**بسمه تعالی**

****

**دانشگاه صنعتی شریف**

**دانشکده مهندسی برق**

**دکتر کربلایی**

**گزارش پروژه سیگنال و سیستم**

**بهار 97**

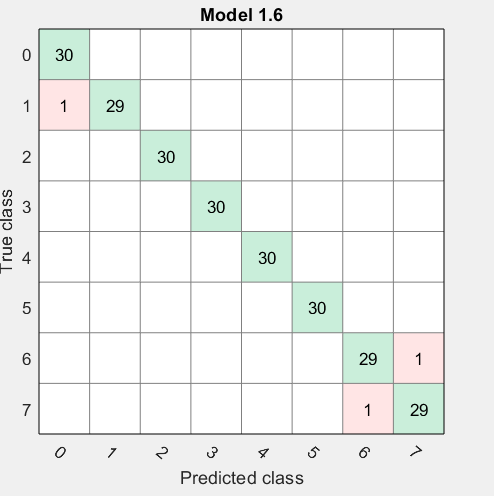
**طاها انتصاری 95101117**

**سوال اول**

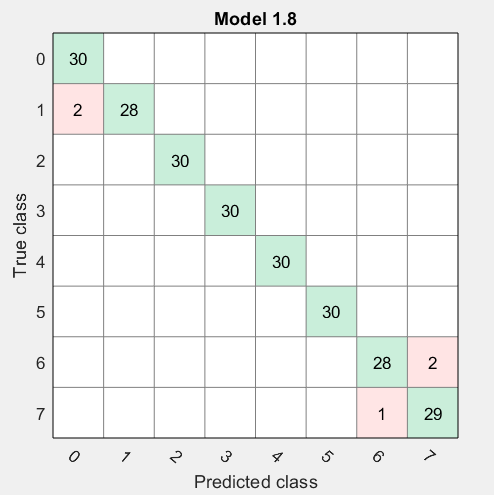
همانطور که در دستور کار گفته­شده­است، سطر اول هر آزمایش داده­ی پرت محسوب می­شوند. بعد از حذف این سطر­ها، با توجه به ذات مسئله که هر 39 سطر باقی­مانده برای یک آزمایش هستند( و نه این که هر یک سطر خود یک آزمایش جداگانه باشند)، هر 39 سطر را به یک سطر تبدیل می­کنیم. در این مرحله، هر یک از 39 سطر را به ستون ها انتقال میدهیم تا در نهایت 12\*39 ستون داشته­باشیم. اکنون درواقع ستون­های 1 تا 12 دارای داده­های لحظه اول سنسورها، ستون­های 13 تا 24 لحظه دوم و الی آخر. اکنون ماتریس ما به جای یک ماتریس 9360\*12 که هر 39 سطر آن یک آزمایش باشند، یک ماتریس 468\*240 داریم که اکنون هر سطر یک آزمایش است و در واقع ستون­ها، ویژگی­های موردنظر ما هستند.

ویژگی های اضافی نیز استخراج و بررسی شدند که در نهایت با توجه به خروجی­های گرفته­شده، خود داده­ها به تنهایی بهترین ویژگی­ها بودند. بعد از الصاق شماره گروه­ها به انتهای ماتریس موردنظر، آن را به برنامه خوشه­بندی­کننده متلب میدهیم. خروجی­های این قسمت همراه با ماتریس خطای آن آمده­اند.

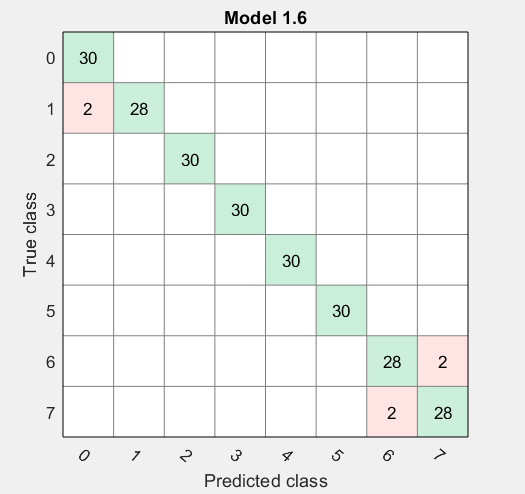
برای حالتی که تنها خود داده­ها را بدون هیچ ویژگی اضافه به طبقه­بندی­کننده بدهیم، به درصد صحت 98.8 با استفاده از SVM خطی میرسیم. ماتریس خطای آن به شکل زیر است:



اگر به ویژگی­های مذکور، مشتق اول و یا همان تغییرات بین لحظه­های متوالی را اضافه کنیم، نتایج بهتر نمیشوند و با استفاده از SVM مکعبی به درصد 97.9 می­رسیم که ماتریس خطای آن هم به صورت زیر است:



برای این حالت، SVM خطی درصد صحت 97.5 و با ماتریس خطای زیر است:



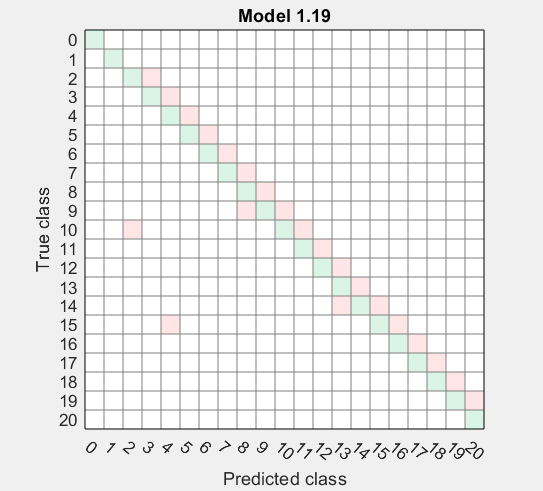
لازم به ذکر است که استفاده از الگوریتم pca برای کاهش تعداد ویژگی و بعد فضا موجب بهتر شدن نتیجه نشد.

**سوال دوم**

برای این سوال مشابه سوال قبل میتوانیم هر 40 سطر را یک آزمایش بگیریم. اما این سوال با سوال قبلی یک فرق ذاتی دارد و آن این که اصولا در این سوال میتوانیم هر یک از 40 سطر را به­عنوان یک آزمایش بگیریم چرا که در طی هر 40 سطر نحوه نشستن فرد تغییر اساسی نمی­کند و میتوانیم با این عمل تعداد داده­های آموزشی برای طبقه­بندی­کننه را افزایش دهیم که نتایج نیز بیانگر همین نکته است. در این سوال نیز نیازی به استخراج هیچ ویژگی نبوده و خود داده­ها به تنهایی کفایت میکنند.

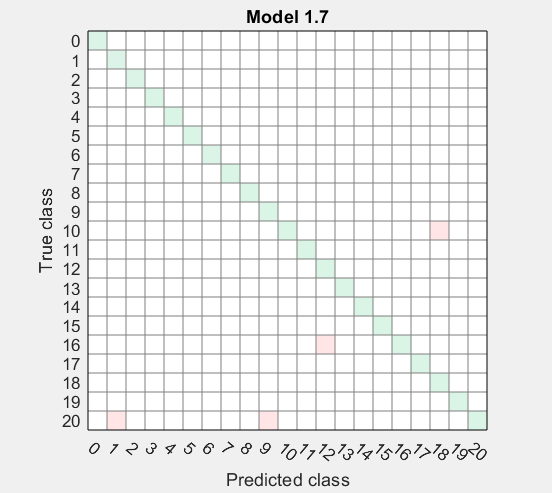
لازم به ذکر است که برخلاف سوال قبل، در این سوال از هر نوع نشستن به تعداد مساوی داده نداریم و برای برخی حالت ها تنها 50 و یا 53 داریم. البته بیشتر حالات 60 بار آزمایش دارند.

اگر ماتریس را مطابق سوال یک تغییر شکل دهیم و هر 39 سطر را تبدیل به یک سطر کنیم و به طبقه­بندی­کننده بدهیم در بهترین حالت به درصد صحت 98.2 با استفاده از الگوریتم Ensemble Bagged Trees میرسیم. شکل ماتریس خطای این حالت به صورت زیر است. البته برای دقیق­تر بودن، ماتریس خطای این قسمت تحت عنوان Ensemble\_confusion به پیوست ارسال­شده­است.



در این حالت 23 مورد از آزمایش­ها درست خوشه­بندی نشده­اند

اکنون اگر به جای تغییر شکل ماتریس همان ماتریس اولیه را بدهیم و با این کار در واقع هر سطر را به عنوان یک آزمایش مستقل ببینیم، مشاهده می­شود که به درصد صحت 100 میرسیم.بهترین نتیجه برای روش SVM مربعی است. البته همچنان تعدادی خطا وجود دارد که به خاطر نگنجیدن در دقت، در عدد 100 ظاهر نشده اند. ماتریس خطای این آزمایش نیزتحت عنوان SVM\_confusion ارسال شده است که شکل آن به صورت زیر است.



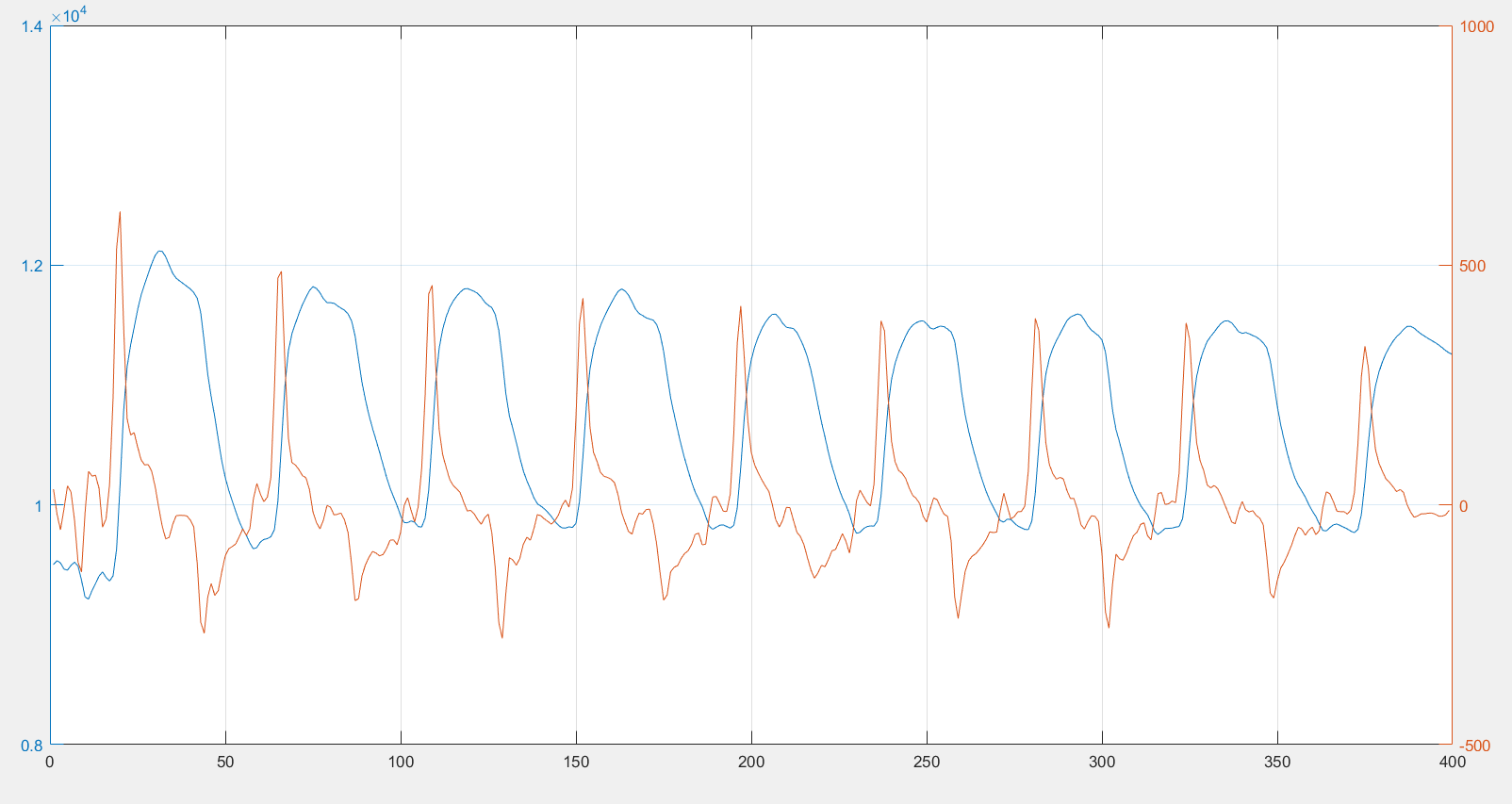
در این حالت 8 مورد از آزمایش­ها اشتباه گروه­بندی شده­اند. البته تعداد آزمایش­ها در این حالت بسیار بیشتر از حالت قبل است. در این حالت ما 47658 آزمایش داریم که 8 مورد را اشتباه تشخیص داده­ایم در حالی که وقتی ماتریس را تغییر شکل داده­بودیم تنها 1222 آزمایش داشتیم که 23 مورد آن اشتباه تشخیص داده­شده­بودند. پس با استفاده از ماتریس اصلی همواره به پاسخ درست دست می­یابیم حتی اگر همه­ی 8 مورد اشتباه تخصیص داده­شده مربوط به یک آزمایش باشد، چرا که در این صورت از 39 سطری که برای یه بار نشستن روی صندلی هستند 21 مورد را هنوز درست تشخیص داده­ایم.

اگر همین روش آخر را به عنوان روش نهایی خود بپذیریم، برای آزمایش­های آتی و استفاده از این مدل آموزش دیده برای پیش­بینی در آینده، بهترین راه میتواند بدین گونه باشد که ما هر چند سطر که داشته باشیم را هر بار جداگانه به مدل بدهیم تا مشخص کند که به کدامین گروه تعلق می­گیرد و پس از این از بین این نتایج، گروهی که بیشینه تعداد اعضا دارد را به عنوان نماینده انتخاب کنیم.

**سوال سوم**

رویکردی اولی که در این سوال پیش­گرفتیم تا به نتیجه برسیم، یافتن پروفایل دم و بازدم و یافتن کوریلیشن آن با داده ها بود. اما این روش چند ایراد داشت. ایراد اول این که در این روش، اگر به نتیجه میرسید، بایستی برای تنفس معمولی و عمیق دو پروفایل جداگانه داشته باشیم. مشکل دوم این بود که در این روش تشخیص پایان دم و پایان بازدم دقت کافی نداشت و بایستی برای دقت بیشتر تعداد دیتای آموزش بسیار بیشتری داشته باشیم. با فرض این­­که مشکلات ذکرشده مانع نبودند و پیش می­رفتیم بدست می­آوردیم که پروفایل دم صعودی و برای بازدم نزولی بود. مشاهده شد که کوریلیشن پروفایل نزولی بازدم نیز در نقاطی بیشینه می­شود که که پروفایل صعودی بیشینه می­شود و در نتیجه به­درد بخور نیست.

رویکرد دوم، که باتوجه به مشاهدات حاصل مانند نمودار زیر حاصل شد، استفاده از مشتق(در واقع تقریب مرتبه اول اویلر مشتق) است. همانطور که در شکل زیر مشاهده می­شود (نمودار زرد نمودار مشتق است) در هنگام شروع دم، مشتق بیشینه می­شود و هنگام شروع بازدم، کمینه می­شود. با توجه به مشاهدات نیز نقطه گذر از صفر بعد از بیشینه یا کمینه شدن به عنوان نقطه پایان دم و یا بازدم انتخاب شد.



پس به دنبال ماکسیمم و مینیمم موضعی در مشتق اول میپردازیم.

در تخمین نقاط شروع و پایان دم و بازدم تنها از کانال­های 1 و 2 استفاده­شده­است.

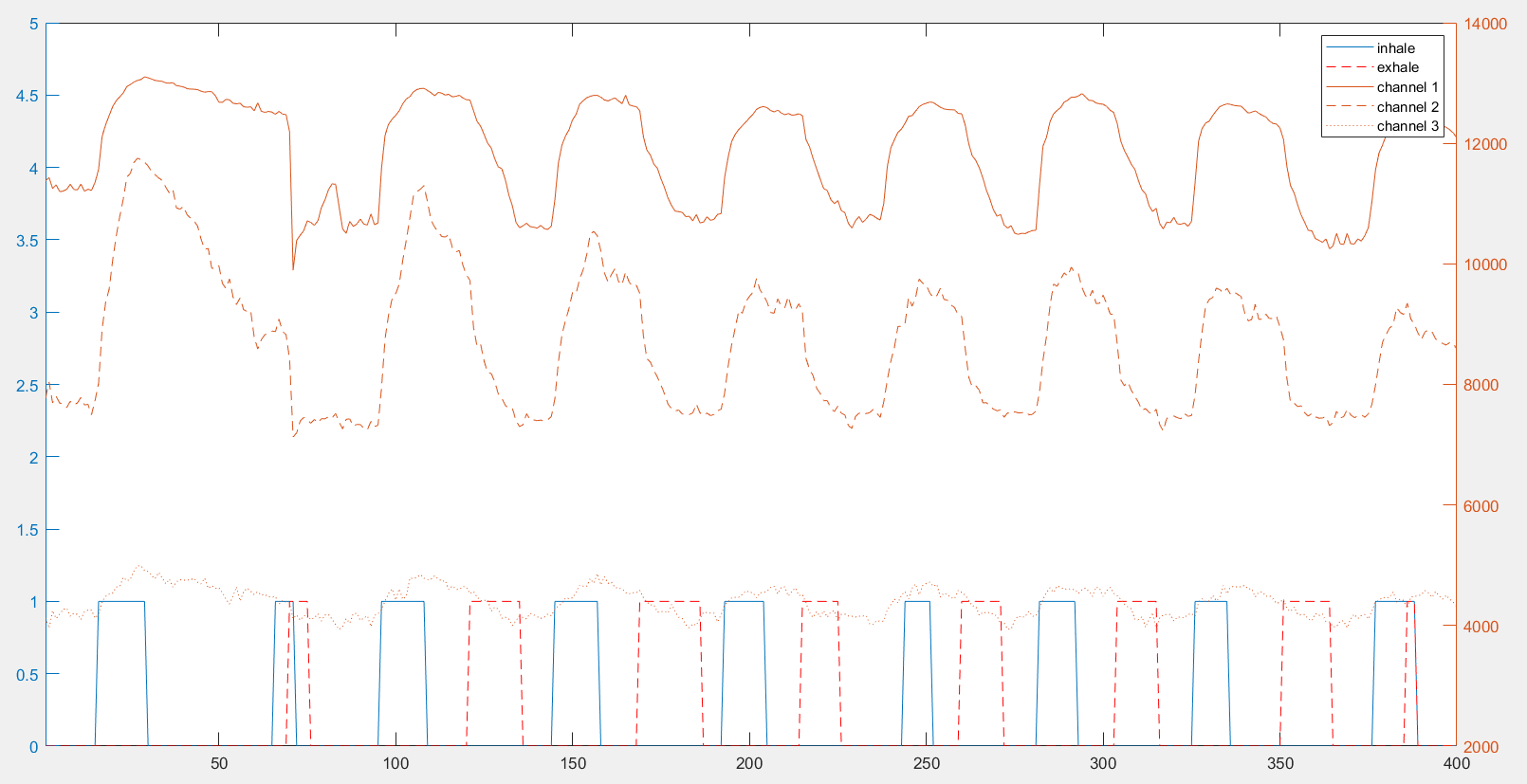
نتایج این الگوریتم برای تنفس عمیق شماره 4 و 5 که داده­های زمانی آن را به طور کامل داریم به شکل زیر است.

البته در داده­های داده شده مشکلاتی به چشم میخورد از جمله این­که مثلا در داده تایمینگ برای deep4 آخرین بازدم در داده­های ما نیست که این مشکل البته در تنفس معمولی بسیار بیشتر است. این داده یا ذکر نشده­اند و یا آخرین زمان ممکن موجود در لیست آن آزمایش به جای آن ذکر شده­است.

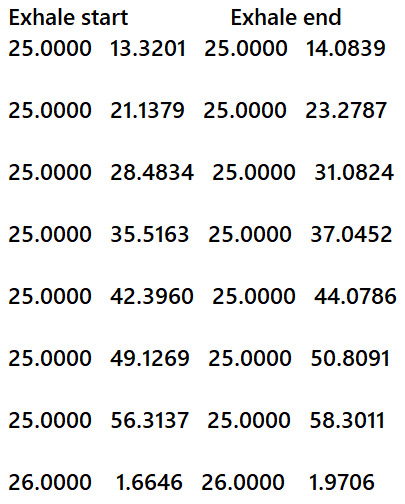
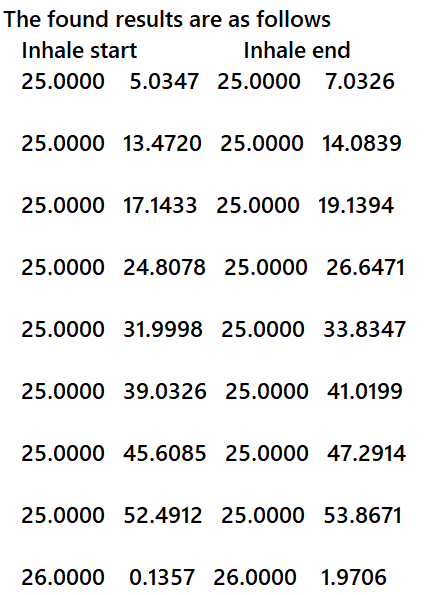
در ادامه نتایج اجرای این الگوریتم بر روی داده­ها، هم آنهایی که زمان دم و بازدم آنها را به صورت دقیق می­دانیم و هم غیر آن، آمده اند. فرمت این داده­ها به صورت اول دقیقه و سپس ثانیه متناظر با عمل موردنظر در آن آزمایش است.

برای داده­های آزمایش تنفس عمیق 4 داریم:

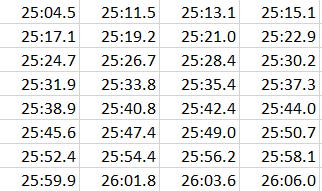
نمودار تطبیقی:



Deep4

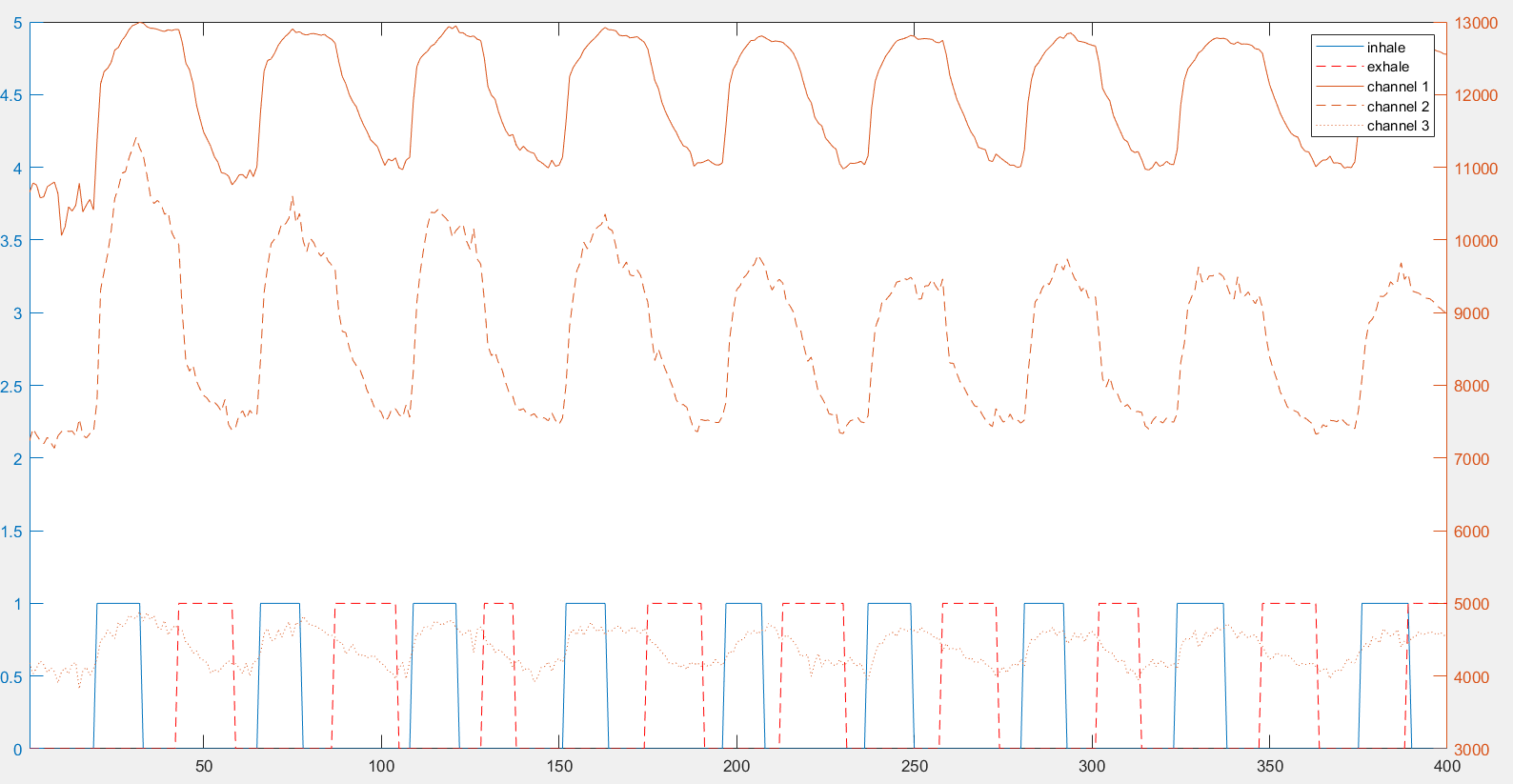


برای مقایسه، مقادیر واقعی این آزمایش به شکل زیر هستند:

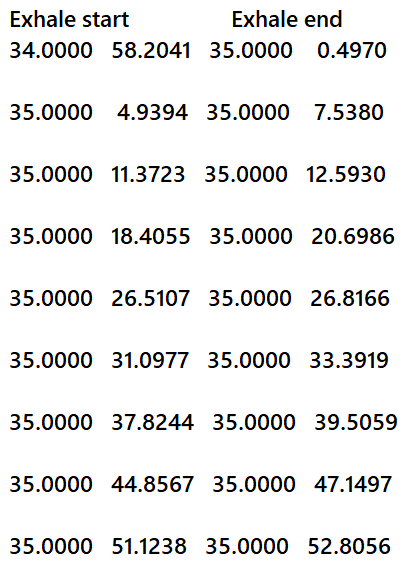


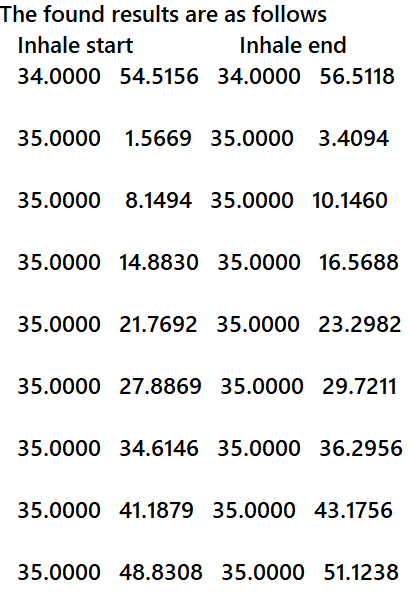
و برای داده های Deep5

نمودار تطبیقی:

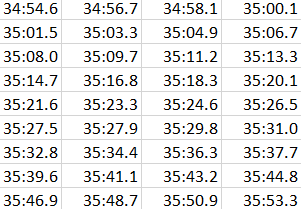


Deep5



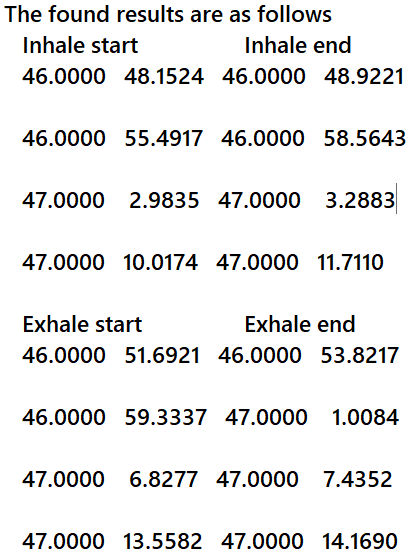
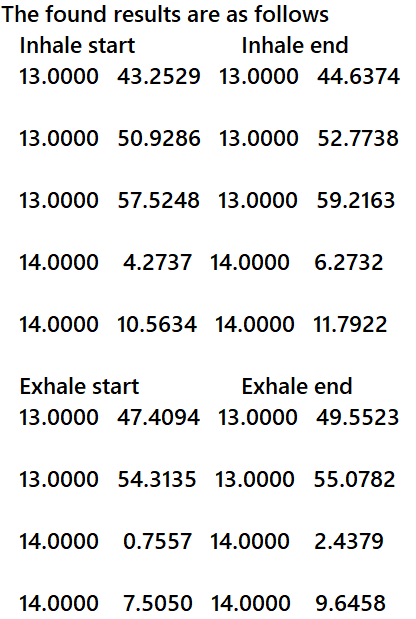


که داده­های واقعی این مرحله نیز به شکل زیر هستند

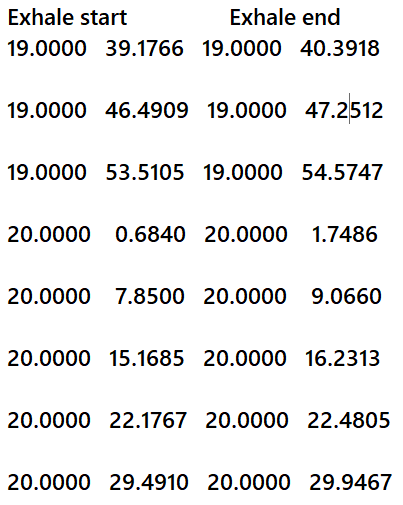


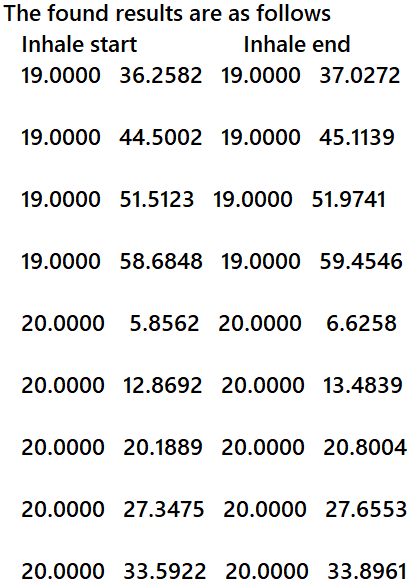
نتایج تنفس عمیق بسیار بهتر از تنفس معمولی بود. زمان این تنفس ها برای سابجکت های 1 تا 3 در زیر آمده­اند:

Deep1: Deep2

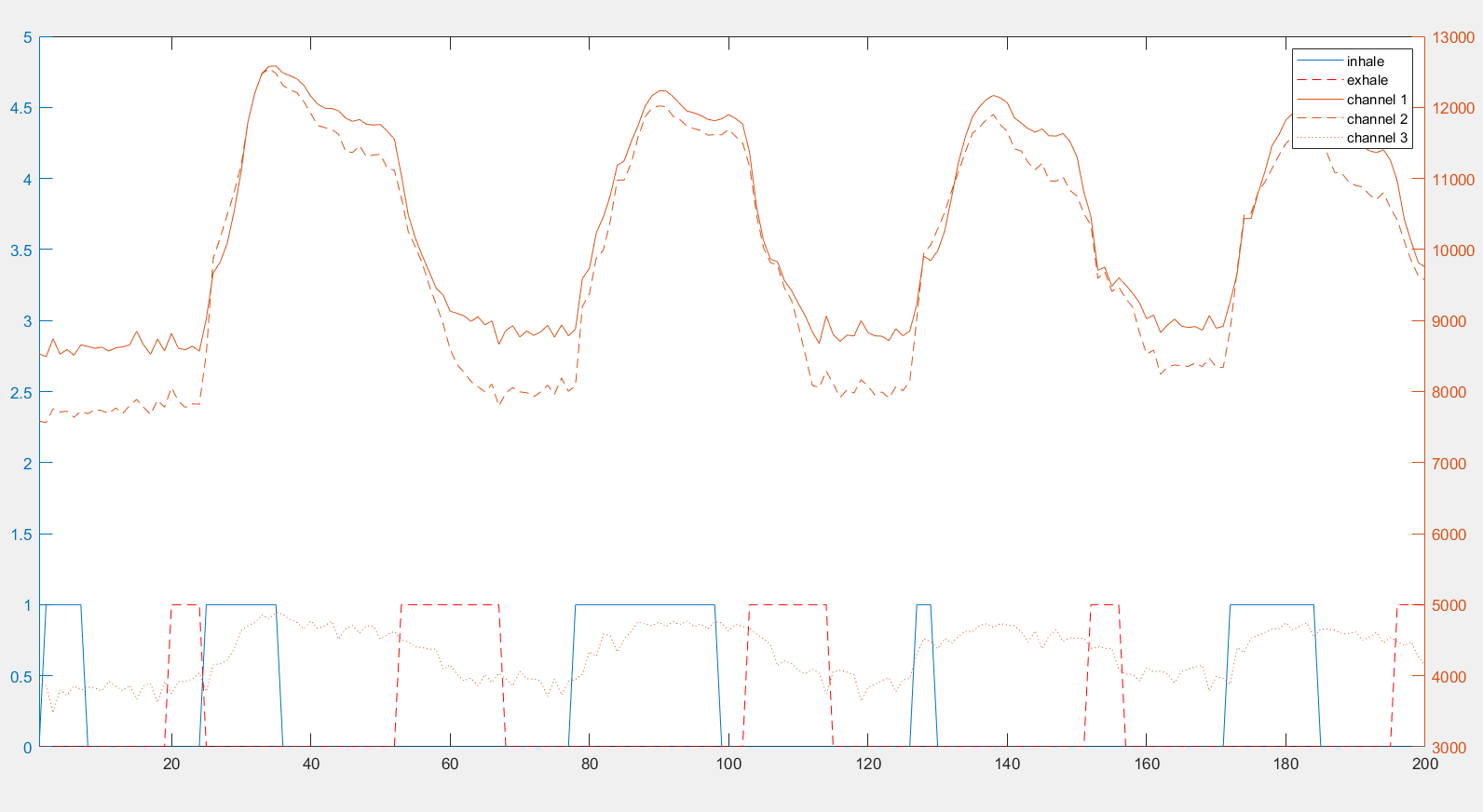


Deep3

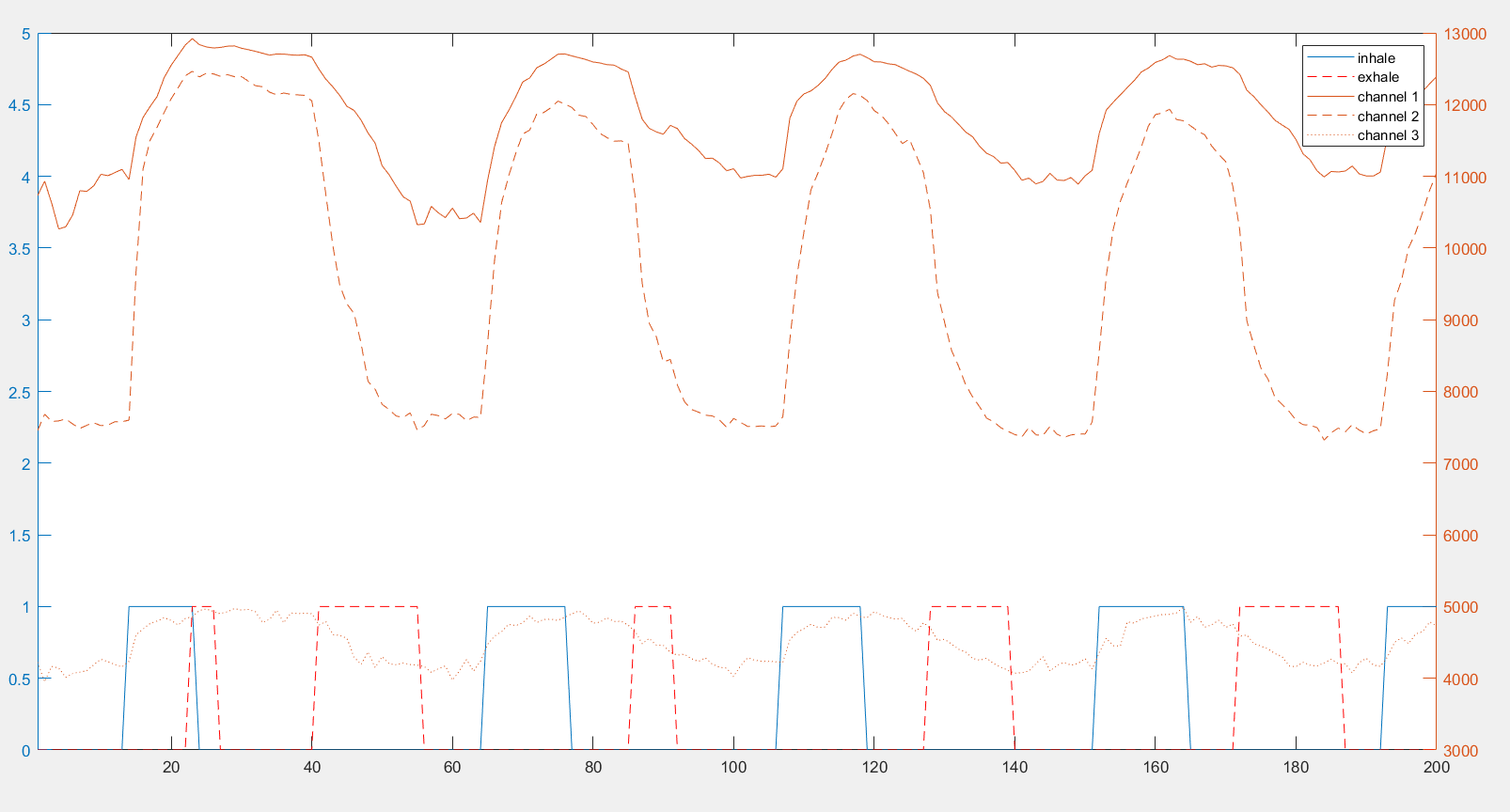




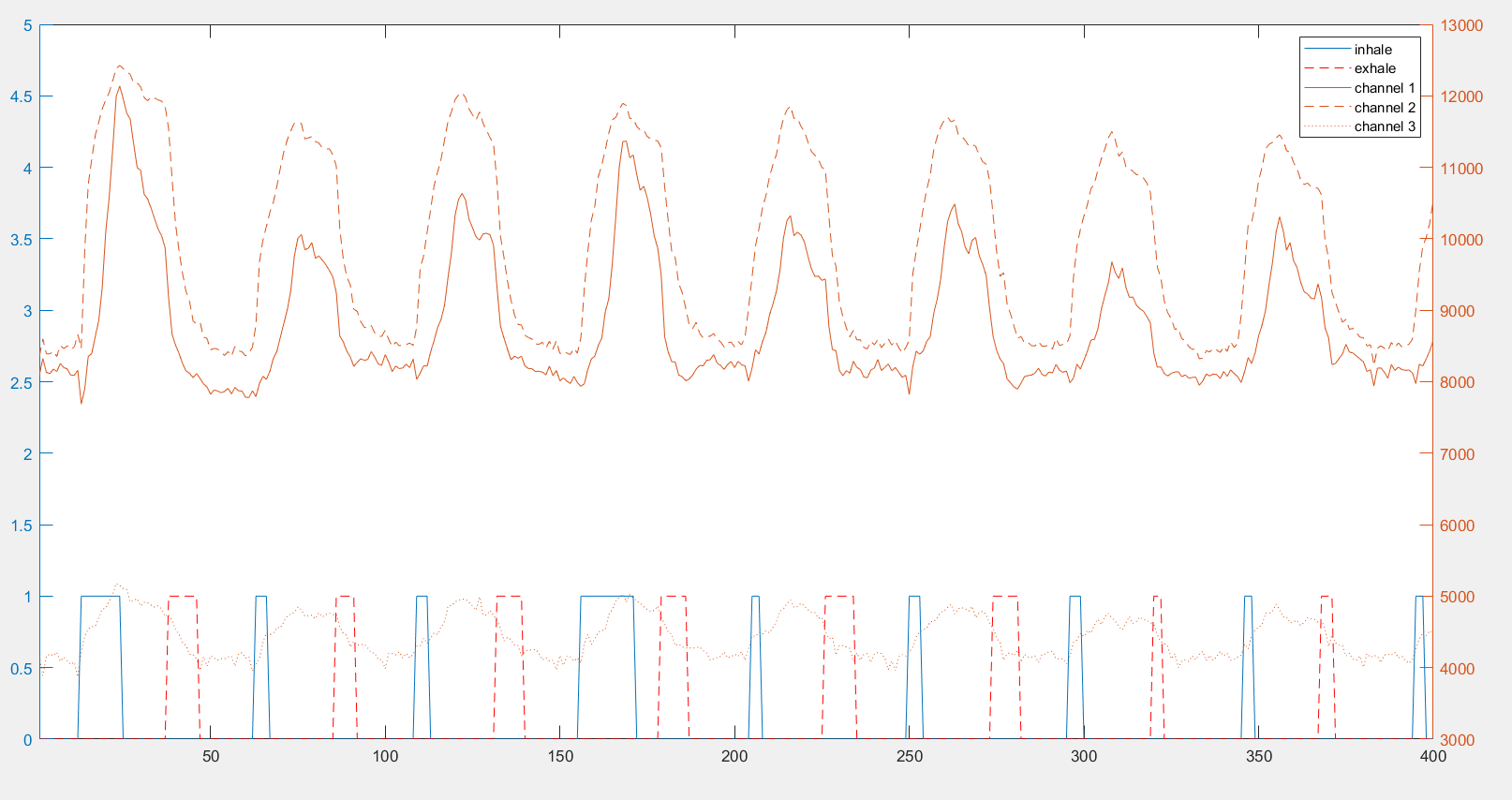
نمودار تطبیقی Deep1



نمودار تطبیقی Deep2:



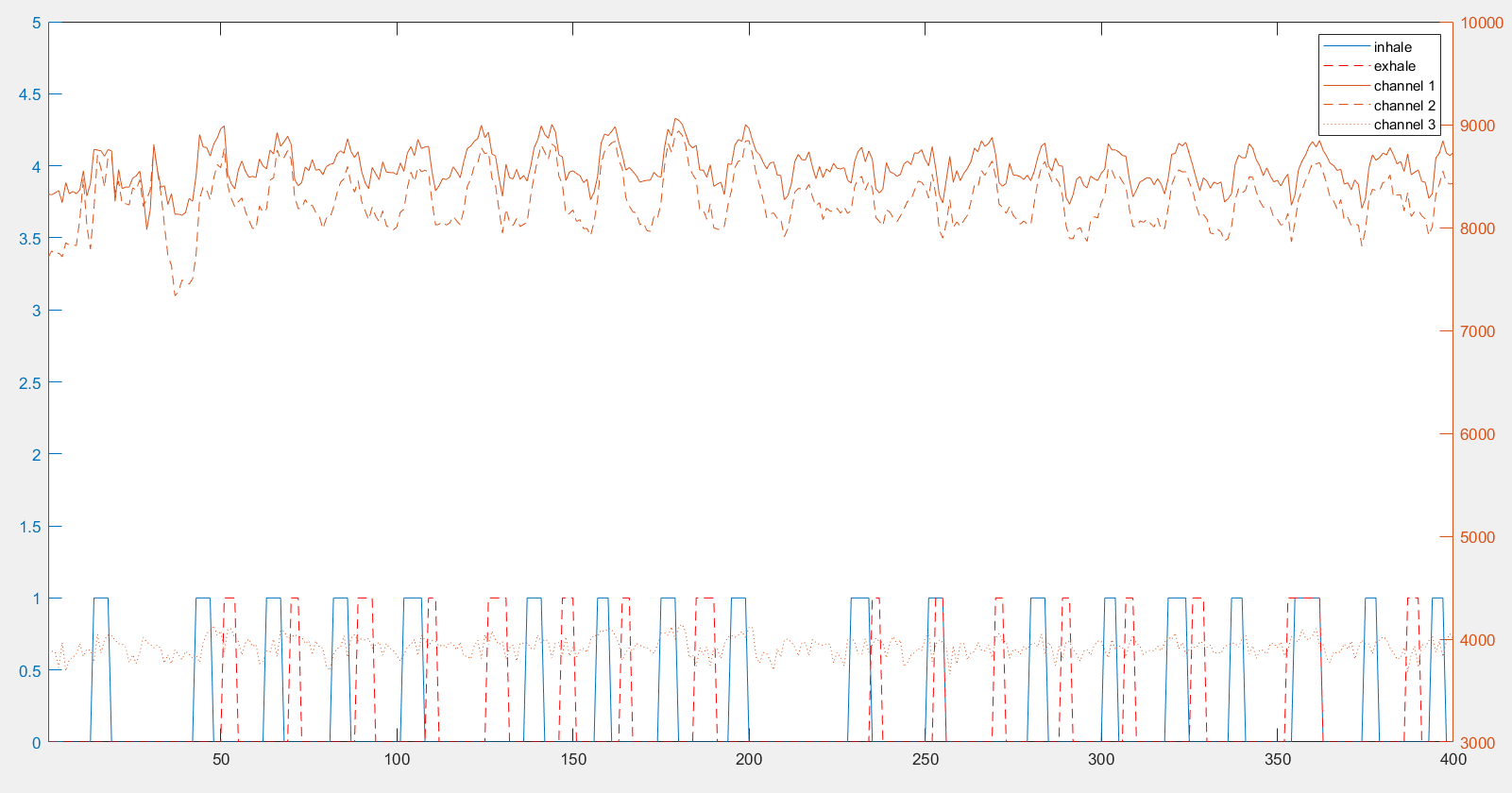
نمودار تطبیقی Deep3:



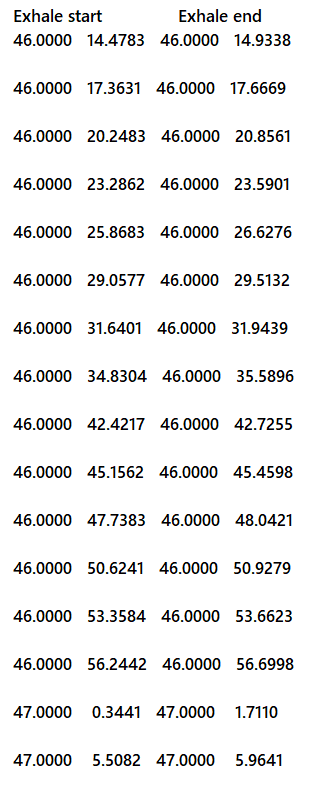
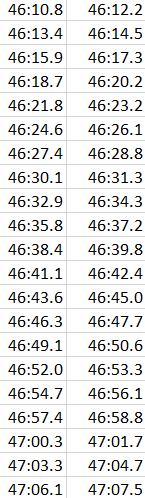
همانطور که گفته­شد برای تنفس معمولی، برخی زمان های داده­شده برای تنفس در بازه آزمایش ما نیستند. این داده ها حذف شده­اند که برای regular4 این برای با 3 تنفس آخر است وبرای regular3 هم تنفس آخر به مدت 10 ثانیه طول کشیده که بازدم آن هم جزو داده­های ما نیست. همچنین داده­های regular3 به نظر می­رسد بسیار نویز دارند و نتایج خوبی حاصل نمیدهند در حالی که همین الگوریتم روی داده­های regular4 بهتر جواب می­دهد.

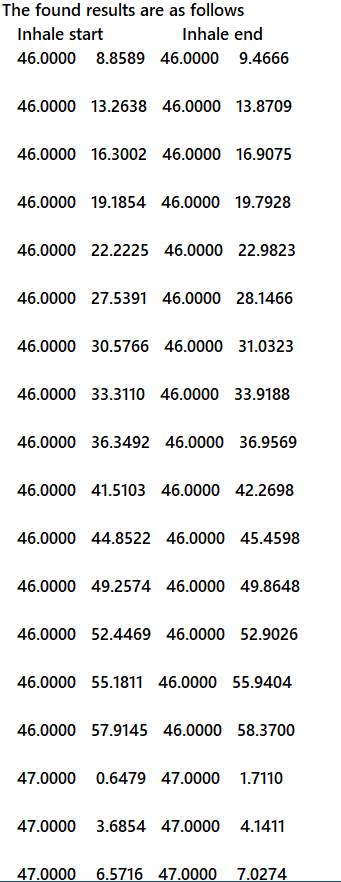
نتایج حاصل برای regular4:

نمودار تطبیقی:



Regular4:

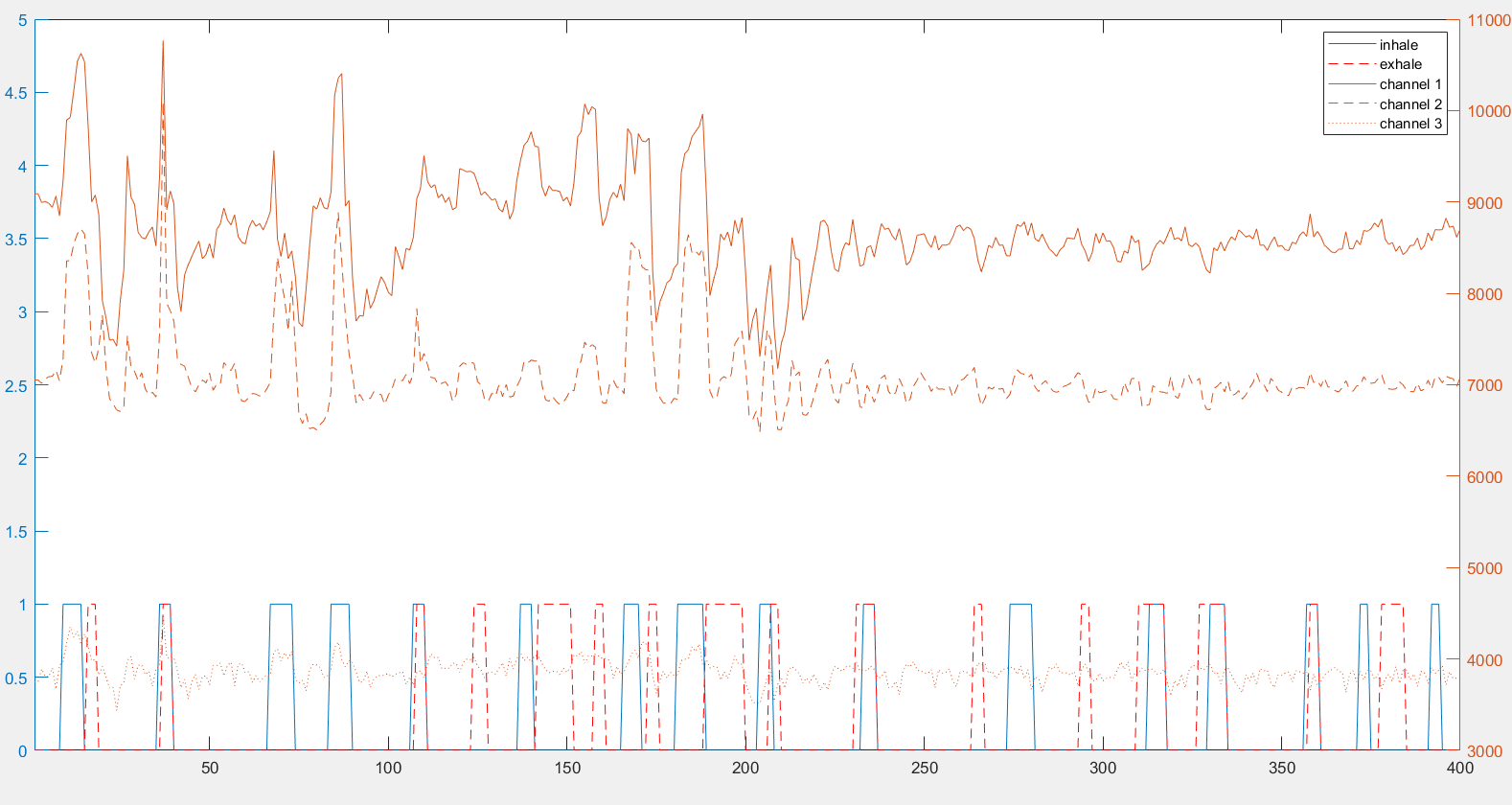




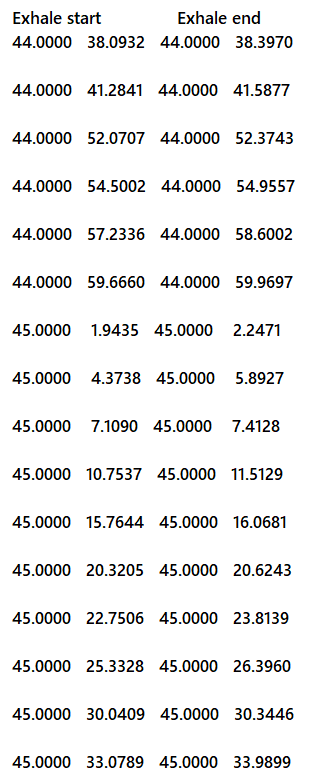
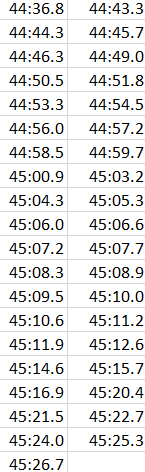
از 21 دم موجود البته تنها 18 مورد شناسایی شده­اند و از 20 بازدم موجود تنها 16 مورد. راست ترین داده­ها، داده ها از فایل اکسل اولیه هستند.

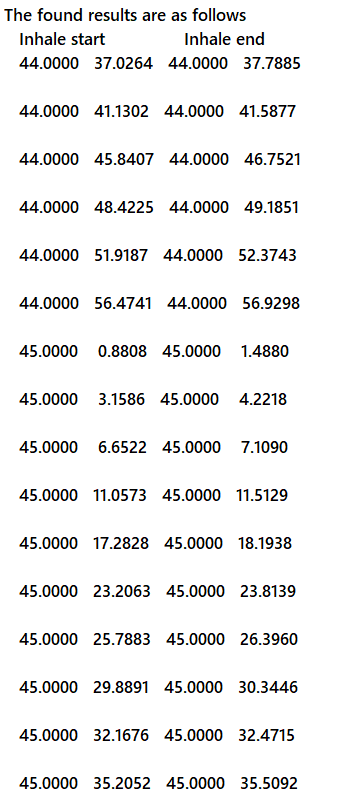
داده­های regular3 همانطور که در ابتدا ذکر شد به خوبی بقیه داده­ها نیستند و همانطور که در شکل زیر دیده می­شود بعد از اندیس 200، داده­ها به نوعی نویز هستند که نمی­توان اطلاعات خاصی از آنها استخراج کرد

نمودار تطبیقی regular3:



Regular3:

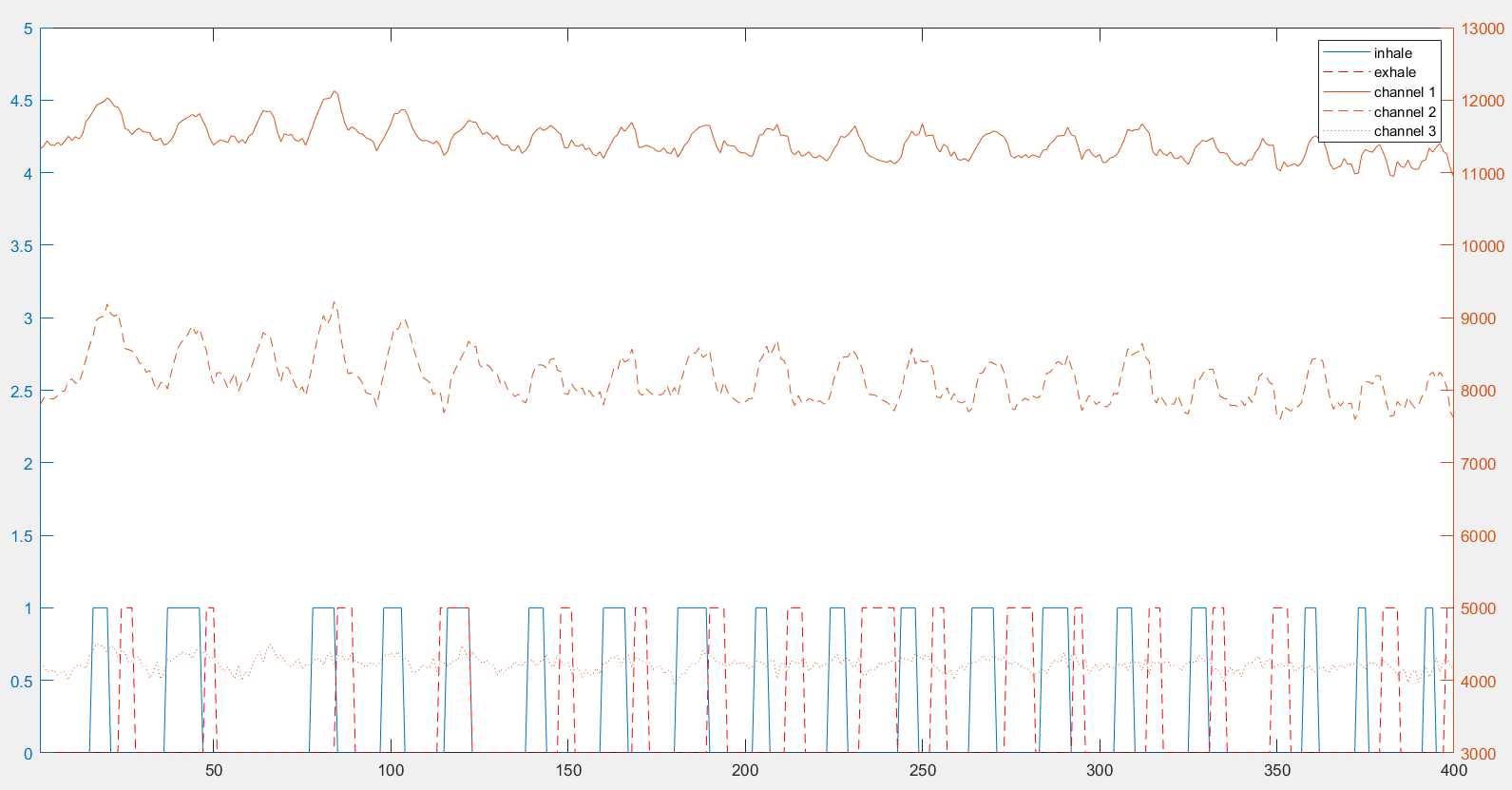


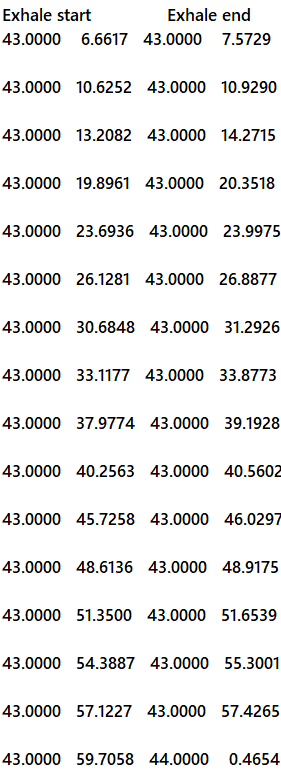


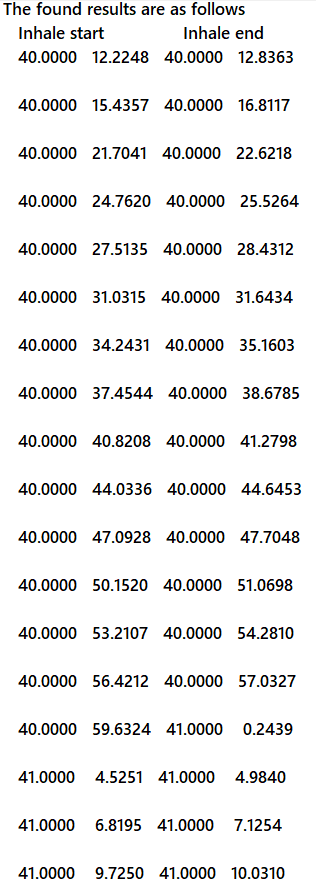
و برای داده­های تست 1 و 2 به ترتیب نتایج زیر به­دست آمده­اند:

Regular1:

نمودار تطبیقی:

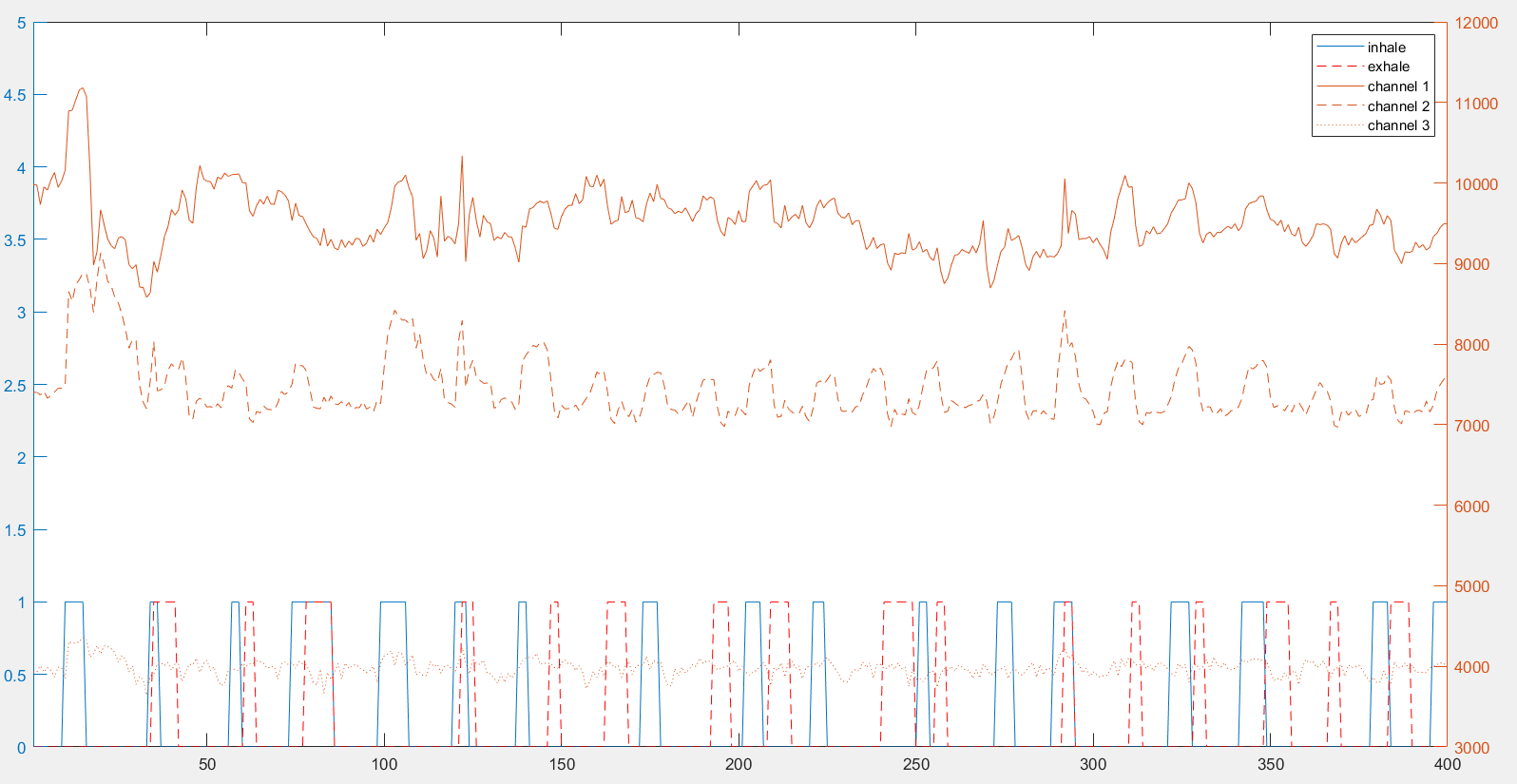


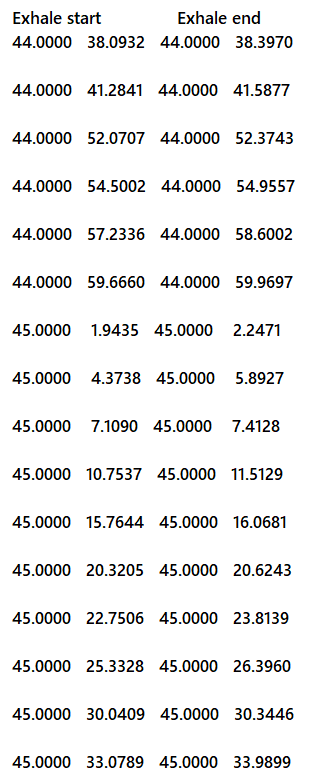


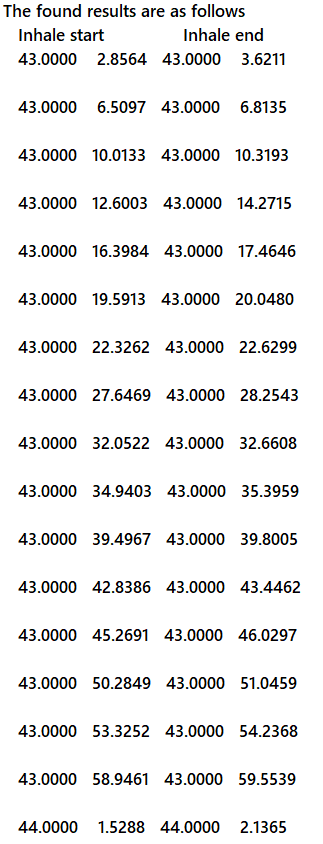


Regular2:

نمودار تطبیقی:







از نتایج و مقایسه­ها معلوم شد که استفاده از کانال 2 به تنهایی نیز برای به­دست دادن همین نتایج کافی بوده و تنها به داده­های این کانال نیاز داریم.

برای تست کردن هرگونه داده اضافی تنها کافی است در کد پروژه در قسمت On tests داده ها را با فرم مناسب داخل deep4 قرار دهید.