



واحد تهران جنوب

فرم شماره ۲

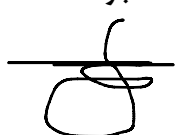
طرح تحقیق پایان نامه کارشناسی ارشد (پروپوزال)

تمامی صفحات طرح تحقیق به صورت تایپ شده تکمیل شود.

عنوان پایان نامه:

فارسی	فاکتورسازی ماتریس غیرمنفی مبتنی بر یادگیری دیکشنری مشترک تبدیل صدا برای بهبود درک گفتار پس از جراحی دهان
انگلیسی	Joint Dictionary Learning-based Non-Negative Matrix Factorization for Voice Conversion to Improve Speech Intelligibility After Oral Surgery

مشخصات دانشجو:

نام:	نرگس	رشته: مهندسی پزشکی	شماره دانشجویی:
نام خانوادگی:	رضایی پیکر	گرایش: بیوالکتریک	۴۰۱۱۴۱۴۰۱۱۱۰۲۷
مجتمع / دانشکده:	دانشکده فنی و مهندسی		
سال تحصیلی اخذ پایان نامه:	<input type="text"/>	ترمه‌های مشروطی: - تعداد واحدهای گذرانده: - معدل دروس گذرانده شده:	امضاء دانشجو: 
نیمسال تحصیلی اخذ پایان نامه:			

کارشناس گروه / مدیر آموزش:

تذکر: اساتید راهنما و مشاور موظف هستند قبل از پذیرش پروپوزال، به سقف ظرفیت راهنمایی و مشاوره خود توجه نموده و در صورت تکمیل نمودن ظرفیت پذیرش، از امضاء این فرم یا در نوبت قرار دادن آن و ایجاد وقفه در کار دانشجویان جدا پرهیز نمایند بدیهی است در صورت عدم رعایت موازین مربوطه، مسولیت تاخیر در ارائه پروپوزال و عواقب کار، متوجه استاد راهنما خواهد بود.

نام و نام خانوادگی استاد راهنما:	نام و نام خانوادگی استاد مشاور (در صورت لزوم):
امضاء	امضاء

تصویب در شورای گروه تخصصی:	تصویب در شورای پژوهشی مجتمع / دانشکده:
تأیید مدیر گروه	تأیید معاون / مدیر پژوهشی مجتمع / دانشکده

امضاء:	امضاء:
تاريخ:	تاريخ:

عنوان فارسی پایان نامه: فاکتورسازی ماتریس غیرمنفی مبتنی بر یادگیری دیکشنری مشترک تبدیل صدا برای بهبود درک گفتار پس از جراحی دهان

۱ - بیان مساله و روش اجرا: (ابعاد مساله، معرفی دقیق مساله، فرضیه ها، جنبه های مجهول، متغیرها و پرسشها و روش های تحقیق)

## ۱- چکیده هدف

این مقاله بر روی تکنیک های تبدیل صوتی مبتنی بر یادگیری ماشین (VC) برای بهبود درک گفتار بیمارانی است که در جراحی قسمت هایی از مفصل هایشان برداشته شده است، تمرکز دارد. به دلیل برداشتن قسمت هایی از آرتیکلاتور، گفتار بیمار ممکن است مخدوش شده و درک آن دشوار باشد. برای غلبه بر این مشکل می توان از روش های VC برای تبدیل گفتار تحریف شده استفاده کرد تا واضح و قابل فهم تر باشد. برای طراحی یک روش مؤثر VC، دو نکته کلیدی باید در نظر گرفته شود: ۱- ممکن است مقدار داده های 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 روش ها

ما یک الگوریتم جدید مبتنی بر یادگیری لغت نامه مشترک فاکتورسازی ماتریس غیرمنفی (JD-NMF) پیشنهاد می کنیم. در مقایسه با تکنیک های VC معمولی، JD-NMF می تواند VC را به طور کارآمد و مؤثر تنها با مقدار کمی از داده های آموزشی انجام دهد.

## ۱-۲ یافته ها

نتایج تجربی نشان می دهد که روش JD-NMF پیشنهادی نه تنها به نمرات قابل توجهی به درک هدف کوتاه مدت<sup>۱</sup> STOL نسبت به روش های به دست آمده با استفاده از گفتار تبدیل نشده اصلی دست می یابد، 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 نمونه های پیشرفته از نظر امتیازات STOL تحت سناریوی مورد نظر بهتر عمل کند.

یک معیار ارزیابی قابل درک استاندارد شده هدف<sup>۱</sup>

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.

## ۲- اصطلاحات فهرست- یادگیری فرهنگ لغت مشترک، فاکتورسازی ماتریس

### غیرمنفی، نمایش پراکنده، تبدیل صدا

استفاده شخصی از این ماده مجاز است. با این حال، اجازه استفاده از این مطالب برای هر هدف دیگری را باید با ارسال درخواستی از [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org) از IEEE دریافت کرد. Szu-Wei Fu با گروه علوم کامپیوتر و مهندسی اطلاعات، دانشگاه ملی تایوان، تایپه، تایوان و مرکز تحقیقاتی نوآوری اطلاعات<sup>۲</sup> (CITI) در Academia Sinica، 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 صدا دار 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 بر VC 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<sup>۳</sup> (JD-GMM) به عنوان یک مدل نقشه برداری مؤثر برای VC شناخته شده است. JD-GMM یک تابع تبدیل خطی را بر

<sup>۲</sup> Center Information Technology Innovation

<sup>۳</sup> Gaussian Mixture Model with Joint Density

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 پیشنهاد شده است. یک شبکه عصبی مصنوعی<sup>۴</sup> (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. هنگامی که 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 وزن دار 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 ما را 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. ریتیم 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،

<sup>4</sup> Artificial Neural Network

اندازه فرهنگ لغت در JD-NM را می توان به طور قابل توجهی نسبت به NMF مبتنی بر نمونه کاهش داد، در نتیجه کارایی تبدیل آنلاین را بهبود می بخشد.

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

## ۴-۱ کار مرتبط

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 مجموعه ای از پایه ها نشان دهد. به این مجموعه از پایگاه ها دیکشنری می گویند. در مدل NMF مبتنی بر نمونه معمولی، هر پایه در ماتریس یک چارچوب گفتاری (نمونه) در داده های آموزشی است. به طور خاص، پایه ها مستقیماً از داده های آموزشی کپی می شوند و هیچ فرآیند یادگیری برای ساخت فرهنگ لغت درگیر نیست. فرض کنید که نمونه ها جمع آوری شده اند، ما یک فرهنگ لغت داریم  $A = [a_1, a_2, \dots, a_I] \in \mathbb{R}^{F \times I}$  که در آن  $a_i$  است  $i^{th}$  نمونه و  $F$  بعد ویژگی است. سپس نمونه گفتار در  $x_1 \in \mathbb{R}^{F \times L}$  را می توان توسط زیر نشان داد:

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

جایی که  $h_1 = [h_1, h_2, \dots, h_{I,l}] \in \mathbb{R}^{I \times l}$  بردار فعال سازی است و  $h_{i,l}$  وزن غیرمنفی است و  $i^{th}$  نمونه است. از آن جایی که هر نمونه گفتاری به طور مستقل مدل می شود، مشخصات هر گفتار را می توان به صورت روبه رو نشان داد:

$$\chi \approx AH$$

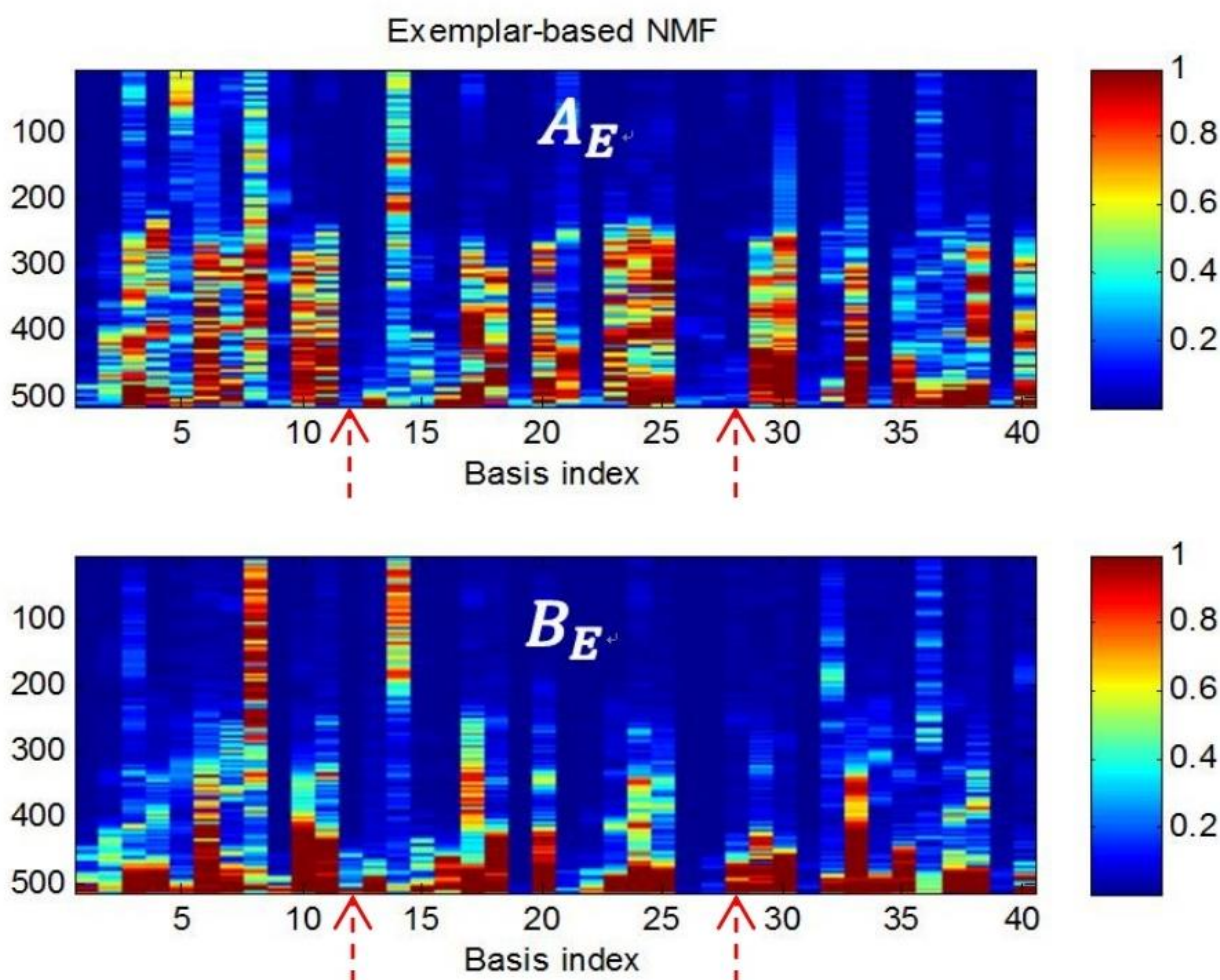
جایی که  $X \in \mathbb{R}^{F \times M}$  طیف نگار است،  $M$  در بیان تعداد فریم ها است، و  $H \in \mathbb{R}^{I \times M}$  ماتریس فعال سازی مربوطه است که بردارستون آن بردار فعال سازی  $h$  است. برای به حداقل رساندن فاصله بین  $X$  و  $AH$ ، قوانینی را برای بهینه سازی متناوب  $A$  و  $H$  با نزول گرادینت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

ب) برای تبدیل صدای مبتنی بر NMF :

(۱) مرحله آفلاین

برای VC دیکشنری های جفت منبع-هدف  $A$  و  $B$  با نمونه های تراز صوتی مورد نیاز است. در NMF های مبتنی بر نمونه، هم دیکشنری منبع و هم دیکشنری هدف مستقیماً از خود داده ها به دست می آید. برای ساخت دیکشنری های جفت شده، یک مجموعه داده موازی (بین گوینده منبع و هدف) جمع آوری می شود. 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 ترازى منبع-هدف بر اساس چارچوب اعمال شوند. شکل ۱ 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 سطوحی 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.



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

(۲) مرحله آنالیز

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:

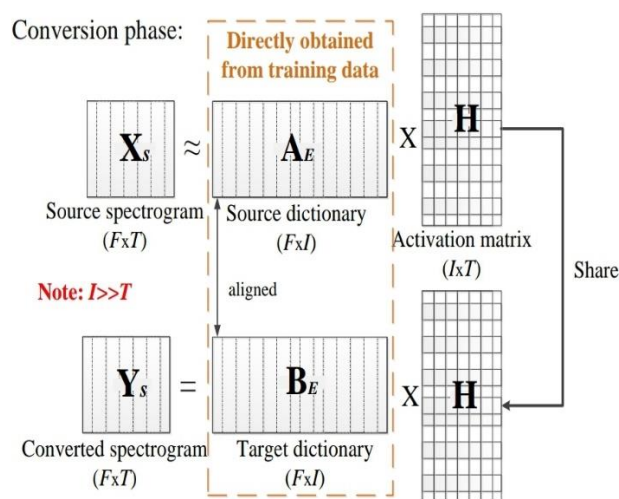
$$Y_s = B_E H$$

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.

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

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 روزرسانی ضربی را به حداقل رساند.

شکل ۲ چارچوب کلی برای VC مبتنی بر نمونه را نشان می‌دهد.



شکل ۱-۲: مرحله آنالیز NMF مبتنی بر نمونه برای VC

## ۱-۵ 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 سناریوی کاربردی 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<sup>5</sup> 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 به جای 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:

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

جایی که 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  $I$  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 ( $Y$  و  $X$ ) با  $H$  مشترک، لغت نامه‌های آموخته شده ( $A$  و  $B$ ) مجبور می‌شوند برای به حداقل رساندن فاصله (واگرایی KL) 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.

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

$$\begin{aligned} & d(X, A_J H) + d(Y, B_J H) \\ &= \sum (X_{fi} \log \frac{X_{fi}}{A_{fi} H_{fi}} - X_{fi} + (A_{fi} H_{fi})) + \sum (Y_{fi} \log \frac{Y_{fi}}{B_{fi} H_{fi}} - Y_{fi} + (B_{fi} H_{fi})) = \\ & \sum (X \log \frac{X}{AH} - X + (AH)) + \sum (Y \log \frac{Y}{BH} - Y + (BH)) \end{aligned}$$

<sup>5</sup> Dynamic Time Warping (DTW)

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

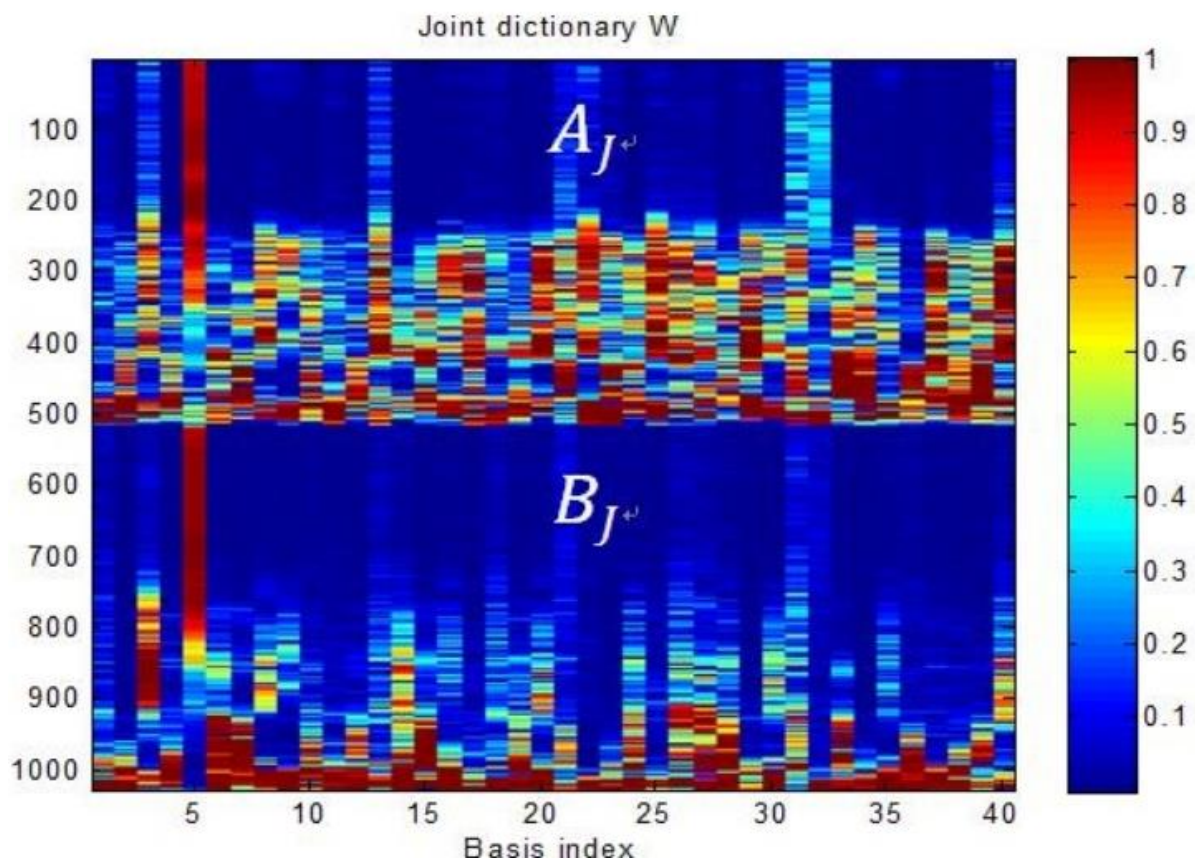
$$\begin{aligned} \Sigma (S \log \frac{S}{WH} - S + (WH) + \lambda ||H|| \\ = d(S, WH) + \lambda ||H|| \end{aligned}$$

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

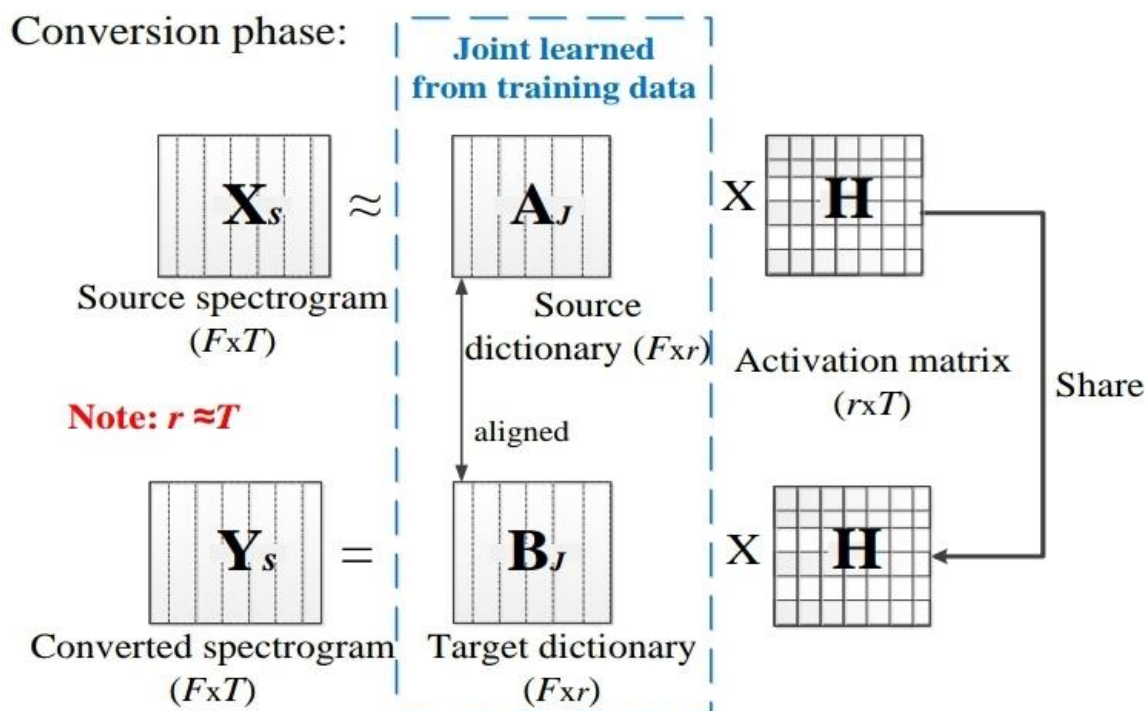
بنابراین، تابع هدف معادل را ساده کردیم. ما به سادگی می‌توانیم قوانین به‌روزرسانی متناوب متداول پیشنهادی را برای تعیین فرهنگ لغت مشترک  $W$  اعمال می‌کنیم.

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

اینجا  $1 \in \mathbb{R}^{2F \times I}$  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 شکل ۳ ما می‌توانیم ببینیم پایه‌های  $A$  و  $B$  به‌طور مشترک تراز شده و یاد می‌گیرند در حین اجرا، "فرآیند تراز  $ac$  DTW است. برای هر دو NMF مبتنی بر نمونه بسیار مهم است (شکل ۱ و JD-NMF). هنگامی که 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 پایه‌های  $B$ ؛ 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.



شکل ۱-۳: فرهنگ لغت منبع و هدف؛ دو دیکشنری را می‌توان با جدا کردن نیمه‌های بالایی و پایینی دیکشنری مشترک  $W$  به ترتیب به‌دست آورد.



شکل ۱-۴: مرحله آنلاین پیشنهادی NMF برای VC در مقایسه با موارد موجود در شکل ۲، دیکشنری‌ها و ماتریس فعال سازی بسیار کوچک تر هستند.

تار، منجر به درک ضعیف گفتار می‌شود. هنگام مقایسه شکل ۱ و شکل ۳ می‌توانیم توجه کنیم که پایه‌های شکل ۱ خیلی معرف نیستند. علاوه بر این، برخی از پایه‌های دو فرهنگ لغت در شکل ۱ 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 شکل ۴ نشان داده شده است. توجه داشته باشید که اندازه دیکشنری‌ها و ماتریس فعال سازی بسیار کوچکتر از اندازه‌های نشان داده شده در شکل ۲ است.

$$Y =$$

$$B_J H$$

جایی که  $Y_s \in \mathbb{R}^{F \times I}$  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:

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{A}_J^T \frac{\mathbf{X}_s}{\mathbf{A}_J \mathbf{H}}}{\mathbf{A}_J^T \mathbf{1} + \lambda}$$

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

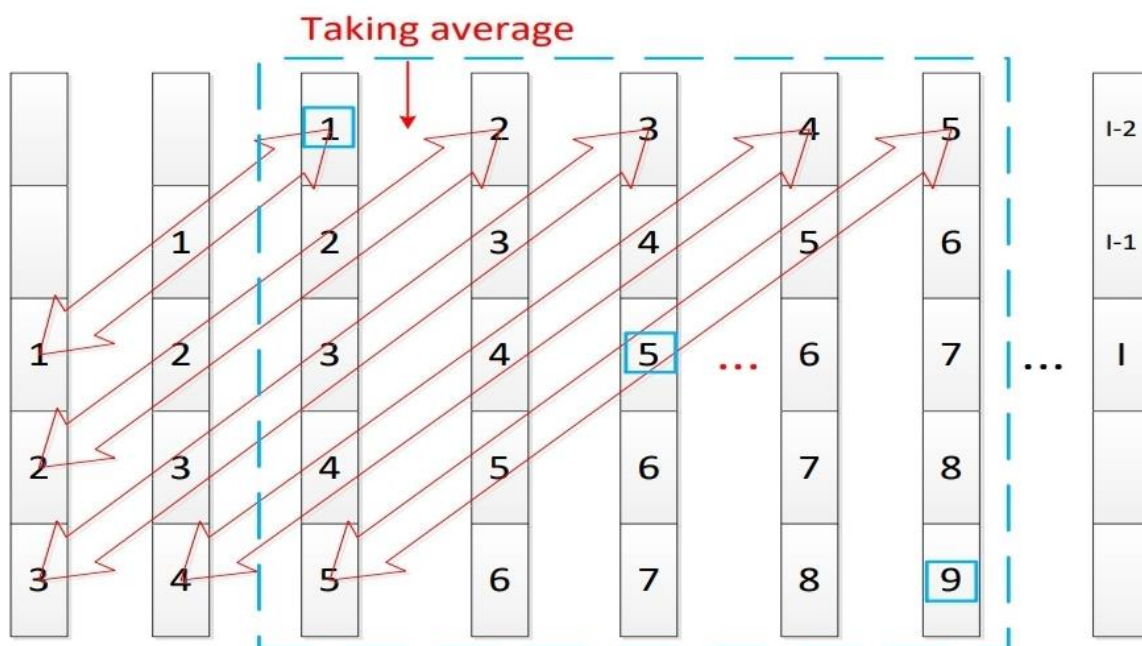
$$2FrT + 2rT + FT$$

که یک تابع خطی از  $r$  است زمانی که  $F$  و  $T$  هر دو ثابت باشند ( $\mathbf{X}$  داده شده است).

در بخش بعدی، ما اطلاعات زمینه ای را برای بهبود بیشتر عملکرد JD-NMF ترکیب می کنیم.

## ۱-۶ اطلاعات متنی

برای در نظر گرفتن اطلاعات زمینه در بسیاری از کاربردهای پردازش گفتار، ویژگی ها به گونه ای آشنایی می شوند که چندین فریم متوالی را در بر می گیرند. 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 نمونه های 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 آشنایی کنیم تا یک فرهنگ لغت مشترک توسعه یافته را آموزش دهیم. بر این اساس، در فاز آفلاین،  $\mathbf{X}$  و  $\mathbf{Y}$  تبدیل به  $\mathbb{R}^{(2q+1)F \times I}$  جایی که  $2q + 1$  است می شوند که به نوبه خود باعث می شود  $\mathbf{A}$  و  $\mathbf{B}$  به  $\mathbb{R}^{(2q+1)F \times I}$  گسترش می یابد. در طول مرحله تبدیل، برای استفاده از فرهنگ لغت توسعه یافته، طیف گرام منبع  $\mathbf{X}$  نیز برای تخمین آشنایی می شود. ماتریس فعال سازی در همین حال، فرهنگ لغت هدف آشنایی همچنین می تواند اطلاعات متنی را برای به دست آوردن مزایای دیگری در نظر بگیرد. شکل ۵ دنباله ای از فریم های گفتار را نشان می دهد. در شکل، زمانی که فریم پنجم قرار است تبدیل شود، پنج بردار چند قاب (در داخل خطوط نقطه چین آبی) را در نظر می گیرد که از فریم اول تا نهم به دست آمده است، بنابراین برای تولید فریم پنجم تولید شده نهایی، می توانیم میانگین را برای ادغام اطلاعات ارائه شده در پنج بردار چند قاب که با فلش های قرمز نشان داده شده اند محاسبه می کنیم. علاوه بر این، عملکرد متوسط می تواند نویز را کاهش دهد و انتقال بین صدهای گفتاری را هموار کند. بنابراین، آشنایی طیف نگار آموزشی می تواند تا حد زیادی کیفیت گفتار تبدیل شده را در چارچوب JD-NMF پیشنهادی بهبود بخشد.



شکل ۵-۱: محاسبه میانگین در بردارهای چندقاب برای کاهش نویز در مرحله تبدیل ( در اینجا در این مثال، اندازه پنجره  $2q + 1 = 5$  )

## ۷-۱ آزمایش

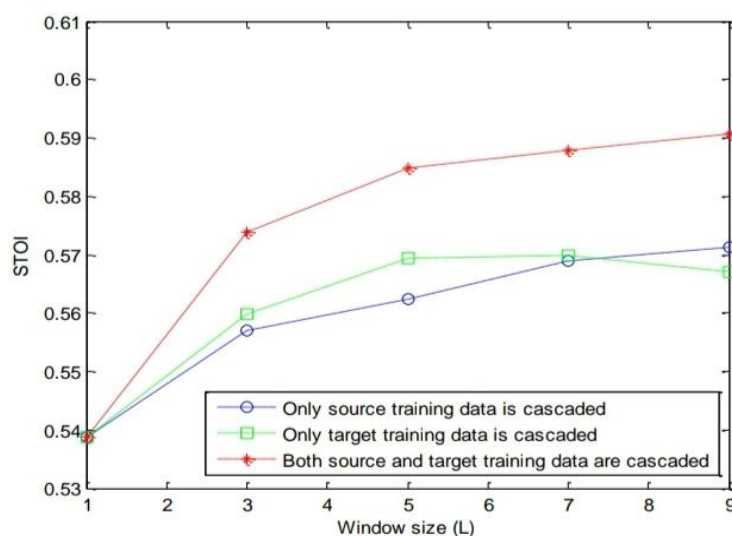
هدف از مطالعه حاضر ارائه یک سیستم VC سریع برای بیماران پس از جراحی دهان می‌باشد. دو ارزیابی عینی در نظر گرفته می‌شود: قابل فهم بودن گفتار تبدیل شده و هزینه محاسباتی تبدیلیها. یک روش ارزیابی استاندارد شده، قابل فهم بودن هدف کوتاه مدت (STOL)، به عنوان معیار قابل فهم عینی ما به کار گرفته شده است. محاسبه 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 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 به دست 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. شکل ۶ نتایج حاصل از STOL (محور Y) از مجموعه‌های توسعه را به عنوان تابع از اندازه پنجره L (محور X) ارائه می‌دهد. در اینجا ما سه مرحله مختلف را بررسی می‌کنیم. ۱- آبشارها تنها داده‌های منبع (فقط یک در شکل ۴) گسترش یافت. ۲- cascading تنها از داده‌های آموزشی هدف نشان می‌دهد. ۳- هنگامی که 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.

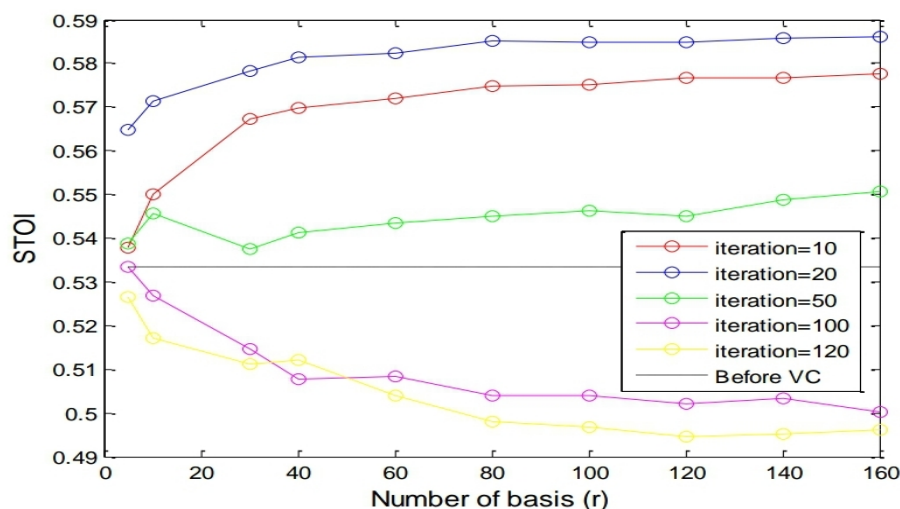


شکل ۶-۱: نتایج STOL برای توسعه تنظیم به عنوان یک تابع از اندازه پنجره

## ب) اثر تعداد پایگاه‌ها و تکرار

دو پارامتر دیگر 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:



شکل ۷-۱: STOL به عنوان یک تابع از تعداد پایگاه‌های R در فرهنگ لغت در زیر تکرار متفاوت است.

#### ۱- اثر تعداد پایگاه‌ها

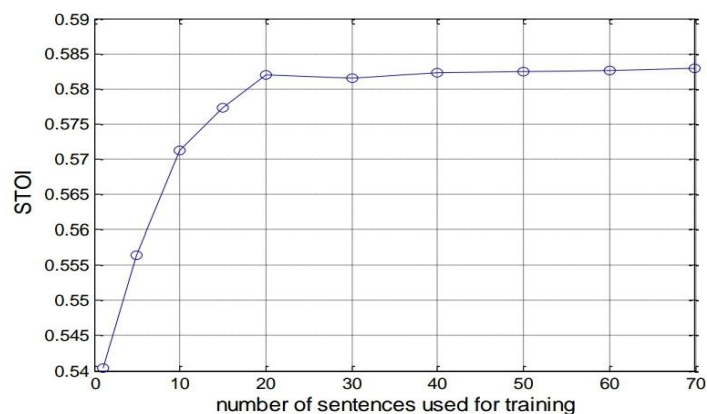
اول ما تغییرات را در STOL برای شمع‌های مختلفی از R مورد بررسی قرار دادیم. به عنوان فرهنگ لغت از داده‌های آموزشی در روش پیشنهادی ما، می‌توانیم اندازه‌های مختلف را برای فرهنگ لغت یاد بگیریم. شکل ۲ نشان می‌دهد که STOL با تعداد زیادی از مبدل زمانی که عدد تکرار کوچک است افزایش می‌دهد. با این حال، اگر الگوریتم 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

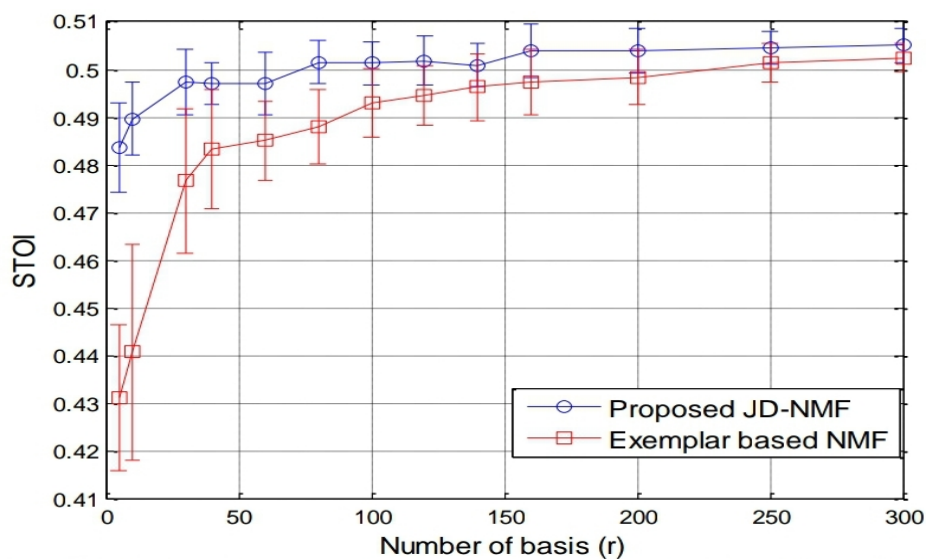
تعداد تکرار معمولاً به صورت تجربی از یک مجموعه توسعه تعیین می‌شود. اگر تعداد توسعه خیلی کم/زیاد باشد، مدل آموخته شده با داده‌های آموزشی کمتر برازش/بیش از حد برازش می‌کند. شکل ۷ نشان می‌دهد که زمانی که الگوریتم چندین بار تکرار می‌شود، مقادیر STOL به دلیل برازش بیش از حد شروع به کاهش می‌کند. اگرچه تفاوت 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 بتواند بر این اساس با اجرای تکرارهای بیشتر بهبود یابد. از این رو، اگر H، B 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 شکل ۷ 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



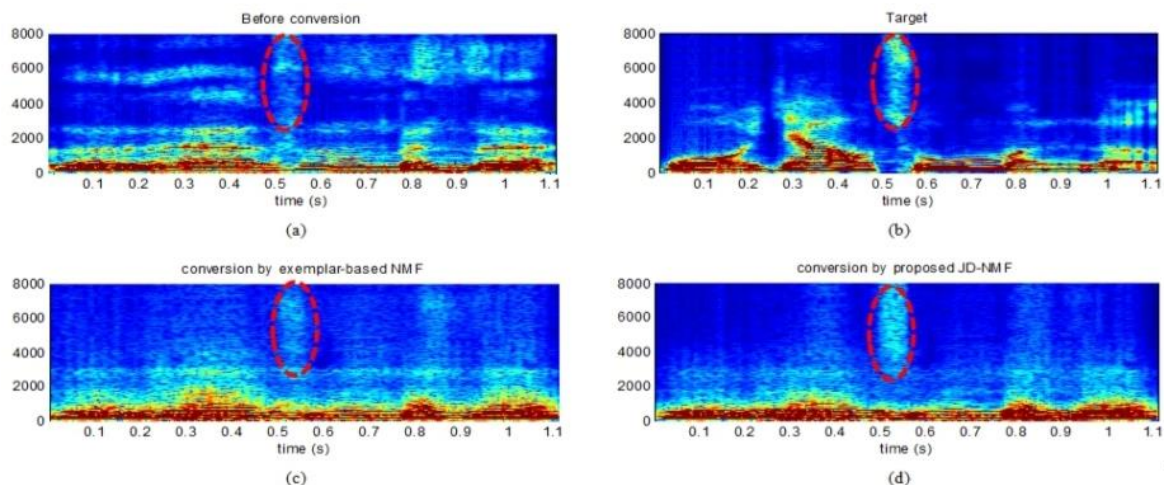
در روش پیشنهادی، اندازه دیکشنری منبع و هدف بر روی  $80 \times (5 \times 513)$  با ۲۰ تکرار در طول تبدیل تنظیم می‌شود.



شکل ۸-۱: STOL به عنوان تابعی از تعداد جملات مورد استفاده آموزش



شکل ۹-۱: STOL به عنوان تابعی از تعداد پایه برای روش‌های پیشنهادی و پایه



شکل ۱-۱۰: 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

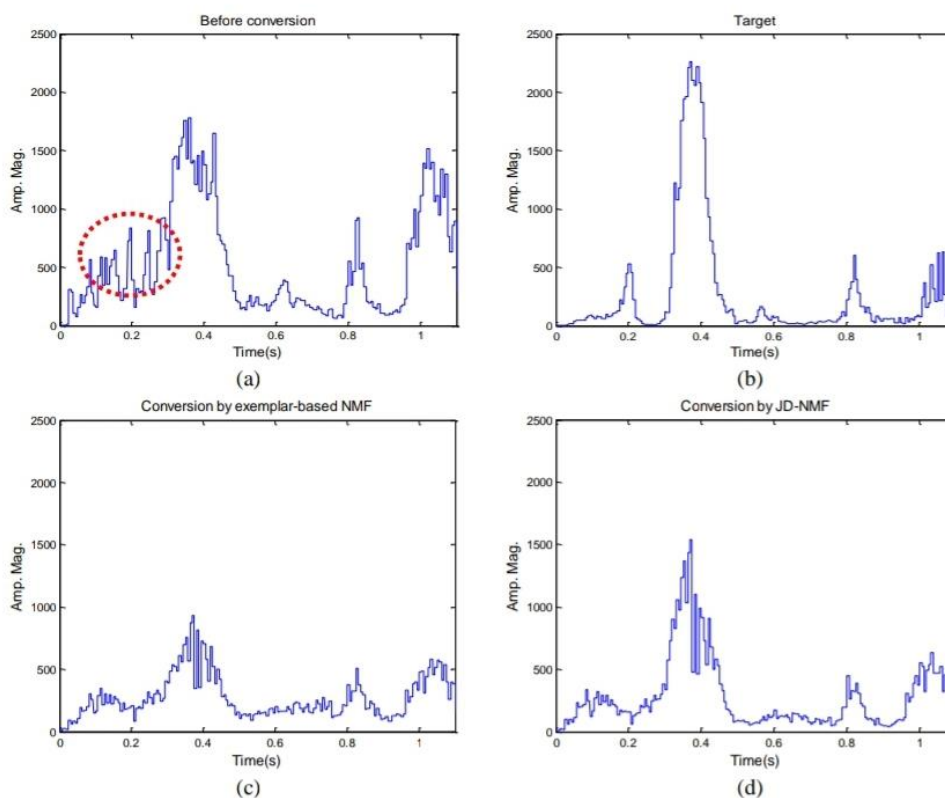
## ج) مقدار داده‌های آموزشی

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. شکل ۸ نشان می‌دهد که بهبود STOL زمانی شروع به اشباع می‌کند که تقریباً 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. شکل ۹ 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 شکل ۹ با میانگین ۱۶۰۰ امتیاز  $(10 \times 4 \times 40)$  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 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، (JD-NMF) STOL 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 به‌طور 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 ۱۱-۱: 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 multiplications 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  $5 \times 513$  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<sup>۶</sup> 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 آنالین 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 نشان 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. شکل ۱۰ یک جفت گفتار منبع و هدف را نشان می‌دهد. شکل ۱۰ (الف) و (ب) به ترتیب با استفاده از NMF مبتنی بر مثال و شکل ۱۰ (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 شده‌اند. علاوه بر این، اجزای فرکانس متوسط برای گفتار قبل از تبدیل نسبتاً پر سر و صدا هستند. این مشاهدات در شکل ۲ 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

<sup>۶</sup> MATLAB

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. شکل ۱۱ پوشش‌های دامنه را از کانال پنجم یک جفت گفتار منبع و هدف پس از تراز از طریق DTW نشان می‌دهد. شکل ۱۱ (الف) و (ب) به ترتیب با تبدیل گفتار توسط JD-NMF های مبتنی بر نمونه و پیشنهادی و شکل ۱۱ (ج) و (د) به ترتیب محورهای  $x$  و  $y$  شاخص زمان و قدر دامنه را نشان می‌دهند. شکل ۱۱ 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 بر این، در حالی که هر دو VC مبتنی بر نمونه و JD-NMF می‌توانند اعوجاج را کاهش دهند، عمق مدولاسیون دومی بسیار بالاتر است. در نهایت 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.

## ۸-۱ 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 واقعی که از بیماران پس از جراحی‌های دهان به دست آمده بود، ارزیابی کردیم. نتایج تجربی ما ابتدا نشان داد که JD-NMF گفتار اصلی را با امتیاز STOL بالا بسیار بهبود بخشید. علاوه بر این، JD-NMF به طور قابل توجهی کارآمدتر و مؤثرتر 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 شفاف‌تری را با عمق مدولاسیون واضح‌تری نسبت به گفتار اصلی تولید می‌کند که توسط NMF مبتنی بر نمونه‌های معمولی تبدیل می‌شود. به طور 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 قرار گرفته‌اند، تا حد زیادی افزایش دهد. در مطالعه 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.

- K. Mády, et al., "Speech evaluation and swallowing ability after intra-oral cancer," *Clinical Linguistics and Phonetics*, vol. 17, pp. 411-420, 2003.
- [2] G. Rentschler and M. Mann, "The effects of glossectomy on intelligibility of speech and oral perceptual discrimination," *Journal of Oral Surgery*, vol. 38, pp. 348-354, 1980.
- [3] B. R. Pauloski, et al., "Speech and swallowing function after anterior tongue and floor of mouth resection with distal flap reconstruction," *Journal of Speech, Language, and Hearing Research*, vol. 36, pp. 267-276, 1993.
- [4] R. Aihara, et al., "Consonant enhancement for articulation disorders based on non-negative matrix factorization," in *Proc. APSIPA 2012*, pp. 1-4.
- [5] R. Aihara, et al., "Individuality-preserving voice conversion for articulation disorders based on non-negative matrix factorization," in *Proc. ICASSP, 2013*, pp. 8037-8040.
- [6] T. Toda, et al., "Voice conversion for various types of body transmitted speech," in *Proc. ICASSP, 2009*, pp. 3601-3604.
- [7] K. Nakamura, et al., "Speaking aid system for total laryngectomees using voice conversion of body transmitted artificial speech," in *Proc. INTERSPEECH, 2006*, pp. 1395-1398.
- [8] Y.-T. Liu, et al., "Nonnegative matrix factorization-based frequency lowering technology for mandarin-speaking hearing aid users," in *Proc. ICASSP, 2016*.
- [9] T. Toda, et al., "Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, pp. 2222-2235, 2007.
- [10] A. Kain and M. W. Macon, "Spectral voice conversion for text-to-speech synthesis," in *Proc. ICASSP, 1998*, pp. 285-288.
- [11] H.-T. Hwang, et al., "A study of mutual information for GMM-based spectral conversion," in *Proc. Interspeech, 2012*, pp. 78-81.
- [12] H.-T. Hwang, et al., "Incorporating global variance in the training phase of GMM-based voice conversion," in *Proc. APSIPA, 2013*, pp. 1-6.
- [13] M. Narendranath, et al., "Transformation of formants for voice conversion using artificial neural networks," *Speech Communication*, vol. 16, pp. 207-216, 1995.
- [14] S. Desai, et al., "Spectral mapping using artificial neural networks for voice conversion," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, pp. 954-964, 2010.
- [15] F.-L. Xie, et al., "Sequence error (SE) minimization training of neural network for voice conversion," in *Proc. Interspeech, 2014*, pp. 2283-2287.
- [16] H.-T. Hwang, et al., "A probabilistic interpretation for artificial neural network-based voice conversion," in *Proc. APSIPA, 2015*.
- [17] S. H. Mohammadi and A. Kain, "Voice conversion using deep neural networks with speaker-independent pre-training," in *IEEE Spoken Language Technology Workshop (SLT), 2014*, pp. 19-23.
- [18] L. Sun, et al., "Voice conversion using deep bidirectional long short-term memory based recurrent neural networks," in *Proc. ICASSP, 2015*, pp. 4869-4873.
- [19] S. H. Mohammadi and A. Kain, "Semi-supervised training of a voice conversion mapping function using a joint-autoencoder," in *Proc. Interspeech, 2015*.

- [20] M. Dong, et al., "Mapping frames with DNN-HMM recognizer for non-parallel voice conversion," in Proc. APSIPA, 2015, pp. 488-494.
- [21] Z. Wu, et al., "Exemplar-based voice conversion using non-negative spectrogram deconvolution," in 8th ISCA Speech Synthesis Workshop, 2013, pp. 201-206.
- Z. Wu, et al., "Joint nonnegative matrix factorization for exemplar-based voice [۲۲] conversion," in Proc. Interspeech, 2014

۳-روش اجرای تحقیق: (شامل روش تهیه داده‌های مورد نیاز، روش تجزیه و تحلیل داده‌ها، مدل‌ها، و نرم‌افزارهای کاربردی)

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

طرح تحقیق پایان نامه کارشناسی ارشد

عنوان فارسی پایان نامه: فاکتورسازی ماتریس غیرمنفی مبتنی بر یادگیری دیکشنری مشترک تبدیل صدا برای بهبود درک گفتار پس از جراحی دهان

۴- زمان بندی / گانت چارت:

ردیف	زمان / ماه نام فعالیت	۱	۲	۳	۴	۵	۶	....	۹
۱	جمع آوری اطلاعات								
۲	بررسی پیشینه								
۳									
۴									
۵									
۶									
۷									
۸									
۹									
۱۰									

نکته: پس از تصویب شورای پژوهشی دانشکده حداقل زمان قابل قبول برای پیش بینی مراحل مطالعاتی و اجرایی پایان نامه کارشناسی ارشد ۶ ماه می باشد.

۵- نظریه شورای گروه تخصصی:

طرح تحقیق پایان نامه خانم / آقای: .....  
 دانشجوی مقطع کارشناسی ارشد رشته ..... در شورای تخصصی گروه مورخ  
 مطرح شد. پس از بحث و تبادل نظر مورد تصویب اکثریت اعضا قرار گرفت □ نگرفت □

ردیف	نام و نام خانوادگی	تخصص	نوع رای	امضاء
۱				
۲				
۳				
۴				
۵				

تاریخ:

امضاء:

مدیر گروه:





واحد تهران جنوب

بسمه تعالی

تعهدنامه حفظ و دفاع از حقوق مادی و معنوی تولیدات علمی دانشگاه آزاد اسلامی و ارائه نتایج آنها  
مرتبط با دانشجویان کارشناسی ارشد

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

مشخصات دانشجو:

نام: نرگس نام خانوادگی: رضایی پیکر شماره دانشجویی:

۴۰۱۱۴۱۴۰۱۱۱۰۲۷

دانشکده: فنی و مهندسی رشته تحصیلی: مهندسی پزشکی گرایش: بیوالکتریک

سال اخذ پایان نامه: - نیمسال تحصیلی: اول

تلفن همراه: ۰۹۱۰۲۴۱۵۴۳۹

پست الکترونیک: rezaeeinarges@gmail.com

تعهدات دانشجو:

- ۱- محتوای پایان نامه کارشناسی ارشد، از آن دیگران نیست (دست اول است)، براساس اصول علمی تهیه شده است و با نام دانشگاه آزاد اسلامی - واحد تهران جنوب ارائه خواهند شد.<sup>۱</sup>
- ۲- به منظور رجوع مناسب و روشن به آثار دیگران، منابع و مأخذ مربوط به نقل قولها، جدولها و نمودارها و یا نتایج تحقیقات دیگران در پایان نامه دقیقاً ذکر خواهد شد؛ همچنین هیچ گونه استفاده‌ای از آثار دیگران بدون ذکر منبع اصلی و به گونه‌ای که قابل تشخیص و تفکیک از متن اصلی نباشد، به عمل نخواهد آمد.
- ۳- بدون ذکر نام دانشگاه آزاد اسلامی - واحد تهران جنوب و در نظر گرفتن حقوق این دانشگاه، در مورد ارائه و انتشار نتایج حاصل از پایان نامه به شکل مقاله، کتاب، اختراع، اکتشاف و ... (در قالب مطالب چاپی یا غیر چاپی) در هر مرحله (قبل و بعد از دفاع از پایان نامه)، اقدامی صورت نخواهد گرفت. بدیهی است که ارسال هر مقاله مستخرج از پایان نامه باید با هماهنگی با استاد راهنما باشد.
- ۴- برای جلوگیری از درج مقاله در نشریات بی اعتبار، قبل از چاپ مقاله، اعتبار نشریه از فهرست نشریات بی اعتبار در سایت معاونت پژوهشی و فناوری دانشگاه آزاد اسلامی به نشانی <http://sp.rvp.iau.ir> بررسی خواهد شد.
- ۵- در صورت هرگونه مغایرت و تخلف از موارد اشاره شده در بندهای ۱ تا ۳ این تعهدنامه، دانشگاه آزاد اسلامی - واحد تهران جنوب مجاز است از ادامه تحصیل و هرگونه فعالیت آموزشی و امکان دفاع از پایان نامه دانشجو در هر مرحله از تحصیل جلوگیری کند. همچنین خسارات مادی و معنوی وارده به دانشگاه آزاد اسلامی و افراد ذی نفع پرداخت خواهد شد.

نام و نام خانوادگی دانشجو: سینا ایوبی

امضاء:

تاریخ:

مقالاتی تحت بررسی قرار خواهند گرفت که طبق بخشنامه‌های سازمان مرکزی باشند.

- ۱- بخشنامه شماره ۷۳/۳۴۵۱۹ مورخ ۹۲/۲/۱۲ باشد. مفاد بخشنامه .... "در صورتی که نام فرد دیگری به غیر از استاد راهنما، مشاور و دانشجو در تیم نویسندگان مقاله مستخرج از پایان نامه و رساله ها قید گردد؛ به مقاله مذکور در مقطع کارشناسی ارشد و دکترای حرفه ای نمره ای اختصاص نمی یابد...."
- ۲- بخشنامه شماره ۷۳/۲۹۹۹۲۰ مورخ ۹۲/۹/۹ باشد. مفاد بخشنامه: ".... در مقاله های مستخرج، در مقاله های مستخرج، نویسنده اول دانشجو و به نام واحد تحصیل دانشجو و استاد راهنما عهده دار مکاتبات است...."
- ۳- بخشنامه شماره ۷۰/۸۱۲۴۸ مورخ ۹۳/۹/۱ باشد. مفاد بخشنامه "نحوه آدرس دهی  
Department of ...., South Tehran Branch, Islamic Azad University, Tehran, Iran: مقاله های انگلیسی:  
\*توجه: تشخیص نشریات بی اعتبار: دو مورد اصلی در تشخیص نشریات بی اعتبار عبارتند از: ۱- تقاضای اخذ وجه توسط ناشر در زمان ارسال یا پذیرش مقاله و ۲-  
آدرس الکترونیکی نشریات بی اعتبار (که اغلب پست های الکترونیکی رایگان نظیر سایت Yahoo و غیره است). همچنین کنترل نشریه در سایت <http://sp.rvp.iaui.ir>



عنوان فارسی پایان نامه:

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

حفظ و دفاع از حقوق مادی و معنوی تولیدات علمی دانشگاه آزاد اسلامی و ارائه نتایج آنها

الف) استاد راهنما:

اینجانب استاد راهنمای آقای / خانم دانشجوی مقطع کارشناسی ارشد دانشگاه آزاد اسلامی - واحد تهران جنوب، از مفاد بخشنامه «حفظ و دفاع از حقوق مادی و معنوی تولیدات علمی دانشگاه آزاد اسلامی و ارائه نتایج آنها»، آگاهی کامل داشته و خود را ملزم به رعایت آن می دانم.

پست الکترونیک:

تلفن:

امضاء:

تاریخ:

ب) استاد مشاور: (در صورت لزوم)

اینجانب استاد مشاور آقای / خانم دانشجوی مقطع کارشناسی ارشد دانشگاه آزاد اسلامی - واحد تهران جنوب، از مفاد بخشنامه «حفظ و دفاع از حقوق مادی و معنوی تولیدات علمی دانشگاه آزاد اسلامی و ارائه نتایج آنها»، آگاهی کامل داشته و خود را ملزم به رعایت آن می دانم.

پست الکترونیک:

تلفن:

امضاء:

تاریخ:



## فرم اطلاعات پایان نامه کارشناسی ارشد

محل درج کد شناسایی پایان نامه (لطفاً در این قسمت چیزی ننویسد).

مشخصات دانشجو:	
نام و نام خانوادگی دانشجو: ..... شماره دانشجویی: .....	
مجتمع/دانشکده: .....	
رشته تحصیلی: ..... گرایش: ..... تعداد واحد پایان نامه: ..... نیم سال تحصیلی اخذ پایان نامه: ..... اول ..... / ..... دوم .....	
امضاء کارشناس آموزشی مجتمع / دانشکده: ..... امضاء رئیس اداره آموزشی مجتمع / دانشکده: .....	
عنوان پایان نامه:	
نام و نام خانوادگی استاد راهنما:	
رشته تحصیلی:	مرتبه علمی:
پایه:	
نوع همکاری: تمام وقت <input type="checkbox"/> نیمه وقت <input type="checkbox"/>	عضو هیات علمی مدعو از سایر واحدهای دانشگاه آزاد اسلامی <input type="checkbox"/>
عضو هیات علمی مدعو از دانشگاه دولتی <input type="checkbox"/>	عضو غیر هیات علمی <input type="checkbox"/>
امضاء استاد:	
نام و نام خانوادگی استاد مشاور:	
رشته تحصیلی:	مرتبه علمی:
پایه:	
نوع همکاری: تمام وقت <input type="checkbox"/> نیمه وقت <input type="checkbox"/>	عضو هیات علمی مدعو از سایر واحدهای دانشگاه آزاد اسلامی <input type="checkbox"/>
عضو هیات علمی مدعو از دانشگاه دولتی <input type="checkbox"/>	عضو غیر هیات علمی <input type="checkbox"/>
امضاء استاد:	
نام و نام خانوادگی مدیر گروه آموزشی - پژوهشی: ..... تاریخ و امضاء:	
تاریخ تصویب پایان نامه در شورای پژوهشی مجتمع/دانشکده: ..... شماره جلسه: .....	

نکته ۱: تمام اطلاعات این فرم صحیح و کامل تایپ شود و به تایید اساتید مربوطه رسانده شود.

نکته ۲: ارسال تصویر کارت ملی (پشت و رو)، آخرین حکم هیئت علمی، رزومه علمی، آخرین مدرک تحصیلی برای کلیه استادان راهنما و مشاور مدعو (عضو هیئت علمی سایر واحدهای دانشگاه آزاد اسلامی و یا وزارتین) برای یک بار الزامی است.

نکته ۳: مسئولین مربوطه می بایست اصل این فرم را به همراه صورتجلسات پروپوزال های تصویب شده در شورای پژوهشی مجتمع / دانشکده و فرم شماره ۱ فایل Excel را بطور همزمان به حوزه معاونت پژوهش و فناوری واحد ارسال نمایند.



واحد تهران جنوب

بسمه تعالی

فرم تصویب (پروپوزال) مربوط به دانشجو ..... به شماره دانشجویی .....

رشته ..... در تاریخ ..... در شورای پژوهشی مجتمع/دانشکده مطرح و

تصویب گردید.

این طرح در تاریخ ..... در شورای پژوهشی مجتمع/دانشکده مطرح گردید ولی به علل زیر مورد موافقت

قرار نگرفت.

علل عدم تصویب طرح تحقیق پایان نامه (پروپوزال):