
بسم الله الرحمن الرحيم

دانشگاه آزاد اسلامی

واحد تهران جنوب

دانشکده فنی و مهندسی

پروژه پایانی

مهندسی پزشکی – بیوالکتریک

عنوان:

فاکتورسازی ماتریس غیرمنفی مبتنی بر یادگیری دیکشنری مشترک تبدیل صدا برای بهبود درک گفتار پس از جراحی دهان

استاد راهنما:

دکتر مهدی اسلامی

نام و نام خانوادگی دانشجو :

نرگس رضایی پیکر

شماره دانشجویی:

۴۰۱۱۴۱۴۰۱۱۱۰۲۷

۱۴۰۱/۷

فصل اول

خلاصه ای از مقاله

مقاله بر روی تکنیک‌های تبدیل صوتی مبتنی بر یادگیری ماشین برای بهتر درک کردن گفتار بیمارانی است که قسمت‌هایی از مفصل‌هایشان برداشته شده، تمرکز دارد.

آرتیکلاتور[1] یک وسیله‌ای است که نقش فکین و مفصل گیجگاهی فکی را در خارج از دهان ایفا می‌کند. این آرتیکلاتورها برمبنای حرکت و آناتومی فک به سه دسته تقسیم می‌شوند؛ یک دسته آرتیکلاتورهای ساده یا لولایی، یک دسته آرتیکلاتورهای از پیش تنظیم یافته و دسته سوم آرتیکلاتورهای قابل تنظیم نام دارند.

به دلیل برداشتن قسمت‌هایی از آرتیکلاتور گفتار بیمار ممکن است ناواضح باشد و درک صحبت‌های بیمار سخت شود برای غلبه بر این مشکل از سیستمی استفاده می‌شود که بتواند صدای ناواضح گفتار بیمار را به گفتار واضح تبدیل کند. این روش VC نام دارد. برای طراحی این روش باید به دو نکته مهم توجه کرد؛ ممکن است مقدار داده‌های آموزشی محدود باشد چون صحبت کردن برای مدت طولانی بعد از عمل برای بیماران مشکل است و نکته بعدی که باید در نظر بگیریم این است که برای بهتر شدن ارتباط این تبدیلات باید سریع انجام شود.

روشی که مقاله پیشنهاد کرده است یک الگوریتم جدید مبتنی بر یادگیری لغت نامه مشترک فاکتورسازی ماتریس غیر منفی[2] (JD-NMF) است. این روش مجموعه ای از الگوریتم‌ها برای تجزیه ماتریس به دو ماتریس است:

$$V \rightarrow W,H$$

فاکتورگیری ماتریس‌ها معمولاً یکتا نیست و روش‌های مختلفی برای انجام آن ارائه شده است. در مقایسه با تکنیک‌های VC معمولی، JD-NMF می‌تواند VC را به طور کارآمد و مؤثر تنها با مقدار کمی از داده‌های آموزشی انجام دهد.

نتایج تجربی نشان داد که این روش JD-NMF یک معیار ارزیابی قابل درک استاندارد شده نسبت به روش‌های گفتارهای تبدیل نشده است و این روش کارآمدتر هم است و نسبت به روش VC مؤثرتر است.

[1] Articulator
[2] Non-negativeMatrixFactorization

۱- چکیده هدف

این مقاله بر روی تکنیک‌های تبدیل صوتی مبتنی بر یادگیری ماشین (VC) برای بهبود درک گفتار بیمارانی است که در جراحی قسمت‌هایی از مفصل‌هایشان برداشته شده است، تمرکز دارد. به دلیل برداشتن قسمت‌هایی از آرتیکلاتور، گفتار بیمار ممکن است مخدوش شده و درک آن دشوار باشد. برای غلبه بر این مشکل می‌توان از روش‌های VC برای تبدیل گفتار تحریف شده استفاده کرد تا واضح و قابل فهم‌تر باشد. برای طراحی یک روش مؤثر VC، دو نکته کلیدی باید در نظر گرفته شود: ۱- ممکن است مقدار داده‌های آموزشی محدود باشد(زیرا صحبت کردن برای مدت طولانی معمولاً برای بیماران بعد از عمل دشوار است.) ۲- تبدیل سریع مطلوب است.(برای ارتباط بهتر)

۱-۱ روش‌ها

ما یک الگوریتم جدید مبتنی بر یادگیری لغت‌نامه مشترک فاکتورسازی ماتریس غیرمنفی (JD-NMF) پیشنهاد می‌کنیم. در مقایسه با تکنیک‌های VC معمولی، JD-NMF می‌تواند VC را به طور کارآمد و مؤثر تنها با مقدار کمی از داده‌های آموزشی انجام دهد.

۱-۲ یافته‌ها

نتایج تجربی نشان می‌دهد که روش JD-NMF پیشنهادی نه تنها به نمرات قابل توجهی به درک هدف کوتاه مدت STOL[1] نسبت به روش‌های به‌دست آمده با استفاده از گفتار تبدیل نشده اصلی دست می‌یابد، بلکه به طور قابل توجهی کارآمدتر است و مؤثرتر از روش معمولی مبتنی بر VC است.

۱-۳ نتیجه گیری

روش JD-NMF پیشنهادی ممکن است از روش VC مبتنی بر نمونه‌های پیشرفته از نظر امتیازات STOL تحت سناریوی مورد نظر بهتر عمل کند.

اهمیت: ما مزایای معیار آموزش مشترک پیشنهادی را برای VC مبتنی بر NMF تأیید کردیم. علاوه بر این ما تأیید کردیم که JD-NMF پیشنهادی می‌تواند به طور مؤثر نمرات درک گفتار بیماران جراحی دهان را بهبود بخشد.

---

[1] یک معیار ارزیابی قابل درک استاندارد شده هدف

۲-اصطلاحات فهرست- یادگیری فرهنگ لغت مشترک، فاکتورسازی ماتریس غیرمنفی، نمایش پراکنده، تبدیل صدا

 Szu-Wei Fu با گروه علوم کامپیوتر و مهندسی اطلاعات، دانشگاه ملی تایوان، تایپه، تایوان و مرکز تحقیقاتی نوآوری فناوری اطلاعات[1] (CITI) در Academia Sini ca، تایپه تایوان کار می‌کند. Pei-Chun Li با گروه شنوایی شناسی و آسیب شناسی زبان گفتار، کالج پزشکی مکی، تایپه، تایوان کار می‌کند. Ying-Hui Li با گروه مهندسی برق دانشگاه یوان‌زه کار می‌کند. Chang-Chien Yang و Li-Chun Hsieh در بیمارستان یادبود مکی، تایپه، تایوان هستند. Yu Tsao با مرکز تحقیقات نوآوری فناوری اطلاعات (CITI) در Academia Sinica ، تایپه، تایوان همکاری می‌کند.

۳-مقدمه

درک گفتار یک فرد پس از جراحی دهان اغلب برای شنوندگان آموزش ندیده دشوار است. بنابراین، چنین بیمارانی ممکن است تمایل به یک سیستم تبدیل صدا داشته باشند که بتواند صدای آن‌ها را به گفتار واضح تبدیل کند. در این مقاله برای بهبود VC ما استفاده از رویکرد درک گفتار بیمارانی که قسمت‌هایی از مفصل آن‌ها در حین جراحی برداشته شده‌اند، بررسی کردیم. وظایف معمولی VC طوری طراحی شده است که گفتار گوینده مبدأ را تغییر می‌دهند تا صدایی شبیه به سخنران دیگر(هدف) شود. اخیراً روش‌های VC برای کاربرد مختلف پزشکی به کار گرفته شده است. Aihara و همکاران یک سیستم VC برای اختلالات بیانی پیشنهاد کردند که تلاش می‌کند فردیت گوینده را بر اساس فرهنگ لغت ترکیبی حاوی حروف صدادار گوینده مبدأ و صامت‌های گوینده هدف حفظ کند. Toda و همکاران سعی کردند VC را برای تبدیل زمزمه‌های غیرقابل شنیدن به گفتار عادی اعمال کند. Lio و همکاران روشی را برای استفاده از فناوری کاهش فرکانس مبتنی برVC برای کاربران سمعک زبان پیشنهاد کردند. روش‌های VC متعددی در گذشته پیشنهاد شده است. یک دسته قابل توجه از روش‌ها از یک مدل پارامتریک برای ترسیم ویژگی‌های صوتی بلندگوی منبع به بلندگوی هدف استفاده می‌کند. مدل مخلوط گاوسی با چگالی مشترک[2](JD-GMM) به عنوان یک مدل نقشه برداری مؤثر برای VC شناخته شده است. JD-GMM یک تابع تبدیل خطی را بر اساس مدل مخلوط گاوسی (GMM) پیاده سازی می‌کند. پارامترهای تبدیل با استفاده از معیارهای حداکثر احتمال حداقل میانگین مربعات خطا یا حداکثر اطلاعات متقابل برآورده می‌شوند. الحاقات متعددی از JD-GMM برای حل مشکل هموارسازی بیش از حد ذاتی ناشی از میانگین گیری آماری

[1] Center Information Technology Innovation
[2] Gaussian Mixture Model with Joint Density

پیشنهاد شده است. یک شبکه عصبی مصنوعی[1] (ANN) مدل قابل توجه دیگری است که برای VC کارایی تأیید شده است. به دلیل ساختار پیچیده خود، یک مدل ANN قادر است رابطه غیرخطی بین گفته‌های سخنرانان مختلف را مشخص کند. از زمان ظهور یادگیری عمیق، VC های مبتنی بر شبکه های عصبی عمیق قابل توجهی را به خود جلب کرده‌اند. اگرچه روش‌های VC مبتنی بر مدل برای کارهای مختلف مؤثر هستند، اما معمولاً به مقدار مشخصی از داده‌های آموزشی نیاز دارند. هنگامی‌که داده‌های آموزشی کافی وجود ندارد، ممکن است مدل‌ها دچار مشکل بیش از حد برازش شوند، به طوری که کیفیت صدای گفتار تبدیل شده ضعیف باشد. برای غلبه بر مشکل بیش از حد برازش احتمالی، چندین روش VC مبتنی بر نمونه غیرپارامتری به عنوان جایگزینی برای چارچوب‌های مبتنی بر مدل پیشنهاد شده است. این دسته از روش‌ها فرض می‌کنند که یک طیف نگار هدف را می‌توان از مجموعه‌ای از طیف‌های هدف پایه(یک فرهنگ لغت)، یعنی نمونه‌ها، از طریق ترکیب‌های خطی وزن‌دار تولید کرد. بر اساس ماهیت غیرمنفی طیف نگار، از روش غیرمنفی کردن عامل ماتریس(NMF) برای تخمین وزن غیرمنفی استفاده می‌شود. در زمان اجرا، فعال سازی‌های هر طیف نگار منبع از طریق فرهنگ لغت منبع تخمین زده می‌شود و سپس به فرهنگ لغت هدف اعمال می‌شود تا طیف نگار هدف مرتبط را تولید کند. بنابراین ances تبدیل شده مستقیماً از نمونه‌های هدف واقعی به جای پارامترهای مدل تولید می‌شود. Wu و همکاران یک چارچوب NMF مشترک برای تخمین مؤثر فعال سازی‌ها با در نظر گرفتن همزمان دو ویژگی صوتی متمایز(یکی با وضوح پایین و یکی با وضوح بالا) پیشنهاد کردند. اگرچه تنها داده‌های آموزشی محدودی برای مدل‌های NMF مبتنی بر نمونه مورد نیاز است، بیشتر داده‌ها به طور خام به عنوان نمونه استفاده می‌شوند، به این معنی است که یک فرهنگ لغت بزرگ ساخته خواهد شد. محدودیت اصلی استفاده از یک فرهنگ لغت بزرگ زمان تبدیل طولانی است که نیاز به تبدیل سریع مارا نقض می‌کند. در این مطالعه، ما توجه خود را بر روی تکنیک‌های VC مبتنی بر NMF برای بیماران جراحی دهان متمرکز کردیم، که برای آن دو نکته کلیدی باید مورد توجه قرار گیرد: ۱- مقدار داده‌های آموزشی ممکن است محدود باشد زیرا صحبت کردن برای مدت طولانی برای بیماران پس از جراحی معمولاً دشوار است. ۲- تبدیل سریع مطلوب است برای تسهیل ارتباطات بهتر با کاربران

برای پرداختن به این دو نکته، ما یک الگوریتم VC مبتنی بر یادگیری فرهنگ لغت مشترک جدید را پیشنهاد می‌کنیم. ریتم الگوی JD-NMF به طور همزمان دیکشنری‌های منبع و مقصد(فرهنگ لغت مشترک) را یاد می‌گیرد. با تعیین تعداد کمی از پایه‌ها با استفاده از تکنیک JD-NMF و NMF می‌تواند مجموعه‌ای از پایه‌ها را بیاموزد که نماینده کل مجموعه نمونه‌ها هستند(تخمین زده شده از داده‌های آموزشی). بر این اساس، اندازه فرهنگ لغت در JD-NM را می‌توان به طور قابل توجهی نسبت به NMF مبتنی بر نمونه کاهش داد، در نتیجه کارایی تبدیل آنلاین را بهبود می‌بخشد.

[1] Artificial Neural Network

بقیه مقاله به صورت زیر سازماندهی شده است: بخش دوم VC مبتنی بر NMF معمولی را بررسی می‌کند. بخش سوم روش پیشنهادی را شرح می‌دهد. نتایج تجربی در بخش چهار ارزیابی شده است. در نهایت، بخش پنج نتیجه گیری‌های ما را ارائه می‌کند.

۱-۴ کارمرتبط

الف) بازنمایی گفتار مبتنی بر NMF :

مفهوم اساسی VC مبتنی بر NMF این است که یک طیف magnitude را به عنوان یک ترکیب خطی از مجموعه‌ای از پایه‌ها نشان دهد. به این مجموعه از پایگاه‌ها دیکشنری می‌گویند. در مدل NMF مبتنی بر نمونه معمولی، هر پایه در ماتریس یک چارچوب گفتاری(نمونه) در داده‌های آموزشی است. به طور خاص، پایه‌ها مستقیماً از داده‌های آموزشی کپی می‌شوند و هیچ فرآیند یادگیری برای ساخت فرهنگ لغت درگیر نیست. فرض کنید که نمونه‌ها جمع‌آوری شده‌اند، ما یک فرهنگ لغت داریم $A = [a_1,a_2,\ldots,a_I] \in R^{F\times I}$ که در آن a_i است i^{th} نمونه و F بعد ویژگی است. سپس نمونه گفتار در $x_l \in R^{F\times L}$ را می‌توان توسط زیر نشان داد:

$$\chi \approx Ah + \sum_{i=1}^{I}(a_i h_{i,l})$$

جایی‌که $h_1 = [\, h_1, h_2, \ldots, h_{I,l}] \in R^{i\times l}$ بردار فعال سازی است و $h_{i,l}$ وزن غیرمنفی است و i^{th} نمونه است. از آن‌جایی که هر نمونه گفتاری به طور مستقل مدل می‌شود، مشخصات هر گفتار را می‌توان به صورت روبه رو نشان داد:

$$\chi \approx AH$$

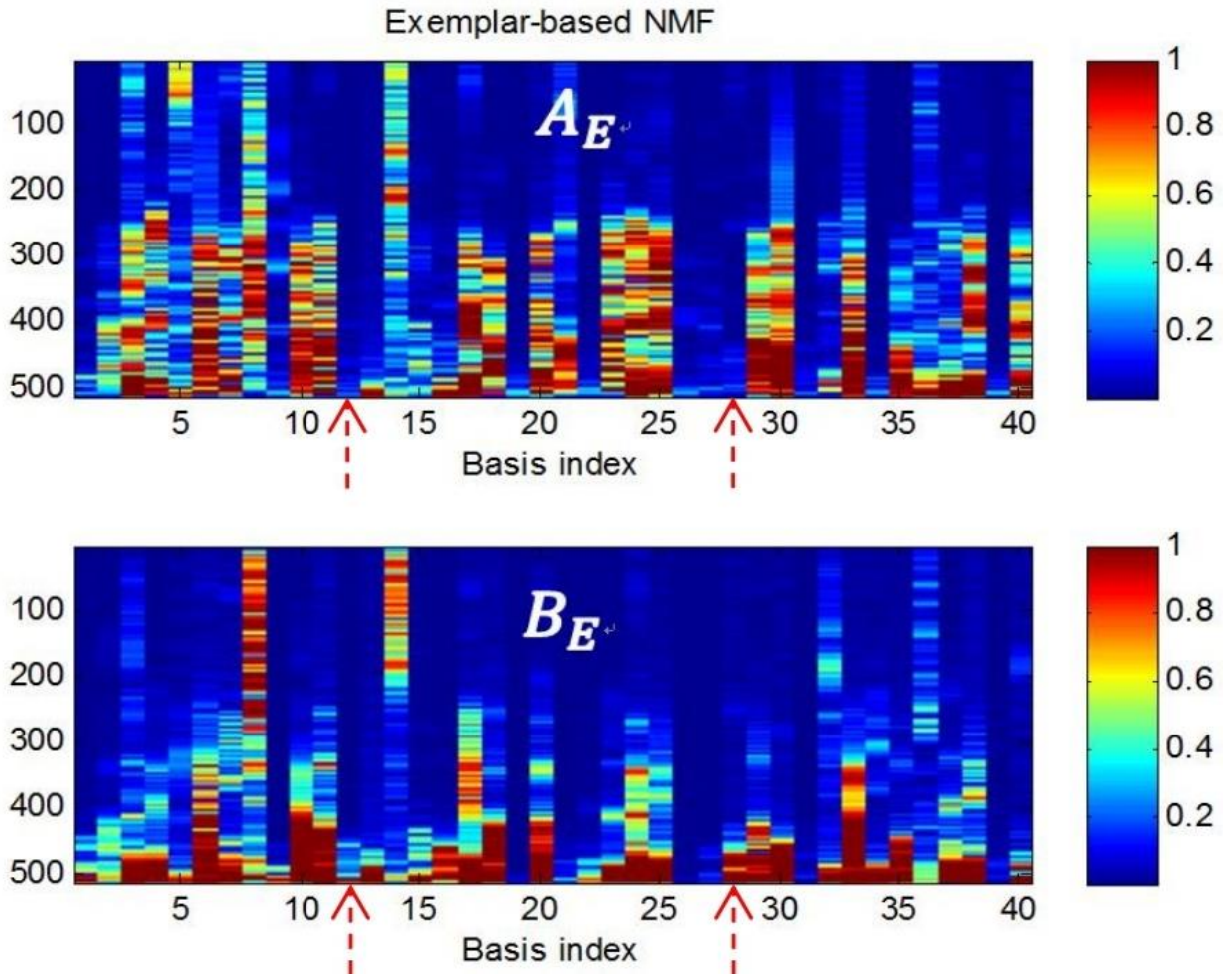
جایی‌که $X \in R^{F\times M}$ طیف نگار است، M در بیان تعداد فریم‌ها است، و $H \in R^{I\times M}$ ماتریس فعال سازی مربوطه است که بردارستون آن بردار فعال سازی h است. برای به حداقل رساندن فاصله بین X و AH، قوانینی را برای بهینه سازی متناوب A و H با نزول گرادیان خاص ارائه کردند.

ب) برای تبدیل صدای مبتنی بر NMF :

۱) مرحله آفلاین

برای VC دیکشنری‌های جفت منبع-هدف A و B با نمونه‌های تراز صوتی مورد نیاز است. در NMF های مبتنی بر نمونه، هم دیکشنری منبع و هم دیکشنری هدف مستقیماً از خود داده‌ها به دست می‌آید. برای ساخت دیکشنری‌های جفت شده، یک مجموعه داده موازی(بین گوینده منبع و هدف) جمع آوری می‌شود. با این حال، به دلیل نرخ گفتار متفاوت، این دو فرهنگ لغت ممکن است با یکدیگر همسو نباشند. بنابراین، تکنیک‌های برنامه

ریزی پویا مانند تاب خوردگی زمانی پویا باید برای به دست آوردن هم ترازی منبع-هدف بر اساس چارچوب اعمال شوند. شکل۱ نمونه‌ای از فرهنگ لغت منبع- هدف را نشان می‌دهد. برای ارائه تصویری، تنها ۴۰ فریم(پایه) به طور تصادفی از داده‌های آموزشی انتخاب شد. محور x شاخص پایه را نشان می‌دهد و محور y نشان دهنده سطل‌های فرکانس است. علاوه بر این، شدت با رنگ‌ها نشان داده می‌شود. در این مثال ما از ۵۱۲ نقطه تبدیل فوریه گسسته برای مشخص کردن صداهای گفتاری ۱۶ کیلوهرتز استفاده کردیم.



شکل۱-۱: فرهنگ لغت منبع و هدف مورد استفاده در NMF مبتنی بر نمونه

۲) مرحله آنلاین

برای تولید طیف نگار گفتاری تبدیل شده، فرض می‌کنیم که دیکشنری‌های منبع و هدف تراز شده می‌توانند یک ماتریس فعال سازی H را به اشتراک بگذارند. بنابراین، طیف گرم تبدیل شده را می‌توان به صورت زیر نمایش داد:

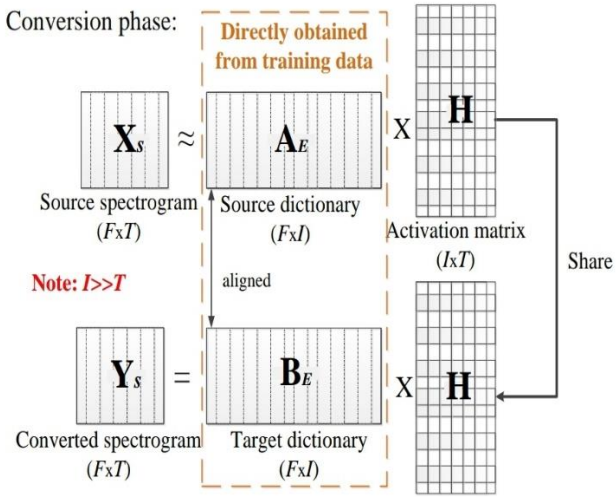
$$Y_s = B_E H$$

جایی که $Y_s \in R^{F\times T}$ طیف نگار تبدیل شده است، $B_E \in R^{F\times I}$ فرهنگ لغت هدف ثابت داده‌های آموزشی نمونه‌ها، و H توسط طیف نگار منبع تعیین می‌شود.$X_s \in R^{F\times I}$ و فرهنگ لغت منبع در فرمول بعدی نشان داده شده است. به دلیل ماهیت غیرمنفی طیف، از تکنیک NMF برای تخمین ماتریس فعال سازی H با به حداقل رساندن تابع هدف استفاده می‌شود.

$$H = \text{argmin } d\,(X_s\,,\, A_E\, H) + \lambda||H||$$

ضریب جریمه پراکندگی کجاست. از آنجایی که تعداد نمونه‌ها معمولاً در NMF های مبتنی بر نمونه زیاد است، محدودیت پراکندگی به گونه‌ای اتخاذ می‌شود که تنها چند نمونه در هر زمان فعال می‌شوند. در یک قاعده به روزرسانی ضربی برای دو معیار(فاصله اقلیدسی و واگرایی) پیشنهاد شد. سایر اقدامات واگرایی و قوانین به روزرسانی را می‌توان یافت با این حال در کاربرد VC واگرایی مناسب تر است. بنابراین می‌توان با اعمال مکرر قانون به روز رسانی ضربی را به حداقل رساند.

شکل۲ چارچوب کلی برای VC مبتنی بر نمونه را نشان می‌دهد.



شکل۱-۲: مرحله آنلاین NMF مبتنی بر نمونه برای VC

۵-۱ پیشنهاد یادگیری دیکشنری مشترک NMF برای صدا

برای تبدیل صدا در سایر کاربردهای NMF به عنوان مثال، تقویت گفتار فرهنگ لغت از داده‌های آموزشی آموخته مبتنی NMF می‌شود. با این حال، در نمونه معمولی، فرهنگ لغت مستقیماً از داده‌های آموزشی کپی می‌شود. به عبارت دیگر، هیچ مرحله آموزشی در مبتنی بر نمونه وجود NMF چارچوب ندارد که روش آموزش آفلاین را ذخیره کند. اما یک اشکال را ایجاد می‌کند: وقتی تعداد پایگاه‌ها زیاد باشد، هزینه محاسباتی در نسخه مخدوش می‌تواند بالا باشد. این به این معنی است که مبتنی بر نمونه ممکن است برای NMF سناریوی‌کاربردی ما مناسب نباشد(تبدیل سریع برای ارتباط بهتر مطلوب است) اگرچه سطح در عملکرد به دست آمده بهتر از سایر روش (JD-GMM) بود. برای تولید JD-NMF برای حل مشکل، ما چارچوب را پیشنهاد می‌کنیم که زمان بیشتری را در مرحله آفلاین (آموزش) صرف استخراج مجموعه‌ای از بازنمایی‌های پایه معنادارتر(یعنی فشرده) می‌کند. در مرحله آنلاین(زمان اجرا) بر اساس مبانی تخمین زده ماتریس فعال سازی را JD-NMF انجام می‌دهد.

الف) مرحله آفلاین

علاوه بر اعمال DTW[1] برای تراز کردن داده‌های آموزشی به روشی مشابه در NMF مبتنی بر نمونه، JD-NMF پیشنهادی شامل یک مرحله آموزشی در مرحله آفلاین است. در مطالعات قبلی، تأیید شده است که هنگام ایجاد تبدیل صوتی مبتنی بر NMF، تهیه یک جفت فرهنگ لغت همراه مهم است زیرا ماتریس فعال سازی توسط ماتریس‌های مبنا و هدف مشترک است. این نشان می‌دهد که این دو فرهنگ لغت به‌جای اینکه به طور مستقل آموزش داده شوند، باید به طور همزمان آموزش داده شوند. ما چارچوب JD-NMF را پیشنهاد می‌کنیم و تابع هدف را برای یادگیری همزمان دو دیکشنری به صورت زیر تغییر می‌دهیم:

$$A_J , B_J = \arg\min d(X, A_J H) + d(Y, B_J H) + \lambda \| H \|_1$$

جایی‌که در آن $X \in R^{F\times I}$ و $Y \in R^{F\times I}$ منبع و هدف جفت شده هستند، داده‌های آموزشی به ترتیب

$A \in R^{F\times I}$ و $B_J \in R^{F\times I}$ هستند. دیکشنری‌های آموخته شده و r تعداد پایه‌هایی است که توسط کاربران قابل تنظیم است. توجه داشته باشید که برای تقریب طیف نگاری منبع و هدف استفاده می‌شود مشروط بر این که همان ماتریس فعال سازی H استفاده شود. به طور خاص، برای بازسازی داده‌های آموزشی جفت شده(X وY) باH مشترک، لغت نامه‌های آموخته شده (A و B) مجبور می‌شوند برای به‌حداقل رساندن فاصله (واگرایی‌KL) با یکدیگر جفت شوند. بنابراین، اگر داده‌های آموزشی منبع و هدف هم‌تراز باشند، مبنای منبع منبع i^{th} آموخته شده همان واحد گفتاری پایه را نشان می‌دهد که مبنای هدف i^{th} است.

[1] Dynamic Time Warping(DTW)

برای حل با استفاده از واگرایی KL به عنوان معیار، دو عبارت اول را می‌توان به صورت زیر فرموله کرد:

$$d(X,A_JH) + d(Y,B_JH)$$

$$= \Sigma\left(X_{fi} \log\frac{X}{AH} - X + (AH)\right) + \Sigma\left(Y \log\frac{Y}{BH} - Y + (BH) =\right.$$

$$\Sigma\left(X \log\frac{X}{AH} - X + (AH) + Y \log\frac{Y}{BH} - Y + (BH)\right)$$

از آنجایی که در عملیات همه عنصر هستند، می‌توانیم X و Y را با A و B، آبشاری کنیم تا تابع هدف را به صورت زیر خلاصه کنیم:

$$\Sigma\left(S \log\frac{S}{WH} - S + (WH) + \lambda \left||H|\right|\right.$$

$$= d(S, WH) + \lambda \left||H|\right|$$

$$S = \begin{bmatrix} x \\ y \end{bmatrix} \in R^{2F\times I} \quad , \quad W = \begin{bmatrix} A \\ B \end{bmatrix} \in R^{2F\times I}$$

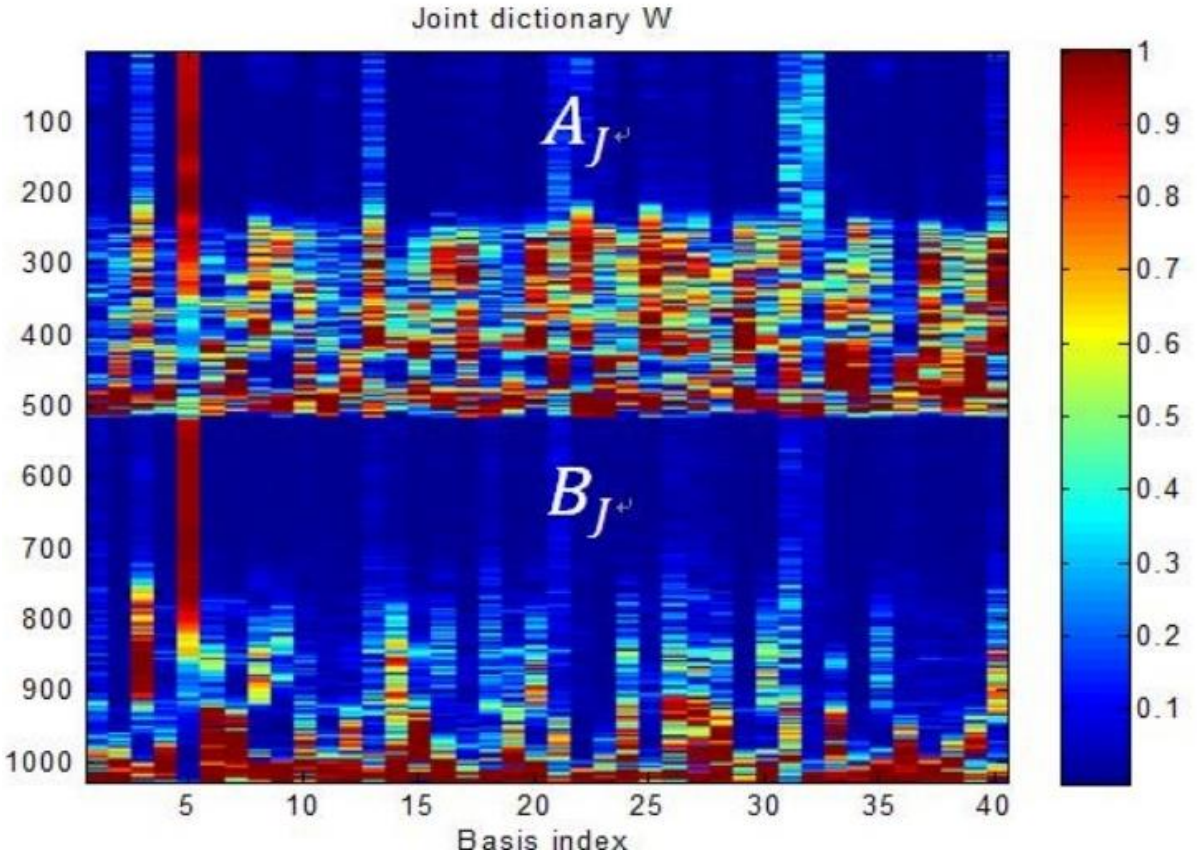
بنابراین، تابع هدف معادل را ساده کردیم. ما به سادگی می‌توانیم قوانین به‌روزرسانی متناوب متداول پیشنهادی را برای تعیین فرهنگ لغت مشترک W اعمال می‌کنیم.

$$W \leftarrow W \otimes \frac{\frac{S}{WH} H^T}{1H^T}$$

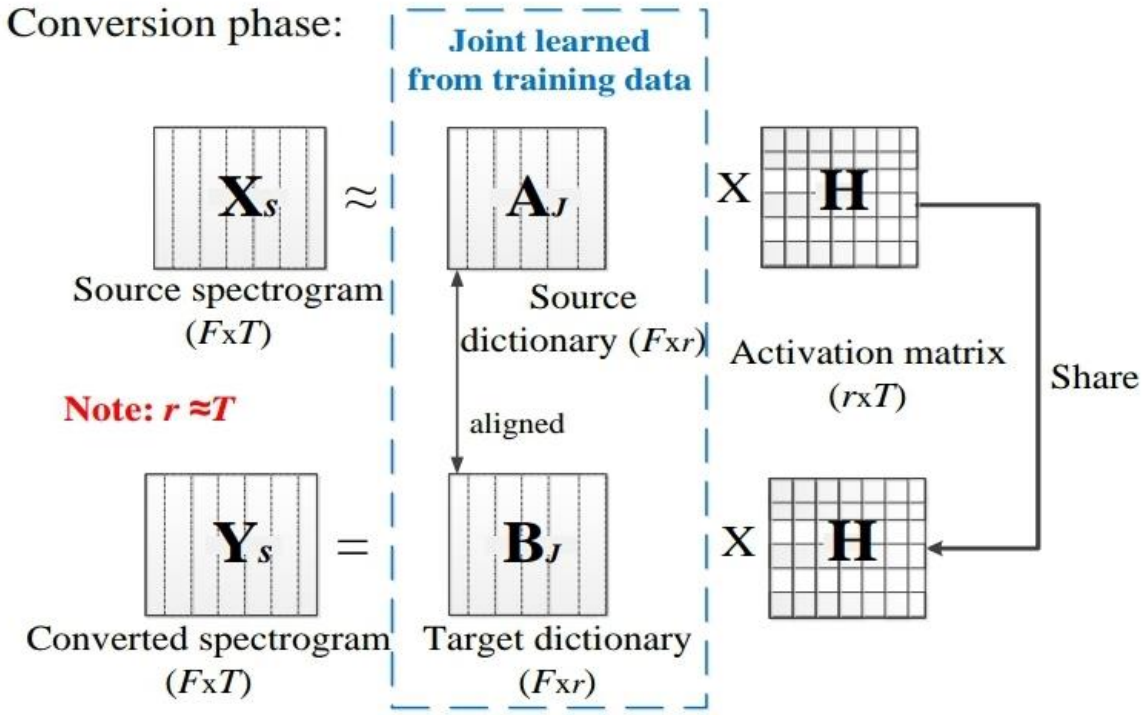
$$H \leftarrow H \otimes \frac{W^T \frac{S}{WH}}{W^T 1 + \lambda}$$

اینجا $1 \in R^{2F\times I}$ یک ماتریس است هدف ما به دست آوردن شکل است. شکل W فرهنگ لغت مشترک نمونه‌ای از لغت نامه آموخته شده را نشان می‌دهد که فرهنگ منبع به ترتیب نیمه (B) و لغت نامه هدف را اشغال می‌کند. در W بالایی و پایینی نقطه تبدیل فوریه ۵۱۲ این مثال ما برای سریع مشخص کردن صداهای گفتاری از کیلوهرتز استفاده کردیم و بنابراین یک ماتریس ۱۰ اینچ و F = 513 است. در شکل۳ ما می‌توانیم ببینیم پایه‌های A و B به‌طور مشترک تراز شده و یاد می‌گیرند در حین اجرا،"فرآیند تراز DTW "ac است. برای هر دو NMF مبتنی بر نمونه بسیار مهم است (شکل۱ و JD-NMF. هنگامی‌که سیگنال‌های گفتاری منبع و هدف دقیقاً در یک راستا قرار نگرفته اند، لغت نامه‌ها ممکن است به خوبی جفت نشوند.) مؤلفه‌های فرکانس متوسط A، نسبتاً پرسر و صدا

هستند در مقایسه با پایه‌هایB؛ و پایه‌های B نسبت به یکدیگر تمایز بیشتری نسبت به پایه‌های A دارند. توجه داشته باشید که A و B از تحریف گفتار(به ترتیب پس از جراحی) و گفتار واضح (قبل از جراحی) آموخته می‌شوند. مشاهده دوم می‌تواند نشان دهد چرا گفتار تحریف شده به گوش می‌رسد.



شکل۱-۳: فرهنگ لغت منبع و هدف؛ دو دیکشنری را می‌توان با جدا کردن نیمه‌های بالایی و پایینی دیکشنری مشترک W به ترتیب به‌دست آورد.



شکل۱-۴: مرحله آنلاین پیشنهادی NMF برای VC در مقایسه با موارد موجود در شکل۲، دیکشنری‌ها و ماتریس فعال سازی بسیار کوچک تر هستند.

تار، منجر به درک ضعیف گفتار می‌شود. هنگام مقایسه شکل۱ و شکل۳ می‌توانیم توجه کنیم که پایه‌های شکل۱ خیلی معرف نیستند. علاوه بر این، برخی از پایه‌های دو فرهنگ لغت در شکل۱ به خوبی جفت نمی‌شوند که دلیل آن ترازهای ناقص DTW است. از سوی دیگر، از آنجایی که فرهنگ لغت مشترک ما از کل داده‌های آموزشی آموخته می‌شود، موضوع ترازهای ناقص را می‌توان کاهش داد. در بخش بعدی، تبدیل گفتار تحریف شده به گفتار واضح را با استفاده از A و B با یک ماتریس فعال سازی مشترک معرفی می‌کنیم. برای کاهش هزینه محاسباتی در مرحله آنلاین، تعداد پایه‌های r باید به حداقل برسد. در بخش چهار نشان می‌دهیم که تنها چند پایه معرف که با استفاده از معیار آموزش مشترک آموخته شده‌اند، برای به‌دست آوردن یک نتیجه رضایت بخش کافی هستند.

ب) مرحله آنلاین

از آنجایی که روش‌های پیشنهادی JD-NMF و روش‌های NMF مبتنی بر نمونه معمولی عمدتاً در مرحله آموزش متفاوت هستند، فرآیند تبدیل می‌تواند به‌طور مشابه ارائه شود، همان‌طور که در شکل۴ نشان داده شده است. توجه داشته باشید که اندازه دیکشنری‌ها و ماتریس فعال سازی بسیار کوچکتر از اندازه‌های نشان داده شده در شکل۲ است.

$$Y = B_J H$$

جایی‌که $Y_s \in R^{F\times I}$ طیف نگار تبدیل شده است توجه داشته باشید که تعداد پایه‌های r در چارچوب ما بسیار کمتر از روش معمولی است. قانون به‌روز رسانی ضربی را می‌توان به صورت زیر تغییر داد:

$$H \leftarrow H \otimes \frac{A_J^T \frac{X_s}{A_J H}}{A_J^T 1 + \lambda}$$

جایی‌که $X_s \in R^{F\times I}$ طیف نگار منبعی است که قرار است تبدیل شود و $1 \in R^{F\times I}$ یک ماتریس همه یک است. می‌توان ببینیم که با کاهش اندازه(تعداد ستون‌ها) A، می‌توانیم مقدار زیادی از زمان محاسباتی را هنگام محاسبه ماتریس فعال سازی H ذخیره کنیم، بنابراین تبدیل سریع را امکان پذیر می‌کنیم. برای تجزیه و تحلیل بیشتر هزینه محاسبات، تعداد ضرب یا تقسیم مورد نیاز برای هر تکرار را می‌توان به صورت زیر تخمین زد:

$2FrT + 2rT + FT$

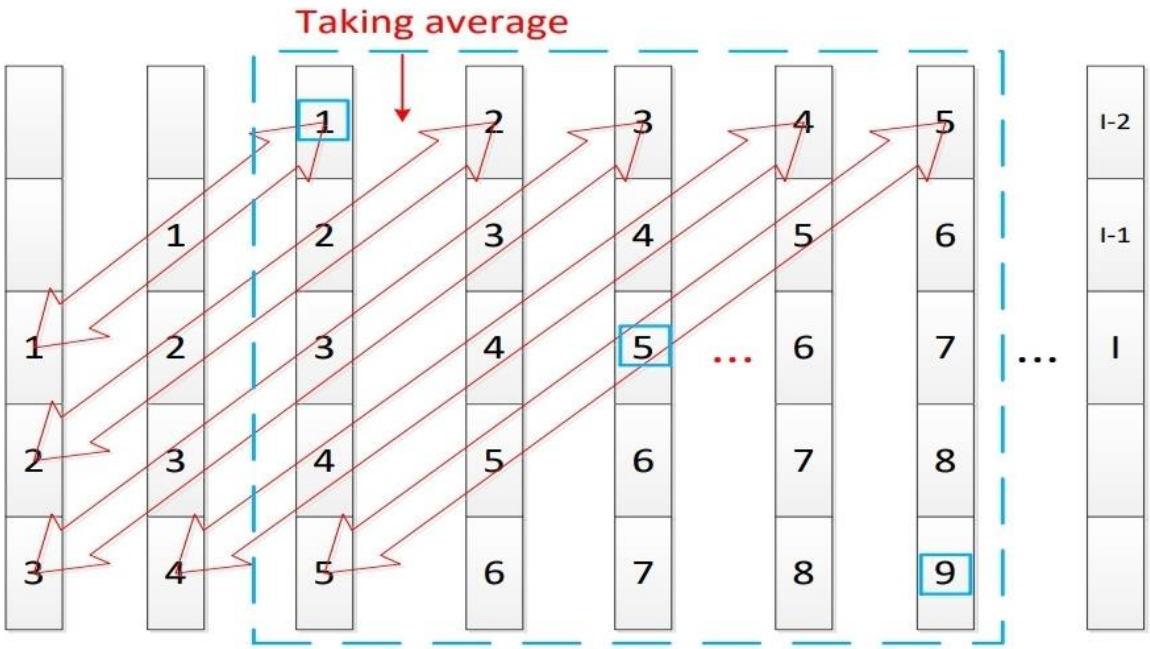
که یک تابع خطی از r است زمانی‌که F و T هر دو ثابت باشند(X داده شده است).

در بخش بعدی، ما اطلاعات زمینه‌ای را برای بهبود بیشتر عملکرد JD-NMF ترکیب می‌کنیم.

۱-۶ اطلاعات متنی

برای در نظر گرفتن اطلاعات زمینه در بسیاری از کاربردهای پردازش گفتار، ویژگی‌ها به گونه‌ای آبشاری می‌شوند که چندین فریم متوالی را در بر می‌گیرند. با این حال هیچ اطلاعات زمانی در نظر گرفته نشده است، یعنی هر فریم به طور مستقل مدل شده است. بنابراین، برای تخمین دقیق تر ماتریس فعال سازی، از نمونه‌های چند قاب در فرهنگ لغت منبع استفاده شد. در چارچوب JD-NMF خود، ما همچنین پیشنهاد دادیم که طیف نگارها را در چندین فریم متوالی آبشاری کنیم تا یک فرهنگ لغت مشترک توسعه یافته را آموزش دهیم. بر این اساس، در فاز آفلاین، X و Y تبدیل به $R^{(2q+1)F\times I}$ جایی که $2q + 1$ است می‌شوند که به نوبه خود باعث می‌شود A و B به $R^{(2q+1)F\times I}$ گسترش می‌یابد. در طول مرحله تبدیل، برای استفاده از فرهنگ لغت توسعه یافته، طیف گرام منبع X نیز برای تخمین آبشاری می‌شود. ماتریس فعال سازی در همین حال، فرهنگ لغت هدف آبشاری همچنین می‌تواند اطلاعات متنی را برای به‌دست آوردن مزایای دیگری در نظر بگیرد. شکل۵ دنباله‌ای از فریم‌های گفتار را نشان می‌دهد.در شکل، زمانی‌که فریم پنجم قرار است تبدیل شود، پنج بردار چند قاب(درداخل خطوط نقطه چین آبی) را در نظر می‌گیرد که از فریم اول تا نهم به‌دست آمده است، بنابراین برای تولید فریم پنجم تولید شده نهایی،

می‌توانیم میانگین را برای ادغام اطلاعات ارائه شده در پنج بردار چند قاب که با فلش‌های قرمز نشان داده شده‌اند محاسبه می‌کنیم. علاوه بر این، عملکرد متوسط می‌تواند نویز را کاهش دهد و انتقال بین صداهای گفتاری را هموار کند. بنابراین، آبشاری طیف نگار آموزشی می‌تواند تا حد زیادی کیفیت گفتار تبدیل شده را در چارچوب JD-NMF پیشنهادی بهبود بخشد.



شکل۱-۵: محاسبه میانگین در بردارهای چندقاب برای کاهش نویز در مرحله تبدیل (در اینجا در این مثال، اندازه پنجره (2q + 1 = 5)

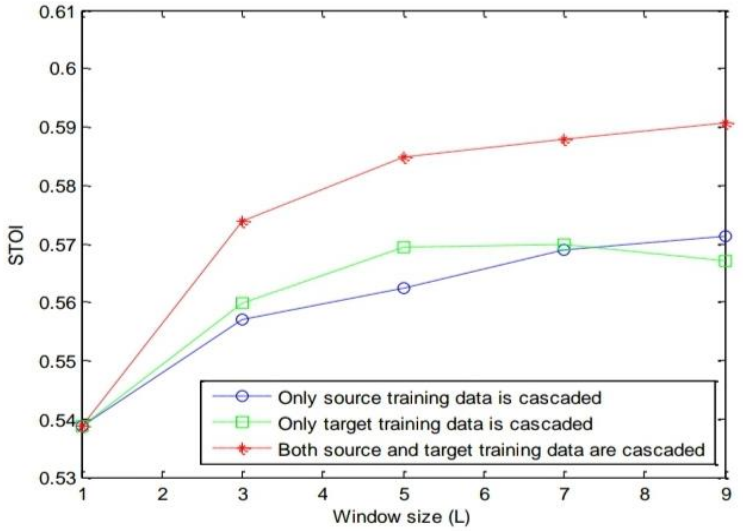
۱-۷ آزمایش

هدف از مطالعه حاضر ارائه یک سیستم VC سریع برای بیماران پس از جراحی دهان می‌باشد. دو ارزیابی عینی در نظر گرفته می‌شود: قابل فهم بودن گفتار تبدیل شده و هزینه محاسباتی تبدیل‌ها. یک روش ارزیابی استاندارد شده، قابل فهم بودن هدف کوتاه مدت (STOL)، به عنوان معیار قابل فهم عینی ما به کار گرفته شده است. محاسبه STOL، براساس همبستگی بین پاکت‌های زمانی هدف و گفتار تبدیل شده برای بخش‌های کوتاه. امتیاز STOL خروجی از ۰ تا ۱ متغیر است و انتظار می‌رود به طور یکنواخت با میانگین قابل فهم بودن گفتار تبدیل شده مرتبط باشد. از این رو، مقدار STOL بالاتر نشان دهنده درک بهتر گفتار است. برای ارزیابی هزینه محاسباتی، از تعداد ضرب یا تقسیم برای هر تکرار استفاده می‌شود. علاوه بر این ما زمان اجرای واقعی فاز آنلاین را برای مقایسه اندازه گیری کردیم. در آزمایشات این تحقیق ۱۵۰ جمله کوتاه به عنوان مجموعه خود تهیه کردیم. از این

میان ۷۰ گفته به صورت تصادفی به عنوان مجموعه آموزشی، ۴۰ گفتار به صورت تصادفی به عنوان مجموعه توسعه و ۴۰ گفتار باقی مانده به عنوان مجموعه ارزیابی انتخاب شدند. یک مرد بدون جفت فیزیکی به عنوان گوینده هدف انتخاب شد. ما ۱۵۰ جمله را که توسط چهار بیمار پس از جراحی دهان و همچنین سخنران مورد نظر بیان شده بود، ضبط کردیم. رویه‌ها توسط کمیته‌های هیئت بررسی نهادی محلی و تصویب قرار گرفت. سیگنال‌های گفتاری با فرکانس ۱۶ کیلوهرتز نمونه برداری شدند و هر ۱۰ میلی ثانیه با یک پنجره ۲۰ میلی ثانیه‌ای نمایش داده شدند. پارامترهای موجود در فرهنگ لغت و ماتریس فعال سازی با اعداد تصادفی از توزیع نرمال(میانگین=۰ و انحراف استاندارد= ۱، با مقدار مطلق) مقدار دهی اولیه داده می‌شوند. با دیکشنری‌های اولیه و ماتریس فعال سازی، آن‌ها را به روز می‌کنیم. برای کاهش اثر اولیه سازی‌های تصادفی ماتریس در NMF ، هر مجموعه آزمایش ۱۰ بار تکرار شد و مقادیر متوسط به‌دست آمد. از آنجایی که JD-NMF پیشنهادی از تعداد پایه‌های بسیار کمتری نسبت به NMF مبتنی بر نمونه استفاده می‌کند، محدودیت پراکندگی اعمال نمی‌شود. در بحث زیر، آزمایش‌های A تا C با استفاده از داده‌های آموزشی و مجموعه توسعه به ترتیب برای مراحل آفلاین و آنلاین انجام شد. سپس از بهترین پارامترها برای آزمایش عملکرد با استفاده از داده‌های مجموعه ارزیابی برای مرحله آنلاین استفاده شد. نتایج در آزمایش D ارائه شد.

الف) واژه نامه‌های آبشار

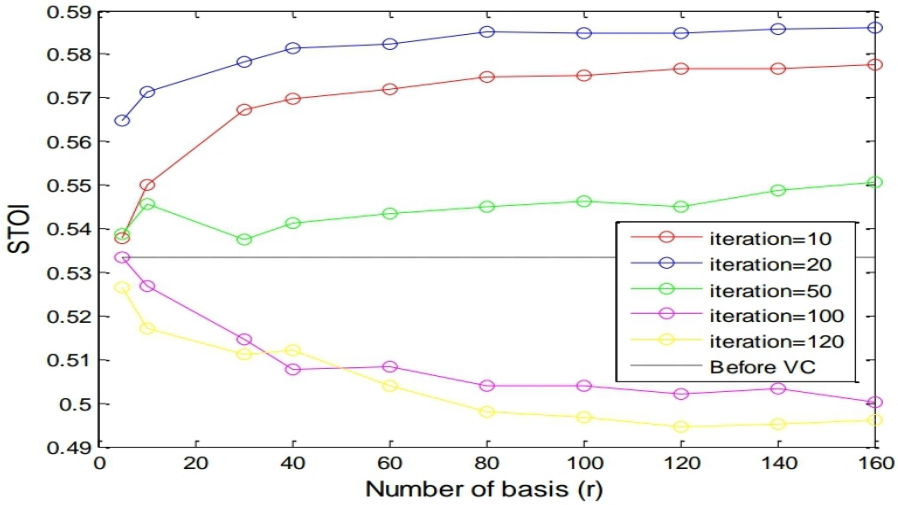
ابتدا اثر استفاده از پایگاه‌های چند قاب را بررسی کردیم که اطلاعات متنی مفید است. شکل۶ نتایج حاصل از STOL (محور Y) از مجموعه‌های توسعه را به عنوان تابع از اندازه پنجره L (X محور) ارائه می‌دهد. در اینجا ما سه مرحله مختلف را بررسی می‌کنیم. ۱- آبشارها تنها داده‌های منبع (فقط یک در شکل۴) گسترش یافت. ۲- cascading تنها از داده‌های آموزشی هدف نشان می‌دهد. ۳- هنگامی‌که تنها یک برنامه فریاد می‌زند، به طوری که پنجره به اندازه کافی افزایش یابد، به طوری که در طول زمان، یک فرآیند بیش از حد بسیاری از فریم‌ها در یک زمان در نظر گرفته می‌شود. عملکرد به طور ممکن است به طور متوسط بهبود یابد، به طوری که اگر یک بار در بسیاری از فریم‌ها در نظر گرفته می‌شود.



شکل۱-۶: نتایج STOL برای توسعه تنظیم به عنوان یک تابع از اندازه پنجره

ب) اثر تعداد پایگاه‌ها و تکرار

دو پارامتردیگر وجود دارد که می‌تواند بر رفتار منفی از لحاظ هوشگیری و هزینه محاسباتی چارچوب JD-NMF تأثیر بگذارد: تعداد پایگاه‌های DIS در تریلی و تعداد تکرار در طول تبدیل. برای تعیین درجه تأثیر آن‌ها، ما را با تنظیمات مختلف برای توسعه محاسبه کردیم. شکل ۷ نمره‌های STOL (محور Y) را به عنوان تابع تعداد دفعات (محور X) در تعداد تکرار متفاوت ارائه می‌دهد. نتایج نشان می‌دهد که تعداد پایگاه‌ها و تکرارها بر یکدیگر تأثیر می‌گذارد، بنابراین ما آن‌ها را به طور جداگانه در جزئیات مورد بررسی قرار دادیم که به شرح زیر است:



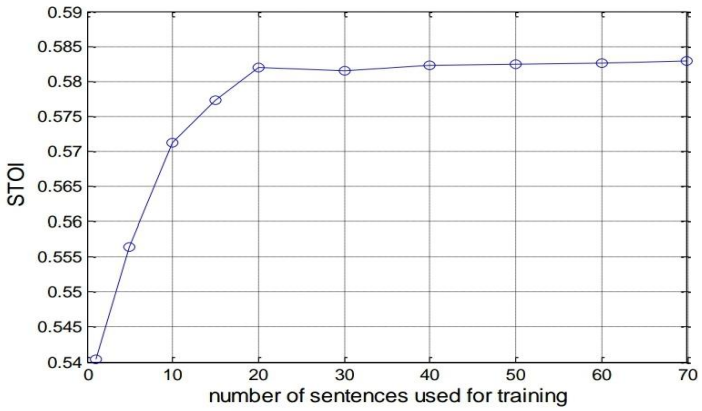
شکل۱-۷: STOL به عنوان یک تابع از تعداد پایگاه‌های R در فرهنگ لغت در زیر تکرار متفاوت است.

۱- اثر تعداد پایگاه‌ها

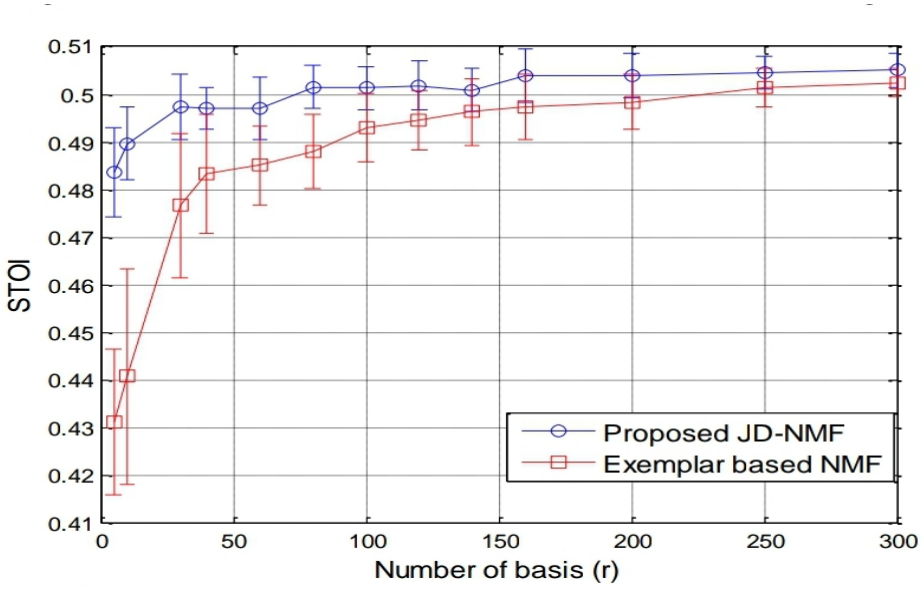
اول ما تغییرات را در STOL برای شمع‌های مختلفی از R مورد بررسی قرار دادیم. به عنوان فرهنگ لغت از داده‌های آموزشی در روش پیشنهادی ما، می‌توانیم اندازه‌های مختلف را برای فرهنگ لغت یاد بگیریم. شکل۲ نشان می‌دهد که STOL با تعداد زیادی از مبدل زمانی که عدد تکرار کوچک است افزایش می‌دهد. با این حال، اگر الگوریتم بیش از حد تکرار شود، منجر به رکورد می‌شود، نشان می‌دهد که هر پایگاه بیشتر باعث می‌شود که قابلیت‌های بیشتر را به تعداد تکرار تبدیل کند. توجه داشته باشید که با ۸۰ پایگاه، بهبود یافته STOL شروع می‌شود، زمانیکه عدد تکرار کوچک است در این مورد، اضافه کردن پایگاه‌های بیشتر بهبود می‌یابند.

۲- اثر تعداد تکرارها

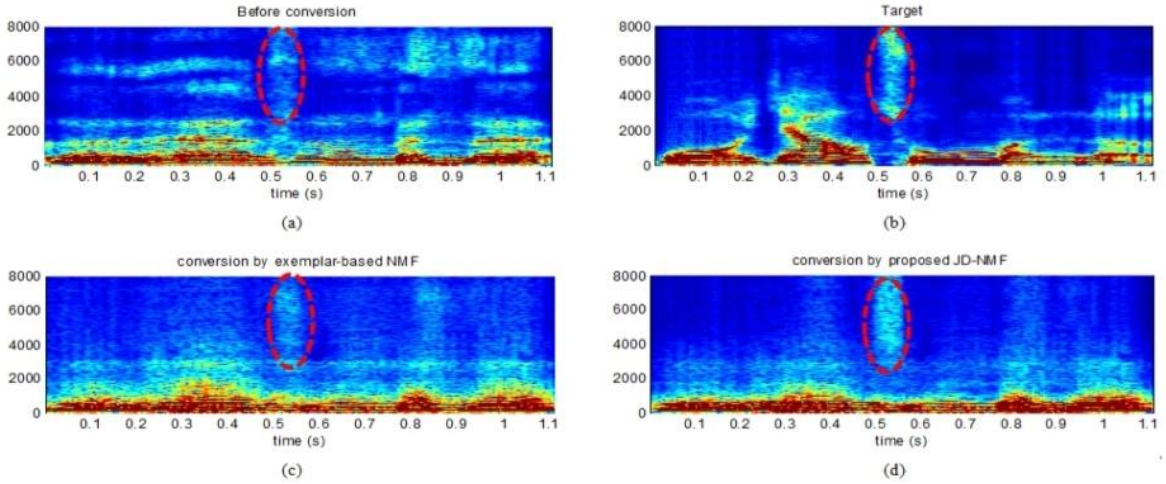
تعداد تکرار معمولاً به صورت تجربی از یک مجموعه توسعه تعیین می‌شود. اگر تعداد توسعه خیلی کم/زیاد باشد، مدل آموخته شده با داده‌های آموزشی کمتر برازش/ بیش از حد برازش می‌کند. شکل۷ نشان می‌دهد که زمانی‌که الگوریتم چندین بار تکرار می‌شود، مقادیر STOL به دلیل برازش بیش از حد شروع به کاهش می‌کند. اگرچه تفاوت بین تروگرام طیف منبع X و طیف نگار مدل شده H، A همیشه تضمین می‌شود که پس از هر تکرار با به‌روزرسانی H کاهش یابد، هیچ تضمین نظری وجود ندارد که گفتار تبدیل شده بتواندبر این اساس با اجرای تکرارهای بیشتر بهبود یابد. از این رو، اگر B، H به هم ریخته و غیرصافی تولید کند. برای غلبه بر این مشکل، می‌توانیم به سادگی تکرار را زودتر متوقف کنیم. این روش منظم سازی، توقف زودهنگام نیز نامیده می‌شود. از شکل۷ با ۲۰ تکرار می‌توانیم بدون صرف زمان محاسباتی زیاد، به بالاترین مقدار STOL برسیم. بنابراین، تعداد تکرار برای مجموعه ارزیابی روی ۲۰ تنظیم شده است. به‌طور خلاصه در روش پیشنهادی، اندازه دیکشنری منبع و هدف بر روی ۸۰ × (۵ ×۵۱۳) با ۲۰ تکرار در طول تبدیل تنظیم می‌شود.



شکل۱-۸: STOL به عنوان تابعی از تعداد جملات مورد استفاده آموزش



شکل۱-۹: STOL به عنوان تابعی از تعداد پایه برای روش‌های پیشنهادی و پایه



شکل۱-۱۰: طیف نگارهای منبع، هدف و گفتار تبدیل شده پس از DTW در مجموعه ارزیابی. الف) گفتار منبع(قبل از تبدیل)، ب)گفتار هدف، ج) گفتار تبدیل شده توسط NMF مبتنی بر نمونه و د) توسط ID-NMF پیشنهادی

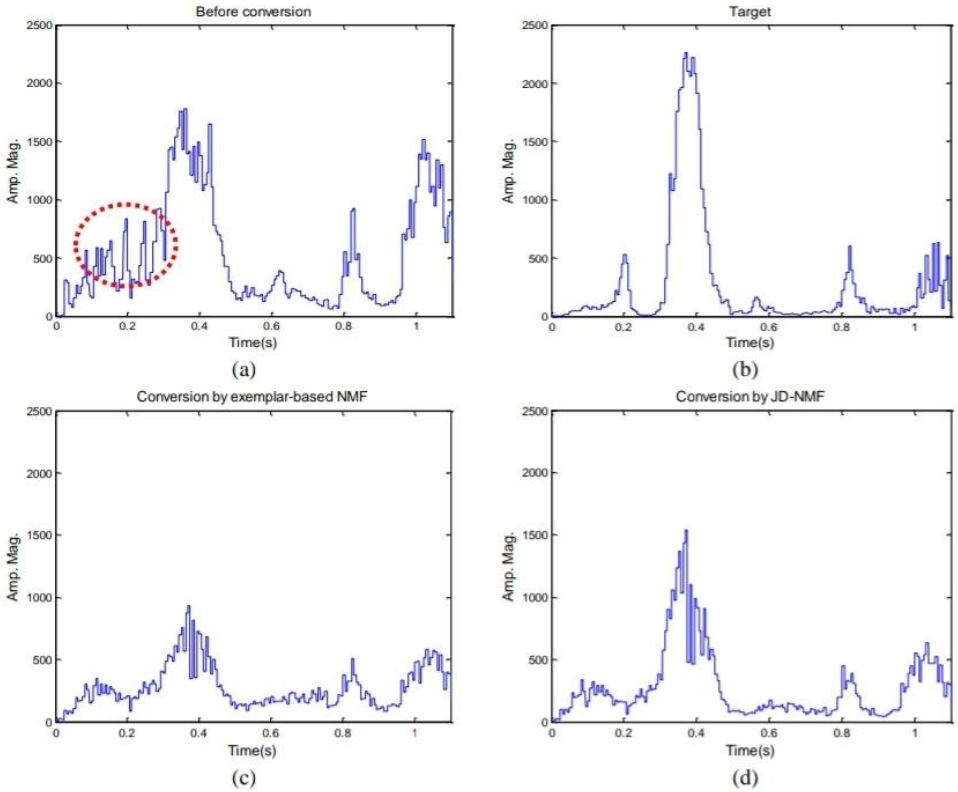
ج) مقدار داده‌های آموزشی

در سناریوی کاربردی ما، جمع آوری حجم زیادی از داده‌های آموزشی دشوار است زیرا صحبت طولانی مدت برای بیماران پس از جراحی دشوار است. بنابراین، ما به بررسی استحکام روش پیشنهادی برای مقادیر مختلف داده‌های آموزشی. محورهای X و Y به ترتیب تعداد جملات مورد استفاده برای آموزش و امتیازات STOL را نشان می‌دهند.

تعداد جملات از ۱ تا ۷۰ متغیر بود. جملات به صورت تصادفی از مجموعه اصلی انتخاب شدند. شکل۸ نشان می‌دهد که بهبود STOL زمانی شروع به اشباع می‌کند که تقریبا ۲۰ جمله برای آموزش استفاده شود. به عبارت دیگر سیستم ما تنها با ۲۰ جمله قابل آموزش است.

د) مقایسه عملکرد کلی

در نهایت، روش‌های پیشنهادی JD-NMF و پایه (NMF مبتنی بر نمونه) را با استفاده از مجموعه ارزیابی مقایسه کردیم. شکل۹ نمرات STOL (میانگین و انحراف معیار) را به عنوان تابعی از پایه‌های عددی برای مقایسه نشان می‌دهد. همان تعداد پایه برای JD-NMF و NMF مبتنی بر نمونه، استفاده شد و پایه‌های مورد استفاده برای NMF مبتنی بر نمونه به طور تصادفی از نمونه‌های تهیه شده از داده‌های آموزشی انتخاب شدند. برای نتایج NMF مبتنی بر نمونه، اندازه پنجره و تعداد تکرار بر اساس مجموعه توسعه بهینه شده است. قابل توجه است، هر نتیجه میانگین در شکل۹ با میانگین ۱۶۰۰ امتیاز (۱۰× ۴× ۴۰) بود: ۴۰ گفتار آزمایشی توسط ۴ بیمار همراه با ۱۰ حرف اول تصادفی ثبت شد تا از مسئله تصادفی اجتناب شود. در همین حال هر انحراف معیار در شکل از ۱۰ نتیجه (به دست آمده از ۱۰ حرف اولیه تصادفی مختلف) تخمین زده می‌شود و هر یک از این ۱۰ نتیجه میانگین ۱۶۰ امتیاز STOL (۴۰ گفتار آزمایشی ثبت شده توسط ۴ بیمار) است. از شکل اشاره می‌شود که وقتی اندازه فرهنگ لغت کوچک است، JD-NMF به طور قابل توجهی بهتر از NMF مبتنی بر نمونه است. به عنوان مثال، STOL (JD-NMF) با ۸۰ پایه تقریباً مشابه NMF مبتنی بر نمونه با ۳۰۰ پایه است. این نشان می‌دهد که پایگاه‌های مشترک آموخته شده اطلاعات معنی دار بیشتری نسبت به نمونه‌های به‌دست آمده به طور مستقیم ارائه می‌دهند. ما می‌خواهیم تأکید کنیم که مطالعه حاضر بر دو الزام اصلی داده‌های آموزشی محدود و تبدیل سریع آنلاین تمرکز دارد. بنابراین، ما فقط نتایج NMF مبتنی بر نمونه و JD-NMF را با پایه‌های کمتر از ۳۰۰ ارائه می‌کنیم.



شکل ۱-۱۱: پاکت‌های دامنه از کانال پنجم منبع، هدف و گفتار تبدیل شده: الف) گفتار منبع(قبل از تبدیل). ب) گفتار هدف و ج) گفتار تبدیل شده از طریق NMF مبتنی بر نمونه و د) توسط JD-NMF پیشنهادی

TABLE I
COMPARISON OF COMPUTATIONAL LOAD

Methods	# of multiplica-tions and divisions (Eq. (15))	Execution time (s)
Proposed JD-NMF (80 bases) (J)	413,125	0.1177
Exemplar-based NMF (300 bases) (E)	1,542,165	0.3332
Ratio (J/E)	0.2679	0.3532

برای تخمین صرفه جویی در هزینه محاسباتی، تعداد ضرب و تقسیم در هر فریم را می‌توان با ابعاد ویژگی F مجموعه ۵ × ۵۱۳ اعمال کرد. برای مقایسه عملی بار محاسباتی، ما همچنین زمان اجرای مورد نیاز برای تولید را مقایسه کردیم. ماتریس فعال سازی در حین تبدیل یک گفته هدف (۱,۲ ثانیه) بر روی یک کامپیوتر ۳,۶

گیگاهرتزی که در نرم افزار متلب[1] پیاده سازی شده است. هر دو نتیجه در جدول فهرست شده‌اند که در آن می‌توان مشاهده کرد که JD-NMF پیشنهادی و NMF مبتنی بر نمونه به ترتیب به 413,125 و 1,542,165 تعداد ضرب و تقسیم نیاز دارند. به عبارت دیگر، نسبت محاسبه دو روش ۰,۲۶۸ است. زمان اجرای JD-NMF و EXEMPLAR بر اساس NMF . نتایج فوق تأیید می‌کند که روش پیشنهادی ما می‌تواند محاسبات آنلاین را در حدود سه ضریب نسبت به روش مرسوم کاهش دهد. در مرحله بعد، JD-NMF ما به صورت بصری اثر را بر روی گفتار تحریف شده با استفاده از نمودارهای طیف نگاری بررسی کردیم. نمودار طیف نگاری تغییرات فرکانس‌های موجود در یک سیگنال گفتاری را مشان می‌دهد. محور y نشان دهنده شاخص زمان است درحالی که نشان دهنده سطح فرکانس با رنگ قرمز نشان داده شده نشان دهنده شدت زیاد و رنگ آبی شدت‌های کم را نشان می‌دهد. شکل۱۰ یک جفت گفتار منبع و هدف را نشان می‌دهد. شکل۱۰ (الف) و (ب) به ترتیب با استفاده از NMF مبتنی بر مثال و شکل۱۰ (ج) و (د) JD-NMF را نشان می‌دهد. شکل‌ها نشان می‌دهد که صدای همخوان (ناحیه در دایره قرمز) قبل از تبدیل نامشخص است زیرا مفصل‌ها حذف شده اند. علاوه بر این، اجزای فرکانس متوسط برای گفتار قبل از تبدیل نسبتاً پر سر و صدا هستند. این مشاهدات در شکل۲ نیز مشاهده شده است. در مرحله بعد، ما متذکر می‌شویم که، اگرچه NMF مبتنی بر نمونه می‌تواند یک صامت را کمی تقویت کند، اما طیف وسیعی از نویز را نیز تولید می‌کند، به خصوص در فرکانس‌های بالا. در مقابل JD-NMF پیشنهادی ما می‌تواند به طور قابل توجهی قسمت صامت را بهبود بخشد و در عین حال یک قسمت با فرکانس بالا را تمیز نگه دارد که بیشتر شبیه به ویژگی گفتار هدف است. در نهایت مقایسه کیفی دیگری از VC های مبتنی بر emplar و JD-NMF از طریق لبه‌های پوششی پردازش شده ارائه می‌کنیم. مطالعات قبلی نشان داد که عمق مدولاسیون نیز عامل مهمی است که بر ادراک گفتار تأثیر می‌گذارد. عمق مدولاسیون بالاتر باعث درک بهتر گفتار می‌شود. در این مطالعه، ما از یک صداگذاری هشت کاناله استفاده کردیم که به عنوان ابزاری برای استخراج پاکت‌ها تحت باندهای فرکانسی مختلف استفاده می‌شود. اشاره شد که باند فرکانسی میانی برای درک گفتار بسیار مهم است. بنابراین، فقط پاکت‌های موجود در کانال پنجم برای مقایسه انتخاب شدند. شکل۱۱ پوشش‌های دامنه را از کانال پنجم یک جفت گفتار منبع و هدف پس از تراز از طریق DTW نشان می‌دهد. شکل۱۱ (الف) و (ب) به ترتیب با تبدیل گفتار توسط JD-NMF های مبتنی بر نمونه و پیشنهادی و شکل۱۱ (ج) و (د) به ترتیب محورهای x و y شاخص زمان و قدر دامنه را نشان می‌دهند. شکل۱۱ نشان می‌دهد که پاکت قبل از تبدیل دارای اعوجاج در حدود ۰,۲ ثانیه(در دایره قرمز) و عمق مدولاسیون کمتری نسبت به گفتار هدف است. علاوه براین، در حالی که هر دو VC مبتنی برنمونه و JD-NMF می‌توانند اعوجاج را کاهش دهند، عمق مدولاسیون دومی بسیار بالاتر است. در نهایت

[1] MATLAB

مقایسه‌ای از شکل ۱۱ (ب) و (د) نشان می‌دهد که پاکت JD-NMF شباهت زیادی به گفتار هدف دارد که به معنی قابل فهم بودن گفتار بهتر است.

۱-۸ نتیجه گیری

ما VC مبتنی بر JD-NMF را برای بیماران جراحی دهان پیشنهاد می‌کنیم. فرآیند کلی JD-NMF را می‌توان به دو مرحله تقسیم کرد: آفلاین و آنلاین. در مرحله آفلاین، JD-NMF ماتریس فرهنگ لغت منبع و هدف جفتی را می‌آموزد. برای اطمینان از همسویی پایه‌های ماتریس فرهنگ لغت منبع و مقصد، این دو ماتریس به طور مشترک یاد می‌گیرند. در فاز آنلاین، هنگام اجرای VC، ماتریس فعال سازی توسط بلندگوهای منبع و هدف به اشتراک گذاشته می‌شوند. ما JD-NMF پیشنهادی را با استفاده از داده‌های گفتاری در دنیای واقعی‌که از بیماران پس از جراحی‌های دهان به‌دست آمده بود، ارزیابی‌کردیم. نتایج تجربی ما ابتدا نشان‌دادکهJD-NMF گفتار اصلی را با امتیاز STOL بالا بسیار بهبود بخشید. علاوه بر این، JD-NMF به طورقابل توجهی کارآمدتر و مؤثر تر از روش متداول مبتنی بر NMFVC است. در نهایت، از طریق آنالیزهای کمی با استفاده از طیف نگار و نمودارهای پوشش گفتار، مشخص شد که JD-NMF پیشنهادی طیف‌های شفاف تری را با عمق مدولاسیون واضح‌تری نسبت به گفتار اصلی تولید می‌کند که توسط NMF مبتنی بر نمونه‌های معمولی تبدیل می‌شود. به‌طور خلاصه، سهم این مقاله دو برابر است. اول، ما اثربخشی معیار آموزش مشترک پیشنهادی را برای VC مبتنی بر NMF تأیید کردیم. دوم اینکه، ما تأیید کردیم که JD-NMF می‌تواند هوش گفتاری بیمارانی را که تحت عمل جراحی دهان قرارگرفته‌اند، تا حدزیادی افزایش دهد. درمطالعه حاضر، ما اثر بخشی روش JD-NMF پیشنهادی را از نظر امتیازات STOL عینی و هزینه محاسباتی آنلاین تأیید کردیم. در آینده قصد داریم موارد زیر را انجام دهیم:

۱- آزمایش‌های تشخیص عینی را برای تأیید بیشتر کاربرد بالینی JD-NMF پیشنهادی انجام دهیم، حتی اگر STOL تأیید شده باشد که قادر به پیش‌بینی دقیق قابل فهم بودن گفتار است.
۲- این مطالعه اثربخشی JD-NMF پیشنهادی در حال اجرا بر روی کامپیوتر را تأیید کرده است. بنابراین، ما قصد داریم آن را به عنوان یک دستگاه الکترونیکی مستقل یا به عنوان یک برنامه برای یک گوشی هوشمند پیاده سازی کنیم.

تصدیق: این تحقیق تا حدی توسط وزارت علوم فناوری تایوان، دانشگاه ملی تایوان پشتیبانی شده است.

منابع:

ارزیابی گفتار و توانایی بلع بعد از سرطان داخل دهان صفحه۱۷. زبان شناسی و آوایی بالینی۴۲۰-۴۱۱، ۲۰۰۳

اثرات گلوسکتومی بر قابل فهم بودن گفتار و تمایز ادراکی دهان صفحه۳۸. ۳۵۴-۳۴۸، ۱۹۸۰

B.R Pauloski و همکاران " عملکرد گفتار و بلع بعد از برداشتن زبان و کف دهان قدامی با بازسازی فلپ دیستال[1] " گفتار، زبان و شنوایی مجله

R. Aihara و همکاران "تقویت همخوان برای اختلالات بیان بر اساس فاکتورسازی ماتریس غیرمنفی" صفحه ۴-۱ ، ۲۰۱۲

R. Aihara و همکاران " تبدیل صدای حفظ کننده فردی برای اختلالات بیان بر اساس فاکتورسازی ماتریس غیرمنفی" ۸۰۴۰-۸۰۳۷ ، ۲۰۱۳

T. Toda و همکاران " تبدیل صدا برای انواع مختلف گفتار منتقل شده توسط بدن" ۳۶۰۴-۳۶۰۱ ، ۲۰۰۹

K. Nakamura و همکاران " سیستم کمک گفتاری برای کل حنجره‌ها با استفاده از تبدیل صوتی گفتار مصنوعی منتقل شده از بدن" ۱۳۹۸-۱۳۹۵ ، ۲۰۰۶

Y-T.Liu و همکاران " فناوری کاهش فرکانس مبتنی بر فاکتورسازی غیرمنفی ماتریس برای کاربران سمعک ماندارین[2]" ۲۰۱۶

T. Toda و همکاران " تبدیل صدا بر اساس تخمین حداکثر احتمال مسیر پارامترهای طیفی، صوت، گفتار و پردازش زبان" صفحه۱۵، ۲۲۲۲-۲۲۳۵، ۲۰۰۷

A.Kain و M.W.Macon " تبدیل صدای طیفی برای سنتز متن به گفتار" ۲۸۸-۲۸۵، ۱۹۸۸

Hwang و همکاران " مطالعه اطلاعات متقابل برای تبدیل طیفی مبتنی بر GMM " صفحه۸۱-۷۸، ۲۰۱۲

Hwang و همکاران " گنجاندن واریانس سراسری در مرحله آموزش تبدیل صدای مبتنی بر GMM" صفحه ۶-۱، ۲۰۱۳

[1] DistalFlap
[2] Mandarin

M.Narendranath و همکاران" تبدیل فرمت‌ها برای تبدیل صدا با استفاده از شبکه‌های عصبی مصنوعی" صفحه ۱۶، ۲۰۷-۲۱۶، ۱۹۹۵

S.Desai و همکاران " با استفاده از شبکه‌های عصبی مصنوعی برای تبدیل صدا " صفحه۱۸، ۹۶۴-۹۵۴، ۲۰۱۰

Xie و همکاران " آموزش کمینه سازی خطای توالی شبکه عصبی برای تبدیل صدا" ۲۲۸۷-۲۲۸۳، ۲۰۱۴

Hwang و همکاران " یک تفسیر احتمالی برای تبدیل صدا مبتنی بر شبکه عصبی مصنوعی" ۲۰۱۵

A.Kain و S.H. Mohammadi "تبدیل صدا با استفاده از شبکه‌های عصبی عمیق با پیش آموزش مستقل از سخنران در کارگاه فناوری زبان گفتاری" صفحه۲۳-۱۹ ، ۲۰۱۴

L.Sun و همکاران " تبدیل صدا با استفاده از شبکه‌های عصبی بازگشتی مبتنی بر حافظه کوتاه مدت دو طرفه عمیق " ۴۸۷۳-۴۸۶۹ ، ۲۰۱۵

A.Kain و S.H. Mohammadi " آموزش نیمه نظارت شده یک تابع نگاشت تبدیل صدا با استفاده از رمزنگاری خودکار مشترک" ۲۰۱۵

M. Dong و همکاران " نقشه برداری فریم‌ها با شناسایی کننده برای تبدیل صدای غیرموازی" ۴۹۴-۴۸۸، ۲۰۱۵

Z.Wu و همکاران" تبدیل صدای مبتنی بر نمونه با استفاده از دکانولوشن غیر منفی طیفگرا" ۲۰۶-۲۰۱، ۲۰۱۳

Z.Wu و همکاران" فاکتورسازی ماتریس غیرمنفی مشترک برای تبدیل صدای مبتنی بر نمونه" ۲۰۱۴

Z.Wu و همکاران" نمایش پراکنده مبتنی بر نمونه با جبران باقی‌مانده برای تبدیل صدا" پردازش صوتی، گفتار و زبان، معاملات IEEE جلد ۲۲، ۱۵۰۶-۱۵۲۱، ۲۰۱۴

K.Masaka و همکاران " تبدیل صدای چندوجهی با استفاده از فاکتورسازی ماتریس غیرمنفی در محیط‌های پر سر و صدا" ۱۵۴۶-۱۵۴۲، ۲۰۱۴

D.D.Lee و H.S.Seung " یادگیری اجزای اشیاء با فاکتورسازی ماتریس غیرمنفی " ۷۹۱-۷۸۸، ۱۹۹۹

Z.Wu و همکاران" تبدیل صدای مبتنی بر مثال با استفاده از فاکتورسازی غیرمنفی مشترک" ابزارها و برنامه‌های چندرسانه‌ای ۹۹۴۳-۹۹۵۸، ۲۰۱۵

D.D.Lee وH.S.Seung " الگوریتم‌هایی برای فاکتورسازی ماتریس غیرمنفی" ۵۶۲-۵۵۶، ۲۰۰۱

M.Muler " تحریف زمان پویا، بازیابی، اطلاعات برای موسیقی و حرکت" صفحه۶۹-۸۴، ۲۰۰۷

P.O.Hoyer " کدگذاری پراکنده غیرمنفی" ۵۶۵-۵۵۷، ۲۰۰۲

P.O.Hoyer " ماتریس غیرمنفی با محدودیت های پراکندگی" صفحه۵، ۱۴۶۹-۱۴۵۷، ۲۰۰۴

R. Peharz و F.Pernkopf " فاکتورسازی ماتریس غیرمنفی پراکنده با محدودیت" صفحه۴۶-۳۸، ۲۰۱۲

A. Cichocki و همکاران" واگرایی‌هایی برای فاکتورسازی ماتریس غیرمنفی: خانواده الگوریتم‌های جدید" در تحلیل مؤلفه‌های مستقل و جداسازی سیگنال کور، صفحه۳۹-۳۲، ۲۰۰۶

J.F.Gemmeke و همکاران" بازنمایی‌های پراکنده مبتنی بر نمونه برای تشخیص خودکار گفتار قوی نویز، پردازش صوتی، گفتار و زبان" ۲۰۸۰-۲۰۶۷، ۲۰۱۱

A.Ozerov و C.Fevotte " فاکتورسازی ماتریس غیرمنفی چند کانالی در مخلوط‌های پیچیده برای جداسازی منبع صوتی پردازش صدا، گفتار و زبان" صفحه۱۸، ۵۶۳-۵۵۰، ۲۰۱۰

K.W.Wilson و همکاران" حذف نویز گفتار با استفاده از فاکتورسازی ماتریس غیرمنفی با پیشین" ۴۰۳۲-۴۰۲۹، ۲۰۰۸

N. Mohammadiha و همکاران" تقویت گفتار تحت نظارت و بدون نظارت با استفاده از فاکتورسازی ماتریس غیرمنفی" ۲۱۴۰-۲۱۵۱، ۲۰۱۳

H.T.Fan و همکاران" تقویت گفتار با استفاده از فاکتورسازی ماتریس غیرمنفی" ۴۴۸۷-۴۴۸۳، ۲۰۱۴

C.H.Taal و همکاران" معیار درک هدف کوتاه مدت برای گفتار پر سر و صدا وزن‌دار با فرکانس زمان" ۴۲۱۴-۴۲۱۷، ۲۰۱۰

C.H.Taal و همکاران" الگوریتمی برای پیش‌بینی قابل فهم بودن گفتار پر سر و صدا با فرکانس زمانی، پردازش صوتی، گفتار و زمان" ۲۱۳۶-۲۱۲۵، ۲۰۱۱

C.Fevotte وهمکاران" فاکتورسازی ماتریس غیرمنفی با واگرایی با کاربرد در تحلیل موسیقی محاسبات عصبی" ۸۳۰-۷۹۳، ۲۰۰۹

P.Saja و همکاران " بازیابی طیف‌های سازنده با استفاده از فاکتورسازی ماتریس غیرمنفی" در علوم و فنون نوری ۳۳۱-۳۲۱، ۲۰۰۳

L.Prechelt " توقف اولیه – اما چه زمانی؟ در شبکه‌های عصبی ترفندهای تجارت" صفحه۵۵-۶۹، ۱۹۹۸

J.L.Flangan " تجزیه و تحلیل گفتار سنتز و ادراک" ۲۰۱۳

S.Haykin "پیشرفت‌ها در تجزیه و تحلیل طیف و پردازش"،۱۹۹۹

Y.H.Lai و همکاران "تأثیرات نرخ سازگاری و سرکوب نویز برقابلیت فهم گفتار مبتنی بر پاکت فشرده" ۲۰۱۵

مؤسسه A.N.S استاندارد ملی آمریکا: روش‌های محاسبه شاخص درک گفتار انجمن آکوستیک آمریکا۱۹۹۷

Szu-WeiFu مدرک B.S در گروه علوم مهندسی اقیانوس و مؤسسه فارغ التحصیل مهندسی ارتباطات از دانشگاه ملی تایوان، تایپه، تایوان، به ترتیب در سال ۲۰۱۲ و ۲۰۱۴. او در حال حاضر در حال پیگیری پرونده دکتری است. مدرک تحصیلی با گروه علوم کامپیوتر و مهندسی اطلاعات دانشگاه ملی تایوان و همچنین دستیار پژوهشی در مرکز تحقیقات نوآوری فناوری اطلاعات، آکادمی سینیکا، تایوان است. علایق تحقیقاتی او شامل پردازش گفتار، تقویت گفتار، یادگیری ماشین و یادگیری عمیق است.

Pei-ChunLi مدرک مهندسی برق از دانشگاه ملی تایوان، در سال ۱۹۹۲و دکتری مدرک مهندسی زیست پزشکی از دانشگاه ملی یانگ مینگ، تایوان در سال ۲۰۰۶. از سال ۲۰۰۹ تا ۲۰۱۲ او عضو هیئت مدیره و مدیر ارشد بود، او درحال حاضر استادیار گروه شنوایی شناسی و آسیب شناسی گفتار و زبان، کالج پزشکی تایوان است. علایق تحقیقاتی او بر فناوری‌های گوش دادن کمکی، روش‌های اندازه‌گیری الکتروآکوستیک و شنوایی شناسی از راه دور متمرکز است.

Ying-Hui Lai مدرک تحصیلی از دپارتمان آموزش صنعتی، دانشگاه ملی تایوان در سال۲۰۰۵ و دکتری در سال۲۰۱۳ از گروه مهندسی پزشکی، دانشگاه ملی یانگ مینگ فارغ التحصیل شد. از سال ۲۰۱۳ تا ۲۰۱۶ پژوهشگر فوق دکتری در مرکز تحقیقات نوآوری فناوری اطلاعات، آکادمی سینیکا بوده است. او در حال حاضر استادیار گروه مهندسی برق است. علایق جستجوی او بر سمعک، کاشت حلزون، نویز تمرکز دارد.

Li-Chun Hsieh مدرک تحصیلی در موسسه علوم مغز دانشگاه ملی یانگ مینگ تایوان در سال ۲۰۱۵ . او در حال حاضردستیار پروفسور در بخش شنوایی و آسیب شناسی زبان گفتار، تایوان است و همچنین پزشک حاضر در بخش گوش و حلق و بینی است.

Yu Tsao مدرک مهندسی برق از دانشگاه ملی تایوان، مدرک مهندسی برق و کامپیوتر از مؤسسه فناوری جورجیا، آتلانتا، ایالات متحده آمریکا در سال۲۰۰۸. از سال ۲۰۰۹ تا ۲۰۱۱، او در مؤسسه ملی فناوری اطلاعات و ارتباطات، کیوتو، ژاپن پژوهشگر بود و در آن‌جا مشغول تحقیق و توسعه محصول در زمینه تشخیص خودکار گفتار برای ترجمه گفتار به گفتار چندزبانه بود. او در حال حاضر همکار پژوهشی مرکز تحقیقات نوآوری فناوری اطلاعات تایوان است. علایق تحقیقاتی او شامل تشخیص گفتار، کدگذاری صوتی، شبکه‌های عصبی عمیق، سیگنال‌های زیستی و مدل سازی آکوستیک است.

فصل دوم

۲-۱ گیت هاب[1]

یکی از بزرگترین انجمن‌های توسعه دهندگان وب در جهان گیت‌هاب است. در واقع گیت‌هاب پلتفرمی است که در آن توسعه دهنگان وب از سراسر جهان در آن گرد هم آمده و با یکدیگر ارتباط و همکاری دارند. در گیت‌هاب شما به عنوان توسعه دهنده وب می‌توانید پروژه‌های خود را با همکارانتان یا هر فرد دیگری که مایل باشید به اشتراک بگذارید و به صورت مشترک روی یک پروژه کار کنید. به این ترتیب به سادگی می‌توانید نسخه‌های قبلی یک نرم افزار را ارتقا دهید بدون این که تغییر یا اختلالی در نسخه‌های فعلی ایجاد شود.

githubکار کردن روی کدها را بسیار ساده کرده است. به کمک این پلتفرم می‌توانید به کوتاه‌ترین و ناپیدا ترین خط کد خود دسترسی پیدا کنید و در صورت لزوم آن را تغییر دهید. اما جذاب‌ترین ویژگی گیت‌هاب این است که به کمک آن می‌توانید با سایر کدنویسان در جهان ارتباط برقرار کنید. تیم بسازید و به‌طور مشترک روی پروژه‌های مختلف کار کنید.

مزایا گیت‌هاب

مزایای گیت‌هاب بسیار زیاد و دلایلی که به خاطر آن از این پلتفرم استفاده می‌کنیم برای هر کدنویسی متفاوت است. اما اولین دلیلی که کد نویسان جهان را مجبور می‌کند به گیت هاب بپیوندند این است که در آن امکان همکاری نرم وجود دارد. همچنین امکان تست و کنترل نسخه دلیل دیگری است که github را برای کد نویسان جذاب کرده است. مزیت دیگر گیت هاب این است که امکان یادگیری مباحث جدید و زبان‌های برنامه نویسی تازه در آن فراهم است. این ویژگی که افراد قادرند نسخه خود را با هر کسی که تمایل دارند به اشتراک بگذارند تا مورد بررسی و تحلیل واقع شده و اگر اشکالی در آن وجود دارد رفع شود، جزو جذابیت‌های غیر قابل انکار گیت‌هاب است. در حال حاضر بسیاری از تیم‌های کد نویسی یا شرکت‌هایی که به‌طور تخصصی در این زمینه کار می‌کنند عضو github هستند و در این پلتفرم پروژه‌های خود را پیش می‌برند.

اصطلاحات رایج در گیت‌هاب

- Repository
- fork
- Pull Request
- commit

[1] GitHub

Repository یا به اختصار Repo به معنای مخزن است. مخزن گیت‌هاب محیطی برای ذخیره سازی پروژه‌های توسعه دهندگان است. در این مخزن می‌توان هر فولدر یا فایلی را با فرمت دلخواه ایجاد کرد.

fork در فارسی به معنای شاخه یا انشعاب است. با این قابلیت شما می‌توانید روی پروژه‌های متن باز موجود در گیت‌هاب کار کنید. اگر پروژه‌ای از قبل وجود داشته باشد، می‌توانید از آن یک انشعاب دریافت و تغییراتی را روی آن اعمال کنید. سپس آن را به عنوان یک پروژه جدید منتشر کنید.

Pull Request یا درخواست ادغام، قلب تپنده‌ی مشارکت در پروژه‌هاست. زمانی استفاده می‌شود که شما از پروژه‌ی اصلی یک شاخه دریافت و درآن تغییراتی اعمال کرده‌اید. حالا با کمک Pull Request می‌توانید به شخص اصلی ایجاد کننده‌ی پروژه، درخواست بدهید تغییرات شما را در پروژه‌ی اصلی اعمال کند.

به هر تغییری در گیت‌هاب یک commit می‌گویند.

مهم‌ترین رقبای github چه چیزهایی هستند؟

گیت‌هاب یکی از بهترین ابزارهای میزبانی کد است که به طور گسترده برای کنترل نسخه مورد استفاده قرار می‌گیرد. همان‌طور که پیش‌ازاین هم بیان کردیم، این سرویس امکان کار بر روی چندین پروژه را به طور هم‌زمان ایجاد می‌کند. بااین‌حال برخی افراد سایر پلتفرم‌ها را ترجیح می‌دهند و معتقدند که کار با پلتفرم‌های رقیب github امکانات بهتری را در اختیار آن‌ها قرار می‌دهد. در ادامه می‌توانید برخی از رقبای گیت هاب را مشاهده کنید:

- AWS CodeCommit
- Beanstalk
- Bitbucket
- Gitbucket
- Gogs
- SourceForge
- TaraVault
- و...

۲-۲ نرم افزار متلب

نرم افزار متلب یک زبان فوق العاده قوی برای محاسبات فنی در رشته های مختلف مهندسی است. نرم افزاری که مسائل و راه حل‌های آنها را با بهره‌گیری از ترکیب کدنویسی، تصویر و محاسبات به راحت‌ترین شکل ممکن ارائه می‌کند.

از نرم افزار متلب برای مقاصد مختلفی استفاده می‌شود:

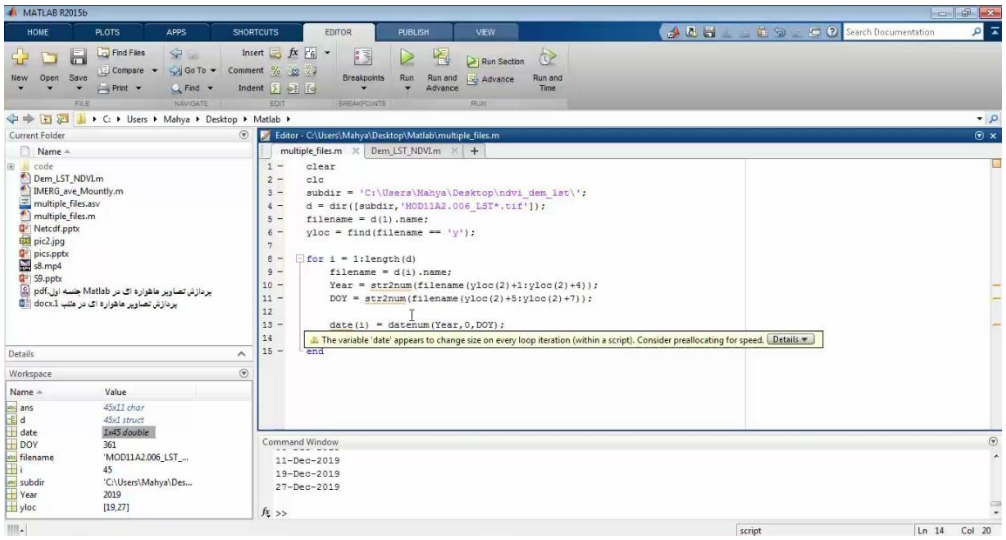
- ریاضی و محاسبات
- توسعه الگوریتم
- مدلسازی و شبیه سازی
- تحلیل اطلاعات و تصویرسازی بصری (تولید نمودار و ...)
- تولید گراف یا نمودارهای علمی و مهندسی

در واقع متلب یک سیستم تعاملی است که عنصر اولیه دیتای آن یک آرایه[1] است که نیازی به اندازه‌گذاری ندارد. این خاصیت به شما امکان حل بسیاری از مسائل فنی محاسباتی را می‌دهد به‌خصوص فرمولاسیون‌های ماتریسی و وکتور. MATLAB -Matrix Laboratory دارای ابزارها (toolbox) های برنامه محوری برای حل مسائل است که برای بسیاری از کاربرانش اهمیت زیادی دارد و به آنها اجازه می دهد تا تکنولوژی به‌خصوصی را یاد بگیرند و اعمال کنند.

برنامه متلب از ۵ بخش عمده تشکیل شده است:

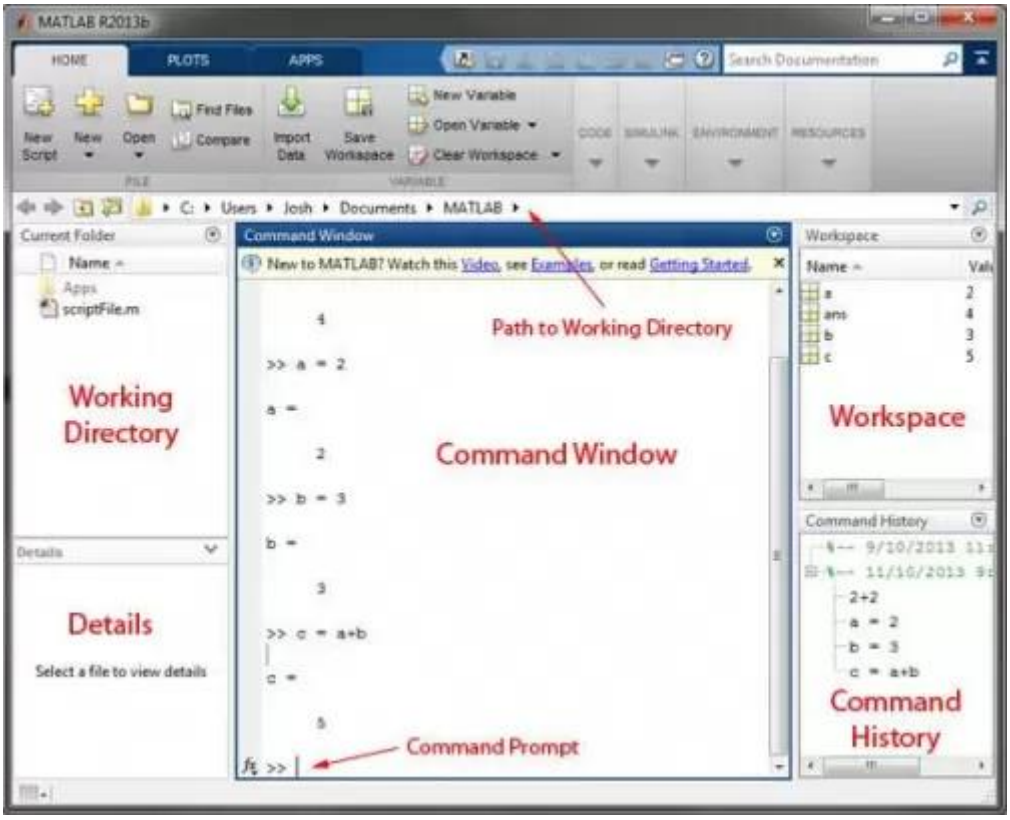
زبان متلب[2] که یک زبان ماتریسی/آرایه ای سطح بالاست و ویژگی های متنوعی دارد:

[1] Array
[2] MATLAB language



شکل۲-۱: زبان متلب

محیط متلب[1] که مجموعه‌ای از ابزارها و امکاناتی است که به شما امکان کار کردن با برنامه را می‌دهد:



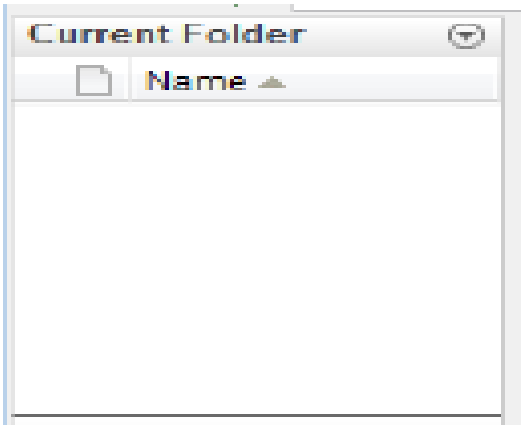
شکل۲-۲: محیط متلب

[1] MATLAB working environment

پنجره ی دسکتاپ متلب شامل پنل های زیر است:

Current Folder

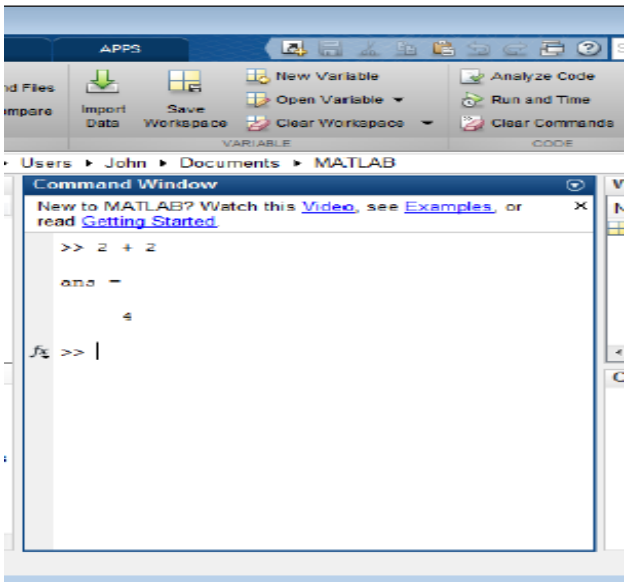
این پنل به شما اجازه‌ی دسترسی به فولدرها و فایل‌های ایجاد شده در متلب را می‌دهد.



شکل۲-۳: محیط current folder

Command Window

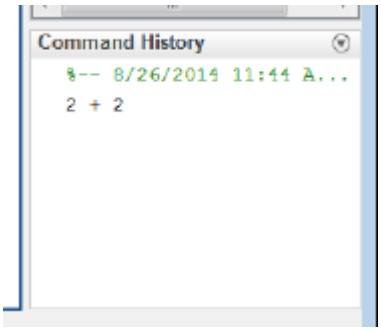
این محیط اصلی برنامه است که می‌توان کدهای متلب را در آن تایپ نموده و اجرا کرد.



شکل۲-۴: محیط command window

Command History

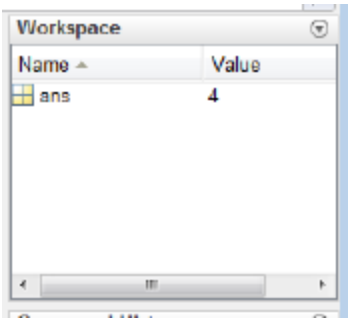
این پنجره ی کدهای اخیری که در محیط command line وارد شده را نمایش می‌دهد.



شکل۲-۵: محیط Command History

Workspace

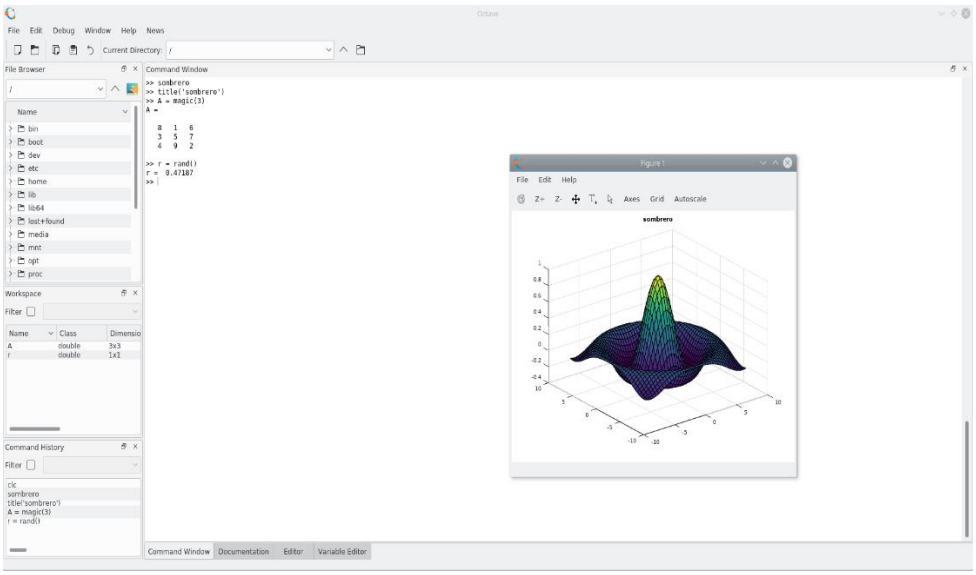
workspace تمام متغیرهایی که در متلب ایجاد شده و یا از فایل های دیگری وارد شده را نمایش می‌دهد.



شکل۲-۶: محیط Workspace

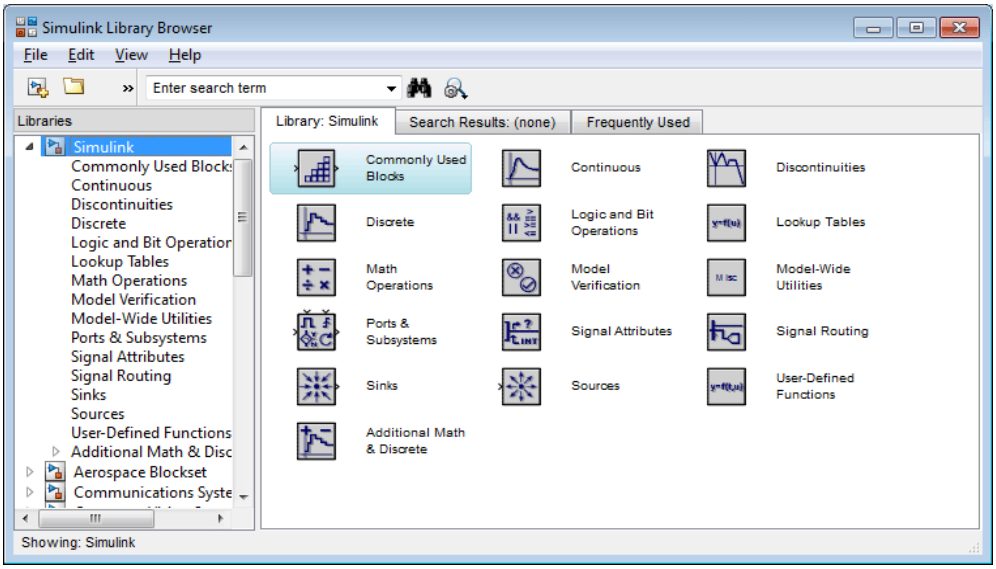
گرافیک‌های کمک رسان[1] که در واقع همان سیستم گرافیک متلب است و دستورات سطح بالایی برای تصویرسازی اطلاعات، بصورت دو بعدی و سه بعدی دارد. پردازش تصویر، انیمیشن و ارائه گرافیکی

[1] Handle graphics



شکل۲-۷: سیستم گرافیک متلب

بخش دستورات ریاضی متلب[1] یک مجموعه ای از الگوریتم های پردازشی و محاسباتی، از عملکردهای اولیه مثل جمع و تفریق تا الگوریتم های پیچیده تری مثل ماتریس معکوس و...



شکل۲-۸: دستورات ریاضی متلب

[1] MATLAB mathematical function library

رابط تصویری برنامه[1] به شما امکان نوشتن برنامه C و فورترن[2] را می‌دهد. در حالیکه متلب مزیت‌های بی شماری نسبت به زبان های مرسوم دیگر مثل فورترن و زبان سی دارد.

مراحل نصب متلب در ویندوز

۱ -ابتدا فایل نصب متلب را از مراجع معتبر دانلود می‌کنیم.

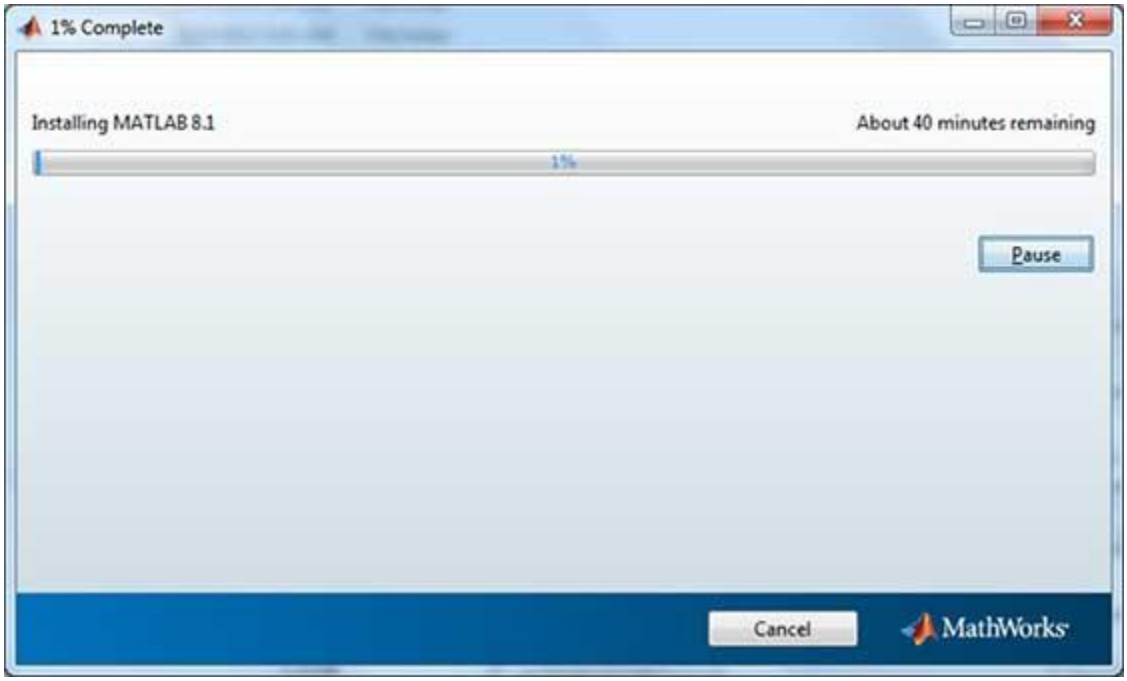
۲ -پس از دانلود روی فایل نصب نرم افزار دوبار کلیک می‌کنیم تا باز شود.

۳ -در اولین پنجره گزینه‌ی Install witout using the Internet را انتخاب کرده و روی Next کلیک می‌کنیم.

شکل۲-۹: نصب متلب

[1] MATLAB application program Interface -API
[2] Fortran

۴- فرآیند کپی و نصب فایل های متلب شروع شده و این فرآیند ممکن است چند دقیقه ای طول بکشد.



شکل۲-۱۰: مرحله نصب متلب

محیط متلب را می‌توان از آیکون ایجاد شده در دسکتاپ ویندوز پس از نصب متلب اجرا کرد.

۲-۳ سورس کد[1]

به طور کلی برنامه ها به دو دسته متن باز[2] و متن بسته[3] تقسیم می شوند. ما تنها می‌توانیم اقدام به مشاهده سورس کد برنامه های متن باز که قالبا هم رایگان می‌باشند نمایید و حتی در آن‌ها تغییرات ایجاد نمایید. معمولاً تمامی پروژه‌ها و برنامه‌های متن باز دنیا درون سایت گیت‌هاب ثبت می‌شوند.

سورس‌کد مهم‌ترین و اساسی‌ترین بخش یک برنامه کامپیوتری است. در واقع خود برنامه است که در انتهای پروژه توسط برنامه نویس در قالب یک پکیج تکمیل می‌شود. سورس کدها توسط برنامه نویس نوشته می‌شود و می‌توان نام آن را نقشه راه آن برنامه گذاشت، این نقشه راه به برنامه نویس کمک می‌کند تا بتواند خیلی سریع و روان به اتفاقاتی که درون برنامه می‌افتد چیره شود.

[1] Source Code
[2] open source
[3] closed source

روش دانلود سورس کد از گیت‌هاب

ما می‌توانیم سورس کد را به صورت فایل zip از گیت‌هاب دانلود کنیم. برای این‌کار ابتدا وارد موضوع پروژه می‌شویم در قسمت code گزینه Download zip را می‌زنیم و سورس کدهای ما دانلود می‌شوند.

۲-۴ ویژوال استودیو کد[1]

یک نرم افزار ویرایشگر کد است که به صورت متن باز برای لینوکس و ویندوز و OS10 می باشد.

VS Code یک ادیتور متن باز است که رایگان از گیت هاب قابل دریافت است و روی سیستم عامل‌های مختلف از جمله ویندوز، مک و لینوکس نصب می شود و زبان‌های برنامه نویسی مختلفی از جمله پایتون، سی پلاس پلاس، جاوا و ... را پشتیبانی می کند. حجم کمتر، برخورداری از یک مخزن بزرگ از افزونه‌ها آن را رقیب جدی برای دیگر ویرایشگرها قرار داده است.

این نرم‌افزار توسط مایکروسافت توسعه داده شده و هم‌اکنون به‌طور رایگان و اپن سورس در دسترس است.

در نظرسنجی سال ۲۰۱۸ وب سایت Stack Overflow ، ویژوال استودیو کد به عنوان محبوب‌ترین ابزار توسعه با رای ۳۴,۹ درصد از ۷۵،۳۹۸ رای انتخاب شد.

ویژگی های VS Code

برخی قابلیت‌های فوق‌العاده این ادیتور شامل موارد زیر است:

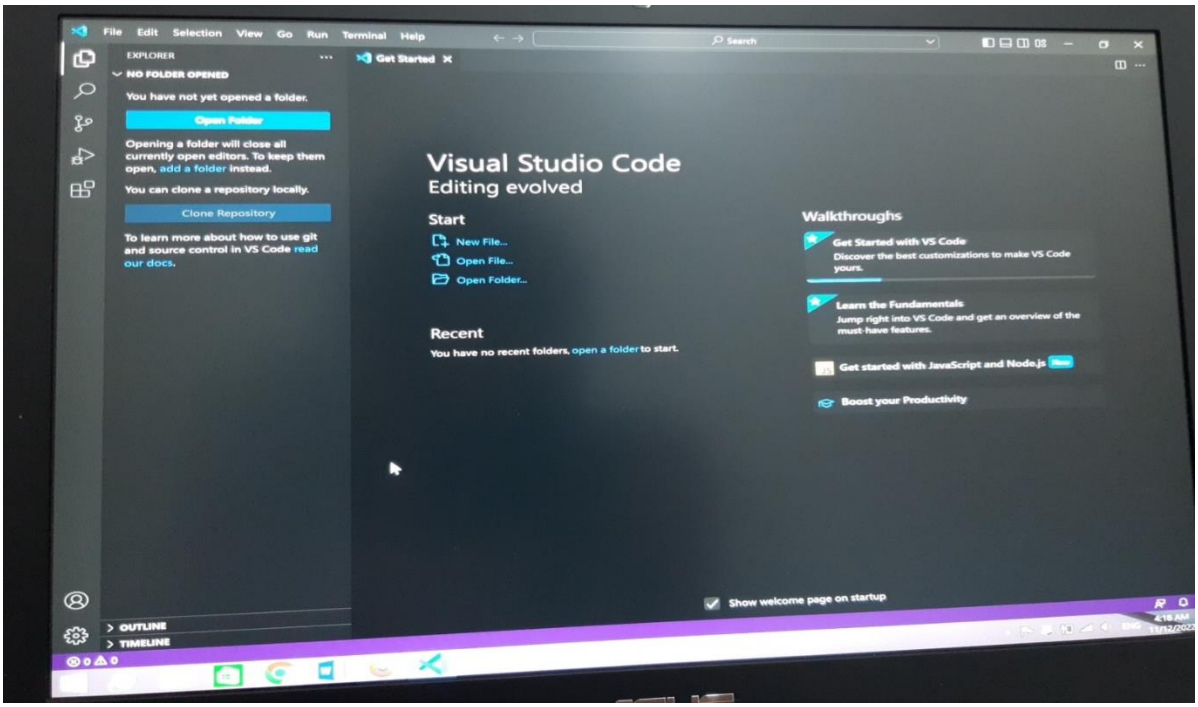
- Syntax highlighting
- Intelligent code completition
- Snippets
- Code refractoring
- Themes
- Extensibility
- Git integration
- ادیتور Visual Studio Code از یونیکد پشتیبانی کرده و قابلیت تایپ فارسی در آن فراهم است.

[1] Visual Studio Code(VS Code)

- VS codeتمام قابلیت‌های یک ادیتور امروزی و مدرن را دارد و استفاده از آن آسان است.

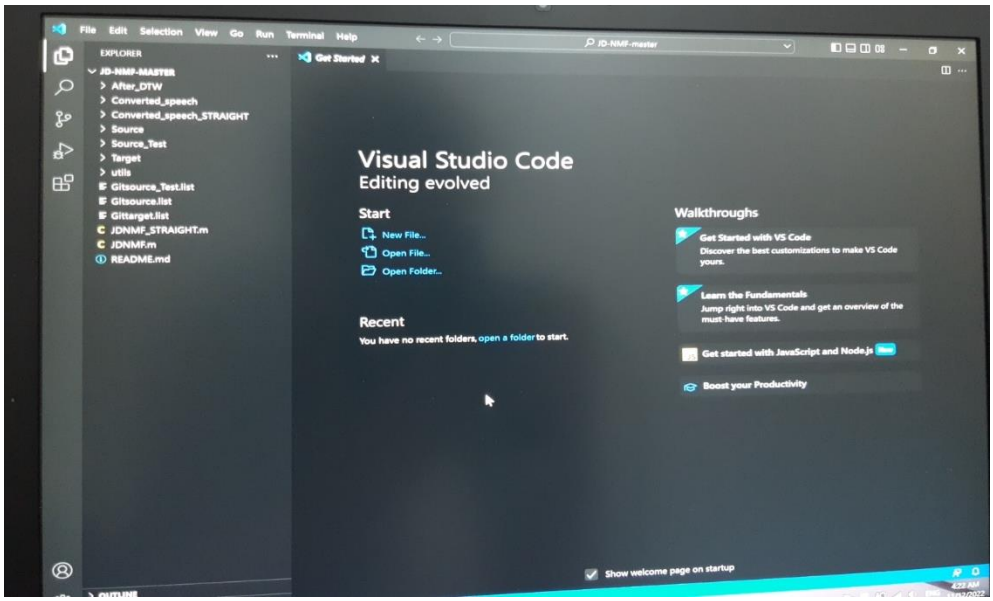
۲-۵ اجرای پروژه روی وی اس کد

در ابتدا برنامه‌های GitHubDesktop و visuall studio code را نصب می‌کنیم. فایل‌های زیپ مربوط به پروژه را دانلود کرده و از حالت زیپ خارج می‌کنیم. سپس وارد محیط VSCode می‌شویم.



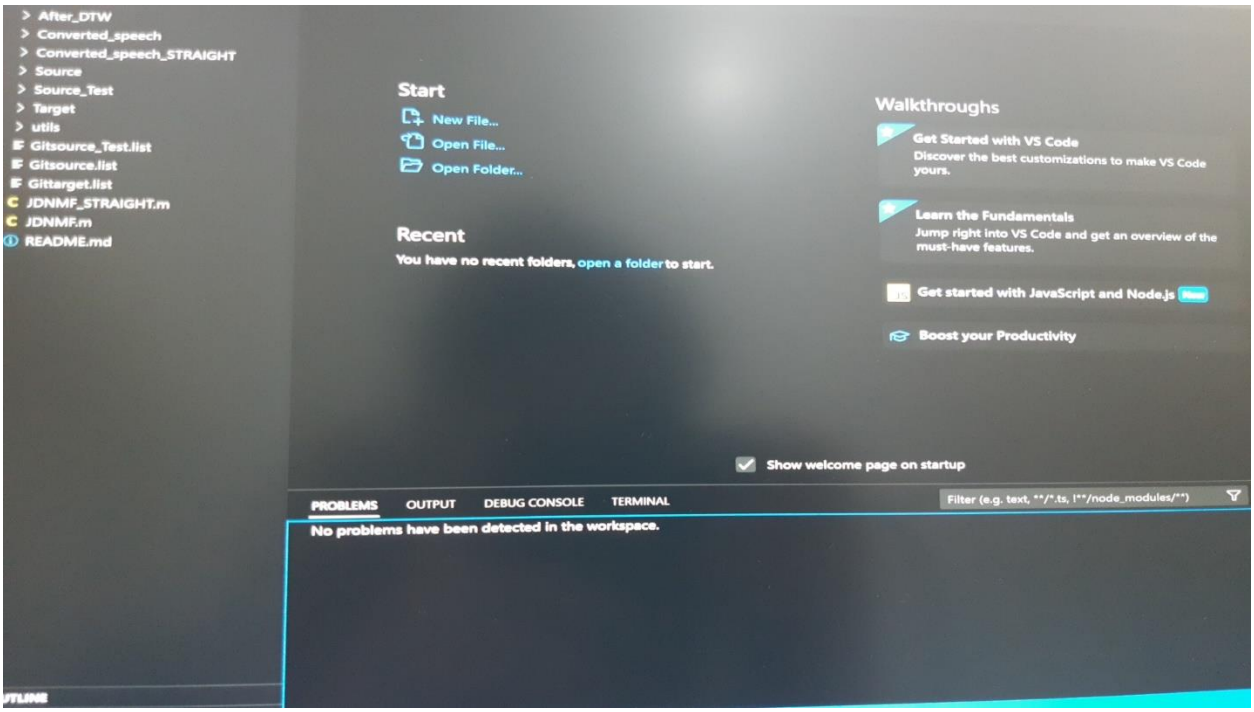
شکل ۲-۱۱: نمایی از VS Code

سپس از روی گزینهopen folder فایل مربوط به پروژه را پیدا کرده و باز می‌کنیم.



شکل ۲-۱۲: باز کردن پروژه در محیط VS Code

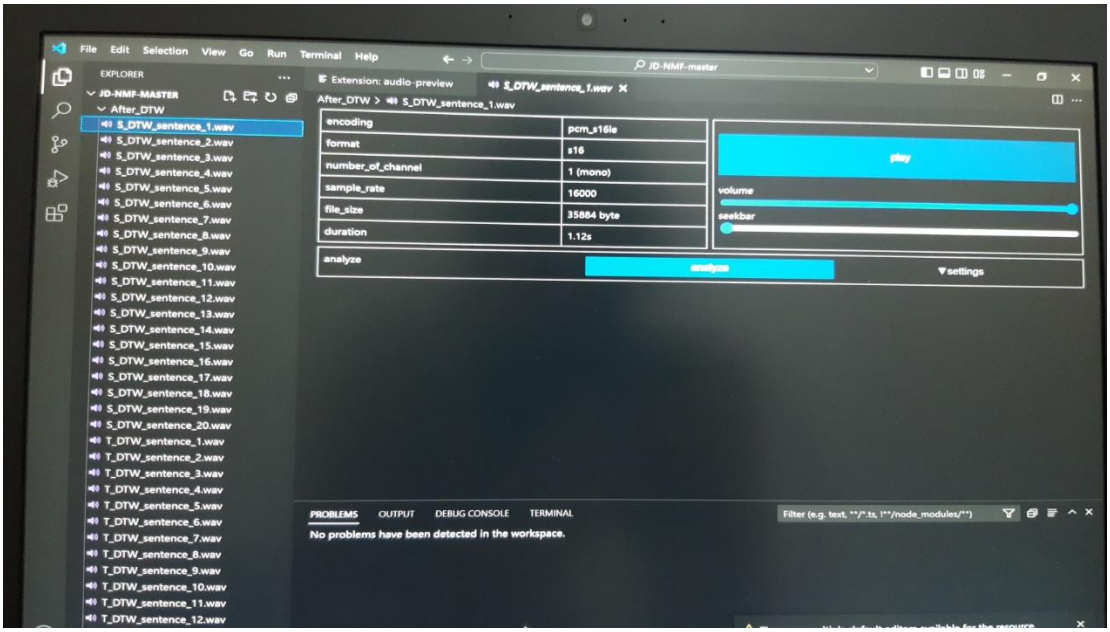
در نوار ابزار بالا یک سری گزینه داریم که شامل file, Eddit,… هستند. ما روی گزینه View عبارت، problem رو انتخاب می‌کنیم.



شکل ۲-۱۳: عدم وجود خطا در ویس ها

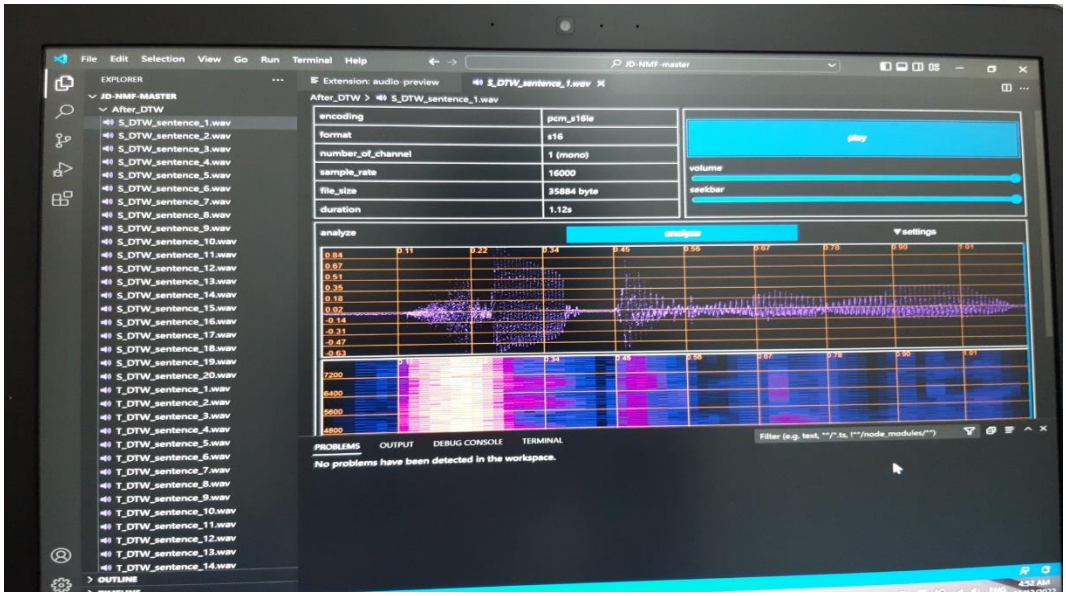
در ادامه برای اجرای فایل‌های ویسی نیاز داریم اکسشنی را نصب کنیم. Audio-preview را نصب می‌کنیم.

بعد از نصب برنامه دوباره به پروژه برمی‌گردیم و فایل‌های ویسی را باز می‌کنیم باتوجه به نصب برنامه مورد نیاز حالا ویس‌ها بدون مشکل پخش می‌شوند.



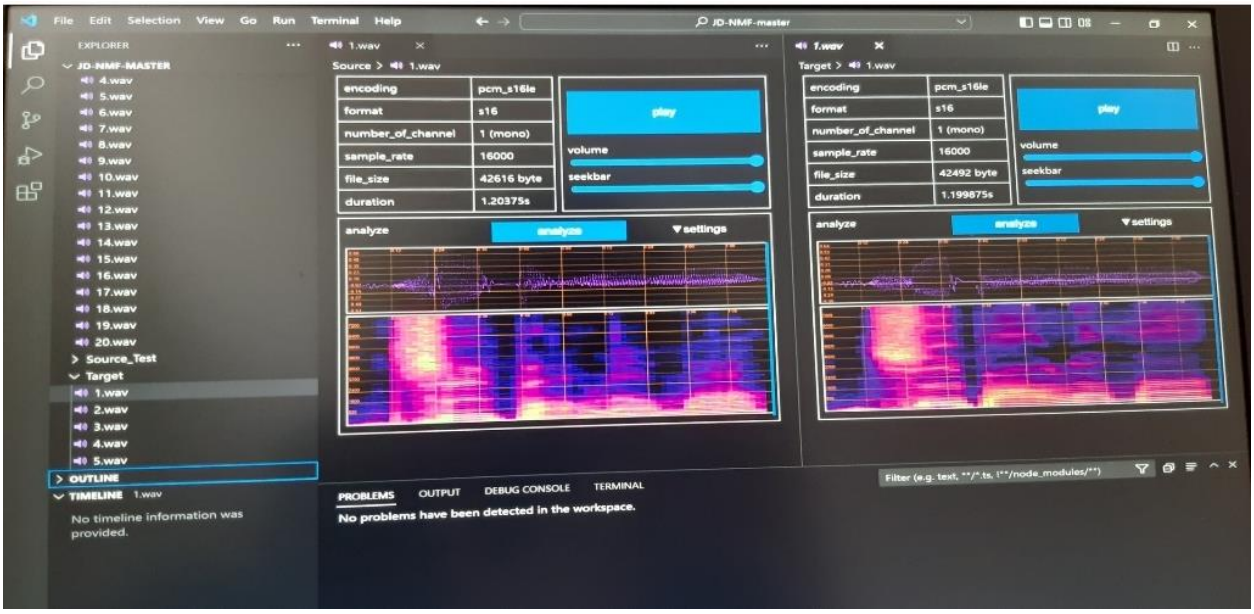
شکل ۲-۱۴: باز شدن فایل‌های ویسی بدون خطا

گزینه آنالیز را انتخاب کرده سپس گزینه play را می‌زنیم.



شکل ۲-۱۵: نمودار آنالیزفایل ویسی

ما در شکل۲-۱۵ علاوه بر اینکه صدا رو داریم دو نمودار نیز داریم که یکی اسپکتوگرام است و طیف گرام را نشان میدهد(نمودار پایین) و در نمودار بالا طیف موجی را داریم. در نمودار موجی یک سری فاصله زمانی وجود دارد که نشان دهنده نقاط سکوت در صدا است و این نقاط سکوت در اسپکتوگرام نیز قابل مشاهده است.



شکل ۲-۱۶: مقایسه ویس ورودی و خروجی

زمانیکه موج خروجی را می‌خواهیم مشاهده کنیم متوجه می‌شویم که این قسمت‌های سکوت کامل حذف شده‌اند و صدای ما سریع‌تر پخش می‌شود. طول موجی که ورودی ما داشته ۱,۲۰ ثانیه بوده اما طول موج خروجی که داریم طول موج ۱,۱۹ ثانیه شده است یعنی بخش‌هایی که اطلاعات مفیدی نداشته را حذف کرده است. این برنامه کمک می‌کند که متوجه شویم چه تغیراتی روی سیگنال اتفاق می‌افتد.

فصل سوم

۳-۱ بررسی مقاله‌های مرتبط

۳-۱-۱ بهبود کیفیت سیگنال گفتار با استفاده از شبکه عصبی عمیق برای کاهش نویز

هدف این پایان‌نامه، بهبود کیفیت سیگنال گفتار با استفاده از شبکه عصبی عمیق برای کاهش نویز می‌باشد. در بیشتر محیط‌های زندگی نویزهای مختلفی حضور دارد که داده‌های صوتی را تخریب می‌کند. یکی از مباحث مهم در پردازش سیگنال ، حذف سیگنال‌های ناخواسته و یا نویز از سیگنال اصلی است. در این پژوهش کاربردی یک فیلتر وفقی براساس روش تفریق طیفی ضرایب موجک (WSS) و شبکه عصبی عمیق برای کاهش نویز پیشنهاد شده است. این مجموعه ترکیبی از تبدیل ویولت، یادگیری وفقی و نگاشت غیرخطی از شبکه‌های عصبی عمیق است. شبکه عصبی عمیق به کمک فیلتر وفقی برای کاهش نویز بیشتر از سیگنال گفتار مورد استفاده قرار می‌گیرد. نتایج پیاده‌سازی این پایان‌نامه بر روی سیگنال‌های مختلف با نسبت سیگنال به نویزهای (SNR) متفاوت بیان و در ادامه کارایی روش پیشنهادی در حضور سیگنال‌های نویزی با نویز غیر ایستان (همهمه) نیز انجام شده است. در این پایان‌نامه برای مقایسه تأثیر این روش‌ها بر روی گفتار از سه مبنای مقایسه‌ای SNR و نسبت سیگنال به نویز قطعه‌ای (SEGSNR) و اندازه‌گیری ارزیابی ادراکی کیفیت گفتار (PESQ) کمک گرفته شده است. نتایج نهایی با استفاده از معیارهای ارزیابی متعددی بررسی و بیانگر عملکرد رضایت‌بخش روش پیشنهادی می‌باشد.

از تبدیل فوریه به تبدیل موجک

تبدیل فوریه از طریق ضرب کردن سیگنال مورد پردازش در قطاری از سیگنال‌های سینوسی با فرکانس‌های مختلف عمل می‌کند. در واقع، از این راه می‌توانیم تعیین کنیم که کدام فرکانس‌ها در سیگنال مورد پردازش وجود دارند. اگر عملگر ضرب نقطه‌ای بین سیگنال مورد نظر و یک سیگنال سینوسی با فرکانس مشخص، برابر با یک عدد با دامنه بزرگ شود، آن‌گاه می‌توان نتیجه گرفت که هم‌پوشانی زیادی بین این دو سیگنال وجود دارد و در نتیجه آن فرکانس مشخص در طیف فرکانسی سیگنال مورد نظر نیز مشاهده خواهد شد. قطعا دلیل این امر از آنجایی ناشی می‌شود که عملگر ضرب نقطه‌ای معیاری برای اندازه‌گیری میزان هم‌پوشانی و شباهت بین دو بردار یا دو سیگنال است.

نکته‌ای که در مورد تبدیل فوریه می‌توان به آن اشاره کرد این است که در حوزه فرکانس دارای رزولوشن بالایی است، در حالی که در حوزه زمان از رزولوشن صفر برخوردار است. به عبارت دیگر، تبدیل فوریه این توانایی را دارد

که به ما بگوید دقیقا چه فرکانس‌هایی در یک سیگنال وجود دارند، اما نمی‌توان با استفاده از آن تعیین کرد که فرکانس مورد نظر در چه لحظه‌ای از زمان در سیگنال اتفاق می‌افتد.

در این تصویر سیگنال‌های اصلی به همراه طیف فرکانسی هر کدام نشان داده شده است.

۳-۱-۲ بهبود قابلیت فهم گفتار در افراد با شنوایی عادی با استفاده از ماسک های شنیداری

بهبود گفتار در بسیاری ازکاربردها مانند بازشناسی گفتار، ارتباطات موبایل و وسایل کمک شنوایی از اهمیت بالایی برخوردار است. اگرچه پیشرفت‌های زیادی در توسعه الگوریتم‌های بهبود کیفیت گفتار حاصل شده است، طراحی الگوریتم‌های بهبود قابلیت فهم گفتار پیشرفت کمتری داشته است. این درحالی است که سیستم شنوایی انسان به خوبی قادر است در شرایط محیطی مختلف، مانند شرایط نویزی، اصوات گوناگون را از هم تشخیص دهد. سیستم‌های آنالیز محاسباتی محیط شنیداری (CASA) در سالهای اخیر به مدل‌سازی این ویژگی شنوایی انسان پرداخته اند. در این پایان‌نامه، یک سیستم بهبود تحت نظارت قابلیت فهم تک گوشی گفتار مبتنی بر ماسک‌های شنوایی برای شنوندگان عادی ارائه شده است. سیستم پیشنهادی ابتدا، از روش ادغام ویژگی آنالیز مولفه های کانونی (CCA) برای ادغام ویژگی‌ها استفاده می‌کند. سپس، ماسک پیشنهادی ادغامی بهینه شده که ترکیبی از دو ماسک شنوایی حوزه گاماتون و بارک است، توسط شبکه عصبی عمیق (DNN) تخمین زده می شود. در انتها، ماسک تخمینی برای ساختن گفتار به کار می رود. شبیه-سازی ها نشان می دهند که قابلیت فهم گفتارهای پردازش شده توسط سیستم پیشنهادی در حضور نویزغیرایستان دیده نشده به طور قابل توجهی بهبود می یابد. همچنین، مقایسه نتایج با یک سیستم مبنا برتری عملکرد سیستم پیشنهادی را نسبت به یک سیستم اخیر نشان می دهد.

همبستگی کانونی

همبستگی کانونی شبیه رگرسیون چند متغیری است، به این معنا که در این روش ترکیبی از متغیرهای پیش بینی کننده به منظور پیش بینی متغیر ملاک به کار برده می‌شود، تفاوت این دو روش در تعداد متغیرهای ملاک است. در رگرسیون چند متغیری فقط یک متغیر ملاک وجود دارد، در صورتی که همبستگی کانونی بیش از یک متغیر ملاک دارد.

تحلیل همبستگی کانونی روی همبستگی بین یک ترکیب خطی از متغیرهای یک مجموعه و یک ترکیب خطی از متغیرهای مجموعه دیگر متمرکز می‌شود. ابتدا هدف ما این است که دو ترکیب خطی با بیشترین همبستگی را

تعیین کنیم سپس دو ترکیب خطی را تعیین می‌کنیم که در میان تمام زوج های ناهمبسته با زوج انتخاب شده اول دارای بیشترین همبستگی باشد و این فرآیند را ادامه می‌دهیم.

مثالی برای همبستگی کانونی؟ فرض کنید متغیرهای پیش بینی کننده ای مانند خانواده، میانگین نمره، علائق شغلی و تیپ شخصیتی و متغیرهای ملاکی مانند مدت فراغت از تحصیل، درآمد سالانه، پرسش های فیزیولوژیکی و روانی و میزان مشارکت در دست داریم. می خواهیم ببینیم کدام دسته از متغیرهای پیش بینی کننده، بهتر از دسته دیگر متغیرهای ملاک را پیش بینی می کنند.

۳-۱-۳ جداسازی سیگنال صحبت بر پایه ICA برای بهبود کیفیت گفتار

تاکنون انواع مختلفی از تکنیک‌های بهبود گفتار مورد مطالعه قرار گرفته‌اند. از آنجایی که نویزهای متنوعی در محیط وجود دارند، هیچ یک از تکنیک‌های بهبود گفتار برای حذف همه انواع نویز مناسب نیستند. علاوه بر نویز پس زمینه در محیط، وجود سیگنال‌های تداخلی صحبت و همچنین انعکاس-های محیط، مسئله بهبود گفتار را پیچیده‌تر می‌کند و لزوم الگوریتم‌های حذف پژواک و تفکیک منابع را برای این منظور فراهم می‌آورد. اخیراً جداسازی کور منبع برای مخلوط‌های کانولوتیو در حوزه فرکانس به عنوان روشی برای تفکیک منابع صوتی معرفی شده است. در این روش از یکی از الگوریتم‌های ICA همچون Infomax به طور جداگانه در هر فرکانس، برای جداسازی مولفه‌های فرکانسی منابع استفاده می‌شود. الگوریتم Infomax با اندازه گام ثابت از نظر همگرایی و پایداری دارای معایبی است. اگر اندازه گام کوچک انتخاب شود، سرعت همگرایی کاهش می یابد و اگر بزرگ انتخاب شود، ممکن است باعث ناپایداری الگوریتم شود. در بخشی از این پایان‌نامه روشی بر پایه تکنیک PSO برای تعیین اندازه گام مناسب در الگوریتم Infomax پیشنهاد می‌کنیم که موجب همگرایی بیشتر الگوریتم می‌شود. از سویی پس از جداسازی مولفه‌های فرکانسی منابع در هر فرکانس، برای بازسازی صحیح سیگنال‌ها از روی مولفه‌های فرکانسی در حوزه زمان باید مولفه‌های فرکانسی مربوط به هر یک از منابع دسته‌بندی شوند. این مسئله، جایگشت در حوزه فرکانس نام دارد و روش‌های متعددی برای حل آن وجود دارد. در بخش دیگری از این پایان‌نامه روشی برای حل مسئله جایگشت با استفاده از همبستگی نسبت توان (power ratio) مولفه‌های فرکانسی در حالت overdetermind، یعنی وقتی که تعداد میکروفون‌ها بیشتر از تعداد منابع باشد، پیشنهاد می‌شود.

تاریخچه الگوریتم PSO

روش PSO ریشه در کارهای Reynolds دارد که یک شبیه سازی ابتدایی از رفتار اجتماعی پرندگان است . در این مدل رفتارهای ساده پیدا کردن نزدیک ترین همسایه ها تنظیم سرعت های پیاده شده است. این مدل برندگان به صورت تصادفی در یک فضای جستجوی جدول پیکسلی قرار داده می شوند و در هر تکرار نزدیکترین همسایه ذره انتخاب شده و سرعت نره با سرعت نزدیکترین همسایه اش جایگزین می شود. این عمل باعث می شود که گروه خیلی سریع به یک جهت حرکت نامعین و بدون تغییر همگرا شوند. جهت رفع این مشکل یک مولفه دیوانگی به صورت تغییر تصادفی در گروه ها استفاده شده است. به منظور توسعه بیشتر این مدل مفهوم سردسته پرندگان نیز به مدل اضافه گردید که به شکل یک حافظه از بهترین موقعیت های هر عضو و همسایگان آن بود . بهترین موقعیت قبلی هر عضو بهترین موقعیتی است که آن عضو از ابتدای حیات خود تا به حال کسب نموده است. بهترین موقعیت همسایگی بهترین موقعیتی است که توسط همسایگان یک عضو ملاقات شده است. این دو بهترین موقعیت به عنوان نقاط جذب عمل می نمایند. با استفاده از یک مجموعه قوانین ساده می توان موقعیت های اعضای گروه را به روز نمود . بدین صورت که عضو به یک نسبت به سمت دو موقعیت بهتر حرکت می نماید . به مرور زمان با تکرار الگوریتم اعضا حول یک هدف جمع می شوند. این رفتار که حتی بدون هماهنگی سرعت ها و فاکتور دیوانگی نتیجه بخش بود . مدل نهایی بهینه سازی گروه ذرات نامیده می شود.

ویژگی های الگوریتم PSO

- محاسبات فضای چند بعدی به صورت یکسری از گام های زمانی انجام می شود که به اصل پوشش معروف است.
- گروه ذرات به فاکتورهای کیفی به صورت بهترین موقعیت های فردی و همسایگی جواب میدهد.
- تخصیص پاسخ ها بین بهترین موقعیت ملاقات شده ذره و بهترین موقعیت ملاقات شده توسط گروه ، تنوع پاسخ ها را تضمین می نماید.
- گروه حالت خود را فقط هنگامی که بهترین موقعیت ملاقات شده توسط ذره و بهترین موقعیت ملاقات شده توسط گروه تغییر می کنند ، تغییر میدهد که به اصل پایداری معروف است.
- در نهایت گروه رفتار تطبیقی از خود نشان میدهد بدین صورت که حالت خود را هنگامی که بهترین موقعیت ملاقات شده توسط ذره و بهترین موقعیت ملاقات شده توسط گروه تغییر می کنند، تغییر می‌دهد.

مزایای الگوریتم ازدحام ذرات

PSO مزایای بسیاری نسبت به دیگر روش های بهینه سازی فرابتکاری دارد. از جمله:

- الگوریتم PSO یک الگوریتم مبتنی بر جمعیت است. این خاصیت باعث می شود که کمتر در مینیمم محلی گرفتار شود
- این الگوریتم براساس قوانین احتمالی عمل می کند نه قوانین قطعی. بنابراین، Pso یک الگوریتم بهینه سازی تصادفی است که می تواند نواحی نامشخص و پیچیده را جستجو کند. این خاصیت، PSO را نسبت به روشهای معمولی انعطاف پذیرتر و مقاومتر می کند.
- PSO با توابع هدف غیر دیفرانسیلی سروکار دارد بدلیل اینکه PSO از نتیجه اطلاعات (شاخص بازدهی یا تابع هدف استفاده می کند تا جستجو را در فضای مسئله هدایت کند.
- کیفیت جواب مسیر پیشنهادی به جمعیت اولیه وابسته نیست. با شروع از هر نقطه در فضای جستجو، الگوریتم جواب مسئله را نهایتا به جواب بهینه همگرا می کند.
- PSO انعطاف پذیری زیادی دارد تا تعادل بین جستجوی محلی و کلی از فضای جستجو را کنترل کند. این خاصیت منحصربفرد PSO به مشکل همگرایی بدموقع غلبه می کند و ظرفیت جستجو را افزایش می دهد که همه این خاصیت‌ها Pso را متفاوت از الگوریتم ژنتیک (GA)و دیگر الگوریتم‌های ابتکاری می کند.

۳-۱-۴ بهسازی سیگنال گفتار در حوزه زمان- فرکانس به روش تفریق طیفی اصلاح شده

هدف اصلی گفتار برقراری ارتباط بین انسان‌ها می‌باشد. زمانیکه سیگنال گفتار در یک محیط نویزی ثبت می‌شود، کیفیت و قابلیت فهم گفتار کاهش پیدا می‌کند. حذف نویز از سیگنال گفتار به منظور بهبود وضوح سیگنال گفتار صورت می‌گیرد. در این پایان نامه، یک روش تفریق طیفی چندباندی اصلاح شده به منظور بهبود سیگنال گفتار پیشنهاد می‌شود. میانگین‌گیری از فریم‌های سیگنال گفتار نویزی به منظور کاهش واریانس سیگنال به عنوان مرحله پیش پردازش اعمال می‌شود و از آنالیز ضرایب پیشگویی خطی (LPC) به منظور تخمین نویز اولیه و تخمین نویز مانده استفاده شده است. سپس روش تفریق طیفی تکرارشونده با سه مرحله تکرار جهت حذف نویز مانده و بهبود کیفیت سیگنال گفتار پردازش شده اعمال شده است. این واقعیت که تاثیر نویزهای رنگی روی طیف گفتار در فرکانس‌های مختلف متفاوت است نیز در نظر گرفته شده است. با بررسی نتایج روش پیشنهادی نشان

داده می‌شود که در بهبود سیگنال گفتار آغشته به نویزکارخانه در آزمون شنیداری با ۶ شنونده، روش پیشنهادی با دو مرحله تکرار تفریق طیفی اصلاح شده در مقایسه با تفریق طیفی چندباندی، به میزان ۰/۸۳ به لحاظ معیار MOS و به میزان ۲/۸ به لحاظ معیار SNR برتری دارد.

روش‌های طیفی (Spectral methods) کلاسی از تکنیک‌هایی هستند که در ریاضیات کاربردی و محاسبات علمی برای حل معادلات دیفرانسیل که به‌طور بالقوه درگیر استفاده از تبدیل سریع فوریه هستند استفاده می‌شوند. ایده این کار، نوشتن راه حلی از معادلات متفاوت به عنوان مجموع «توابع پایه» و سپس انتخاب ضریب در مجموع برای ارضای معادلات متفاوت تا حد ممکن است. این روش از جمله پردقت‌ترین شیوه‌های عددی برای حل معادلات دیفرانسیل با مشتقات جزئی می‌باشد.

تخمین طیفی (Spectral estimation) زمینه‌ای است در پردازش آماری سیگنال‌ها که به تخمین چگالی طیف توان (PSD) سیگنال‌های تصادفی با استفاده از دنباله‌ای از نمونه‌های زمانی آن‌ها می‌پردازد.

۳-۱-۵ استفاده از فیلتر کالمن برای بهبود سیگنال گفتار

یکی از موضوعات مهم پردازش سیگنال، بهبود گفتار است. بهبود گفتار بخشی از پردازش گفتار است که هدف آن افزایش قابلیت فهم و یا خوشایندی از یک سیگنال گفتار می باشد. رایج ترین رویکرد در بهبود گفتار، حذف نویز است، که در آن با تخمین مشخصه های نویز، اجزای نویز را از بین می برند و تنها سیگنال گفتار تمیز را حفظ می کنند. در یک دید کلی می توان سیستم های بهسازی گفتار را به دو نوع عمده سیستم تک کاناله و سیستم های چند کاناله تقسیم بندینمود .در این پایان نامه سیستم های تک کاناله مورد بررسی قرار گرفته است. در سیستم های تک کاناله بهسازی گفتار فقط یک میکروفون در دسترس بوده که سیگنال حاصل از آن، گفتار آغشته به نویز خواهد بود. در سیستم های مرسوم تک کاناله حذف نویز، با شناسایی بخش های سکوت گفتار، نویز موجود در این قسمت ها را استخراج نموده و به عنوان نویز کلی سیستم می شناسند. در میان روش های بهسازی گفتار، فیلتر کالمن یکی از موثرترین آن ها است که به دلیل برخورداری از توانایی بالا در حذف نویز و همچنین سرعت بالای الگوریتم از اهمیت بالایی برخوردار می باشد.مهمترین چالش در روشفیلتر کالمن نداشتن تخمین درست از پارامترهای نویز است.که به دلیل تخمین نادقیق، مقداری نویز باقی می‌ماند. در این پایان نامه، با استفاده از روش آنالیز LPC، به تخمین نویز از روی سیگنال نویزی پرداخته شده است. دیده شده است که پارامترهای نویز تخمینی به پارامترهای نویز واقعی نزدیک تر است و به همین دلیل سبب بهبود روش فلیتر کالمن شده است. این گفته‌ها با معیارهای ارزیابی کمی و کیفی، PESQو MOSتایید می‌شوند.

فیلتر کالمن (Kalman Filter) یک تخمین‌گر است که از تخمین حالت قبل و مشاهده فعلی برای محاسبه تخمین حالت فعلی استفاده می‌کند و یک ابزار بسیار قوی برای ترکیب اطلاعات در حضور نامعینی‌ها است. در برخی موارد، توانایی فیلتر کالمن برای استخراج اطلاعات دقیق خیره کننده است. فیلتر کالمن مدت‌هاست که به عنوان راه‌حل بهینه برای بسیاری از کارهای ردیابی و پیش‌بینی داده‌ها مورد استفاده قرار می‌گیرد.

فیلترهای EKF و UKF هر کدام برای مقاصد خاصی به کار برده می‌شود. در واقع، در کاربردهای زیادی می‌توان هر دو فیلتر را به طور موثر استفاده کرد و از تفاوت موجود بین آن‌ها چشم‌پوشی کرد. اما در سیستم‌های با میزان غیرخطی بالا، استفاده از UKF ارجحیت دارد. یکی دیگر از مزایای UKF این است که از نظر پیاده سازی نسبت به EKF راحت‌تر است که این ویژگی آن را به یک ابزار قدرتمند در مسایل فیلترسازی تبدیل کرده است.

۳-۱-۶ یک روش ترکیبی برای بهبود کیفیت صحبت

همزمان با رشد و توسعه ارتباطات و به ویژه ارتباطات مخابراتی ، نیاز به بهبود کیفیت سیگنال های گفتار بیش از پیش گردیده است . منظور از بهبود سیگنال گفتار، کاهش یا حذف نویز از سیگنال صحبت برای دستیابی به کیفیت مناسب به لحاظ شنیداری است. نویز از اصلی ترین عوامل محدود کننده در سیستم های مخابراتی است بنابراین حذف نویز چه از نظر تئوری و چه در کاربردهای عملی در مرکزیت پردازش سیسگنال و علوم مخابراتی قرار دارد. در این پایان نامه، ابتدا روش های مختلف بهسازی سیگنال های گفتار که تاکنون مورد استفاده قرار گرفته اند از قبیل روش‌های تفریق طیفی،تجزیه مقادیر منفرد(SVD)، بهسازی گفتار با استفاده از فیلتر وینر یا فیلتر کالمن، روش‌های استفاده کننده از آرایه میکروفن و ... معرفی شده است. روش‌های تفریق طیفی و SVD به دلیل پیاده سازی ساده‌تر و محاسبات کمتر، در سیستم‌هایی که حساس به زمان هستند بهتر عمل می کنند، لذا تمرکزمان روی این دو روش است . اساس کار تفریق طیفی به این صورت می باشد که نویز توسط یکی از الگوریتم های LPC، VADو Waveletتخمین زده می شود و نویز تخمینی از سیگنال آغشته به نویز تفریق می شود. امروزه بحث "تجزیه مقادیرتکین" یا SVD به عنوان یکی از قدرتمندترین ابزار برای تفکیک زیرفضاهای سیگنال و نویز، یاد شده است. در این پایان نامه از روشی نوین بمنظور حذف نویز از سیگنال های آلوده به نویز استفاده شده است. ایده پیشنهادی در این پایان نامه بر مبنای استفاده از SVD و تفکیک زیرفضاهای نویز و سیگنال از یکدیگر بوده و بمنظور بهینه سازی پارامترهای آن از LPC بهره برده ایم. آنچه که روش حذف نویز پیشنهادی را نسبت به متدهای دیگر متمایز می کند، همانا توانایی آن در حذف انواع نویزهای سفید و رنگی از سیگنال های ایستا، ناایستا و صوتی و گفتاری می باشد. معیار ارزیابی ما در این پایان نامه، دو معیار کمی و کیفی

است که معیار کمی، مقدار افزایش SNR(نسبت سیگنال به نویز) را محاسبه می کند و معیار کیفی، شامل تست شنوایی MOS است. با این دو معیار می توان به مقایسه خوب و قابل قبولی بین روش ها پرداخت و نتایج بدست آمده نشان می دهد که ایده ی پیشنهادی ما بهبود بهتری را برای صحبت فراهم می آورد.

فیلتر وینر به صورت یک فیلتر بهینه پایین گذر به منظور برطرف کردن اثرات نویزهای جمع شونده طراحی شده است. این فیلتر با تخمین محلی مقادیر میانگین و واریانس پیرامون هر پیکسل، فاصله میانگین مربعی بین سیگنال مشاهده شده و سیگنال بدون نویز را حداقل می‌کند.

فیلتر وینر کاربردهای مختلفی در پردازش سیگنال ، پردازش تصویر ، سیستم های کنترل و ارتباطات دیجیتال دارد. این کاربردها معمولاً در یکی از چهار دسته اصلی قرار می گیرند:

- شناسایی سیستم
- دکانولوشن
- کاهش نویز
- تشخیص سیگنال

به عنوان مثال ، فیلتر وینر را می توان در پردازش تصویر برای از بین بردن نویز از یک تصویر استفاده کرد.

معمولاً قبل از تشخیص گفتار ، از این فیلتر به منظور کاهش نویز استفاده می‌شود.

۳-۱-۷ بهبود تشخیص گفتار با ارتقای پایش صدا

در سالهای اخیر حوزه مراقبت درمانی توسعه یافته است طوری که منابع پزشکی و بیماران مستقیما به روشهای هوشمند مجهز می‌شوند که مراقبت درمانی هوشمند را ارائه می‌دهد. توسعه طراحی و سیستم اتوماسیون گفتاری خدمات یاررسان را در محیط هوشمند مراقبت درمانی فراهم می‌آورد. در اتوماسیون سیستم گفتاری، تشخیص گفتار یکی از مراحل پایه برای درک شناخت انسان و رفتارهای مربوطه می-باشد. سیستم شناخت یا تشخیص گفتار برای افرادی که از دیزآرتری رنج می‌برند، ناتوانایی نرولوژیکی که کنترل ماهیچه‌های حرکت گفتار را مورد حمله قرار می‌دهد قابل دسترسی می‌باشد. در این تحقیق هدف اصلی توسعه واحد تشخیص گفتار یا صدا براساس ساختار کمک رسانی ارتباطات برون ده صدا و بازده صدا (VIVOCA) می‌باشد. که می‌تواند برای افراد مبتلا سودمند واقع شود. در کل هفت ویژگی از هر داده سیگنال گفتار جداگانه دو زبانه تخمین زده شده وجود دارد که توسط فرد در جاهای مختلف زبانهای تامیل و انگلیسی ادا شده است دفترچه کد الگوریتم ژنتیکی برای بردار

سنجی اتخاذ می‌شود که برای مدلسازی تشخیص مورد استفاده قرار می‌گیرد. بهینه سازی مدل مارکو پنهان (HMM)براساس روش بهینه سازی ذرات پراکنده (PSO) برای ارتقاء دقت تشخیص در مقایسه با HMM قدیمی انجام می‌شود. نتایج آزمایش واحد مذکور ۹۵ درصد دقت و درستی را نشان می‌دهند. واحد پیشنهادی برای توسعه سیستم تشخیص گفتار سودمند است که کار بیماران و افراد را با سازمان ویژه برای برقراری ارتباط آسان می‌سازد واحد یا مدل پیشنهادی براساس پیچیدگی ارزیابی می‌شود که برای معرف انرژی پایین کارآمد خواهد بود.

مدل پنهان مارکوف می‌تواند فرایندهای پیچیده مارکوف را که حالتها بر اساس توزیع احتمالی مشاهدات را نتیجه می‌دهند، مدل کند. به‌طور مثال اگر توزیع احتمال گوسین باشد در چنین مدل مارکوف پنهان خروجی حالتها نیز از توزیع گوسین تبعیت می‌کنند. علاوه بر این مدل پنهان مارکوف می‌تواند رفتارهای پیچیده‌تر را نیز مدل کند. جایی که خروجی حالت‌ها از ترکیب دو یا چند توزیع گوسین پیروی کند که در این حالت احتمال تولید یک مشاهده از حاصلضرب گوسین انتخاب شده اولی در احتمال تولید مشاهده از گوسین دیگر به دست می‌آید.