علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۸، شماره ۲، ۱۳۹۹
صفحات ۹۴- ۱۰۴
مقاله عادی

کلاسه‌بندی چند کلاسه‌ی تصور کلمات فارسی در سیگنال‌های مغزی

محمدرضا اصغری بجستانی[۱]، غلامرضا محمدخانی[۲*]، وحیدرضا نفیسی[۳]

[*]نویسنده مسئول، دریافت: ۱۴۰۰/۰۲/۲۱، بازنگری: ۱۴۰۰/۰۳/۰۸، پذیرش: ۱۴۰۰/۰۳/۳۰

[۱] دانشجوی دکتری مخابرات، پژوهشکده برق و فناوری اطلاعات، سازمان پژوهشهای علمی و صنعتی ایران، تهران، ایران
[۲] استادیار، پژوهشکده برق و فناوری اطلاعات، سازمان پژوهشهای علمی و صنعتی ایران، تهران، ایران
[۳] دانشیار، پژوهشکده برق و فناوری اطلاعات، سازمان پژوهشهای علمی و صنعتی ایران، تهران، ایران

چکیده

گفتار یا مکالمه خاموش فرایندی است که فرد بدون حرکت دادن عضلات خود برای گفتن کلمه، صدای کلمات را تصور می‌کند. استفاده از سیگنال‌های مغزی شخصی که در حال گفتار خاموش می‌باشد برای تشخیص کلمات واقعی در نظر گرفته شده، می‌تواند گامی بزرگ در جهت کمک به افراد دارای ناتوانی جسمی (مانند سندرم قفل شدگی) برای برقراری ارتباط مؤثر با دیگران باشد. پیشرفت‌های اخیر در فن آوری‌ها و دستگاه‌های ضبط سیگنال‌های مغز، به ویژه الکتروانسفالوگرام (EEG)، این تحقیق را در شناخت گفتار خاموش ممکن کرده است. در حالی که این تحقیقات در گام‌های اولیه است، مطالعات منتشر شده نتایج امیدوار کننده‌ای را نشان داده‌اند.

اگرچه در سطح بین المللی فعالیت‌های متعددی برای شناسایی اصوات در مکالمه خاموش صورت گرفته، ولی تا کنون برای کلمات فارسی کار زیادی انجام نشده است. در این پژوهش سعی کرده‌ایم پس از طرح آزمایش و ثبت داده‌ها، با روش‌های کلاسه بندی گوناگون و با استناد به ویژگیهای متنوع، تشخیص بین چند کلمه فارسی انجام گردد. ۶ کلمه پر کاربرد فارسی و سکوت برای این منظور انتخاب شده‌اند. دقت‌های بدست آمده شامل متوسط ۶۰٪ برای تفکیک بین دو کلمه، متوسط ۴۰٪ برای تفکیک بین دوکلمه به همراه سکوت و ۳۳٪برای ۶ کلمه و سکوت روی دادگان دراز مدت است. این مقادیر بالاتر از متوسط احتمال موفقیت در انتخاب تصادفی (به ترتیب ۵۰ و ۳۳ و ۱۴درصد) هستند. بعلاوه نتایج حاصل نشان دهندهٔ تاثیر بالای عامل "زمان" بر دقت کلاسه‌بندی است بطوری که سیگنال‌های با فاصله زمانی زیاد براحتی و با دقت بسیار بالا از یکدیگر تفکیک می‌شوند، حتی اگر مربوط به یک فعالیت باشند. برعکس تمایز بین سیگنال‌های تقریباً همزمان، حتی هنگام تصور دو کلمهٔ متفاوت بسیار مشکل است. همچنین ملاحظه شد که در تمام موارد قدرت تفکیک با افزایش تعداد کانال‌های داده افزایش می‌یابد.

کلمات کلیدی: مکالمه خاموش، تصور گفتار ، رابط مغز-رایانه، سیگنال‌های مغزی، EEG، کلاسه‌بندی.

۱- مقدمه

در سال‌های اخیر توسعهٔ رابط‌های مغز-رایانه[۱] (BCI) با استفاده از سیگنال‌های مغزی و به خصوص الکتروانسفالوگراف[۲] (EEG) مورد توجه بسیاری از محققین قرار گرفته‌است[۱-۳]. کدگشایی از سیگنال‌های مغزی در طی فعالیت‌هایی مانند تصور حرکت[۳]، درک و دریافت تصاویر[۴] و تولید یا فهم گفتار از زمینه‌های اصلی توسعه رابط‌های مغز-رایانه می‌باشند.

در تصور حرکت، سیگنال مغزی مربوط به تصور حرکت دست، پا، چشم، زبان یا ماهیچه‌های دیگر مورد بررسی قرار می‌گیرد. در این حالت ممکن است هیچ حرکتی فیزیکی وجود نداشته باشد؛ یا خواست خود فرد و یا به خاطر ناتوانی‌های جسمی-حرکتی (مثل فلج یا قطع نخاع و رشته‌های عصبی) و یا حتی عدم وجود (قطع) عضو موردنظر. نتایج تشخیص این نوع تصورات حرکتی از روی سیگنال مغزی را می‌توان برای انجام عمل مناسب- مثل حرکت صندلی چرخدار یا حرکت یک بازوی مکانیکی- استفاده کرد[۴و۵].

در کاربردهای درک تصویر از سوژه خواسته می‌شود که به تعدادی تصاویر نشان داده شده نگاه کند. تصاویر ممکن است شامل عکس اشیا یا حیوانات، مناظر طبیعی، تصاویر هندسی یا حتی حروف و کلمات باشد. خروجی رابط انسان-کامپیوتر در این مورد می‌تواند بسیار متنوع باشد، مثل تشخیص نوع تصویر (مثلاً حیوان یا گیاه)، نوع شکل (دایره یا مربع)، اسم یا مشخصه‌ای از فرد (مثل جنسیت، آشنا/ناآشنا یا دوست/دشمن بودن). یامی‌تواند کلمه یا حرفی باشد که شخص در

نظر گرفته است. پتانسیل وابسته به رخداد[۵] ERP-P300 یکی از مفیدترین و شناخته شده‌ترین ویژگی‌های سیگنال‌های مغزی برای این نوع کاربردهاست[۶و۷].

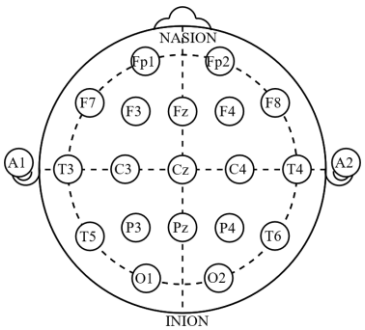
در تصور گفتار[۶] که به نام گفتار خاموش[۷] و مکالمهٔ خاموش[۸] هم شناخته می‌شود، شرکت کنندگان[۹] برای یک فاصلهٔ زمانی مشخص، تلفظ یک مصوّت[۱۰] [۱، ۸-۱۰]، سیلاب[۱۱] [۱۱-۱۴] یا کلمهٔ کامل[۱۲] [۱۵-۱۸] را در ذهن خود تصور می‌کنند، بدون آن که صدایی تولید کنند یا لب و دهان خود را حرکت دهند. سیگنال EEG مربوط به این فاصلهٔ زمانی برای تشخیص واژهٔ مورد نظر پردازش می‌شوند.

در این پژوهش، ما سیگنال‌های مغزی متناظر با تصور گفتار ۶ کلمه کامل فارسی را مورد بررسی قرار داده‌ایم. هدف اصلی رسیدن به یک ساختار طبقه‌بند مناسب است که بتواند تعداد متفاوتی از کلمات (کلاس‌ها) را با حداقل توان محاسباتی و قاعدتاً بیشترین دقت ممکن برای مصارف زمان-حقیقی طبقه‌بندی نماید. در این مقاله روش تحقیق، ساختار طبقه‌بند نهایی و بعضی نتایج مفید را بیان کرده‌ایم.

۲- روش تحقیق:

۲-۱- ابزار ثبت سیگنال

در این تحقیق، برای ثبت سیگنال‌های EEG از یک سیستم ثبت سیگنال EEG-3840 که توسط شرکت نگار اندیشگان ساخته شده است، استفاده کرده‌ایم. این سیستم، یک سیستم کلینیکی است و از اجزای مختلف سخت‌افزاری و نرم‌افزاری تشکیل شده‌است. تعداد کانال‌های EEGی دستگاه، ۲۱ عدد است که بر اساس استاندارد۱۰-۲۰ (شکل ۱) نام گذاری شده‌اند. توضیحات کامل این دستگاه در سایت شرکت نگار اندیشگان [۲۱] قابل مشاهده است.

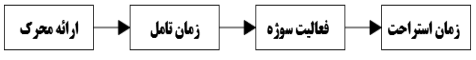


شکل ۱- نام و موقعیت قرارگیری کانال‌های EEG در استاندارد ۱۰-۲۰ [۱۹]

۲-۲- شرح آزمایش

مغز انسان به طور دائم و بدون توقف در حال انجام فعالیت‌های پیچیده و متعددی است. با اطلاعات فعلی ما در مورد ساختمان و نحوهٔ عملکرد مغز و همچنین ابزارها و الگوریتم‌های پردازشی در دسترس، امکان پردازش و تحلیل عملکرد مغز در زمان فعالیت عادی و کنترل نشدهٔ آن وجود ندارد. بنابراین برای پردازش مؤثرتر سیگنال‌های مغزی باید محدودیت‌هایی برای آن درنظر گرفته شود.

بر اساس مطالعات انجام شده روی پژوهش‌های مشابه و تجربه شخصی مجریان از روش انجام کار[۲۲]، برای ثبت هدفمند و زمانبندی شدهٔ سیگنال EEG برای کاربرد مکالمه‌ی خاموش، آزمایشی به صورت کلی شکل ۲ در نظر گرفتیم :



شکل ۲- روند کلی آزمایش برای ثبت سیگنال مغزی

در این آزمایش مراحل زیر به ترتیب صورت می‌پذیرد :

ارائه محرک :

شامل تولید هرگونه محرکی است که سوژه با دریافت آن آماده‌ی اجرای فعالیت شود. ارائه محرک می‌تواند شامل نمایش یک علامت روی صفحه، پخش یک صدا، روشن شدن یک چراغ و . . . باشد.

زمان تامل[۱۳]:

معمولاً ارائه محرک به سوژه، خود باعث برانگیخته شدن یک فعالیت مغزی در او می‌شود. اگر بخواهیم به غیر از این فعالیت، سیگنال مربوط به فعالیت دیگری را ثبت کنیم، ابتدا لازم است قبل از شروع فعالیت جدید مدتی صبر کنیم تا اثر محرک در فعالیت مغزی از بین برود. بسته به نوع محرک این زمان می‌تواند از چند دهم ثانیه تا چند ثانیه متغیر باشد.

فعالیت سوژه[۱۴]:

در این مرحله سوژه فعالیتی را که برای او تعریف کرده‌ایم انجام می‌دهد. فرض بر این است که سیگنالی که در این زمان از سوژه ثبت می‌شود، سیگنالی مرتبط با آن فعالیت باشد. البته این فرض همیشه به طور کامل درست نیست. زیرا اولا مغز همیشه فعالیت‌های زمینه‌ای دیگری نیز دارد و ثانیا ممکن است فعالیت انجام شده از طرف سوژه کاملاً منطبق بر آنچه تعریف شده نباشد. اگر به هر دلیل صحت فرض اولیه زیر سؤال برود، قاعدتاً سیگنال ثبت شده در آزمایش نیز نامعتبر شناخته خواهدشد. در غیر این صورت سیگنال دریافتی متناظر با فعالیت خواسته‌شده تلقی می‌شود.

زمان استراحت:

اغلب با یک بار اخذ سیگنال نمی‌توان به درستی و با دقت بالا ویژگی‌های مفید سیگنال را شناسایی و استخراج کرد. به همین دلیل معمولا آزمایش طرح‌شده در یک جلسه چندین بار پشت سر هم تکرار می‌شود (همچنین ثبت سیگنال در چند جلسه با فاصله زمانی نیز رایج است). به منظور جلوگیری از خستگی سوژه لازم است بین هر دو تکرار آزمایش، زمانی برای استراحت در نظر گرفته شود، تا در آن سوژه بتواند آزادانه بعضی فعالیت‌ها که در حین ثبت مجاز نیست (مثل حرکت دست، تغییر جزئی وضعیت نشستن، قورت دادن آب دهان و ...) را انجام دهد. این فاصله زمانی در حدود چند ثانیه است.

برای ایجاد توالی و زمان بندی لازم در آزمایش، نرم افزار مخصوص با استفاده از زبان دلفی[۱۵] طرح و پیاده‌سازی شد. بعلاوه این برنامه محرک‌ها را به سوژه ارائه و اطلاعات زمان بندی[۱۶] لازم برای پردازش داده‌ها را به آن اضافه می‌کند.

۲-۳- دادگان

در کاربرد حاضر، هدف، تفکیک بین کلمات است. ۶ کلمه را در قالب سه جفت دو کلمه‌ی برای این منظور در نظر گرفته شده‌اند. این سه جفت عبارتند از: "بله-خیر" ، "بالا-پایین" و "چپ-راست". در هر آزمایش یکی از این کلمات یا سکوت به صورت تصادفی برای سوژه پخش می‌شود. از سوژه خواسته شده که هر کلمه‌ای را که شنید، سه مرتبه به صورت شمرده، بدون آنکه بر زبان بیاورد، تکرار کند. در صورت پخش سکوت، کاربر هیچ فعالیت خاصی انجام نمی‌دهد. دلیل دسته بندی کلمات محدودیت زمان ثبت پیوستهٔ سیگنال به خاطر جلوگیری از خستگی سوژه‌ها بوده است.

در یک نشست هر یک از سه محرک (دو کلمه بعلاوه‌ی سکوت) ۲۰ بار به صورت شبه تصادفی تکرار می‌شوند. عمل ثبت در ۵ نشست (هر نشست در یک روز با فواصل چند روزه) برای هر سوژه انجام می‌شود. در هر روز نیز هر یک از سه ترکیب، دوبار مورد آزمایش قرار گرفته و سیگنال مغزی سوژه ثبت شد. به این ترتیب در یک روز هرکلمه به تعداد ۴۰ بار و سکوت جمعا ۱۲۰ بار مورد ثبت قرار گرفته‌اند.

برای ثبت سیگنال‌های مغزی متناظر با مکالمه‌ی خاموش، پخش صوتی کلمه‌ی مورد نظر را به عنوان محرک در نظر گرفتیم. دلیل استفاده از این محرک(به جای محرک‌های دیداری رایج در آزمایش‌های مشابه) آن است که می‌توانیم از کاربر بخواهیم در حین آزمایش چشم خود را بسته نگه دارد. این امر به طور محسوسی تمرکز سوژه روی آزمایش را افزایش می‌دهد و باعث حذف بسیاری از فعالیت‌های ناخواسته‌ی مغز ناشی از توجه به پیرامون می‌شود. به علاوه با این کار یکی از مهم‌ترین سیگنال‌های مزاحم در EEG که مربوط به پلک زدن یا حرکت چشم است نیز به طور کامل حذف خواهد شد.

با توجه به ساده و کوچک بودن کلمات در آزمایش، زمان پخش محرک حدود ۵۰۰ تا ۷۰۰ میلی ثانیه است. فعالیتی که کاربر باید انجام دهد، سه بار تکرار کلمه‌ی شنیده شده به صورت شمرده در ذهن خود است. از کاربر خواسته می‌شود پس از شنیدن کامل صدا، تکرار کلمات را انجام دهد. این کار حدود ۲ ثانیه طول می‌کشد. پس از ۴ ثانیه از شروع محرک اول، محرک دوم به سوژه داده می‌شود. تعداد دفعات تکرار هر کلمه در هر ثبت ۲۰ در نظر گرفته شده است. بنابر این زمان کل هر ثبت پیوسته[۱۷] برابر ۲۴۰ ثانیه خواهد بود.

تعداد افراد شرکت کننده در آزمایش ۵ نفر است که همگی مرد، سالم و راست دست بوده‌اند. از سوژه خواسته می‌شد که روی یک صندلی ثابت، مقابل یک میز، به صورت آرام بنشیند. پس از نصب کلاه و الکترودها روی سر، از طریق نرم افزار ثبت، کیفیت اتصال الکترودها به سر مشاهده و در صورت لزوم تا رسیدن به حالت مناسب اصلاح می‌شد.

هنگام ثبت، سوژه با چشمان بسته، با تمرکز روی آزمایش و کمترین فعالیت جانبی به محرک‌های صوتی پخش شده توسط نرم افزار زمان‌بندی آزمایش گوش داده و فعالیت خواسته شده (تکرار سه مرتبه کلمه شنیده شده به صورت خاموش و شمرده) را انجام می‌دهد کیفیت ثبت سیگنال و فعالیت سوژه در حین ثبت مشاهده و در فاصله بین دو ثبت یادآوری‌های لازم به سوژه داده می‌شد. در صورت رضایت بخش نبودن کیفیت سیگنال یا رفتار سوژه در حین ثبت، مثلاً خستگی خواب آلودگی یا فعالیت اضافه، در صورت لزوم آزمایش تکرار شده است.

۳- پردازش داده‌ها

داده‌های ذخیره شده توسط سیستم ثبت، به صورت فایل‌های رقومی به نرم‌افزارهای پردازش صادر می‌شود. این داده‌ها به صورت خام برای پردازش مناسب نیستند و لازم است ابتدا پیش پردازش‌هایی روی آن‌ها صورت گیرد. از مهم ترین این پیش پردازش‌ها می‌توان فیلتر کردن باندهای فرکانسی مناسب، برش ناحیه‌ی مورد نظر، نمونه برداری مجدد با فرکانس مناسب، تغییرسطوح سیگنال (مثلاً نرمال کردن) ، اعمال فیلترهای خاص در بازه‌های زمان یا فرکانس، را نام برد.

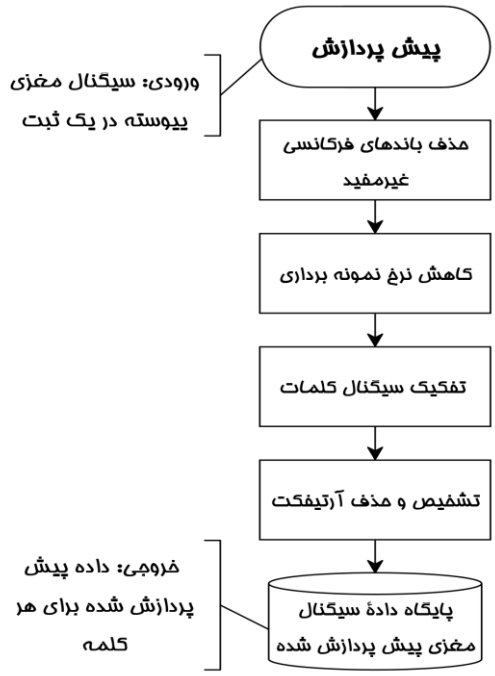
پس از پیش پردازش سیگنال، باید ویژگی‌های مناسب را از آنها استخراج کنیم. این ویژگی‌ها اطلاعات موجود در داده‌ها را که برای کاربرد خاص مورد نظر مفید هستند بسته‌بندی می‌کنند. استخراج ویژگی مناسب علاوه بر کاهش حجم داده‌های پردازش شده توسط سیستم، با بالا بردن کیفیت عملکرد الگوریتم‌ها رسیدن به هدف را بسیار سریعتر و آسانتر می‌کند.

در این قسمت ابتدا پیش پردازش‌های انجام شده روی سیگنال‌های EEG توضیح داده شده اند. در ادامه انواع ویژگی‌های سیگنال‌های EEG که در این پژوهش مورد استفاده قرار گرفتند، معرفی شده‌اند. در انتها طبقه‌بند و روش طبقه‌بندی مورد استفاده توضیح داده شده است.

کلیه برنامه‌های پردازشی توسط نرم‌افزار متلب[۱۸] پیاده سازی شده‌اند.

۳-۱- پیش پردازش

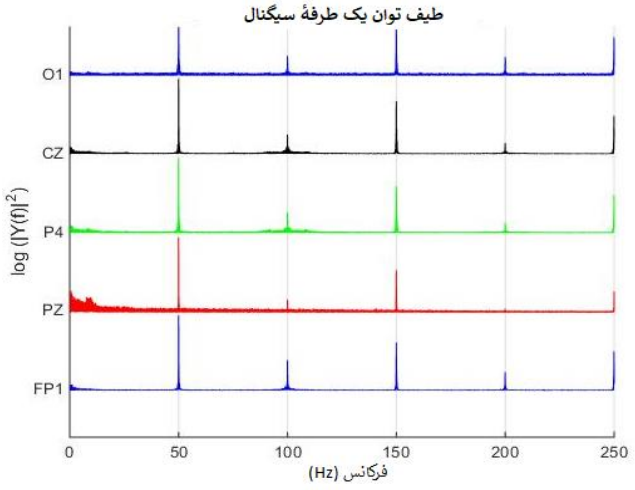
نوع و ترتیب پیش پردازش‌های انجام شده روی داده‌ها در شکل ۳ نشان داده شده و در ادامه توضیح داده شده‌اند.



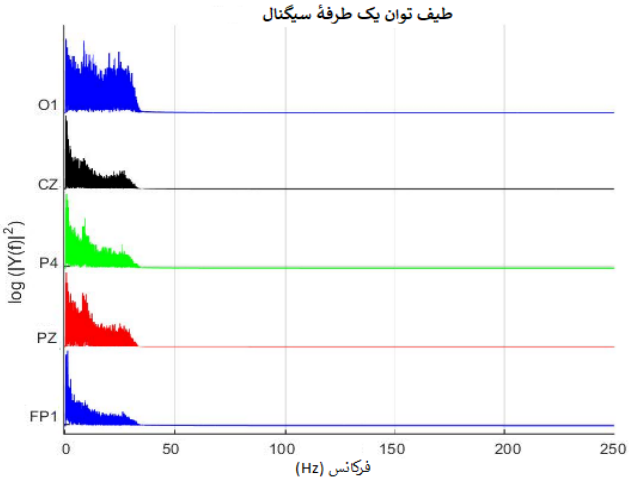
شکل ۳- پیش پردازش‌های انجام شده روی داده‌ها

۳-۱-۱- فیلتر کردن باندهای فرکانسی غیرمفید

در تحقیق حاضر فرکانس نمونه برداری سیگنال‌های EEG، Fs=500Hz بوده است. بنابر این حداکثر فرکانس موجود در داده‌های خام ذخیره شده ۲۵۰Hz است که علاوه بر آنکه بسیار بالاتر از فرکانس‌های مفید است، شامل محدوده فرکانس برق شهر (۵۰ هرتز) و ضرایب آن نیز هست که با توجه به قدرت زیاد آن، سیگنال اصلی EEG را تحت الشعاع قرار می‌دهد. با توجه به آنکه سیگنال مورد نظرما، تا اوایل باند گاما، حداکثر فرکانس ۳۲ هرتز را لازم دارد و همچنین برای حذف موثر مولفه‌ی ۵۰ هرتز، داده‌های خام را از یک فیلتر پایین گذر با فرکانس قطع ۳۲هرتز گذراندیم. شکل ۴-الف طیف سیگنال اولیه و شکل ۴ Error! Reference source not found.-ب طیف توان سیگنال حاصل را نمایش می‌دهد. واضح است که نویز ۵۰ هرتز حذف و سیگنال‌های فرکانس پایین EEG آشکار شده‌اند. قله موجود در فرکانس حدود ۱۰ هرتز مربوط به امواج آلفا است که هنگام بسته بودن چشم‌ها، بخصوص در کانال‌های P و O ظاهر می‌شوند.



(الف)



(ب)

شکل ۴- طیف توان سیگنال‌های EEG. الف- داده‌ی خام ب- پس از اعمال فیلتر پایین گذر ۳۲ هرتز

۳-۱-۲- تغییر نرخ نمونه برداری[۱۹]

نرخ نمونه برداری اولیه داده‌ها ۵۰۰ هرتز است که با توجه به وجود حداکثر فرکانس ۳۲ هرتز در سیگنال فیلتر شده، طبق قضیه نایکویست، نرخ نمونه برداری حداقل ۶۴ برای آن کافیست. بنابراین برای کاهش حجم و زمان پردازش داده‌ها، آن‌ها را به نرخ ۱۰۰ هرتز تبدیل کردیم. دلیل انتخاب این فرکانس، سادگی پیاده سازی کاهنده‌ی نرخ[۲۰]، با حذف ۴ داده از هر ۵ عدد داده‌ی متوالی است. در واقع هر ۵ داده‌ی متوالی را با میانگین‌شان جایگزین کردیم. با این کار، حجم داده‌ها به ۲۰ درصد مقدار اولیه کاهش می‌یابد، بدون آنکه تغییر محسوس در شکل و کیفیت داده‌ها ایجاد شود.

۳-۱-۳-تفکیک سیگنال‌های مربوط به هر کلاس

همانطور که دیدیم هر آزمایش شامل سه محرک است که در یک ثبت پیوسته، هر کدام از آنها ۲۰ بار با ترتیب شبه تصادفی تکرار می‌شوند. به عبارت دیگر هر ثبت دارای ۲۰ دورهٔ[۲۱] ۱۲ ثانیه‌ای شامل سه محرک (جمعاً ۲۴۰ ثانیه) است.

به علاوه در هر نشست چند ثبت با کلمات مختلف گرفته شده‌اند. تعداد کل کلاس‌های داده‌ی موجود ۷ عدد -شامل شش کلمه و سکوت- است. هنگام کلاسه‌بندی برای آموزش و ارزیابی سیستم پردازش باید داده‌های هر کلمه به صورت جداگانه به سیستم داده شوند. برای راحتی در انجام این پروسه، داده‌های مربوط به هر کلاس در هر ثبت تفکیک شده و به صورت جداگانه ذخیره می‌شوند. هنگام پردازش می‌توان هر کدام از داده‌ها را انتخاب و به دلخواه با سایرین مقایسه کرد. برای سادگی هر کلمه یا کلاس را در برنامه با یک شماره ارجاع می‌دهیم. نام و شماره کلمات در جدول ۱ آمده است.

جدول ۱ : نام و شماره کلمات (کلاس‌ها)

شماره (به تفکیک ثبت)	شماره (کل دادگان)	کلمه
۱۰ تا ۱۹	۱	"بالا"
۲۰ تا ۲۹	۲	"بله"
۳۰ تا ۳۹	۳	"چپ"
۴۰ تا ۴۹	۴	"خیر"
۵۰ تا ۵۹	۵	"پایین"
۶۰ تا ۶۹	۶	"راست"
۷۰ تا ۷۹	۷	سکوت

بعلاوه دو کانال A1 و A2 را نیز که اغلب به عنوان مرجع در نظر گرفته شده و به لاله‌ی دو گوش متصل می‌شوند، استفاده نکرده‌ایم. برای کاهش حجم داده‌های ذخیره شده‌ی میانی و افزایش سرعت خواندن و نوشتن آن‌ها، داده کانال‌های اضافی را در فایل‌های داده‌ی پیش پردازش شده حذف و فقط داده‌های ۱۹ کانال EEG را ذخیره کرده‌ایم. ترتیب ذخیره یا شماره کانال‌های EEG در جدول ۲ آمده است.

جدول ۲ : نام و شماره اندیس کانال‌های EEG

سمت راست سر		بالای سر		سمت چپ سر	
شماره	نام کانال	شماره	نام کانال	شماره	نام کانال
۲	FP2	۱۷	FZ	۱	FP1
۴	F8	۱۸	CZ	۳	F7
۶	F4	۱۹	TZ	۵	F3
۸	T4			۷	T3
۱۰	C4			۹	C3
۱۲	T6			۱۱	T5
۱۴	P4			۱۳	P3
۱۶	O2			۱۵	O1

۳-۱-۴- تشخیص و حذف آرتیفکت[۲۲]

به نویزهای موثر بر سیگنالهای مغزی آرتیفکت گویند. آرتیفکت‌ها انواع و اقسام بسیار گوناگونی دارند. برخی از آنها به راحتی قابل تشخیص هستند اما بعضی به گونه‌ای با سیگنال واقعی حاصل از فعالیت مغزی شباهت دارند که تشخیص آن‌ها نیاز به تبحر بسیار بالایی دارد و البته تشخیص صحیح آن‌ها به صورت اتوماتیک نیز تقریبا غیر ممکن است.

در تحقیق حاضر اولین رویکرد ما در مقابله با آرتیفکت‌ها اجتناب بوده است. یعنی شرایط را به گونه‌ای فراهم کنیم که آرتیفکت کمتری تولید شود. این اقدامات عبارتند از:

• طرح آزمایش با استفاده از محرک شنیداری به جای دیداری طوری که سوژه هنگام ثبت چشم‌ها را می‌بندد. این کار اصلی‌ترین آرتیفکت چشمی را که پلک زدن است حذف می‌کند و بقیه انواع آرتیفکت چشمی را نیز بسیار کاهش می‌دهد.

• آموزش اولیه سوژه‌ها برای انجام کمترین فعالیت جانبی (بخصوص حرکتی و عضلانی) حین آزمایش

برای حذف اثر سیگنال برق شهر از روش تصحیح استفاده کرده‌ایم که شامل بکارگیری یک فیلتر دیجیتال پایین گذر مناسب می‌شود.

با حذف عوامل اصلی خطا قسمت‌هایی که دارای آرتیفکت قابل تشخیص یا تصحیح باشند بسیار کاهش داده شده اند. در نهایت با بازبینی ثبت‌ها- حین اجرا و پس از آن- آنهایی که کیفیت خوبی نداشتند، از فرآیند پردازش حذف کرده‌ایم. در هنگام پردازش داده‌ها نیز واریانس مقطعی سیگنال (در محدوده زمانی یک دوره) محاسبه و درصورت تفاوت عمده با مقادیر پیش فرض یا محاسبه شده، آن دوره را از چرخه‌ی محاسبات خارج می‌کنیم.

با این وجود، متاسفانه برخی خطاها مثل جابجایی الکترودها، اتصالات ضعیف متناوب، سیگنال‌های مزاحم محیط اطراف، تاثیر رفت و آمد و سر و صدا در محیط و ... بطور کامل قابل پیشگیری، تصحیح و حتی تشخیص و حذف نبوده است.

۳-۲- استخراج ویژگی

در روش‌های کلاسه‌بندی، علاوه بر خود ابزار کلاسه‌بند، انتخاب ویژگی‌های مناسب که بر اساس آن بتوان یک کلاسه‌بندی با کیفیت داشت، اهمیت زیادی دارد.

پس از آزمایش چندین نوع ویژگی شناخته شده برای سیگنال‌های مغزی، در نهایت دامنه ضرایب تبدیل فوریه‌ی سیگنال به عنوان ویژگی استفاده شد. برای استخراج این ویژگی، تبدیل فوریهٔ سیگنال‌های هر کلمه پس از پیش پردازش محاسبه می‌گردد. با توجه به مختلط و متقارن بودن تبدیل فوریهٔ سیگنال‌های حقیقی، دامنه یا قدر مطلق ضرایب محاسبه شده و از نیمهٔ دوم آن صرف نظر می‌کنیم. به‌علاوه برای کاهش ابعاد و همچنین آسان‌تر شدن پروسه انتخاب محدودهٔ فرکانس مورد نظر، مقادیر به دست آمده به گونه‌ای نمونه برداری مجدد می‌شوند که به ازای هر یک هرتس طیف فرکانس، یک ضریب موجود باشد. در این صورت عدد اول ضرایب مقدار DC، عدد دوم دامنه محدوده فرکانسی یک هرتز، عدد سوم دامنه محدوده فرکانسی ۲ هرتز و ... خواهد بود. همچنین این کار سبب می‌شود که طول بردار ویژگی ثابت و مستقل از طول سیگنال EEG باشد. با توجه به آن که هنگام پیش پردازش، سیگنال‌ها با فیلتر پایین گذر صفر تا صد هرتز فیلتر شده و به نرخ ۱۰۰ نمونه در ثانیه نمونه برداری مجدد شده‌اند، طبیعی است که حداکثر طول تبدیل فوریه در این حالت ۵۰ و حداکثر فرکانس موجود در آن ۳۲ هرتز است. بنابراین اگر کل محدوده فرکانسی سیگنال را انتخاب کنیم، طول بردار ویژگی برای هر کانال ۳۲ خواهد بود. در انتها برای از بین بردن اثر دامنهٔ سیگنال در ویژگی‌ها، آن را با تقسیم بر مجموع مقادیر مؤلفه‌ها، نرمال می‌کنیم.

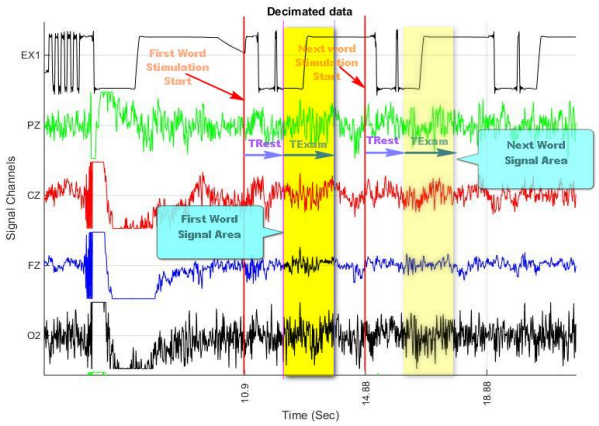
لازم به ذکر است تبدیلات مجدد روی این ویژگی، مثل تبدیل به طیف توان یا استخراج تقریب AR، تقریب نمایی و ... بررسی شده و بهبودی در دقت طبقه بندی نداشتند.

قبل از استخراج ویژگی لازم است ابتدا محدوده زمانی سیگنال مربوط به هر کلاس در آزمایش را مشخص کرده و آن را برش دهیم. برای تعیین محدوده سیگنال مورد نظر از دو پارامتر استفاده شده است (شکل ۵):

الف- T_{Rest} : طول محدوده زمانی مربوط به پخش سیگنال صوتی و پاسخ مغز به آن است. شروع این محدوده، زمان شروع محرک صوتی است که توسط سیگنال‌های همزمانی مشخص شده است.

ب- T_{Exam} : طول زمانی محدوده مربوط به فعالیت سوژه و تکرار کلمه به صورت خاموش است. شروع این محدوده پس از پایان T_{Relax} در نظر گرفته می‌شود.

این دو پارامتر برای تمام کلمات ثابت بوده و در برنامه تنظیم و به تابع استخراج ویژگی منتقل می‌شود.



شکل ۵ : پارامترهای انتخاب محدوده سیگنال مربوط به هر کلمه برای استخراج ویژگی

۳-۳- روش طبقه‌بندی

در این قسمت ساختار نهایی طبقه‌بند مورد استفاده معرفی و چگونگی ارزیابی عملکرد آن توضیح داده می‌شود.

۳-۳-۱- ساختار طبقه‌بند

برای کلاسه بندی بین دو کلاس، به سادگی از یک ماشین بردار پشتیبان[۲۳] (SVM) استفاده کردیم. این ماشین توسط دو گروه ویژگی مربوط به نمونه‌های دو کلاس آموزش داده می‌شود. ماشین آموزش دیده با ارائه یک یا چند مجموعه از ویژگی‌ها می‌تواند مشخص کند که هر کدام از ویژگی‌ها مربوط به چه کلاسی بوده است.

برای کلاسه بندی چند کلاسه، یک کلاسه‌بند مبتنی بر ماشین‌های بردار پشتیبان دو کلاسه و رای اکثریت طرح شده و مورد استفاده قرار گرفت. در این کلاسه‌بند برای N کلاس، تعداد $N(N\text{-}1)/2$ کلاسه‌بند باینری وجود دارد که هر یک برای تفکیک دو کلاس از هم آموزش داده می‌شود.

فرض کنید $SVM_{i,j}$، ماشینی باشد که برای تفکیک دو کلاس i و j آموزش داده شده است. هر کدام از i و j می‌توانند از ۱ تا N تغییر کنند. مسلماً این ماشین هر بردار ویژگی ورودی x را به یکی از دو کلاس i یا j نسبت خواهد داد؛ هرچند ممکن است در واقع x متعلق به هیچکدام از آنها نباشد. واضح است که:

$$SVM_{i,j} = SVM_{j,i}$$

پس بدون تغییر در مسأله می‌توانیم فرض کنیم:

$$i = 1..N\,, \qquad j = i+1..N$$

به عبارت دیگر هر چند برای $j<=i$ عملا ماشینی وجود ندارد اما برای هر کلاس، تعداد N-1 کلاسه‌بند وجود دارد که هر کدام آن را، از یکی از کلاس‌های دیگر، تشخیص می‌دهد. هنگام استفاده از این کلاسه‌بند، مجموعه ویژگی با کلاس مجهول x به تمام کلاسه‌بندها داده می‌شود. هر یک رای خود را در مورد کلاس آن ارائه می‌دهد و کلاسی که بیشترین رأی را آورده باشد به عنوان کلاس برنده شناخته شده و مجموعه ویژگی به آن کلاس تعلق می‌گیرد:

$$l_{i,j} = SVM_{i,j}(x) \quad i = 1..N, \qquad j = i+1..N$$

$$x = \mathrm{Max}_{k=1}^{N}\{\,\mathrm{count}[(l_{i,j} == k)]\}$$

نحوه عملکرد این کلاسیفایر را می‌توان به این صورت توضیح داد که هر یک از کلاسه‌بندهای $SVM_{i,j}$ فقط برای تشخیص بین دو کلاس i و j آموزش داده شده و بنابراین خروجی آن یکی از دو مقدار i یا j خواهد بود. اگر مجموعه ویژگی داده شده به این کلاسه‌بند متعلق به یکی از دو کلاس i یا j باشد، انتظار می‌رود خروجی کلاسه‌بند، حداقل با احتمالی بزرگتر از ۵۰ درصد، مقدار صحیح کلاس باشد. اگر مجموعه ویژگی متعلق به هیچ یک از این دو کلاس نباشد، باز هم مقدار خروجی، این بار با احتمال حدود ۵۰ درصد، یکی از دو مقدار i یا j خواهد بود. با توضیح فوق مشخص است که اگر کلاس ویژگی مجهول x باشد، تعداد رای برای تمام مقادیر $k \neq x$ مقداری در حدود $(N-1)/2$ است و تنها برای $k = x$ تعداد رای بیش از این مقدار و حداکثر N-1 خواهد شد.

در طبقه‌بندی چند کلاسه، موقع رای گیری ممکن است دو یا چند کلاس دارای تعداد رای اکثریت نسبی یکسان باشند. در چنین حالتی می‌توان قواعد خاصی برای تصمیم گیری اتخاذ کرد. در کلاسه‌بند فعلی اگر دو مقدار، مثلاً k1 و k2 دارای تعداد رای حداکثر و برابر باشند، کلاس خروجی برابر خروجی کلاسه‌بند بین این دو کلاس در نظر گرفته خواهد شد($k = l_{k1,k2}$) و اگر بیش از دو مقدار دارای رای اکثریت باشند، کلاس خروجی نامعلوم فرض شده و مقدار k=0 به آن تعلق می‌گیرد. بنابراین کلاسه‌بند N کلاسه (N>2)، دارای N+1 حالت محتمل برای خروجی است.

۳-۳-۲- تخمین کیفیت طبقه‌بندی

با توجه به تعداد نسبتاً کم داده‌ها، برای ارزیابی کلاسه‌بندها از روش ارزیابی متقابل مونته کارلو[۲۴] استفاده کرده‌ایم. در این روش در چندین مرحله هر بار مجموعه‌ی داده‌ها به صورت تصادفی به دو گروه آموزش و آزمایش تقسیم شده و پس از

آموزش کلاسه‌بند با گروه آموزش، نتیجه عملکرد آن روی گروه آزمایش محاسبه می‌شود. میانگین دقت آزمایش‌ها، به عنوان دقت نهایی کلاسه‌بند در نظر گرفته می‌شود.

به صورت اختیاری با در نظر گرفتن ملاحظاتی چون زمان پردازش، عمل آموزش و آزمایش حداقل ۱۰۰ بار تکرار شده است. در هر مرحله ۷۰ درصد دادگان برای آموزش و ۳۰ درصد بقیه برای آزمایش استفاده شده‌اند. میانگین و واریانس این محاسبات، مرجع ارزیابی قرار گرفته است. دادگان مربوط به هر سوژه جداگانه ارزیابی شده‌اند.

۴- نتایج آزمایش

در تمام آزمایش‌های این بخش، پارامترهای استخراج ویژگی، در مقادیری که نتایج کلی بهتری تولید می‌کرده‌اند ثابت شده‌اند: مقادیر طیف از دو ثانیه میانی محدودهٔ زمانی مربوط به هر کلمه استخراج (T_{Rest}=1 sec, T_{Exam}=2 sec,)و از تمام باند فرکانسی، جز مؤلفهٔ DC استفاده شده است(فرکانس ۱ تا ۳۲ هرتز).

لازم به ذکر است که در اینجا منتخبی از نتایج بیان شده‌اند. در بیشتر موارد، نمونهٔ معرفی شده، نمایانگر طیفی از حالات مشابه با نتایج نزدیک به هم است که بیان (یا حتی آزمودن) همهٔ آنها چندان مفید به نظر نمی‌رسد. در مواردی نیز میانگین یا برآیندی از همهٔ حالات ممکن محاسبه و ارائه شده‌اند.

۴-۱- کلاسه‌بندی دو کلاسه

۴-۱-۱- کلاسه‌بندی با یک کانال

ابتدا به بررسی میزان توانمندی هر کدام از کانال‌ها، به تنهایی، در تفکیک دو کلاس می‌پردازیم. چندین حالت برای این منظور در نظر گرفته شده‌اند. در تمام موارد نتایج داده‌های سوژه شماره ۱ گزارش شده است.

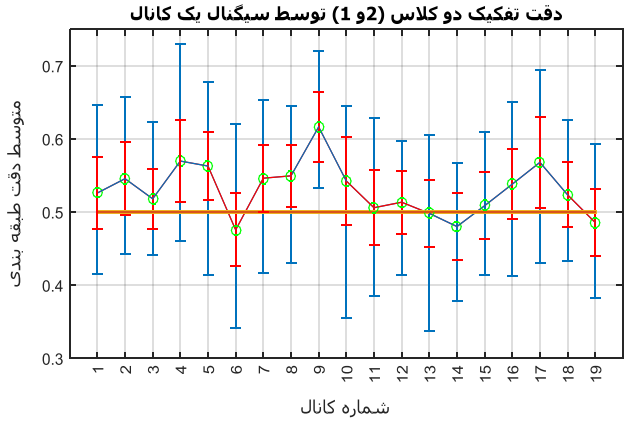
الف- دو کلمه "بالا" و "بله" روی کل دادگان

در اولین بررسی، میزان دقت تفکیک بین دو کلمه "بالا" و "بله" روی تمام نمونه‌های[۲۵] این دوکلمه (کلاس‌های ۱ و ۲)، فقط توسط داده‌های یک کانال EEG را محاسبه کردیم. جدول ۳ نتایج این بررسی را نشان می‌دهد.

جدول ۳: دقت تفکیک دو کلمه "بالا" و "بله" توسط هر یک از کانال‌ها

شماره کانال	دقت طبقه‌بندی(٪)	شماره کانال	دقت طبقه‌بندی(٪)
۱	۵۲/۶ ± ۵/۰	۱۱	۵۰/۶ ± ۵/۱
۲	۵۴/۶ ± ۵/۰	۱۲	۵۱/۴ ± ۴/۴
۳	۵۱/۸ ± ۴/۱	۱۳	۴۹/۸ ± ۴/۷
۴	۵۷/۰ ± ۵/۶	۱۴	۴۸/۰ ± ۴/۶
۵	۵۶/۳ ± ۴/۷	۱۵	۵۰/۹ ± ۴/۷
۶	۴۷/۶ ± ۵/۰	۱۶	۵۳/۸ ± ۴/۸
۷	۵۴/۶ ± ۴/۶	۱۷	۵۶/۸ ± ۶/۳
۸	۵۵/۰ ± ۴/۳	۱۸	۵۲/۴ ± ۴/۵
۹	۶۱/۷ ± ۴/۹	۱۹	۴۸/۶ ± ۴/۶
۱۰	۵۴/۳ ± ۶/۱		

شکل ۶ مقادیر دقت را بر حسب شماره کانال نمایش داده است. در این شکل محور افقی شماره کانال و محور عمودی دقت طبقه‌بندی است. علامت دایره نمایانگر میانگین ، خطوط قرمز[۲۶] حول میانگین نشان دهنده انحراف معیار و خطوط آبی نشان دهنده محدوده مقادیر بدست آمده برای دقت، در تکرارهای مختلف ارزیابی متقابل مونته-کارلو است. خط افقی قهوه‌ای رنگ نیز سطح شانس[۲۷] یا دقت تصمیم‌گیری تصادفی است.

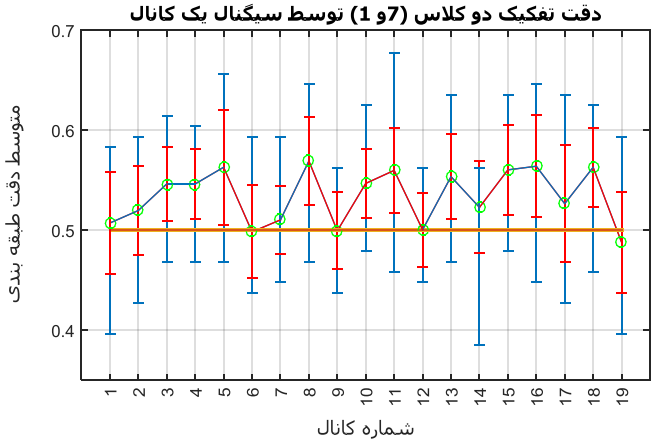


شکل ۶ : دقت تفکیک دو کلمه "بالا" و "بله" توسط هر یک از کانال‌ها در کل دادگان

بر اساس این نتایج فقط کانال شماره ۹ (C3) قادر است به تنهایی این دو کلاس را، به صورت ضعیف، از هم تفکیک کند و بقیه‌ی کانال‌ها عملکردی در سطح شانس دارند. محل این کانال (قسمت مرکزی سمت چپ مغز سوژه) تقریباً قسمتی از مغز است که مسئولیت پردازش صوت و گفتار را بر عهده دارد.

ب- دو کلمه "بالا" و "سکوت" روی کل دادگان

دقت تفکیک بین کلمه "بالا" و "سکوت" توسط هر یک از کانال‌ها اعمال شده روی کل دادگان ثبت شده(کلاس‌های ۱ و ۷) در شکل ۷ نشان داده شده است. مشاهده می‌شود که نسبت به حالت قبل کانال‌های بیشتری توانسته‌اند بین این دو کلاس تفکیک ایجاد کنند؛ هرچند دقت این تفکیک هنوز بسیار ناچیز است.

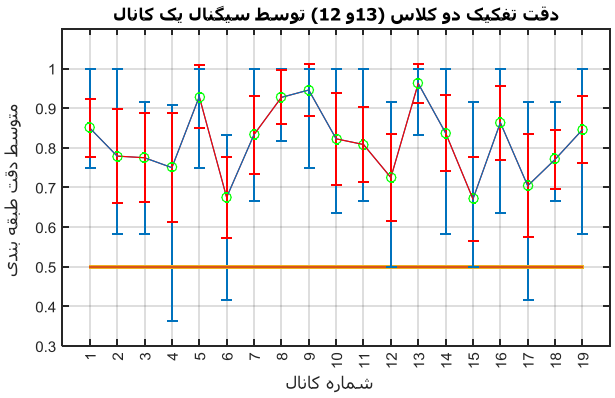


شکل ۷: دقت تفکیک کلمه "بالا" و "سکوت" توسط هر یک از کانال‌ها در کل دادگان

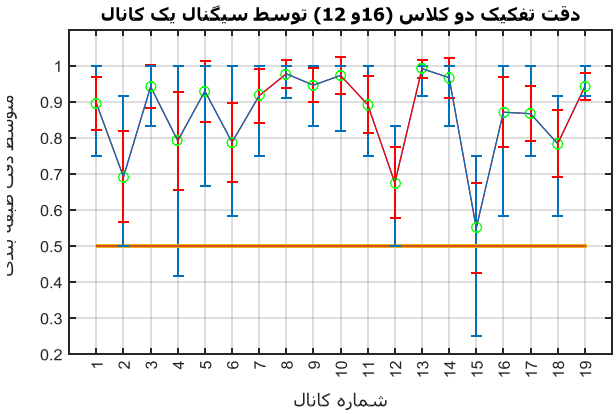
۴-۱-۲- معرفی اثر زمان بر دقت کلاسه‌بندی

یکی از مشکلات اساسی هنگام کار با سیگنال‌های EEG، ایستا نبودن[۲۸] این سیگنال است که باعث می‌شود ویژگی‌های سیگنال با زمان تغییر می‌کند. برای بررسی این واقعیت قابلیت تفکیک داده‌های یک کلاس (کلمه) بین جلسات مختلف را در نظر گرفتیم.

شکل ۸، نتیجه تفکیک کلاس‌های (۱۲و۱۳)- کلمه "بالا" در ثبت‌های دوم و سوم، با فاصله زمانی چند دقیقه- و (۱۲و۱۶) - همان کلمه در دو ثبت با فاصله زمانی چند روز -توسط هر یک از کانال‌ها را نشان می‌دهد.



(الف)



(ب)

شکل ۸: نتیجه کلاسه‌بندی یک کلمه توسط یک کانال روی داده‌های جلسات مختلف الف- با فاصله چند دقیقه ب- با فاصله چند روز

این نتایج نشان می‌دهد که اولاً داده‌های غیر همزمان، حتی از یک کلاس و با فاصله زمانی نسبتاً کوتاه، براحتی از یکدیگر قابل تفکیک هستند؛ ثانیاً با افزایش فاصله زمانی، اختلاف و در نتیجه تفکیک‌پذیری آنها افزایش می‌یابد. این مساله سبب می‌شود ادغام داده‌های زمان‌های مختلف به عنوان یک کلاس، کار گروه‌بندی و تفکیک آنها را مشکل کرده و کیفیت نهایی سیستم را کاهش دهد. بنابراین بهتر است کلاسه‌بندها را برای کار روی داده‌های با فاصله زمانی کم آموزش داده و آزمایش کنیم.

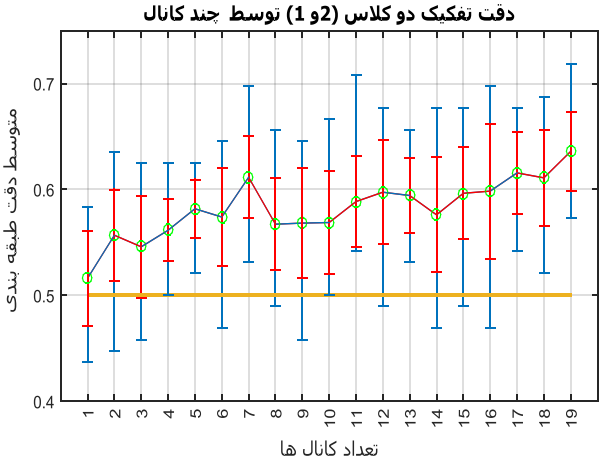
لازم به ذکر است که در برخی کاربردها برای کاهش اثر زمان، خصوصیاتی از سیگنال EEG بلافاصله قبل (یا بعد) از سیگنال مورد پردازش را به عنوان داده بایاس[۲۹]، پایه، زمینه یا امثال آن در نظر گرفته و در محاسبات وارد می‌کنند. مثلا هنگام محاسبه ERP-P300 متوسط سیگنال هر کانال در یک فاصله زمانی کوتاه قبل از ارائهٔ محرک، از سیگنال بعد از آن کم می‌شود. در پژوهش حاضر، ما هم با توجه به نوع ویژگی در نظر گرفته شده یعنی دامنهٔ مؤلفه‌های فرکانسی، مقادیر همین مؤلفه‌ها در یک فاصلهٔ زمانی کوتاه قبل از سیگنال اصلی را به عنوان فعالیت زمینه در گرفتیم که البته تأثیر چندانی در نتایج نداشت.

۴-۱-۳- کلاسه‌بندی با چند کانال

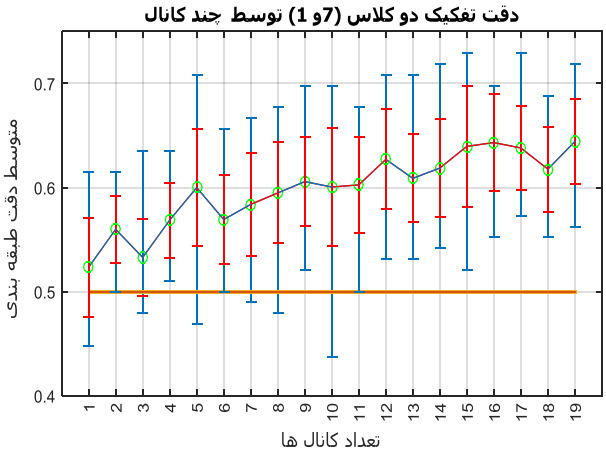
حال اثر افزایش تعداد کانال‌ها روی دقت تفکیک دو کلمه را بررسی می‌کنیم. مثل قسمت قبل ابتدا میزان دقت تفکیک بین دو کلمه "بالا" و "بله" و سپس کلمهٔ "بالا" و سکوت را روی کل دادگان یک سوژه، منتها این بار توسط داده‌های چند کانال EEG با هم محاسبه کردیم. شکل ۹ نتایج این بررسی را نشان می‌دهد. دقت کنید که در اینجا محور افقی تعداد کانال‌ها را نشان می‌دهد نه شمارهٔ کانال. به عبارت دیگر، مقدار iام نمودار، دقت بدست آمده توسط تمام کانال‌های ۱ تا i است نه فقط کانال i.

ملاحظه می‌شود که در هر دو مورد، با افزایش تعداد کانال‌ها، دقت تفکیک تقریبا به طور ثابت افزایش یافته است و در نهایت به طور معناداری از سطح تصادف بیشتر شده است. البته دقت نهایی باز هم خیلی زیاد نیست (حدود ۶۵ درصد) اما بوضوح از حالت شبه تصادفی فاصله گرفته است. همچنین در بیشتر موارد دقت تفکیک سکوت از یک کلمه (در واقع گفتار از غیر گفتار) بیش از دقت تفکیک دو کلمه از هم است.

مسلماً ترکیبات بسیار زیادی برای انتخاب چند کانال از میان ۱۹ کانال موجود است و پیدا کردن بهترین ترکیب با در نظر گرفتن طول بردار ویژگی خود مسأله‌ای پیچیده است ولی روند کلی تغییرات نشان می‌دهد که بکارگیری همه کانال‌ها برای طبقه‌بندی، حداقل از حیث دقت، نزدیک به بهینه است. به همین خاطر در ادامه همیشه از تمام کانال‌ها استفاده خواهیم کرد. استفاده از تمام مولفه‌های فرکانسی نیز با روندی مشابه استنتاج شده است.



(الف)

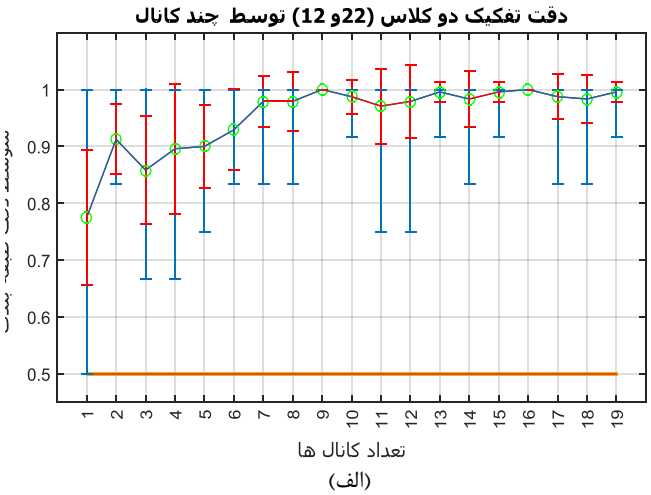


(ب)

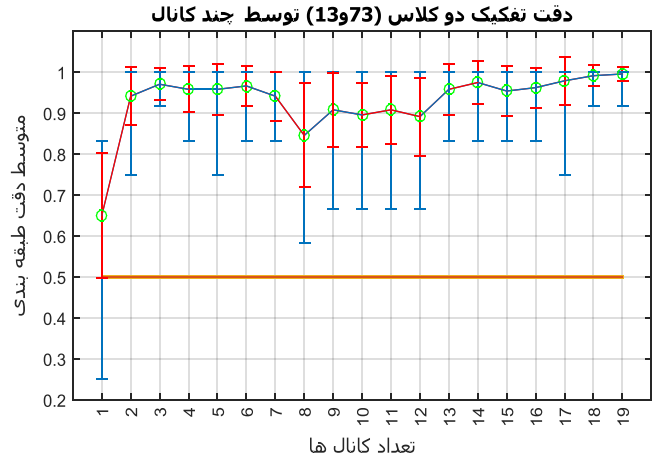
شکل ۹ : دقت تفکیک دو کلمه در کل دادگان توسط چند کانال: الف-"بالا" و "بله"، ب- "بالا" و سکوت

۴-۱-۴- کلاسه‌بندی با چند کانال روی داده‌های یک نشست

باتوجه به اثر زمان، می‌توان انتظار داشت که با کلاسه‌بندی روی داده‌های با فاصله زمانی کم، به دقت بالاتری برسیم. با این فرض، قابلیت طبقه‌بندی نمونه‌های دو کلمه با فواصل زمانی کوتاهتر را در نظر گرفتیم. نتیجه تفکیک کلاس‌های ۱۲و۲۲ (کلمه "بالا" و "بله" در ثبت دوم که با فاصله زمانی چند دقیقه از هم اخذ شده‌اند) و همچنین ۱۳و۷۳ ("بالا" و سکوت در ثبت سوم) در شکل ۱۰ آورده شده است.



(الف)



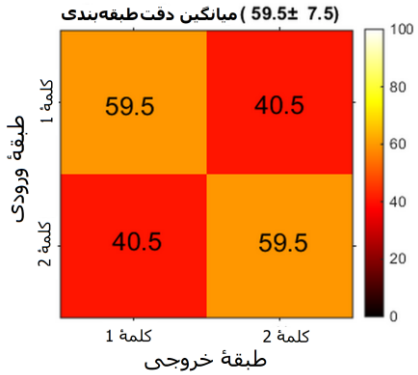
(ب)

شکل ۱۰: دقت تفکیک دو کلمه در یک ثبت توسط چند کانال: الف-"بالا" و "بله" در ثبت دوم، ب- "بالا" و سکوت در ثبت سوم

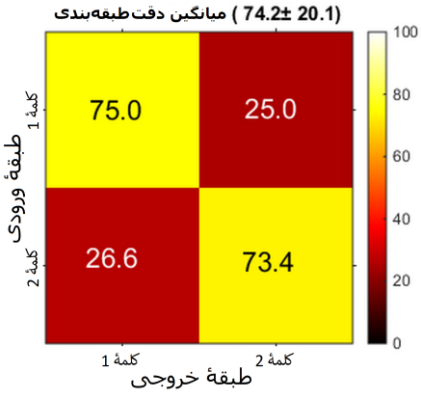
این نمودارها نشان می‌دهند که دقت تفکیک به مقدار قابل ملاحظه‌ای نسبت به حالت استفاده از کل دادگان(شکل ۹) افزایش یافته است بطوری که با بکارگیری بیش از ۹ یا ۱۰ کانال، جدا سازی تقریبا به صورت کامل انجام شده و افزایش دقت به حالت اشباع رسیده است.

۴-۱-۵- نتیجه کامل کلاسه‌بندی دو کلمه‌ای

نتایج بیان شده در قسمت‌های قبل همگی فقط روی نمونه‌های خاص از کلمات محاسبه شده بودند. برای ارزیابی کامل کلاسه‌بند، در این قسمت نتایج کامل کلاسه‌بندی دو کلاسه ارائه می‌شود. این نتایج متوسط دقت کلاسه بندی روی تمام کلمات، ثبت‌ها و سوژه‌هاست. به عبارت دیگر دقت طبقه بندی برای تمام ترکیب‌های دوتایی ۷ کلاس (۲۱=۶/۲*۷ حالت)، برای هر سوژه محاسبه شده و ماتریس تداخل[۳۰] آن استخراج شده است. سپس متوسط تمام این نتایج به عنوان نتیجه نهایی محاسبه شده است. شکل ۱۱ این نتایج را نشان می‌دهد. در قسمت الف تمامی داده‌های پنج نشست هر سوژه مورد استفاده قرار گرفته است. در قسمت ب کلاسه‌بندی فقط روی داده‌های یک نشست انجام شده است. ملاحظه می‌شود که نتایج کلی با نتایج جزئی بیان شده در قسمت‌های قبل تطابق دارد. متوسط دقت کلاسه‌بندی بر روی داده‌های نشست‌های روزانه ۲۰/۱ ± ۷۴/۲ و روی تمامی دادگان ۷/۵ ± ۵۹/۵ درصد بوده‌است.



(الف)



(ب)

شکل ۱۱: ماتریس تداخل کلاسه‌بندی دو کلاسهٔ کامل الف- کل دادگان، ب- متوسط نشست‌ها

۴-۲- کلاسه‌بندی چند کلاسه

در این بخش به توصیف نتایج کلاسه بندی بر روی بیش از دو کلاس می‌پردازیم. با توجه به کیفیت بهتر نتایج هنگام بکارگیری تمام کانال‌ها، در اینجا نتایج با سیگنال هر ۱۹ کانال محاسبه شده است. در این حالت با توجه به ثابت بودن تمام پارامترهای پردازش، برای نمایش نتیجه از ماتریس تداخل استفاده کرده‌ایم. در این ماتریس عدد نوشته شده در سطر i و ستون j مشخص کننده‌ی احتمال آن است که یک داده از کلاس i، به عنوان کلاس j شناسایی شود. ستون آخر احتمال آنکه داده به هیچکدام از کلاس‌ها طبقه‌بندی نشود را نشان می‌دهد.

مسلماً عناصر قطری احتمال تشخیص درست و مقادیر غیر قطری احتمال تشخیص غلط یک داده هستند. متوسط دقت طبقه‌بندی به صورت میانگین اعداد قطر اصلی تعریف می‌شود.

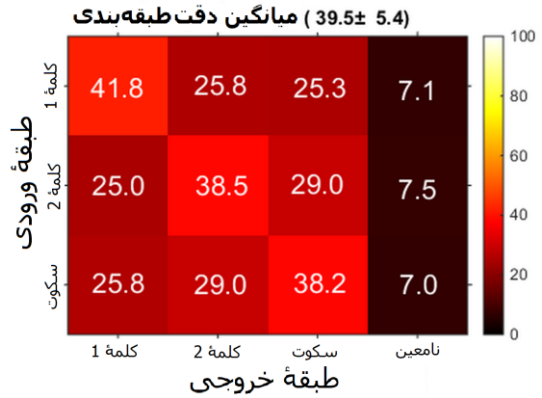
۴-۲-۱- کلاسه‌بندی سه کلاسه

در این قسمت متوسط دقت کلاسه‌بندی بین سه کلاس شامل دو کلمه و سکوت ارائه شده است. مجدداً دو حالت استفاده از کل دادگان یا استفاده از داده‌های یک نشست برای محاسبه دقت در نظر گرفته شده است. نتایج به دست آمده در شکل ۱۲ نشان داده شده‌اند. متوسط دقت روی کل داده‌ها حدود ۵/۴ ± ۳۹/۵ است. در حدود ۴۰ درصد از موارد کلاسه‌بندی به درستی انجام شده است. در حدود ۵۲٪ شناسایی غلط بوده است و در حدود ۷ درصد موارد کلاسه‌بند قادر به شناسایی (درست یا نادرست) نشده است.

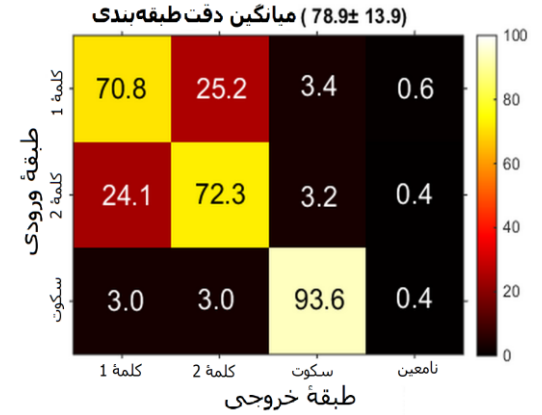
برای حالت طبقه بندی داده‌های روزانه، همانند گذشته دقت بالاتری به دست آمده است. متوسط دقت ۱۳/۹ ± ۷۸/۹ است. بیش از ۷۵ درصد موارد تشخیص درست صورت گرفته و فقط در کمتر از یک درصد موارد کلاسه‌بند قادر

به تشخیص نشده است. بعلاوه دقت تشخیص سکوت به‌مراتب بیش از تشخیص کلمات از یکدیگر بوده است.

لازم به ذکر است که دقت کلاسه‌بندی تصادفی در این حالت ۳۳/۳۳ درصد است و بنابر این عملکرد نهایی طبقه بند بهتر از نرخ شانس است؛ هر چند نتایج آن اصلاً قابل استفاده در یک سیستم رابطه مغز-رایانهٔ واقعی نیست.



(الف)



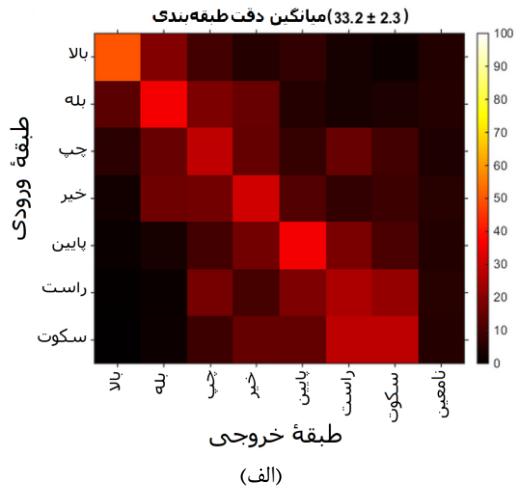
(ب)

شکل ۱۲: ماتریس تداخل طبقه‌بندی سه کلاسه بین دو کلمه و سکوت. الف- متوسط کل دادگان، ب- متوسط داده‌های روزانه

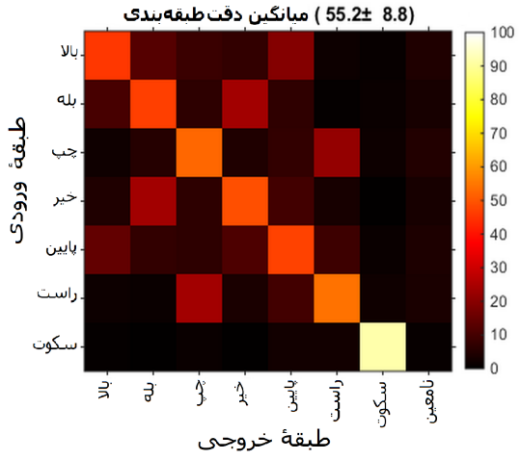
۴-۲-۲- کلاسه‌بندی هفت کلاسه

در این قسمت نتیجهٔ کلاسه‌بندی همهٔ ۷ کلاس دادهٔ موجود، شامل ۶ کلمه و سکوت ارائه شده است. ماتریس تداخل برای هر دو حالت استفاده از کل دادگان یا داده‌های روزانه در شکل ۱۳ دیده می‌شود. دقت روی کل داده‌ها حدود ۲/۳ ± ۳۳/۲ و متوسط دقت با داده‌های روزانه ۸/۸ ± ۵۵/۲ است. نرخ کلاسه‌بندی تصادفی در این حالت حدود ۱۴/۳ درصد است و بنابر این عملکرد نهایی طبقه بند باز بسیار بهتر از نرخ شانس است.

نکتهٔ قابل تأمل در شکل ۱۳-ب آن است که بیشترین میزان تداخل بین کلماتی است که با هم در یک آزمایش قرار دارند، مثل "بالا" و "پایین" یا "چپ" و " راست". فاصلهٔ زمانی نمونه‌های این کلاس‌ها بسیار کوتاه و در حد چند ثانیه است. احتمالاً هنگام طبقه‌بندی این کلاس‌ها وابستگی زمانی بالای آنها، مانع تفکیک مناسب می‌شود. مجدداً سکوت با دقت بالایی از بقیهٔ کلمات تفکیک شده‌است.



(الف)



(ب)

شکل ۱۳: ماتریس تداخل طبقه‌بندی ۷ کلاسه بین ۶ کلمه و سکوت. الف- متوسط کل دادگان، ب- متوسط داده‌های روزانه

۵- مقایسه و نتیجه‌گیری

موضوع تحقیق در این پژوهش، طراحی و پیاده‌سازی یک کلاسه‌بند انعطاف پذیر برای طبقه‌بندی سیگنال‌های مغزی EEG طی تصور کلمات است که بتواند تعداد مختلفی کلمه را تشخیص دهد. معیارهایی مثل سادگی، عدم نیاز به آموزش مجدد با تغییر تعداد کلمات، نیاز محاسباتی و حافظهٔ کم و البته دقت حداکثر برای سیستم مد نظر بوده است.

تعدادی زیادی مقاله مرتبط با مکالمه خاموش در کنفرانس‌ها و مجلات مختلف ارائه شده است. در فصل اول تعدادی از این مقالات بررسی و معرفی شدند. کارهای اولیه اغلب روی کلاسه‌بندی زیرکلمات و آواها متمرکز شده بودند اما در چند سال اخیر، تقریباً همزمان با پژوهش ما، نتایج بعضی کارهای نزدیک که آنها نیز به کلاسه‌بندی کلمات پرداخته‌اند منتشر شده است.

در این فصل دو نمونه از نتایج منتشر شده از پژوهش‌های مشابه را برای مقایسه با نتایج بدست آمده از فعالیت حاضر –که در فصل قبل ذکر شد- آورده‌ایم. مورد اول از کارهای نسبتاً قدیمی ولی معروف در زمینه گفتار خاموش است. مورد دوم نیز از جدیدترین کارهای منتشر شده است که علاوه بر وجود چند کلاس، بررسی روی کلمات کامل صورت گرفته است. مسلماً با توجه به عدم دسترسی کامل به دادگان و روش پردازش این مقالات نمی‌توانیم به صورت کامل نتایج آنها را برهم یا با نتایج حاصل از کار خودمان تطبیق دهیم اما می‌توانیم مقایسه‌ای نه چندان دقیق بین آنها داشته باشیم.

وجه اشتراک تمام این مقالات آن است که همه آنها به گونه‌ای برای رسیدن به نتایج مناسب امیدوارکننده هستند اما هیچکدام نتوانسته‌اند به نتیجه قابل اطمینان و کاربردی دست یابند. البته بعضی کارها نیز درصد موفقیت بسیار بالایی را گزارش کرده اند که اغلب روش آزمایش و پردازش آنها مبهم و حتی در مواردی تردید برانگیز است.

یکی از اولین و پرارجاع‌ترین مقالات منتشر شده در زمینه مکالمه خاموش مرجع [۱۳] است. این پژوهش به منظور کلاسه بندی دو سیلاب /ba/ و /ku/ در آزمایشی روی 7 نفر با استفاده از دستگاه ثبت EEGی 128 کاناله و نرخ نمونه برداری 1024 هرتز، انجام شده‌است. آزمایش طراحی شده در این مقاله نیز از رایجترین روش‌های ثبت سیگنال مغزی در مکالمه خاموش است. ایده اولیه آزمایش مورد استفاده ما هم از همان سبک گرفته شده است. نتایج گزارش شده در منبع [۱۳] در جدول ۴ آورده شده است.

جدول ۴: نتایج کلاسه‌بندی بین دو آوای /ba/ و /ku/ برای هریک از هفت سوژه‌ی مورد آزمایش در [۱۳]

تعداد آزمایش	نرخ کلاسه‌بندی	سوژه
۱۵	۵۶	سوژه ۱
۱۱	۸۸	سوژه ۲
۲	---	سوژه ۳
۱۳	۴۶	سوژه ۴
۲۴	۷۵	سوژه ۵
۱۲	۸۱	سوژه ۶
۸	۶۷	سوژه ۷
۸۳	۶۸/۸۴	متوسط تمام سوژه‌ها

همانطور که دیده می‌شود نرخ کلاسه‌بندی کمی بالاتر از سطح شانس (۵۰ درصد برای تفکیک بین دو کلاس) است. این آزمایش نشان داد که هدف مکالمه خاموش دست یافتنی خواهدبود.

در یکی از جدیدترین تحقیقات مشابه، میزان دقت کلاسه‌بندی بین پنج کلمه (کلمات UP، DOWN، LEFT، RIGHT و SELECT) روی داده‌های گرفته شده از ۹ سوژه‌ی مرد ۱۸ تا ۳۶ ساله توسط دو نوع آزمایش مختلف بررسی شده است[۲۰].

دو نوع آزمایش برای ثبت داده‌ها در نظر گرفته است. در آزمایش اول که فشردن موس (MC[۳۱]) نامیده شده‌است، سوژه‌ها هر کدام از پنج کلمه را به ترتیب در ذهن خود تصور و زمان شروع و پایان تصور هر کلمه را با فشردن دکمه ماوس مشخص می‌کردند. از این آزمایش هر کلمه ۶۰ مرتبه در دو جلسه ۴۰ و ۲۰ تایی ثبت گردیده است. در آزمایش دوم به نام پنجرهٔ زمانی ثابت(TF[۳۲])، سوژه‌ها با دیدن یک علامت + به مدت ۴ ثانیه کلمه را تصور کرده و سپس با خاموش شدن آن ۲ ثانیه استراحت داشتند. از هر کلمه ۴۰ تکرار نیز با این آزمایش ثبت شده است. وسیله‌ی ثبت مورد استفاده، دستگاه ۱۴ کاناله Emotiv Epoch با نرخ نمونه برداری ۱۲۸ بوده است که ۱۲ کانال آن برای استخراج ویژگی استفاده شده است.

با دقت در روش ثبت داده‌های این مرجع، به خصوص در آزمایش TF، مشخص است که از یک طرف نمونه‌های مربوط به هر کلمه از نظر زمانی بسیار نزدیک به هم هستند و از طرف دیگر نمونه‌های کلمات مختلف در ثبت‌های جداگانه و با فاصلهٔ زمانی از هم ثبت شده‌اند. بنابراین با توجه به مطالب ذکر شده در مورد نقش زمان در دقت طبقه‌بندی (بخش ۴-۱-۲)، افزایش دقت قابل پیش‌بینی است.

داده‌ها با فیلتر نیم تا ۵۰ هرتز فیلتر شده‌اند. ویژگی در نظر گرفته شده میانگین و انحراف معیار ۶ گروه ضریب حاصل از تبدیل موجک پنج مرحله ای (ضرایب D1 تا D5 و A5) بوده است و برای کلاسه‌بندی ۴ نوع کلاسه‌بند SVM، NB[۳۳]، RF[۳۴] و LDA[۳۵] بکار رفته‌اند.

نتایج کلاسه‌بندی به صورت جدول ۵ خلاصه شده است. این نتایج با دقت تفکیک روی داده‌های روزانه در این مقاله قابل مقایسه است. بهترین نتایج این مقاله در آزمایش TF و با طبقه‌بند RF به دست آمده‌است. در صورت استفاده از طبقه‌بند RF لازم است با تغییر تعداد یا نوع کلمات، یک طبقه‌بند جدید آموزش داده شود و از این نظر قابلیت گسترش و استفاده مجدد طبقه‌بند استفاده شده در پژوهش ما را ندارد.

جدول ۵: نتایج کلاسه‌بندی اعلام شده توسط [۲۰] روی پنج کلمه

سوژه	فشردن موس (MC)				پنجرهٔ زمانی ثابت (TF)			
	SVM	NB	RF	LDA	SVM	NB	RF	LDA
S1	۶۸/۸	۷۳/۱	۸۷/۲	۵۸/۷	۶۱/۳	۷۴/۳	**۸۶/۴**	۴۹/۷
S2	۴۱/۸	۵۲/۹	۵۷/۱	۴۵/۵	۶۸/۸	۸۲/۹	**۸۴/۴**	۶۷/۳
S3	۵۰/۳	۶۴/۱	۶۹/۸	۵۵	۶۰/۸	۷۲/۴	**۸۸/۹**	۵۸/۳
S4	۶۱/۳	۷۸/۹	۷۹/۳	۵۳/۹	۶۸/۳	۹۱	**۹۸/۵**	۷۴/۳
S5	۳۷	۴۴/۴	۵۴/۶	۳۳/۸	۵۵/۴	۷۶/۹	**۸۰/۴**	۵۱/۸
S6	۶۷/۳	۵۳	۷۰/۴	۵۱/۳	۸۷/۴	۸۲/۹	**۹۳/۹**	۷۶/۵
S7	۴۸/۶	۵۴/۵	۶۰/۹	۴۶/۱	۶۸/۴	۶۶/۹	**۸۳/۵**	۵۴/۹
S8	۵۰/۲	۶۷/۲	۷۲	۴۶	۸۳/۹	۹۵	**۹۷**	۸۳/۹
S9	۴۹/۸	۶۷/۲	۷۳/۱	۵۶/۶	۴۰/۲	۵۹/۸	**۷۳/۸**	۴۰/۲
میانگین	۵۲/۷	۶۱/۷	۶۹/۳	۴۹/۶	۶۶	۷۸	**۸۷/۴**	۶۱/۸

۶- مراجع

[1] C. DaSalla, H. Kambara, Y. Koike, M. Sato, "Spatial filtering and single-trial classification of EEG during vowel speech imagery", *Proceedings of the International Convention on Rehabilitation Engineering and Assistive Technology (ICREAT), pp. 1-4,* 2009.

[2] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, F. E. Abd El-Samie, "EEG seizure detectionand prediction algorithms: A survey", *EURASIP Journal on Advances in Signal Processing*, pp.183 ,2014.

[3] C. Wang, W. Ding, J. Shan, B. Fang, "A Review of Research on Brain-Computer Interface Based on Imagined Speech". In: Sun F., Liu H., Fang B. (eds) Cognitive Systems and Signal Processing. ICCSIP 2020. *Communications in Computer and Information Science*, vol 1397, pp 370-378, 2021.

[4] L. Mondada, M. E. Karim, and F. Mondada, "Electroencephalography as implicit communication channel for proximal interaction between humans and robot swarms", *Swarm Intelligence*, vol. 10, no. 4, pp.1–19, 2016.

[5] K. Tanaka, K. Matsunaga, and H. O. Wang, "Electroencephalogram based control of anelectric wheelchair", *IEEE Transactions on Robotics,* vol. 21, no. 4, pp. 762–766, 2005.

[6] F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D.J. Krusienski, and T. M. Vaughan, "A P300-based brain–computer interface for people with amyotrophic lateral sclerosis", *Clinical neurophysiology*, vol. 119, no. 8, pp. 1909-1916, 2008.

[7] E. Donchin and Y. Arbel, "P300 based brain computer interfaces: a progress report,Foundations of Augmented Cognition", *Neuroergonomics and Operational Neuroscience*, pp. 724–731, 2009.

[8] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns", *Neural networks*, vol. 22, no. 9, pp. 1334–1339 (2009)

[9] Beomjun Min, Jongin Kim, Hyeong-jun Park, Boreom Lee, "Vowel Imagery Decoding toward Silent Speech BCI Using Extreme Learning Machine with Electroencephalogram", *BioMed Research International*, vol. 2016, Article ID 2618265, 11 pages, 2016.

[10] B. M. Idrees and O. Farooq, Vowel classification using wavelet decomposition during speech imagery, 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, pp. 636- 640 (2016)

[11] S. Deng, R. Srinivasan, T. Lappas, and M. D'Zmura, Eeg classification of imagined syllable rhythm using hilbert spectrum methods, Journal of neural engineering, vol. 7, no. 4, p. 046006 (2010)
[12] J. Kim, S.-K. Lee, and B. Lee, Eeg classification in a single trial basis for vowel speech perception usingmultivariate empirical mode decomposition, Journal of neural engineering, vol. 11, no. 3, p. 036010 (2014)
[13] K. Brigham and B. V. Kumar, Imagined speech classification with eeg signals for silent communication:a preliminary investigation into synthetic telepathy, iCBBE, 2010. IEEE, 2010, pp. 1–4 (2010)
[14] M. DZmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, Toward eeg sensing of imagined speech,HCI 2009. Springer, pp. 40–48. (2009) https://doi.org/10.1007/978-3-642-02574-7 5
[15] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, Analysis and classification of speech imagery eeg for bci, Biomedical signal processing and control, vol. 8, no. 6, pp. 901–908 (2013)
[16] E. F. Gonz´alez-Casta˜neda, A. A. Torres-Garc´ıa, C. A. Reyes-Garc´ıa, and L. Villase˜nor-Pineda, Sonification and textification: Proposing methods for classifying unspoken words from eeg signals, Biomedical Signal Processing and Control, vol. 37, pp. 82–91 (2017)
[17] K. Mohanchandra and S. Saha, A communication paradigm using subvocalized speech: translating brain signals into speech, Augmented Human Research, vol. 1, no. 1, p. 3 (2016)
[18] P Saha, M Abdul-Mageed, S Fels, "SPEAK YOUR MIND! Towards Imagined Speech Recognition With Hierarchical Deep Learning", *arXiv preprint,* 2019. https://arxiv.org/abs/1904.05746
[19]Neupsy Key, https://neupsykey.com/eeg-instrumentation, February 2017.
[20] M. Alsaleh, R. Moore, H. Christensen and M. Arvaneh, "Examining Temporal Variations in Recognizing Unspoken Words Using EEG Signals," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 976-981, 2018.
[۲۱] شرکت طراحی نگاراندیشگان، http://www.negand.com
[۲۲] م. اصغری بجستانی، غ. محمدخانی، س. گرگین، و. نفیسی، غ. فراهانی، "طبقه‌بندی سیگنال‌های مغزی EEG برای تشخیص بین دو واژه در گفتار خاموش،" *پردازش علائم و داده‌ها*، دوره ۱۷، شماره ۲، ص ۱۲۰-۱۱۳، ۱۳۹۹.
http://jsdp.rcisp.ac.ir/article-1-843-fa.html

محمدرضا اصغری بجستانی تحصیلات خود در مقطع کارشناسی و کارشناسی ارشد مهندسی برق مخابرات را به ترتیب از دانشگاه شیراز و تربیت مدرس تهران در سال‌های ۱۳۶۹ و ۱۳۷۳ به اتمام رسانده و هم اکنون دانشجوی دکترای تخصصی همان رشته در سازمان پژوهشهای علمی و صنعتی ایران است. وی از سال ۱۳۷۰ تاکنون به فعالیت‌های آموزشی و پژوهشی در زمینه‌های مختلف مرتبط با رشته تحصیلی خود اشتغال داشته است. او به پژوهش و توسعه سامانه‌های نرم افزاری پردازش سیگنال‌های دیجیتال (صوت، تصویر، بیولوژیک و ...) علاقه‌مند است. آدرس پست الکترونیکی ایشان عبارت است از:
bejestani@irost.ir

غلامرضا محمدخانی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد مهندسی برق از دانشگاه صنعتی شریف تهران به ترتیب در رشته‌های الکترونیک و مخابرات و در سال‌های ۱۳۷۲ و ۱۳۷۶ به اتمام رساند. وی از سال ۱۳۷۸ تا سال ۱۳۸۱ تحصیلات خود در مقطع دکترا را در مدرسه عالی مهندسی ENSIL از دانشگاه لیموژ (Ecole National Supérieur d'Ingénieur de Limoges) در کشور فرانسه ادامه داد. همچنین ایشان از سال ۱۳۸۱ تا ۱۳۸۵ به عنوان محقق پسا دکتری در مرکز تحقیقاتی LSS (Laboratory of Signals and systems) از CNRS در Supelec پاریس مشغول بودند. وی از سال ۱۳۸۵ تاکنون عضو هیئت علمی پژوهشکده برق سازمان پژوهشهای علمی و صنعتی ایران می‌باشد. زمینه‌های مورد علاقه تحقیق او پردازش سیگنال در مخابرات دیجیتال و سیستم‌های بی سیم مانند مخابرات سیار یا ماهواره‌ای می‌باشد. اخیراً نیز به تحقیق در زمینه‌های یادگیری ماشین و پردازش سیگنال با کاربردهای مهندسی پزشکی پرداخته اند و مسئولیت آزمایشگاه رابط مغز - رایانه BCI (Brian Computer Interface) پژوهشکده برق و فناوری اطلاعات سازمان پژوهش‌ها را به عهده دارند. آدرس پست الکترونیکی ایشان عبارت است از:
mohammadkhani@irost.ir

وحیدرضا نفیسی تحصیلات خود در مقطع دکتری رشته مهندسی پزشکی گرایش بیوالکتریک دانشگاه صنعتی امیرکبیر را در سال ۱۳۸۴ به اتمام رسانده و از سال ۱۳۸۱ به عنوان عضو هیات علمی در سازمان پژوهشهای علمی و صنعتی ایران مشغول به کار است. حوزهٔ فعالیت او ابزار دقیق (تجهیزات دیالیز و دستگاه‌های تنفس مصنوعی) و پردازش سیگنال‌های حیاتی است. آدرس پست الکترونیکی ایشان عبارت است از:
vr_nafisi@irost.org

---

[1] Brain-Computer Interface
[2] Electroencephalograph
[3] Motor-imagery
[4] Image perception
[5] Event Related Potential
[6] Speech imagination
[7] Silent-speech
[8] Silent-talk
[9] Participant
[10] Vowel
[11] Syllable
[12] Word
[13] Settling time
[14] Subject
[15] Borland Delphi Ver. 7.0
[16] Timing signals
[17] Continuous Record
[18] MATLAB® R2014b
[19] Resampling
[20] Decimator
[21] Epoch
[22] Artifact
[23] Support Vector Machine
[24] Monte Carlo Cross Validation
[25] Instance
[26] Error bar
[27] Chance Level
[28] Non-Stationary
[29] Bias
[30] Confusion matrix
[31] Mouse-Click
[32] Time Frame
[33] Naïve Bayes
[34] Random-Forest
[35] Linear Discriminant Analysis

The CSI Journal on Computing Science and Information Technology
Vol. 18, No.2, 2021

Multi-class classification of Persian Imagines words

Mohammad Reza Asghari Bejestani[1], Gholam Reza Mohammadkhani[2], Vahid Reza Nafisi[3]

[1, 2, 3] Electrical & IT department, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran

Abstract

Silent talk is a process that one imagines the sound of words in his/her mind without moving his/her muscles to say the word. Using human brain signals to distinguish between considered words during silent talk, could be a great progress in helping people with physical disabilities to communicate effectively with others. This approach can also be useful in situations where visual or auditory communication is undesirable, as in military situations. Recent advances in technologies and devices for brain signal recording, especially electroencephalogram (EEG), have transformed the silent talk recognition to a feasible task. While these researches are in their primary steps, published studies have shown promising results in this particular area of science. Although there are many published works on this subject for some world-wide languages, but there have been few activities on Persian words. Therefore, in this project, it was planned to record required data and try to distinguish between several Persian words using different classification methods and various features of signals. Our main goal was to develop a classifier that can be easily reconfigured for different number of classes while having the best performance with acceptable feature extraction and computational complexities. After trying several approaches, we finally selected normalized frequency spectrum amplitudes as features and a bank of binary SVM classifiers with the majority selection rule as the classifier. Classification accuracies have been examined with Monte-Carlo Cross-Validation and were about 60% for binary, 40% in three classes, and 30% for six words plus silence classification for long time span features. Furthermore, the results showed the significant effect of the time difference between inter and intra-class samples on classification performance.

Keywords: Silent Talk, Imagined speech, Brain Signals, EEG, Classification, Persian words.