم بازشناسی گفتار پیوسته فارسی مبتنی بر مدل مخفی مارکوف؛ سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران، ١٣٨٦.

[7] Elenius D., Blomberg M., Adaptation and Normalization Experiments in Speech Recognition for 4 to 8 Year old Children; Department of Speech Music and Hearing KTH, Stockholm, Sweden, INTERSPEECH, 2005.

[5] Sanand D.R., Kurimo M., A Study on Combining VTLN and SAT to Improve the Performance of Automatic Speech Recognition; Adaptive Informatics Research Center, Aalto University, Finland, Interspeech, 2011.

[9] Elenius D., Adaptation techniques for children's speech recognition; KTH/TMH, 2004.

[10] Young S., Evermann G., et al., "The HTK book", Cambridge University Engineering Department, 2006.

[11]Evandro B., Gouvêa, Acoustic-feature-based Frequency Warping for Speaker Normalization; Department of Electrical and Computer Engineering Carnegie Mellon University, Pittsburgh, Pennsylvania December, 1998.

[12] Feng H., Yuan C., Li Y., Speaker Normalization Method Based On the Piece-Wise Linear Frequency Warping; dept. computer and information engineering, 2009 International Conference on E-Learning, E-Business, Enterprise Information Systems, and E-Government.