# طرح پژوهش پایاننامه کارشناسی ارشد



واحد تهران جنوب

# تمامی صفحههای طرح پژوهش به صورت تایپ شده تکمیل شوند.

# عنوان پایاننامه

تبدیل صدای تک شات با جداسازی نمایشهای بلندگو و محتوا با استفاده از راهنمای فعال سازی و عادی سازی نمونه تطبیقی	فارسى
One-shot audio conversion by separating speaker representations and content using	انگلیسی
activation guidance and adaptive sample normalization	

#### مشخصات دانشجو

زهرا	<b>رشته:</b> مهندسی پزشکی	شماره دانشجویی:
		41414.111.44
کربلایی محمدی	<b>گرایش:</b> بیوالکتریک	
فنی و مهندسی		
، پایاننامه: ۱٤٠٠ - ۱٤٠١	ترمهای مشروطی: ۰	امضاء دانشجو:
اخذ پایاننامه: اول/دوم	تعداد واحدهای گذرانده:	
	معدل درسهای گذراندهشده:	LIPZ
کرب فنی <b>, پای</b>	الایی محمدی و مهندسی ان نامه: ان نامه:	الایی محمدی گرایش: بیوالکتریک و مهندسی ان امه: از اینامه: از اینامه: اول/دوم تعداد واحدهای گذرانده:

# كارشناس گروه/ مدير أموزش:

ام و نام خانوادگی استاد راهنما: دکتر مهدی اسلامی	نام و نام خانوادگی استاد مشاور (در صورت لزوم):
مضاء:	امضاء:

تصویب در شورای گروه تخصصی:

تا یید مدیر گروه

تأیید مدیر / معاون پژوهشی دانشکده
امضاء:

امضاء:

تاریخ:

#### طرح پژوهش پایاننامه کارشناسی ارشد

عنوان فارسی پایان نامه: تبدیل صدای تک شات با جداسازی نمایشهای بلندگو و محتوا با استفاده از راهنمای فعال سازی و عادی سازی نمونه تطبیقی

## 1 - بیان مساله و معرفی آن: ابعاد مساله، معرفی دقیق موضوع و مبانی نظری، تعریف، پشتوانه نظری و پیشینه متغیرها و رابطه آنها

#### بيان مساله:

اخیراً، تبدیل صدا (VC) بدون دادههای موازی با موفقیت با سناریوی چند هدف سازگار شده است که در آن یک مدل واحد برای تبدیل صدای ورودی به بسیاری از بلندگوهای مختلف آموزش داده شده است. با این حال، چنین مدلی از این محدودیت رنج می برد که فقط می تواند صدا را به بلندگوهای داده های آموزشی تبدیل کند، که سناریوی قابل اجرا VC را محدود می کند. در این مقاله، ما یک رویکرد VC یک شات جدید را پیشنهاد کردیم که قادر است VC را تنها با یک بیان مثال از بلندگوی منبع و هدف به ترتیب انجام دهد و سخنران مبدأ و هدف حتی نیازی به دیده شدن در طول آموزش ندارند. این امر با جدا کردن نمایشهای بلندگو و محتوا با عادی سازی نمونه (IN) به دست می آید. ارزیابی عینی و ذهنی نشان می دهد که مدل ما قادر به تولید صدایی مشابه گوینده هدف است. علاوه بر اندازه گیری عملکرد، ما همچنین نشان می دهی مدل میتواند نمایشهای معنی دار سخنران را بدون هیچ نظارتی بیاموزد [۱].

ایجاد یک سیستم تبدیل صوتی (VC) برای یک بلندگوی هدف جدید معمولاً به مقدار زیادی داده گفتاری از بلندگوی هدف نیاز دارد [۲]. سیگنال های گفتاری ذاتاً هم اطلاعات ساکن و هم اطلاعات زبانی را حمل می کنند. بخش ایستا مانند گوینده، شرایط آکوستیک مستقل از زمان است و صرفاً در طول کل بیان تغییر می کند، در حالی که بخش زبانی ممکن است در هر چند فریم به طور چشمگیری تغییر کند [۱]. الگوریتم AdaIN-VC از یک رمزگذار خودکار متغیر متشکل از یک رمزگذار بلندگو، یک رمزگذار محتوا و یک رمزگشا استفاده می کند. با عادی سازی نمونه تطبیقی (AdaIN) ، اطلاعات سخنران و اطلاعات محتوا به خوبی قابل تفکیک هستند [۳].

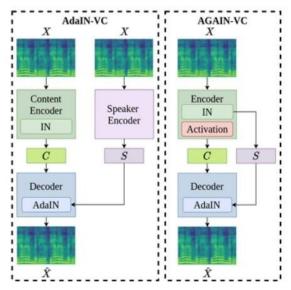
یک محدودیت آشکار دارد که رمزگذار بلندگوی از پیش آموزش دیده صرفاً برای تأیید بلندگو آموزش دیده است. AdaIN-VC از دو رمزگذار مستقل برای استخراج تعبیههای بلندگو و جاسازی محتوا به ترتیب استفاده می کند. با این وجود، ما معتقدیم که رمزگذار بلندگو در اینجا تا حدودی اضافی است. یعنی وظایف آنها می تواند توسط یک رمزگذار انجام شود [۳]. برخی از کارها سعی می کنند سیستم ASR را برای اجرای VC بدون نظارت ترکیب کنند. با ترجمه گفتار به دنبالههای واجی پسین و سپس سنتز گفتار با ترکیب کننده دامنه هدف، VCبدون نظارت میتوان به دست آورد. با این حال، عملکرد این نوع رویکردها به شدت به دقت سیستم ASR بستگی دارد و اگر سیستم ASR به خوبی کار نکند، خراب خواهد شد. برخی از کارهای دیگر سعی می کنند از مدل مولد عمیق مانند VAB برای انجام VC بدون نظارت استفاده کنند، این کارها VC را به عنوان یک مشکل نگاشت دامنه، با هدف یادگیری شبکه هایی که می توانند گفته ها را بین دامنه های مختلف منتقل کنند، فرموله می کنند. این آثار قادر به تولید گفتار با کیفیت خوب هستند و می توانند ویژگی های گوینده را با موفقیت تبدیل کنند. اما محدودیت عمده این آثار این است که نمی توانند صدای گویندگانی را که هرگز در مرحله آموزشی دیده نشده اند ترکیب کنند. سیگنال های گفتاری ذاتاً هم اطلاعات ساکن و هم اطلاعات زبانی را حمل می کنند. بخش ایستا مانند گوینده، شرایط آکوستیک مستقل از زمان است و صرفاً در طول کل بیان تغییر می کند، در حالی که بخش زبانی ممکن است در هر چند فریم به طور چشمگیری تغییر کند [۱]

الگوریتم AdaIN-VC از یک رمزگذار خودکار متغیر متشکل از یک رمزگذار بلندگو، یک رمزگذار محتوا و یک رمزگشا استفاده می کند. با عادی سازی نمونه تطبیقی(AdaIN) ، اطلاعات سخنران و اطلاعات محتوا به خوبی قابل تفکیک هستند. یک محدودیت آشکار دارد که رمزگذار بلندگوی از پیش آموزش دیده صرفاً برای تأیید بلندگو آموزش دیده است. از این رو، استحکام تولید گفتار مشکوک است [۳]. AdaIN-VC از دو رمزگذار مستقل برای استخراج تعبیههای بلندگو و جاسازی محتوا به ترتیب استفاده می کند. با این وجود، ما معتقدیم که رمزگذار بلندگو در اینجا تا حدودی اضافی است. یعنی وظایف آنها می تواند توسط یک رمزگذار انجام شود. الگوریتم AdaIN-VC از یک رمزگذار محتوا و یک رمزگذار بلندگو استفاده می کند، در حالی که AGAIN-VC تنها از یک رمزگذار و یک فعال سازی برای هدایت آموزش استفاده می کند [۳].

-

<sup>&</sup>lt;sup>1</sup> voice conversion

<sup>&</sup>lt;sup>2</sup> adaptive sample normalization



شکل ۱: AGAIN-VC و AdaIN-VC از یک رمزگذار محتوا و یک رمزگذار بلندگو استفاده می کند، در حالی که AGAIN-VC تنها از یک رمزگذار و یک فعال سازی برای هدایت آموزش استفاده می کند [۳].

#### اهمیت و ضرورت پژوهش:

به تازگی، تبدیل صدا (VC) به طور گسترده مورد مطالعه قرار گرفته است. بسیاری از سیستمهای VC از تکنیکهای یادگیری مبتنی بر تفکیک برای جدا کردن گوینده و اطلاعات محتوای زبانی از سیگنال گفتار استفاده میکنند. سپس با تغییر اطلاعات بلندگو به بلندگوی هدف، صدا را تبدیل میکنند. برای جلوگیری از نشت اطلاعات گوینده به درون جاسازیهای محتوا، کارهای قبلی یا ابعاد را کاهش میدهند یا تعبیه محتوا را به عنوان یک گلوگاه اطلاعاتی قوی تعیین میکنند. این مکانیسمها به نوعی به کیفیت سنتز لطمه می زند. الگوریتم AGAIN-VC را پیشنهاد شده است، یک سیستم VC ابتکاری با استفاده از راهنمای فعال سازی و عادی سازی نمونه تطبیقی. -VC یک مدل مبتنی بر رمزگذار خودکار است که از یک رمزگذار و یک رمزگشا تشکیل شده است. با فعالسازی مناسب به عنوان گلوگاه اطلاعاتی در تعبیههای محتوا، مبادله بین کیفیت سنتز و شباهت سخنران گفتار تبدیل شده به شدت بهبود مییابد. این سیستم یک شات VC بدون در نظر گرفتن ارزیابیهای ذهنی یا عینی بهترین عملکرد را به دست میآورد [۳].

### هدف و نوآوری:

هدف کلی:

برخی از کارهای قبلی با استفاده از اَموزش خصمانه برای حذف ویژگی های خاص از یک گفته پیشنهاد شدهاند که در این پژوهش این مشکل برطرف خواهد شد.

هدف جزئی:

اطلاعات جهانی مورد نیاز در رمزگشا توسط رمزگذار بلندگو کنترل می شود. با معماری طراحی شده، مدل ما تشویق می شود تا نمایش های فاکتورسازی شده را بیاموزد.

هدف کاربردی:

مدل ما در طول فرآیند آموزش نیازی به برچسب بلندگوی گفتهها ندارد که جمع آوری دادهها را آسان تر می کند. جالب اینجاست که رمزگذار بلندگو حتی اگر برچسب بلندگو ارائه نکنیم، تعبیههای بلندگوی معنیداری را یاد می گیرد.

#### پرسش (های)/فرضیه (های) پژوهش:

- ۱. أيا مى توان ميان AdaIN-VC و AGAIN-VC ارتباط برقرار كرد؟
  - ۲. أيا مشخصات صوتى در الگوريتم AdaIN-VC تاثير دارند؟
  - ۳. استفاده از AGAIN-VC تا چه حد در تبدیل صدا کمک می کند؟

#### متغیرهای پژوهش:

- ۱. مشخصات صوت
- ۲. محتوای صوتی
- ۳. زمان صوت ورودی و خروجی

تعریف مفهومی و عملیاتی متغیرها:

Y - پیشینه پژوهش و فهرست منابع: سابقه نتایج به دست آمده در خارج و داخل کشور (بدون جداکردن آنها) و نظریههای علمی مرجع در مقالهها، کتابها، طرحهای پژوهشی و پایان نامههای اخیر درباره موضوع پژوهش (فقط براساس جمع بندی یافتهها بدون اشاره به عنوان، هدف، ابزارها، جامعه و نمونه پژوهش ها)

#### پیشینه:

در [۴] Hiroyuki Miyoshi و همکاران تبدیل صدا (VC) با استفاده از یادگیری توالی به دنباله احتمالات پسین زمینه پیشنهاد کردهاند. VC معمولی با استفاده از احتمالات پسین زمینه مشترک، پارامترهای گفتار هدف را از احتمالات پسین زمینه برآورد شده از پارامترهای گفتار منبع پیش بینی می کند. اگرچه VC معمولی می تواند از دادههای غیر موازی ساخته شود، تبدیل فردیت گوینده مانند ویژگی آوایی و سرعت گفتار موجود در احتمالات پسین دشوار است؛ زیرا احتمالات خلفی منبع مستقیماً برای پیش بینی پارامترهای گفتار هدف استفاده می شود. در این کار، فرض شده است که دادههای آموزشی تا حدی شامل دادههای گفتاری موازی است و یادگیری ترتیب به دنباله را بین احتمالات پسین منبع و هدف پیشنهاد می کند. مدلهای تبدیل تبدیل تبدیل غیرخطی و با طول متغیر را از توالی احتمال منبع به هدف انجام می دهند. علاوه بر این، یک الگوریتم آموزشی مشترک برای ماژولها پیشنهاد می شود. برخلاف VC معمولی، که به طور جداگانه تشخیص گفتار را که احتمالات پسین را تخمین می زند و سنتز گفتاری که پارامترهای گفتار هدف را پیش بینی می کند، آموزش می دهد، روش پیشنهادی این مقاله به طور مشترک این ماژولها را همراه با ماژولهای تبدیل احتمال پیشنهادی آموزش می دهد. نتایج تجربی نشان می دهد که این رویکرد از VC معمولی بهتر عمل می کند.

در [۵] Mingjie Chen و همکاران به منظور بهبود کارایی داده، مدلی پیشنهاد کردند که از یک رمزگذار بلندگو برای استخراج تعبیههای بلندگو و لایههای عادی سازی نمونه تطبیقی وزن (W-AdaIN) استفاده میکند. آزمایش ها با ۱۰۹ سخنران در دو موقعیت کم منابع انجام میشود، که در آن تعداد نمونه های آموزشی ۲۰ و ۵ برای هر سخنران است. یک ارزیابی عینی نشان میدهد که مدل پیشنهادی به طور قابل توجهی از روشهای پایه بهتر عمل میکند. علاوه بر این، یک ارزیابی ذهنی نشان میدهد که هم برای طبیعی بودن و هم شباهت، مدل پیشنهادی از روش پایه بهتر عمل میکند.

<sup>&</sup>lt;sup>3</sup> weight adaptive instance normalization

:	منابع	ست	فهو
	L		$\boldsymbol{\mathcal{F}}$

- [1] Ju-chieh Chou, Cheng-chieh Yeh, "One-shot Voice Conversion by Separating Speaker and Content Representations with Instance Normalization", arXiv, 2019.
- [2] Hui Lu, Zhiyong Wu, "One-shot Voice Conversion with Global Speaker Embeddings", INTERSPEECH, pp. 15-19, 2019.
- [3] Yen-Hao Chen, Da-Yi Wu, "Again-VC: A One-Shot Voice Conversion Using Activation Guidance and Adaptive Instance Normalization", IEEE, 2021.
- [4] Hiroyuki Miyoshi, Yuki Saito, "Voice Conversion Using Sequence-to-Sequence Learning of Context Posterior Probabilities", arXiv, 2017.
- [5] Mingjie Chen, Yanpei Shi, "Towards Low-Resource Stargan Voice Conversion Using Weight Adaptive Instance Normalization", IEEE, 2021.

۳-روش اجرای پژوهش

طرح پژوهش:

جامعه آماری:

نمونه آماری، حجم نمونه و روش نمونهبرداری:

ابزارهای پژوهش: معرفی، روش اجرا، نمره گذاری، اعتبار و روایی مؤلف ابزار، پژوهشهای خارجی و داخلی

روش تحليل دادهها:

طرح پژوهش پایاننامه کارشناسی ارشد

عنوان فارسى پاياننامه:

۴ - زمان بندی / گانت چارت

۶	۵	۴	٣	۲	•	زمان/ماه نام فعالیت	ردیف
						جمع اًوری اطلاعات و بررسی پیشینه	١
						جستجوی پرسشنامهها و فراهم سازی انها	۲
						ارائه گزارش اول	٣
						اجرای پرسشنامهها و تدوین پیشینه	۴
						ارائه گزارش دوم	۵
						ورود و تحلیل دادهها	۶
						جمعبندی و تدوین پایاننامه	٧
						تدوين مقاله	٨

**نکته:** پس از تصویب شورای پژوهشی دانشکده **حداقل زمان** قابل قبول برای پیش بینی مراحل مطالعاتی و اجرایی پایاننامه کارشناسیارشد ۶ ماه است.

:	A. É	-1	. 4		1::	۸
تخصصي	ىروە	رای	سو	ريه	بط	–ω

		طرح پژوهش پایاننامه خانم / اَقای:
مطرح شد.	در شورای تخصصی گروه مورخ	دانشجوی مقطع کارشناسیارشد رشته
	اعضاء قرار گرفت 🗆 نگرفت 🗅	پس از بحث و تبادل نظر مورد تصویب اکثریت

امضاء	نوع رای	تخصص	نام و نام خانوادگی	ردیف
				١
				۲
				٣
				۴
				۵

تاريخ:	امضاء:	مدیر گروه :
تاريخ:	امضاء:	دير کروه :

فرم شماره ۱

باسمه تعالى



واحد تهران جنوب

تعهدنامه حفظ و دفاع از حقوق مادی و معنوی تولیدات علمی دانشگاه ازاد اسلامی و ارائه نتایج انها مرتبط با دانشجویان کارشناسیارشد

نامه:	پایا	ن	عنوا
-------	------	---	------

	جو:	شخصات دانش
شماره دانشجویی:	نامخانوادگی:	نام:
:گرایش:گرایش:	رشته تحصیلی	دانشكده:
مسال تحصیلی اول □ دوم □	۱٤٠١ _ ١٤٠٠	سال اخذ پایاننامه:
پست الکترونیک:	تلفن همراه:	تلفن:
و مآخذ مربوط به نقلقولها، جدولها و نمودارها و یا نتایج پژوهشهای دیگران ادهای از آثار دیگران بدون ذکر منبع اصلی و به گونهای که قابل تشخیص و تفکیک	قيقاً ذكر خواهد شد؛ همچنين هيچگونه استف نباشد، بهعمل نخواهد آمد.	پایان نامه د متن اصلی
، و در نظر گرفتن حقوق این دانشگاه، در مورد ارائه و انتشار نتایج حاصل از پایاننامه لب چاپی یا غیرچاپی) در هر مرحله (قبل و بعد از دفاع از پایاننامه)، اقدامی صور، ج از پایاننامه باید با هماهنگی با استاد راهنما باشد.	،، كتاب، اختراع، اكتشاف و (درقالب مطا	شكل مقاله
ی از چاپ مقاله، اعتبار نشریه از فهرست نشریههای بیاعتبار در سایت معاونت پژوهش .http://sp بررسی خواهد شد.	بری از درج مقاله در نشریههای بیاعتبار، قبا انشگاه اَزاد اسلامی به نشانی rvp.iau.ir.	۴– برای جلوگ و فناوری د
در بندهای ۱ تا ۳ این تعهدنامه، دانشگاه آزاد اسلامی – واحد تهران جنوب مجاز اساع از پایان نامه دانشجو در هر مرحله از تحصیل جلوگیری کند. همچنین خسارتها:  نفع داخت خوادد شد		از ادامه تح
ی نفع پرداخ <i>ت</i> خواهد سد. نام و نامخانوادگی دانشجو:	نوی وارده به دانسخاه اراد اسالامی و افراد ده	مادی و مع

امضاء:

تاريخ:

فرم شماره ۲

باسمه تعالى



واحد تهران جنوب

	نامه:	عنوان فارسی پایان

<sup>&</sup>lt;sup>۴</sup> مقالههای برگرفته از پایاننامه و رساله باید به نام دانشجو باشد و نام افراد مرتبط با پایاننامه (دانشجو، استاد راهنما یا استاد مشاور) در مقاله درج شود. درج نام افراد غیرمرتبط، تخلف محسوب می شود مگر آنکه طبق قراردادی با مؤسسههایی خارج از دانشگاه و تایید معاونت پژوهش و فناوری دانشگاه، این مجوز با سهم معینی در مقاله به افراد داده شده باشد. (تبصره ۲ بخشنامه شماره ۸۷/۴۰۴۸۲۲ مورخ ۹۰/۱۰/۲۹)

<sup>\*</sup>توجه: تشخیص نشریههای بی اعتبار: دو مورد اصلی در تشخیص نشریههای بی اعتبار عبارتند از: ۱- تقاضای اخذ وجه توسط ناشر در زمان ارسال یا پذیرش مقاله و ۲- آدرس الکترونیکی نشریههای بی اعتبار (که اغلب پستهای الکترونیکی رایگان نظیر سایت Yahoo و غیره است). همچنین بازبینی نشریه در سایت http://sp.rvp.iau.ir

# 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 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 میدانم.

تلفن: پست الكترونيك:

امضاء:

تاريخ:

### ب) استاد مشاور (در صورت لزوم)

اینجانب استاد مشاور آقای/ خانم 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 میدانم.

تلفن: پست الکترونیک:

امضاء:

تاريخ:

# فرم الف

# بسمه تعالى





محل درج کد شناسایی پایاننامه (لطفا در این قسمت چیزی ننویسید.)

		مشخصات دانشجو:	
شماره دانشجویی:		نام و نام خانوادگی دانشجو:	
تعداد واحد پایان نامه:	گرایش:	رشته تحصيلى:	
		_	
امضاء كارشناس أموزش مجتمع/ دانشكده:			
		عنوان پایاننامه:	
		O O	
		نام و نام خانوادگی استاد راهنما:	
صادره:	تاريخ تولد:	شماره شناسنامه:	
مرتبه علمی: پایا			
عضو هیات علمی مدعو از سایر واحدهای	نيمەوقت □	نوع همکاری: تماموقت □	
عضو غيرهيات علمي 🏻	ناه دولتی 🗆	عضو هیات علمی مدعو از دانشگاه دولتی 🗆	
		1	
1		نام و نام خانوادگی استاد مشاور:	
صادره:	تاریح تولد:	شماره شناسنامه:	
		رشته تحصیلی:	
عضو غیرهیات علمی 🏻	ناه دولتی □	عضو هیات علمی مدعو از دانشگ	
تاریخ تصویب پایان نامه در شورای پژوهشی مجتمع/دانشکده :			
	تعداد واحد پایان نامه:  امضاء رئیس اد مرتبه علمی: عضو هیات علمی مدعو از سایر واحدهای عضو غیرهیات علمی □ مرتبه علمی:  مرتبه علمی: عضو هیات علمی مدعو از سایر واحدهای	گرایش: تعداد واحد پایاننامه:  مع / دانشکده:	

- نکته 1: 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 هیات علمی سایر واحدهای دانشگاه آزاد اسلامی و یا وزارتین) برای یک بار الزامی است.
- نکته ۳: مسئولین مربوطه میبایست اصل این فرم را به همراه صورتجلسات پروپوزالهای تصویب شده در شورای پژوهشی مجتمع/ دانشکده و فرم شماره ۱ (یک نسخه فایل اکسل آن) را به طور همزمان به حوزه معاونت پژوهش و فناوری واحد ارسال نمایند.

امضاء معاونت پژوهشی واحد:

امضاء رئيس مجتمع/دانشكده: