



واحد تهران جنوب

فرم شماره 2


طرح تحقیق پروژه پردازش سیگنال دیجیتال (Fast VC)

تمامی صفحات طرح تحقیق به صورت تایپ شده تکمیل شود.

عنوان پروژه :

فارسی	افزایش سرعت مدت زمان گفتار بدون اینکه محتوای گفتار از دست برود (تبدیل صدای سریع و کارآمد)
انگلیسی	Fast VC

مشخصات دانشجو:

نام:	مرضیه	رشته: مهندسی پزشکی	شماره دانشجویی:
نام خانوادگی:	علی آتشی	گرایش: بیوالکتریک	40114140111030
مجتمع / دانشکده:	دانشکده فنی و مهندسی		
سال تحصیلی ارائه پروژه :	1401	نرمهای مشروطی: - تعداد واحدهای گذرانده: 0 معدل دروس گذرانده شده:	امضاء دانشجو:
نیمسال تحصیلی ارائه پروژه :	اول		

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، مسوولیت 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:
تاریخ:	تاریخ:

افزایش سرعت مدت زمان گفتار بدون اینکه محتوای گفتار از دست برود (تبدیل صدای سریع و کارآمد)

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

جدول بررسی مزایا و معایب پروژه های بررسی شده

Attention, I'm Trying to Speak CS224n Project: Speech Synthesis

Akash Mahajan
Management Science & Engineering
Stanford University
akashmjn@stanford.edu

مقاله اول

مزایا
تبدیل گفتار به موسیقی و دیگر داده های سیگنال مانند امکان فشرده سازی یک منبع فشرده در حد بسیار بالاتر
تفاوت ها
ورودی ها در نیمه پایینی یک فریم طیف منفرد هستند تفاوت میان پیچیدگی های سببی و غیر سببی و همچنین افزودن رمزگذاری های موقعیتی برجسته شده است در سیستم واژگان با فاصله جایگزین می شوند و اعداد و حروف اختصاری هستند به راحتی می توان این را برای پشتیبانی از اختصارات و اعداد و تاریخ ها و موارد دیگر هم گسترش داد

Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis

Ye Jia* Yu Zhang* Ron J. Weiss* Quan Wang Jonathan Shen Fei Ren
Zhifeng Chen Patrick Nguyen Ruoming Pang Ignacio Lopez Moreno Yonghui Wu
Google Inc.
{jiaye,ngyuzh,ronw}@google.com

مقاله دوم

از معایب آن آموزش اسپیکر بلندگو دیگر است که زمان زیادی را صرف جمع آوری داده ها باید کرد
افزایش تعداد بلندگو های آموزش دیده شده به بهبود سیستم کمک میکند
در این سنتز فقط طیف مورد بررسی قرار میگیرد

در نهایت صدای تولید شده از صدای اصلی کمی متفاوت بوده و شوخه ها ی خودشان را به شکل جدا
تشکیل می دهند و همپوشانی کاملی هنوز نشان می دهد

wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations

Alexei Baevski

Henry Zhou

Abdelrahman Mohamed

Michael Auli

{abaevski, henryzhou7, abdo, michaelauli}@fb.com

مقاله سوم

یادگیری بازنمایی از صدای گفتار به تنهایی به دنبال تنظیم دقیق گفتار رونویسی شده می تواند بهترین عملکرد را داشته باشد روش های نیمه نظارتی از نظر مفهومی ساده تر هستند ورودی گفتار در فضای پنهان حل می کند و کمی سازی های بازنمایی های نهفته به طور مشترک آموخته می شود

از معایب این سیستم توانایی برچسب گذاری محدود است به دست آوردن داده های برچسب دار سخت تر از داده های بدون برچسب است

FastVC: Fast Voice Conversion with non-parallel data

Oriol Barbany^{1,2}, Milos Cernak¹

¹Logitech Europe S.A., 1015, Lausanne, Switzerland

²École Polytechnique Fédérale de Lausanne (EPFL), 1015, Lausanne, Switzerland

milos.cernak@ieee.org

مقاله چهارم

مزیت سخنرانی با طول دلخواه مبتنی بر چند بلندگوی منبع و چند بلندگو هدف داده های غیر موازی آموزش داده شده و نیاز به هیچ گونه اطلاعاتی ندارند مستقل از گوینده هستند از این سیستم برای اجرای سیستم های دیگر مثل تبدیل گفتار به گفتار در زبان دیگری به شکل مستقیم استفاده شده

ESPnet-ST: All-in-One Speech Translation Toolkit

Hirofumi Inaguma¹ Shun Kiyono² Kevin Duh³ Shigeki Karita⁴

Nelson Yalta⁵ Tomoki Hayashi^{6,7} Shinji Watanabe³

¹ Kyoto University ² RIKEN AIP ³ Johns Hopkins University

⁴ NTT Communication Science Laboratories ⁵ Waseda University

⁶ Nagoya University ⁷ Human Dataware Lab. Co., Ltd.

inaguma@sap.ist.i.kyoto-u.ac.jp

مقاله پنجم

باعث کاهش خطر انتشار خطا در قسمت منبع ترجمه گفتار و متن هدف می شود
برای محققان پژوهشگران آسان تر شده و به گونه ای طراحی شده که برای هدف قرار دادن و توسعه
طراحی امکان پذیر است

PAC-Bayes, MAC-Bayes and Conditional Mutual Information: Fast rate bounds that handle general VC classes

Peter Grünwald*

Thomas Steinke[†]

Lydia Zakyntinou[‡]

June 18, 2021

مقاله ششم

پردازش روی خطاها تعمیم یافته و قسمت خطاها بهبود پیدا کرده است

A FAST AND EFFICIENT VOICE CONVERSION TOOL

Luigi Attorresi, Federico Miotello, Eugenio Poliuti

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano

Piazza Leonardo Da Vinci 32, 20122 Milano, Italy

[luigi.attorresi, federico.miotello, eugenio.poliuti]@mail.polimi.it

مقاله هفتم

کیفیت صدای خروجی حد زیادی بهبود بخشیده شده یادگیری و آموزشی گسترده امکان پذیر است
این سیستم بر اساس رمزگذاری و رمزگشایی طراحی شده احتمال خطا در تبدیل رمز گذاری ها وجود
دارد
بر اساس تبدیل یک دنباله دنباله دیگر بازبینی شده
امکان اضافه کردن بلندگوی هدف و بلندگوی ورودی وجود دارد
نیاز به داده های زیادی دارند زمان بر است و بسیار هزینه زیاد و هزینه زیادی دارد