

تحلیل سیگنال گفتار بیماران آلزایمری فارسی زبان توسط سیستم مدل شنوایی

مریم مومنی^۱، مهدیه رحمانی^۲

^۱ استادیار، دانشکده فنی و مهندسی، دانشگاه اراک، m-momeni@araku.ac.ir

^۲ کارشناسی مهندسی برق مخابرات، دانشگاه اراک، matie.rah@gmail.com

چکیده

تشخیص و نظارت بر بیماری آلزایمر امروزه توجه محققین بسیاری را در دهه اخیر جلب کرده است و لذا روش‌های باز شناخت سیگنال گفتار کمک شایان توجهی در شناخت بهنگام بیماری آلزایمر داشته‌اند. هدف این مقاله تشخیص سیگنال گفتار بیماران آلزایمری از افراد سالم با استفاده از سیستم مدل شنوایی انسان است که به منظور دستیابی به این هدف پایگاه داده‌ای متشکل از سیگنال گفتار بیماران آلزایمری و افراد سالم فارسی زبان جمع‌آوری شده است. سیستم شنوایی انسان یکی از بهترین تشخیص دهنده‌های صوت است لذا می‌تواند بعنوان مدل مناسبی در شناسایی سیگنال گفتار استفاده شود و حاصل آن خروجی چهاربعدي در راستای زمان، نرخ، فرکانس و مقیاس است. از خروجی بدست آمده در راستای زمان میانگین‌گیری و از داده سه‌بعدي حاصله بیشینه مقدار ویژگی‌های فرکانسی، مدولاسیون طیفی و زمانی به ترتیب در راستای فرکانس، مقیاس و نرخ برای بیماران آلزایمری و افراد سالم بعنوان بردار ویژگی استخراج می‌شود. بعلاوه، بیشینه مقدار انرژی اسپکتروم نیز به بردار ویژگی حاصله اضافه می‌شود. طبقه‌بندی نتایج استخراج شده از ۱۰۰ داده از افراد سالم و بیماران آلزایمری توسط ماشین بردار پشتیبان در مقایسه با روش‌های زمانی و فرکانسی همچون بسته تبدیل موجک، ضرایب کپسترال فرکانس مل، نرخ عبور از صفر و چرخش طیف نتیجه شایان توجهی را دربرداشت.

کلیدواژه

بیماری آلزایمر، سیگنال گفتار، سیستم مدل شنوایی انسان، مدولاسیون طیفی

مقدمه

در تحلیل سیگنال گفتار بیماران آلزایمری دیده شده است که گفتار بیماران آلزایمری آرام‌تر و به همراه تعداد مکث‌های زیادی است. آن‌ها مدت زمان زیادی را صرف پیدا نمودن کلمه مناسب می‌کنند و اغلب جملات خود را ناتمام رها می‌کنند که این موضوع تاثیر مستقیم در استفاده از سیگنال گفتار و استخراج ویژگی‌های آکوستیک آن دارد [۳-۷]. بعلاوه، ویژگی‌های نوایی مثل فرکانس پایه، فرمونت‌ها [۸، ۹]، ویژگی‌های زبان، فرکانس مثل ضرایب کپسترال فرکانس مل (MFCC^۱) در حوزه پردازش سیگنال گفتار بیماران آلزایمری و افراد سالم بسیار پرکاربرد است [۱۰-۱۲]. مرجع [۱۳] ویژگی‌هایی را که وابسته به تغییرات زمانی سیگنال گفتار در بیماران آلزایمری است مورد ارزیابی قرار داده است. مرجع [۱۴] اهمیت حروف بی‌صدا در دستیابی به نتایج مطلوب را در تشخیص بیماری

آلزایمر یک نوع اختلال عملکرد مغزی است که به تدریج توانایی‌های ذهنی بیمار تحلیل می‌رود و از علائم اولیه این بیماری فقدان حافظه، اختلال در تصمیم‌گیری و اشتباه در انتخاب واژگان درست است. قسمتی از قشر مغز که توانایی پردازش صحبت را دارد، یکی از اولین بخش‌هایی است که تحت تاثیر این بیماری قرار می‌گیرد [۱] بنابراین پردازش سیگنال گفتار این بیماری می‌تواند در تشخیص آن مفید واقع شود. نشان داده شده است که برخی درمانگاه‌های غیرتخصصی، بیماری آلزایمر را در مراحل اولیه نادرست تشخیص می‌دهند [۲] لذا ارائه روشی کارا، آسان، غیرتهاجمی و بهینه از نظر اقتصادی مثل پردازش سیگنال گفتار برای تشخیص این بیماری بسیار سودمند خواهد بود و تاثیرات متقابل آن مورد اهمیت است.

^۱ Mel Frequency Cepstral Coefficients

نشان می‌دهد. در مرجع [۱۵]، شدت و سطح بیماری آلزایمر با استفاده از روان صحبت کردن و تشخیص احساس بیماران بررسی شده که برای شناسایی احساس از ویژگی‌های پروژودیک و زبان‌شناختی سیگنال گفتار استفاده شده است. از سوی دیگر، تشخیص بیماری آلزایمر با استفاده از سیگنال گفتار به فرهنگ، زبان، محتوای گفتار، جنسیت، سن گوینده، لهجه و بسیاری از عوامل دیگر وابسته است [۳]. از این رو سیگنال گفتار بیماران آلزایمری در زبان‌های فارسی [۴]، اسپانیایی [۵]، انگلیسی [۱۶]، فرانسوی [۶]، کاتالان، چینی، باسک، عربی، پرتغالی [۱۳]، روسی [۷]، آلمانی [۱۷] و مجارستانی [۱۱] مورد بررسی قرار گرفته است. مرجع [۱۶] سیگنال گفتار بیماران انگلیسی زبان را از لحاظ ویژگی‌های عاطفی و احساس از قبیل خشم، نفرت و آرامش مورد بررسی قرار داده است و مراجع [۱۷، ۱۸] ویژگی‌های آکوستیک سیگنال صحبت را استخراج و تحلیل نموده‌اند.

از آنجاییکه سیستم مدل شنوایی (AMS^۲) انسان بعنوان بهترین تشخیص‌دهنده سیگنال گفتار عمل می‌کند لذا می‌تواند بعنوان بهترین مدل برای شناسایی سیگنال گفتار استفاده شود. AMS انسان در شناسایی و تحلیل سیگنال گفتار، شبیه‌سازی، کد کردن و اندازه‌گیری کیفیت صدا کاربرد دارد [۱۹]. در مراجع [۲۰-۲۲] به کمک AMS، واج و سیگنال‌های گفتار مورد بررسی و تحقیق قرار گرفته‌اند. در [۲۳] با توجه به ساختار شنوایی انسان بررسی‌های روانشناسی و نوروفیزیولوژیک یک مدل محاسباتی برای تجزیه و تحلیل صوت ارائه شده است. این مدل ساختاری یکپارچه از نمایش ویژگی‌های طیفی و زمانی سیگنال را ارائه می‌دهد. مدل ریاضی کاملی ارائه می‌شود که چگونگی تغییر سیگنال‌های پیچیده در سطوح مختلف مدولاسیون طیفی-زمانی (MTF^۳) را نشان می‌دهد و آن را با مدل‌های موجود در پردازش شنوایی مقایسه می‌کند. در برخی مطالعات از MTF برای تحلیل سیگنال گفتار استفاده و کارایی آن بیان شده است [۲۴-۲۶].

از آنجاییکه یکی از عوامل تاثیر گذار در پردازش سیگنال گفتار زبان است، زبان بومی فارسی در این مطالعه مد نظر قرار گرفته شد. به این منظور، پایگاه داده بومی متشکل از بیماران آلزایمری در سطوح ۵ و ۶ بیماری آلزایمر و افراد سالم جمع-آوری شد. کارهای دیگری نیز در زبان فارسی صورت گرفته است که اکثر این کارها مرتبط با تحلیل سیگنال گفتار بیماران آلزایمری فارسی‌زبان در حوزه زبان‌شناسی و استخراج ویژگی-های مرتبط به صورت دستی و تنها با تحلیل روش‌های آماری بررسی شده است [۲۷ و ۲۸]. هدف اصلی از این مقاله پردازش

پیشینه تحقیق

جدول ۱، خلاصه مطالعات انجام شده در زمینه تحلیل سیگنال گفتار بیماران آلزایمری را نشان می‌دهد.

همانطور که در جدول ۱ قابل مشاهده است، مراجع [۵، ۳۳] ویژگی‌های زبانی و آماری سیگنال گفتار را مورد پردازش قرار داده‌اند و [۱۷، ۳۴] به تحلیل ویژگی‌های گفتاری و زبانی پرداخته‌اند. در [۱۶] نیز گفتار از لحاظ احساس و عواطف و در [۱۳] زبان‌های مختلف مورد بررسی قرار گرفته است. مراجع [۴، ۱۷، ۱۸] ویژگی‌های آکوستیک و [۱۱، ۱۲] ویژگی‌های فرکانسی مانند MFCC را استخراج و ارزیابی کرده‌اند. ویژگی - آکوستیک، گفتاری و زبانی به صورت دستی استخراج می‌شوند که بسیار زمان بر است و دقت بالایی ندارد. مرجع [۴] علاوه بر عملکرد خوب دارای پیچیدگی محاسباتی است. مقاله حاضر به استخراج خودکار ویژگی‌های سیگنال گفتار بیماران آلزایمری توسط AMS پرداخته که دارای دقت بسیار بالایی است. نویسندگان این پایگاه داده را در کار قبلی خود مورد ارزیابی قرار داده و به استخراج روش‌های متداول در پردازش سیگنال گفتار همچون بسته موجک در کنار ضرایب کپسترال فرکانس مل، نرخ عبور از صفر، چرخش طیف، پهنای باند، انرژی سیگنال و فرکانس مرکز طیفی سیگنال گفتار بیماران آلزایمری و افراد سالم پرداخته‌اند و طبقه‌بندی نتایج استخراج شده توسط ماشین بردار پشتیبان دقت ۹۶٪ را در برداشت [۳۰].

روش

روش کلی مقاله به این صورت است که ابتدا داده‌های موجود در پایگاه داده برای استفاده در مراحل بعدی پیش‌پردازش و سپس با استفاده از AMS ویژگی‌های مورد نظر استخراج شده و توسط ماشین بردار پشتیبان طبقه‌بندی می‌شوند.

از آنجاییکه یکی از عوامل تاثیر گذار در پردازش سیگنال گفتار زبان است، زبان بومی فارسی در این مطالعه مد نظر قرار گرفته شد. به این منظور، پایگاه داده بومی متشکل از بیماران آلزایمری در سطوح ۵ و ۶ بیماری آلزایمر و افراد سالم جمع-آوری شد. کارهای دیگری نیز در زبان فارسی صورت گرفته است که اکثر این کارها مرتبط با تحلیل سیگنال گفتار بیماران آلزایمری فارسی‌زبان در حوزه زبان‌شناسی و استخراج ویژگی-های مرتبط به صورت دستی و تنها با تحلیل روش‌های آماری بررسی شده است [۲۷ و ۲۸]. هدف اصلی از این مقاله پردازش

^۲ Auditory System Model

^۳ Modulation Transform Function

(جدول-۱): خلاصه مطالعات انجام شده در زمینه تحلیل سیگنال گفتار بیماران آلزایمری

مراجع	سال پژوهش	ضرایب ویژگی مورد استفاده	زبان	روش تشخیصی گفتار	دقت (درصد)
[۱۳]	۲۰۱۵	فرکانس مرکز طیفی، انرژی کوتاه مدت، حروف صدادار و بی صدا، بعد فرکانس	فرانسوی، انگلیسی، کاتالان، اسپانیایی، چینی، باسک، عربی، پرتغالی	نزدیکترین همسایه	۹۰
[۵]	۲۰۱۶	فعل، اسم، حروف اضافه و ربط، ویژگی‌های آماری	اسپانیایی	SVM	۸۸
[۴]	۲۰۱۶	ویژگی‌های آکوستیک و استنتاج فازی و بهینه سازی	فارسی	تحلیل کمی سطوح میانگین مربع خطا و نمودارهای کانتور	۹۹/۹۶
[۱]	۲۰۱۵	نرخ کلمات پرسشی، تعداد فعل، اسم و حروف اضافه، فرکانس فعل، فرکانس اسم	انگلیسی	SVM، بیزساده، درخت دی، نزدیکترین همسایه	-
[۱۶]	۲۰۱۳	در این مقاله سیگنال گفتار از لحاظ احساسی و عاطفی مورد بررسی قرار گرفته است	انگلیسی	SVM شبکه عصبی پروسپترون چند لایه درخت دی نزدیکترین همسایه بیز ساده	۹۳/۴۷ ۹۳/۰۲ ۹۳/۴۷ ۸۷/۵۹ ۵۹/۸۷
[۳۴]	۲۰۱۶	ویژگی‌های گفتاری و زبانی	آلمانی	F-score	۸۰
[۳۳]	۲۰۱۷	استخراج خطاهای گفتاری، ویژگی‌های آماری	آلمانی	نزدیکترین همسایه تحلیل خطی SVM	۶۲/۲۳
[۱۸]	۲۰۱۸	ویژگی‌های آکوستیک	انگلیسی	مدل مخفی مارکوف مدل ترکیبی گوسی	۶۰/۶
[۱۹]	۲۰۱۸	و جیترو و شیمر محلی MFCC	انگلیسی	شبکه عصبی پیچشی	۷۳/۶
[۱۷]	۲۰۱۸	ویژگی‌های آکوستیک (نرخ گفتار، سکوت و بردار) و ویژگی‌های زبانی	آلمانی	-	۸۱/۹
[۱۱]	۲۰۱۹	MFCC	مجارستانی	شبکه عصبی	۷۲/۳
[۳۰]	۲۰۲۰	بسته موجک، ضرایب کیسترال فرکانس مل، نرخ عبور از صفر، چرخش طیف، پهنای باند، انرژی سیگنال و فرکانس مرکز طیفی	فارسی	SVM	۹۶
مقاله حاضر		مدولاسیون زمانی و طیفی و انرژی اسپکتروم	فارسی	SVM	۹۹/۹۹

پیش پردازش

اساس مکانیزم ادراک انتخابی و دقیق صدای طبیعی کاملاً شناخته شده نیست، ممکن است سیگنال گفتار به صورتی تنظیم شود که پاسخ AMS با آکوستیک بیولوژیکی سیگنال صدای طبیعی محیط همبستگی داشته باشد [۲۵]. بنابراین نویز موجود در داده‌های جمع‌آوری شده حذف و سیگنال گفتار نرمالیزه شده است. با توجه به ماهیت غیرایستاد سیگنال گفتار، هر سیگنال توسط پنجره همینگ به قسمت‌های ۳۰-۵۰ میلی ثانیه با همپوشانی ۵۰٪ تقسیم شده‌اند و سکوت در سیگنال گفتار توسط عبور از صفر و انرژی سیگنال استخراج و حذف شده‌اند.

سیستم مدل شنوایی

ابتدا تبدیل موجک سیگنال $s(t)$ حاصل می‌شود و سپس تجزیه و تحلیل طیفی توسط فیلتر کوچلر ۴ با ۱۲۸ فیلتر میان‌گذر دارای همپوشانی و Q ثابت با فرکانس‌های مرکزی که به صورت یکنواخت در محدوده فرکانس لگاریتمی توزیع شده‌اند انجام می‌شود که پاسخ فرکانسی هر فیلتر با $h(t; x)$ مشخص شده است. خروجی این فیلتر کوچلر با $y_{coch}(t, x)$ نمایش داده می‌شود [۲۵].

$$y_{coch}(t, x) = s(t) \otimes_t h(t; x), \quad (1)$$

که \otimes_t عملگر کانولوشن نسبت به زمان است. خروجی فیلتر کوچلر $y_{coch}(t, x)$ در یک الگوی عصبی-شنوایی $y_{AN}(t, x)$ که شامل یک فیلتر بالاگذر، فشرده‌ساز غیرخطی $g(\cdot)$ و فیلتر پایین‌گذر $w(t)$ است، قرار می‌گیرد [۲۵].

$$y_{AN}(t, x) = g(\theta_t y_{coch}(t, x)) \otimes_t w(t), \quad (2)$$

مرحله آخر عملکرد یک شبکه مهارکننده جانبی^۴ است که اساس کوچلر است. این شبکه به سادگی توسط مشتق مرتبه اول با توجه به محور تونوتپیک^۶ و یکسوساز نیم‌موج اعمال می‌شود [۲۵].

$$y_{LIN}(t, x) = \max(\partial_x y_{AN}(t, x), 0), \quad (3)$$

$$y_{final}(t, x) = y_{LIN}(t, x) \otimes_t \mu(t; x), \quad (4)$$

خروجی نهایی توسط انتگرال‌گیری در طول یک پنجره کوتاه به دست می‌آید [۲۵].

$$\mu(t; x) = e^{-t/\tau_u(t)}, \quad (5)$$

در این مرحله طیف زمانی و فرکانسی سیگنال گفتار از طریق یک بانک فیلتر که دارای پارامترهای طیفی-زمانی مدولاسیون

است و تغییرات زمانی آن‌ها از نرخ آهسته به سریع $(\omega = [1, 2, 4, 8, 16, 32]Hz)$ و تغییرات طیفی آن‌ها از مقیاس‌های باریک به گسترده است $(\Omega = [0.5, 1, 2, 4, 8]cyc/oct)$ تخمین زده می‌شود. حوزه دریافت فرکانس-زمان این فیلترها در فرکانس‌های مختلفی در امتداد محور تونوتپیک قرار دارند. پاسخ طیفی-زمانی rc^v پایین رونده و بالا رونده به شکل زیر خواهد بود [۳۱].

$$r_{c\downarrow}(t, x; \omega_c, \Omega_c, \theta_c, \phi_c) \quad (6)$$

$$= y(t, x) \otimes_{tx} [(h_t h_s - \widehat{h}_t \widehat{h}_s) \cos(\theta_c + \phi_c) + (\widehat{h}_t h_s + h_t \widehat{h}_s) \sin(\theta_c + \phi_c)]$$

$$r_{c\uparrow}(t, x; \omega_c, \Omega_c, \theta_c, \phi_c) \quad (7)$$

$$= y(t, x) \otimes_{tx} [(h_t h_s + \widehat{h}_t \widehat{h}_s) \cos(\theta_c - \phi_c) + (\widehat{h}_t h_s - h_t \widehat{h}_s) \sin(\theta_c - \phi_c)]$$

که در آن \otimes_{tx} نشان دهنده کانولوشن زمان و مکان است و h_t و h_s به ترتیب نشان دهنده مدولاسیون زمانی و فرکانسی است. θ و ϕ نشان دهنده ویژگی‌های فاز هستند.

فرمول مناسبی از rc در صورتیکه خروجی دامنه و فاز، تابعی از تبدیل موجک باشند را می‌توان به شکل زیر نیز نشان داد [۳۱].

$$r_{c\downarrow}(t, x; \omega_c, \Omega_c, \theta_c, \phi_c) \quad (8)$$

$$= R\{z_{\downarrow}\} \cos(\theta_c + \phi_c) + I\{z_{\downarrow}\} \sin(\theta_c + \phi_c) = |z_{\downarrow}| \cos(\psi_{\downarrow} - \theta_c - \phi_c)$$

$$r_{c\uparrow}(t, x; \omega_c, \Omega_c, \theta_c, \phi_c) \quad (9)$$

$$= R\{z_{\uparrow}\} \cos(\theta_c - \phi_c) - I\{z_{\uparrow}\} \sin(\theta_c - \phi_c) = |z_{\uparrow}| \cos(\psi_{\uparrow} + \theta_c - \phi_c)$$

$R\{\cdot\}$ و $I\{\cdot\}$ به ترتیب نشان‌دهنده قسمت حقیقی و مختلط تابع است. توابع z_{\uparrow} و z_{\downarrow} که به صورت تابعی از تبدیل موجک زمانی h_{TW} و موجک فرکانسی h_{SW} است در زیر نوشته شده است [۳۱]

$$z_{\downarrow}(t, x; \omega_c, \Omega_c) \quad (10)$$

$$= y(t, x) \otimes_{tx} [h_{TW}(t; \omega_c) h_{SW}(x; \Omega_c)] = |z_{\downarrow}(t, x; \omega_c, \Omega_c)| e^{j\psi_{\downarrow}(t, x; \omega_c, \Omega_c)}$$

$$z_{\uparrow}(t, x; \omega_c, \Omega_c) \quad (11)$$

$$= y(t, x) \otimes_{tx} [h_{TW}^*(t; \omega_c) h_{SW}(x; \Omega_c)] = |z_{\uparrow}(t, x; \omega_c, \Omega_c)| e^{j\psi_{\uparrow}(t, x; \omega_c, \Omega_c)}$$

بلوک دیاگرام مراحل پردازش سیگنال در AMS مطرح شده در این مقاله در شکل (۱) نمایش داده شده است. طیف زمانی-فرکانسی سیگنال گفتار (اسپکتروگرام اولیه (شکل (۲)، رابطه (۵) حاوی اطلاعات زمان و فرکانس است. با پردازش اطلاعات زمان و فرکانس به ترتیب اطلاعات نرخ و مقیاس قابل حصول است که در این مقاله تغییرات حوزه نرخ و مقیاس به ترتیب تحت عناوین مدولاسیون زمانی و طیفی مطرح شده است.

⁷ Cortical Representation

⁴ Cochlear

⁵ Lateral Inhibitory Network

⁶ Tonotopic

نتیجه

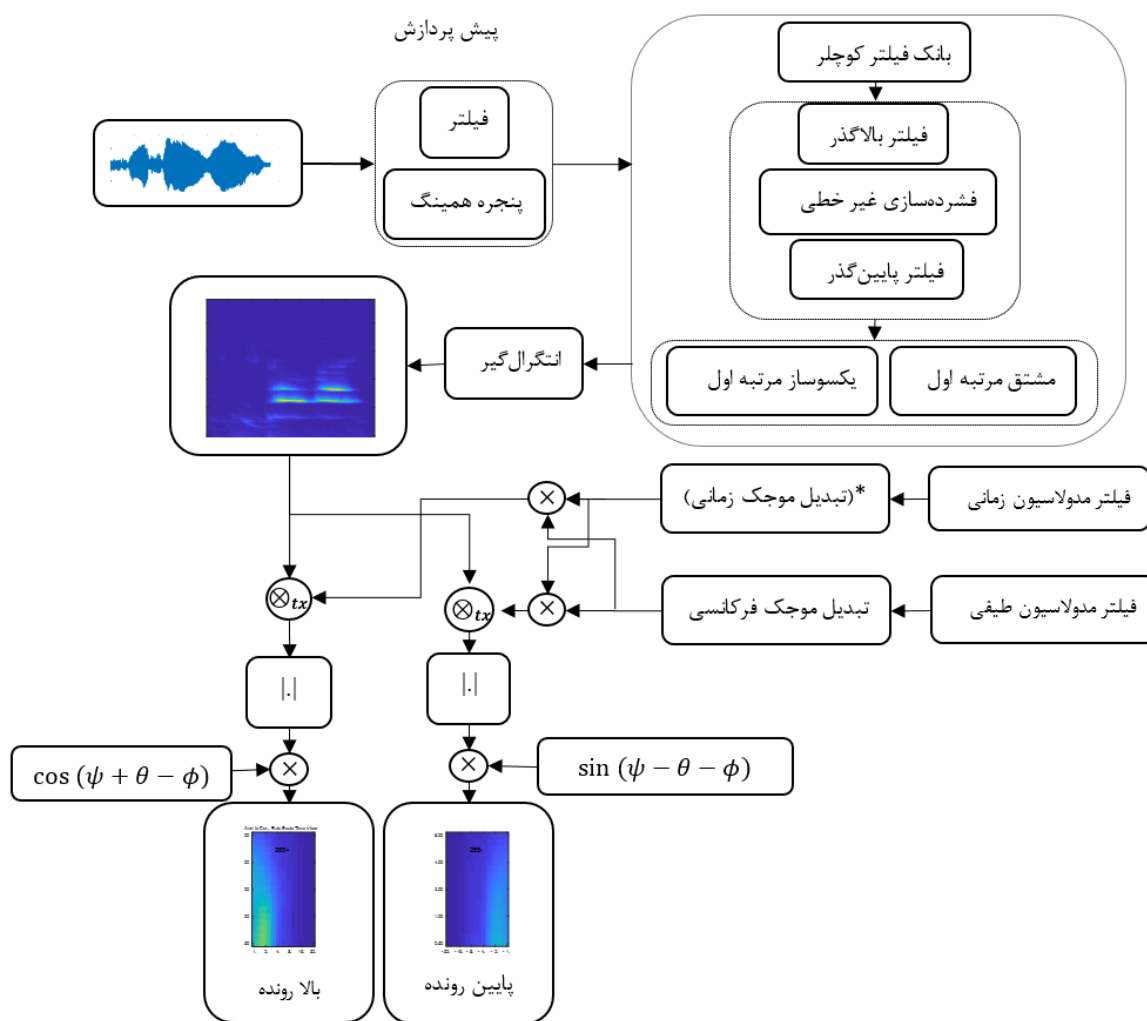
پایگاه داده

صدای ضیط شده از بیماران آلزایمری در مرکز نگهداری بیماران آلزایمری شایستگان اراک، توسط میکروفون SWD\MMDEVAPI HP با فرکانس نمونه برداری ۴۴/۱ KHz

در فرمت wav. جمع آوری شده‌اند. دادگان شامل ۱۰۰ داده از افراد سالم و بیماران آلزایمری در فاصله سنی ۵۰-۸۵ سال بوده‌اند. داده‌های مورد پردازش در محیطی دوستانه جمع آوری و از بیماران سوالاتی مشخص در آرامش پرسیده شده است.

(جدول-۲): مشخصات پایگاه داده جمع آوری شده

تعداد افراد	جنسیت	بازه سنی	سطح بیماری	میانگین طول مکالمه هر فرد	تعداد سیگنال گفتار استخراج شده از مکالمه	میانگین طول هر سیگنال
۱۱	مذکر	۸۵-۷۱	۶-۵	۶۵ دقیقه	۵۰	۲۰۰۰ (ms)
۱۰	مذکر	۷۹-۵۰	-	۳۰ دقیقه	۵۰	۱۶۰۰ (ms)

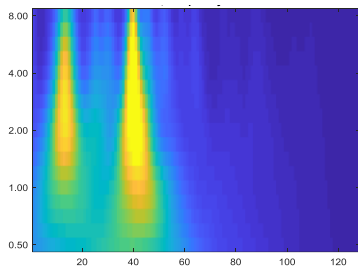


(شکل-۱): دیاگرام AMS

نتایج تجربی

با استفاده از روش بیان شده در بخش ۲، سیگنال گفتار فرد سالم و آلزایمری مورد بررسی قرار گرفتند. شکل (۲) از سیگنال گفتار دو گوینده آقا (فرد سالم و آلزایمری) با گفتن جمله /من یک

سیگنال‌های جمع آوری شده در این پژوهش دارای مجوز اخلاقی از مرکز مربوطه می‌باشد. داده های اخذ شده توسط پردازشگر i5-5200U CPU 2.20GHz با نرم افزار Matlab2018 مورد پردازش قرار گرفته‌اند. در جدول (۲) مشخصات پایگاه داده جمع آوری شده مورد استفاده در این پژوهش مشاهده می‌شود.

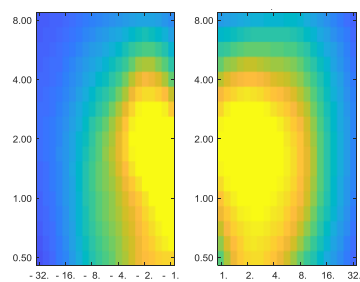


(ب)

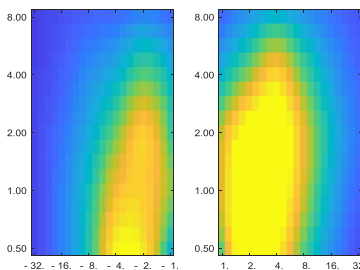
(شکل-۳): نمایش فرکانس-مقیاس (الف) بیمار آلزایمری و (ب) فرد سالم.

سپس پاسخ r_c نسبت به فرکانس میانگین‌گیری و تغییرات نرخ-مقیاس مشاهده می‌شود (شکل (۴)) که نشان‌دهنده تغییرات مدولاسیون طیفی در راستای مقیاس و مدولاسیون زمانی در راستای نرخ در اسپکتروگرام است. هر دو شکل (الف) و (ب) از دو قسمت بالارونده (سمت راست) و پایین‌رونده (سمت چپ) تشکیل شده است که به تغییرات موجک و تغییرات فاز سیگنال بستگی دارد (فرمول‌های (۸) و (۹)).

در ادامه میانگین‌گیری پاسخ r_c در راستای فرکانس انجام شد و تغییرات مقیاس-نرخ سیگنال گفتار در طول زمان در فرد سالم و آلزایمری مورد توجه قرار گرفت؛ نتیجه تغییرات نهایی مقیاس-نرخ-فرکانس در شکل (۵) قابل مشاهده است که با افزایش مقیاس وابستگی بین نرخ و فرکانس کاهش یافته و نرخ در فرکانس‌های بالاتر پاسخ ضعیف‌تری نسبت به افزایش مقیاس داشته است که نمایانگر تاثیرات بانک فیلترهای انتخابی (۱،۲،۴،۸،۱۶،۳۲) در نمایش مدولاسیون زمانی سیگنال است.



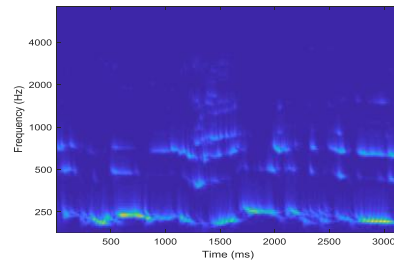
(الف)



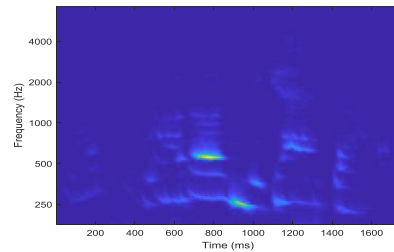
(ب)

(شکل-۴): نمایش نرخ-مقیاس (الف) بیمار آلزایمری و (ب) فرد سالم.

کشاورز هستم/ به دست آمده است (سایر شکل‌ها نیز مربوط به این جمله می‌باشند). شکل (۲) اسپکتروگرام سیگنال گفتار بیمار آلزایمری (الف) و فرد سالم (ب) که نشان‌دهنده شدت فرکانس در زمان معلوم است را نشان می‌دهد.



(الف)

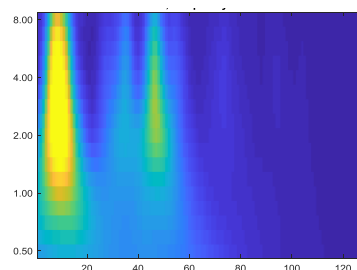


(ب)

(شکل-۲): اسپکتروگرام (الف) بیمار آلزایمری و (ب) فرد سالم.

سپس اسپکتروم بدست‌آمده از فرمول (۴) از بانک فیلترهای انتخابی مدولاسیون طیفی و زمانی عبور کرده و سپس خروجی این بانک فیلترها در طول زمان میانگین‌گیری شده تا نمایشی از مدولاسیون وابسته به مقیاس (Ω) ، نرخ (ω) و فرکانس (f) حاصل شود. از طرفی، بخاطر چندبعدی بودن پاسخ r_c (فرمول-های (۸) و (۹))، علاوه بر تفسیر مشکل این پاسخ، نمایش آن نیز آسان نمی‌باشد، از اینرو در راستای نمایش بهینه‌تر، میانگین-گیری در راستای تغییرات فرکانس و یا نرخ انجام شد.

شکل (۳) نمایش فرکانس-مقیاس بعد از میانگین‌گیری در راستای نرخ است؛ این شکل بررسی مدولاسیون طیفی را برعهده دارد و با توجه با رخداد بیشینه مقدار مقیاس، فرکانس مرکزی را در مدولاسیون طیفی مشخص می‌کند.

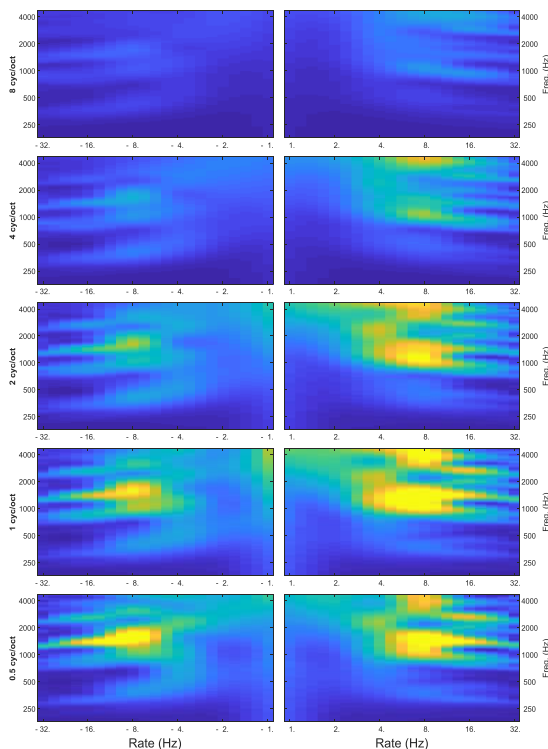


(الف)

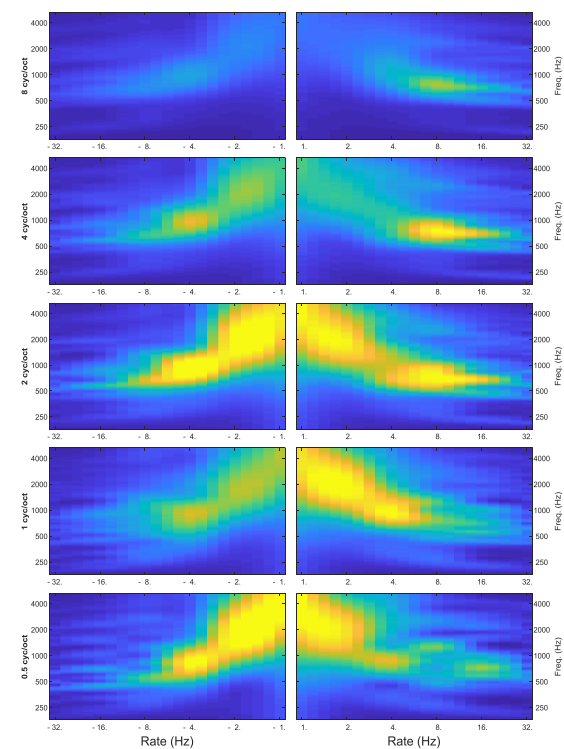
تحلیل نتایج

با بررسی تغییرات فرکانس-مقیاس در شکل (۳) مشاهده می‌شود که در هر دو گروه افراد سالم و آلزایمری، در فرکانس‌های پایین-تر وابستگی به مقیاس بیشتر است. با توجه به تغییرات قابل مشاهده در شکل (۴) و سایر شکل‌های حاصل مشاهده شد که وابستگی نرخ و مقیاس به یکدیگر بیش از وابستگی سایر مولفه‌ها است که این تفسیر توأمان نرخ و مقیاس تنها در استفاده از AMS قابل حصول است و تغییرات نرخ-مقیاس به‌عنوان تابعی از زمان مطرح می‌باشد که نشان‌دهنده تغییرات ریپل طیف گفتار است. با افزایش نرخ وابستگی بین مدولاسیون‌های زمانی و طیفی کاهش می‌یابد. این نتایج برای سایر بیماران آلزایمری و افراد سالم بررسی شد و نتایج مشابهی به دست آمد.

یکی از یافته‌های مدل AMS، اسپکتروگرام (تغییرات زمان-فرکانس) سیگنال گفتار است (شکل (۲))؛ همانطور که در این شکل قابل مشاهده است، بطور مثال، در فرکانس ۷۵۰ هرتز، تغییرات فرکانس در طول زمان برای دو شکل (الف) و (ب) کاملاً متفاوت است که حاکی از تفاوت سیگنال گفتار فرد آلزایمری و سالم است. از دیگر یافته‌های این مدل، نتیجه چهاربعدی مقیاس، نرخ، زمان و فرکانس است که اطلاعات مفید و با ارزشی را در اختیار ما قرار می‌دهد که با میانگین‌گیری نسبت به هر کدام از مولفه‌های خروجی این مدل می‌توان به شکل‌های جدید قابل تفسیری دست یافت.



(ب)



(الف)

(شکل-۵): نمایش سه بعدی نرخ-فرکانس-مقیاس (الف) بیمار آلزایمری و (ب) افراد سالم.

به منظور بررسی ویژگی‌های استخراج شده در ۵۱ داده از افراد دارای بیماری آلزایمر و ۴۲ داده از افراد سالم و با توجه به نتیجه سه‌بعدی حاصل در شکل (۵)، ویژگی‌های مدولاسیون زمانی (CTM^۸) (شکل (۶-الف)) بر حسب Hz، توسط بیشینه مقدار مولفه مقیاس در راستای دو مولفه نرخ و فرکانس در نظر گرفته شده است؛ به طور مشابه ویژگی‌های مدولاسیون طیفی (CSM^۹) (شکل (۶-ب)) و ویژگی‌های فرکانسی (CF^{۱۰}) (شکل (۶-ج)) به ترتیب بر حسب Hz و cyc/oct و توسط بیشینه مقدار مولفه زمان

با بررسی خروجی‌های نرخ-مقیاس (شکل (۴)) برای سایر افراد سالم و بیماران آلزایمری مشاهده شد که میانگین بیشینه تغییرات نرخ برای بیماران آلزایمری در قسمت بالارونده در بازه [۱،۶] و در قسمت پایین‌رونده در بازه [-۳،-۱] و تغییرات نرخ در افراد سالم در قسمت بالارونده در بازه [۱،۸] و در قسمت پایین‌رونده در بازه [-۸،-۱] می‌باشد. با بررسی تغییرات پاسخ-های قوی در شکل (۵)، سطح پاسخ در بیماران آلزایمری بسیار بیشتر از افراد سالم هستند و در محدوده متفاوت فرکانسی مشاهده می‌شوند.

^۸ Characteristic Temporal Modulation

^۹ Characteristic Spectral Modulation

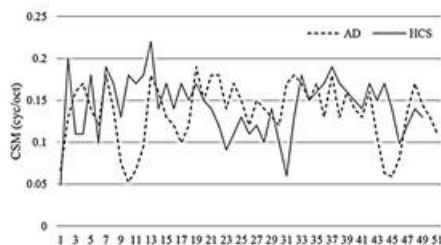
^{۱۰} Characteristic Frequency

در راستای دو مولفه دیگر در نظر گرفته شده است؛ بطور مثال، بیشینه مقدار فرکانس به ازاء هر نرخ و مقیاس برای هر داده استخراج و بعنوان ویژگی طیفی برای سیگنال گفتار بیماران آلزایمری و افراد سالم حاصل شده است. همانطور که در شکل ۶ نشان داده است با توجه به عدم در اختیار داشتن رابطه خطی بین داده، بررسی داده ها توسط روش های خطی معمول قابل انجام نمی باشد لذا با توجه به پراکندگی غیرخطی ویژگی های استخراج شده به منظور انجام مقایسه بر روی ویژگی های مورد نظر از طبقه بندی SVM با کرنل غیرخطی استفاده شده است. سپس دقت طبقه بندی تغییرات بیشینه نقاط در ویژگی های فرکانس، مدولاسیون طیفی، زمانی بعنوان یک بردار ویژگی و انرژی اسپکتروم بعنوان بردار ویژگی دوم توسط SVM به طور جداگانه به ترتیب ۹۵٪ و ۶۳٪ و به صورت همزمان ۹۹/۹۹٪ به دست آمد.

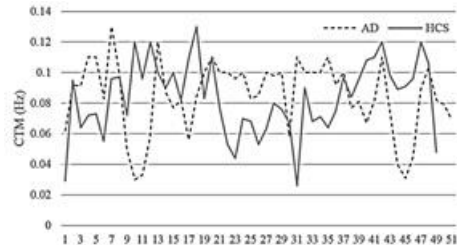
نسبت داده های آموزش و تست، ۶۰ به ۴۰ درصد انتخاب، سپس با استفاده از طبقه بندی SVM با توابع کرنل مختلف داده ها طبقه بندی شده و به کمک رابطه (۱۲) دقت عملکرد طبقه بندی ارزیابی شده است:

$$\text{دقت} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (12)$$

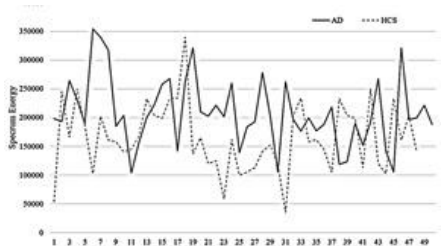
که در آن، TP مثبت درست، TN منفی درست، FP مثبت نادرست و FN منفی نادرست است.



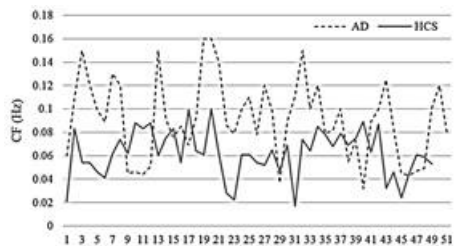
(ب)



(الف)



(د)



(ج)

(شکل ۶-): تغییرات بیشینه نقاط در ویژگی های مدولاسیون زمانی بر حسب Hz (الف)، ویژگی های مدولاسیون طیفی بر حسب cyc/oct (ب)، ویژگی های فرکانسی بر حسب Hz (ج) و انرژی اسپکتروم (د) برای بیماران آلزایمری (AD) و افراد سالم (HCS).

بحث و نتیجه گیری

علاوهم تفاوت پایگاه داده به کار گرفته شده در این مقاله نسبت به سایر مقالات مورد بررسی (جدول-۱)، نتایج آزمایش-ها نشان می‌دهد که الگوریتم پیشنهادی عملکرد قابل قبولی در مقایسه با سیستم‌های تشخیص بیماری آلزایمری موجود ارائه می‌دهد.

برای دستیابی به نتایج عمومی‌تر در نتایج حاصله می‌توان پایگاه داده‌ای مشتمل بر افراد آلزایمری در سطوح مختلف بیماری تشکیل داد همچنین به کارگیری این سیستم با استفاده از یک پایگاه داده قوی با تعداد بالای داده به بهتر شدن کارایی سیستم کمک شایان توجهی می‌کند. در حال حاضر پیش‌بینی-هایی که در علم پزشکی صورت می‌گیرد بر اساس عوامل ژنتیکی، تحصیلات، حرفه افراد، منطقه آب و هوایی، تاثیرات محیطی و سن افراد است لذا با در نظر گرفتن تاثیر این عوامل می‌توان پایگاه داده جامع‌تری تشکیل داد. از سوی دیگر، می‌توان از دیگر الگوریتم‌های داده کاوی برای طبقه‌بندی بهره جست. بعلاوه، با توجه به تفاوت تصاویر خروجی AMS در دو گروه افراد سالم و آلزایمری، بکارگیری روش‌های یادگیری عمیق امکانپذیر خواهد بود.

همانطور که گفته شد در بیماری آلزایمر یکی از ابتدایی‌ترین قسمت‌هایی که آسیب می‌بیند قشر مغز می‌باشد که بخش گفتار نیز شامل این قسمت است. از طرفی، گفتار ابزاری ساده، بهینه و غیرتهاجمی برای تشخیص به‌هنگام بیماری آلزایمر است و به دلیل ماهیت و تفاوت آن در افراد سالم و بیماران آلزایمری می‌تواند سودمند و کارا باشد. در این مقاله از AMS (با استفاده از جعبه ابزار NSL [۳۲]) استفاده و سیگنال گفتار بیماران آلزایمری و افراد سالم فارسی‌زبان با هم مقایسه شد. از آنجایی که AMS در تشخیص سیگنال گفتار مفید است، از این مدل می‌توان برای تشخیص گفتار بیماران آلزایمری هم بهره جست. با توجه به کارایی بالای این روش در شناسایی سیگنال گفتار و عدم استفاده از این روش در سیگنال گفتار بیماران آلزایمری تا به امروز، این مقاله به بررسی تاثیر استفاده از AMS در استخراج ویژگی و تحلیل سیگنال گفتار بیماران آلزایمری پرداخته است. نتایج بدست آمده بسیار رضایت بخش بوده است و نتایج امیدوارکننده‌ای برای تشخیص زودهنگام این بیماری دربرداشته است.

Computers in Human Behavior, vol.61, no.1, pp. 364-371, Aug. 2016.

- [5] L. Hernandez-Dominguez, E. Garcia-Cano, S. Ratte, G. Sierra-Martinez, "Detection of Alzheimer's disease based on automatic analysis of common objects descriptions," Berlin, Allemagne, 7th Workshop on Cognitive Aspects of Computational Language Learning, pp.10-15, 11 Aug 2016.
- [6] A. Konig, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert, R. David, "Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease," Diagnosis, Assessment and Disease Monitoring, vol. 1, no.1, pp. 112-124, 2015.
- [7] K. C. Fraser, J. A. Meltzer, F. Rudzicz, "Linguistic Features Identify Alzheimer's Disease in Narrative Speech," Journal of Alzheimer's Disease, vol. 49, no.2, pp. 407-422, oct. 2016.
- [8] D. Gharavian, M. Sheikhan, F. Ashoftedel, "Using neutralized formant frequencies to improve emotional speech recognition," IEICE

مراجع

- [1] S. J. Chaudhari, R. M. Kagalkar, "Automatic SpeakerAge Estimation and Gender Dependent Emotion Recognition," International Journal of Computer Applications, vol. 117, no. 17, pp. 5-10, May 2015.
- [2] A. Connolly, E. Gaehl, H. Martin, J. Morris, N. Purandare, "Underdiagnosis of dementia in primary care: Variations in the observed prevalence and comparisons to the expected prevalence," Aging & Mental Health, vol. 15, no. 8, Nov. 2011.
- [3] M. El Ayadi, M. S. Kamel, F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, vol. 44, no. 3, pp. 978-984, Mar. 2011
- [4] M. Nasrolahzadeh, Z. Mohammadpoori, J. Haddadnia, "Analysis of mean square error surface and its corresponding contour plots of spontaneous speech signals in Alzheimer's disease with adaptive wiener filter."

- Temperature,” *Cognitive Computation*, vol. 7, no. 1, pp.44-55 2015.
- [17] J. Weiner, T. Schultz, “Selecting Features for Automatic Screening for Dementia Based on Speech,” *Lecture Notes in Computer Science*, pp. 747–756, 2018.
- [18] B. Mirheidari, D. Blackburn, T. Walker, A. Venneri, M. Reuber, H. Christensen, “Detecting Signs of Dementia Using Word Vector Representations,” *Interspeech 2018*, pp. 1893-1897, sep. 2018.
- [19] M. Karjalainen, "Auditory models for speech processing", *Proc. of Int. Congr. of Phonetic Sciences*, 1987.
- [20] N. Mesgarani, M. Slaney, S. A. Shamma, “Discrimination of speech from nonspeech based on multiscale spectro-temporal Modulations,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 3, pp. 920–930, May 2006.
- [21] P. K. Ghosh, L. M. Goldstein, S. S. Narayanan, “Processing speech signal using auditory-like filterbank provides least uncertainty about articulatory gestures,” *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 4014–4022, Jun. 2011.
- [22] A. Klapuri, “Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 255–266, Feb. 2008.
- [23] T. Chi, P. Ru, S. A. Shamma, “Multiresolution spectrotemporal analysis of complex sounds,” vol. 118, pp. 887-906, May 2005.
- [24] S. M. N. woolley, T. E. fremouw, A. hsu, F. E. Theunissen, “Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds,” *nature neuroscience*, September 2005, doi:10.1038/nn1536.
- [25] T. Chi, Y. Gao, M. C. Guyton, P. Ru, S. Shamma (1999), “Spectro-Temporal Modulation Transfer Functions and Speech Intelligibility,” *The Center for Auditory and Acoustic Research (CAAR)*, doi: 10.1121/1.428100.
- Electronics Express*, vol. 8, no. 14, pp.1155-1160, 2011.
- [9] A. Potamianos, P. Maragos, “Speech formant frequency and bandwidth tracking using multiband energy demodulation,” *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp.3795-3806, Jun. 1996.
- [10] B. J. Mohan, R. Babu, “Speech recognition using MFCC and DTW,” *2014 International Conference on Advances in Electrical Engineering (ICAEE)*, pp. 1-14, Jan. 2014.
- [11] G. Gosztolya, V. Vincze, L. Tóth, M. Pákáski, J. Kálmán, I. Hoffmann, “Identifying Mild Cognitive Impairment and mild Alzheimer’s disease based on spontaneous speech using ASR and linguistic features,” *Computer Speech & Language*, vol. 53, pp. 181–197, Jan. 2019.
- [12] T. Warnita, N. Inoue, K. Shinoda, “Detecting Alzheimer’s Disease Using Gated Convolutional Neural Network from Audio Data,” *Interspeech 2018*, pp.1706-1710, Sep. 2018.
- [13] K. López-de-Ipina, J. Solé-Casals, H. Eguiraun, J. B. Alonso, C. M. Travieso, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martinez-Lage, B. Beitiaf, “Feature selection for spontaneous speech analysis to aid in Alzheimer’s disease diagnosis: A fractal dimension approach”, *Computer Speech and Language*, vol. 30, no.1 pp. 43-60, 2015.
- [14] S. Ahmed, A. M. F. Haigh, C. A. de Jager, P. Garrard, “Connected speech as a marker of disease progression in autopsy-proven Alzheimer’s disease,” *Brain*, vol. 136, no. 12, pp. 3727-3737, Oct. 2013.
- [15] J. J. G. Meilán, F. Martínez-Sánchez, J. Carro, J. A. Sánchez, E. Pérez, “Acoustic Markers Associated with Impairment in Language Processing in Alzheimer’s Disease,” *The Spanish journal of psychology*, vol. 15, no. 02, pp.487-494, Jul. 2012.
- [16] K. Lopez-de-Ipiña, J. B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C. Travieso, M. Ecay-Torres, P. Martinez-Lage, H. Egiraun, “On Automatic Diagnosis of Alzheimer’s Disease based on Spontaneous Speech Analysis and Emotional

[26] T. M. Elliott, F. E. Theunissen, "The Modulation Transfer Function for Speech Intelligibility," *PLoS Comput Biol*, vol. 5, no. 3, 2005, doi:10.1371/journal.pcbi.1000302.

[۲۷] ع. آهنگر، م. جعفرزاده فدکی، الف. صحتی، "بررسی کاربرد ابزار روابط واژگانی در گفتگوهای سالمندان آلزایمری و عادی"، *مجله اصول بهداشت روانی*، سال هجدهم، شماره اول، صفحه ۲۲-۸، دی و بهمن ۹۴.

[۲۸] غ. مالک زاده، الف. گلفام، م. شهابی، "بررسی تطبیقی ابزار انسجام در گفتار سالمندان عادی و مبتلا به آلزایمر در زبان فارسی"، *مجله علوم پزشکی دانشگاه آزاد اسلامی مشهد*، دوره ۵، شماره ۳، پیاپی ۱۹، صفحه ۱۵۳-۱۶۱، پاییز ۸۸.

[29] M. R. Scha'dler, B. T. Meyer, B. Kollmeier, "Spectro-temporal modulation subspace-spanning filter bank features for robust automatic speech recognition," *Acoustical Society of America*, vol. 131, no. 5, pp. 4134-51, February 2012.

[۳۰] م. رحمانی، م. مومنی، "تحلیل سیگنال گفتار بیماران آلزایمری"، *هوش محاسباتی در برق*، سال یازدهم، شماره اول، صفحه ۸۱-۹۵، بهار ۹۹.

[31] R. Santoro, M. Moerel, F. De Martino, R. Goebel, K. Ugurbil, E. Yacoub, E. Formisano, "Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex," *PLoS Comput Biol*, vol. 10, no. 1, pp. 1-14, 2014.

[32] <https://isr.umd.edu/Labs/NSL/Software.html>

[33] J. Weiner, M. Engelbart, T. Schultz, "Manual and Automatic Transcriptions in Dementia Detection from Speech," *Interspeech 2017*, Aug. 2017.

[34] J. Weiner, C. Herff, and T. Schultz, "Speech-Based Detection of Alzheimer's Disease in Conversational German," *Interspeech 2016*, Sep. 2016.