



بسمه تعالی



دانشگاه صنعتی امیر کبیر

(پلی تکنیک تهران)

دانشکده مهندسی برق

پایان نامه دکترای مهندسی برق - مخابرات (سیستم)

عنوان:

# افزایش کیفیت و بهبود عملکرد سیستم های تبدیل گفتار فارسی

استاد راهنما:

دکتر ابوالقاسم صیادیان

استاد مشاور:

دکتر حمید شیخ زاده

نگارش:

مهدی اسلامی



فرم اطلاعات پایان نامه  
کارشناسی ارشد و دکترا

دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)

معاونت پژوهشی  
فرم پروژه تحصیلات تکمیلی ۷

مشخصات دانشجو

نام و نام خانوادگی: مهدی اسلامی  دانشجوی آزاد  بورسیه  معادل

شماره دانشجویی: ۷۹۲۲۳۹۱۹ دانشکده: مهندسی برق رشته تحصیلی: مخابرات سیستم

نام و نام خانوادگی استاد راهنما: دکتر ابوالقاسم صیادیان

عنوان به فارسی: افزایش کیفیت و بهبود عملکرد سیستم تبدیل گفتار فارسی

عنوان به انگلیسی: Quality Improvement of Farsi Voice Conversion System.

نوع پروژه دکترا:  کاربردی  بنیادی  توسعه ای  نظری

تاریخ شروع: ۸۲/۱/۱۱ تاریخ خاتمه: ۸۵/۱۲/۱۴ تعداد واحد: .....

سازمان تامین کننده اعتبار: مرکز تحقیقات مخابرات ایران

واژه های کلیدی به فارسی: تبدیل گفتار با کیفیت بالا- مدل مخلوط گوسی GMM- مدل تعمیم یافته هارمونیکي GHM- تبدیل طیفی

واژه های کلیدی به انگلیسی: High Quality Voice Conversion, GMM Model, GHM Model, Spectral Conversion, Statistical Modelling

نظرها و پیشنهادهای به منظور بهبود فعالیت های پژوهشی دانشگاه:  
افزایش امکانات و بودجه تحقیقاتی  
استاد راهنما: دکتر ابوالقاسم صیادیان

دانشجو: مهدی اسلامی

امضاء استاد راهنما: تاریخ:

نسخه ۱: معاونت پژوهشی

نسخه ۲: کتابخانه و به انضمام دو جلد پایان نامه به منظور تسویه حساب با کتابخانه و مرکز اسناد و مدارک علمی

تقدیم به:

پدر بزرگوارم ...

مادر فداکارم ...

همسر عزیزم ...

و همه شهیدان راه وطن...

## تشکر و قدردانی:

با تشکر از استاد محترم راهنما، آقای دکتر صبیادیان که در تمامی لحظات انجام پروژه دکتری، راهنمایی‌ها و دلگرمی‌های ایشان مشوق اصلی اینجانب در انجام این مهم بوده است و تشکر از پدر گرامی که حمایت‌ها و تشویق‌های بی‌دریغ ایشان سبب ایجاد محیطی آرام و صمیمی در خارج از محیط دانشگاه بود و همچنین تشکر و قدردانی از الطاف و صبر خانواده گرامی خصوصاً خواهران و برادر بزرگوار که سبب رفع خستگی‌ها بود و تشکر از آقای دکتر شیخ زاده که به دفعات از نظرات ارزشمند ایشان در انجام پروژه دکتری بهره‌مند گردیدم و همچنین تشکر و قدردانی از تمامی اساتید بزرگوار دانشگاه خصوصاً: آقایان دکتر فائز، دکتر آقایی نیا، دکتر احدی، دکتر طاهری، دکتر الماس گنج، دکتر مرادی، دکتر عارف، دکتر محامدپور، دکتر احمدیان، دکتر کلانتری، دکتر جبه دار، دکتر جمالی، دکتر احمدی، دکتر مشیری، دکتر فرهنگی، دکتر آرزم و .... و تمامی معلمین بزرگوار در طول دوران تحصیلی، آقایان افتخاری، معصومی، احمدی، تفضیلیان، اقبالپور، خادمی، بطحایی، اکبری، امینی، رفیعی، پوستچی، صالحی، فارسی، داوودآبادی و ... که افتخار شاگردی ایشان را داشته‌ام و تشکر از مدیریت محترم کارخانجات مخابراتی ایران بالاخص آقای مهندس مطلع و مدیریت‌های قبلی، آقای مهندس واحدی و استاد بزرگوار، آقای مهندس دانشمند که حمایت آنها را همواره در انجام این مهم در کنار خود احساس می‌نمودم و تشکر و قدردانی از الطاف اساتید بزرگوار: آقایان دکتر عابدی، دکتر مغانی، دکتر محمدی، دکتر کوهساری و دکتر عابدی پور که محیطی صمیمی، پر نشاط و مجهز را در دانشکده مهندسی برق جهت انجام فعالیت‌های آموزشی/پژوهشی دانشجویان فراهم نموده‌اند و تشکر از حمایت مرکز تحقیقات مخابرات ایران در انجام این پروژه و در نهایت تشکر از تمام کسانی که اینجانب را در تدوین این رساله یاری نمودند.

## چکیده:

در این رساله به مطالعه و پیاده سازی سیستم تبدیل گفتار با کیفیت بالا پرداخته شده و روش های افزایش کیفیت و بهبود عملکرد آنها در زبان فارسی مورد بررسی قرار گرفته است. در سیستم های تبدیل گفتار، گوینده A (مبدا) عباراتی را بیان می کند و هدف از آن عبارت است از تغییر متکلم جملات بیان شده از گوینده A به گوینده B (مقصد). کاربرد سیستم های تبدیل گفتار در ساخت پایگاه دادگان گفتاری جهت کاربرد در سیستم های تبدیل متن به گفتار و بازشناسی گفتار است. همچنین این سیستم قابل استفاده در صداگذاری فیلم ها و ... می باشد. در این قبیل کاربردها، صدای هر گوینده در محیط استودیو ضبط شده است و نیازی به پردازش بلادرنگ ندارد.

روش آماری مبتنی بر GMM بهترین کارایی را در مقایسه با روش های دیگر دارد. این روش بعثت دخالت دادن تمامی خوشه ها در تولید یک بردار برای گوینده جدید، دارای حالت بلورشدگی (کاهش وضوح) در صدای بازسازی شده می باشد به گونه ای که بازسازی صدا (با تغییر گوینده) توسط روشهای مذکور نسبت به حالت کاملاً طبیعی، فاصله زیادی دارد. در روش GMM(2) از مدل های متعدد GMM برای مدل سازی هر واج استفاده نموده ایم. همچنین در مرحله متناظر کردن خوشه های هر حالت، قبل از اعمال الگوریتم DTW از یک تبدیل LMR برای انطباق بیشتر پارامترهای دو حالت متناظر از دو گوینده استفاده می شود. در روش GMM(3) به منظور ارائه تخمین دقیق تر سیگنال گفتاری و کیفیت بالاتر سیگنال تبدیل شده، از مدل GHM استفاده می شود که از کارکردی بهتر نسبت به روش قبل برخوردار است. در الگوریتم GMM(4) ویژگی های گفتار بطور پیوسته با استفاده از همبستگی میان ویژگی های گفتار گوینده مبدا و مقصد، تغییر می نمایند. به منظور غلبه بر هموار شدگی طیفی ناشی از متوسط گیری آماری، از واحدهای آوایی نیمه هجا، به عنوان کوچکترین واحدهای آوایی شامل اطلاعات عروضی گفتار استفاده می شود. همچنین با توجه به مزایای GHM، از آن به عنوان آنالیز و سنتز کننده استفاده می شود.

در این رساله به ارائه روشی می پردازیم که علاوه بر ویژگی های درون قابی، از ویژگی های برون قابی (دینامیکی) برای یافتن بردار تبدیل یافته از گفتار گوینده A به گفتار گوینده B استفاده می کند. این روش مبتنی بر VQ بوده و در آن از یک ساختار شبکه برای یافتن یکی از بهترین مسیرها جهت نگاشت دنباله ای از قاب های گفتاری مربوط به کتاب کد گوینده A به کتاب کد گوینده B استفاده می شود. معیار بهینگی در یافتن مسیر عبارت است از: حفظ ویژگی های دینامیکی گفتار گوینده بعلاوه پیوستگی طیفی در گفتار تبدیل یافته. نوآوری دیگر ارائه شده، استفاده از نیمه هجا به عنوان کوچکترین واحد گفتاری در برگیرنده اطلاعات عروضی از گفتار گوینده است که متناسب با ساختار زبان فارسی می باشد. در نهایت به کمک اصلاحات مناسب دیگری که در روش یادگیری و طراحی تبدیلهای خطی مورد نیاز انجام شده است، به عملکرد بسیار مناسبی در تبدیل گفتار در مقایسه با روشهای رایج نائل شده ایم.

**ABS** Analysis By Synthesis.  
**ANN** Artificial Neural Networks.  
**C** Consonant.  
**CQ** Communication Quality.  
**DFW** Dynamic Frequency Warping.  
**DTW** Dynamic Time Warping.  
**EM** Expectation Maximization.  
**GHM** Generalized Harmonic Model.  
**GMM** Gaussian Mixture Model.  
**HMM** Hidden Markov Model.  
**HNM** Harmonic plus Noise Model.  
**LAR** Log Area Ratio.  
**LMR** Linear Multivariate Regression.  
**LPC** Linear Prediction Coding.  
**LP-PSOLA** Linear Prediction Pitch Synchronous Overlap and Add.  
**LSF** Line Spectral Frequency.  
**MAP** Maximum a Posteriori.  
**MFCC** Mel Frequency Cepstral Coefficient.  
**ML** Maximum Likelihood.  
**MOS** Mean Opinion Score.  
**MVF** Maximum Voiced Frequency.  
**NCC** Normalized Cross Correlation.  
**NUU** Non Uniform Units.  
**OLA** Overlap and Add.  
**PM** Periodicity measure.  
**RMS** Root Mean Square.  
**SD** Spectral Distortion.  
**SM** Sinusoidal Model.  
**SSM** Stochastic Segment Modeling.  
**TD-PSOLA** Time Domain Pitch Synchronous Overlap and Add.  
**TSVQ** Trellis-Structured Vector Quantization.  
**TTS** Text-to-Speech.  
**V** Vowel.  
**VC** Voice Conversion.  
**VQ** Vector Quantization.  
**VT** Voice Transformation.

## فهرست عناوین:

### فصل ۱.

#### معرفی و تشریح کاربرد سیستم های تبدیل گفتار و

#### اهداف رساله

- ۱-۱- پیش گفتار..... ۲
- ۲-۱- تعریف سیستم های تبدیل گفتار..... ۳
- ۳-۱- کاربردهای تبدیل گفتار..... ۳
- ۱-۳-۱- کاربرد در سیستم های TTS..... ۳
- ۲-۳-۱- ترجمه خودکار گفتار به گفتار..... ۴
- ۳-۳-۱- آموزش..... ۴
- ۴-۳-۱- یاری رسانی پزشکی..... ۵
- ۵-۳-۱- سرگرمی..... ۵
- ۴-۱- اجزای سیستم تبدیل گفتار..... ۶
- ۵-۱- اهداف رساله..... ۸
- ۶-۱- ترتیب مطالب ارائه شده..... ۹

### فصل ۲.

#### طراحی مجموعه گفتاری جهت آموزش و ارزیابی

- ۱-۲- پیش گفتار..... ۱۱
- ۲-۲- متن گفتار..... ۱۳
- ۳-۲- ضبط صدا..... ۱۴
- ۱-۳-۲- ضبط صدا به روش تقلید..... ۱۴
- ۲-۳-۲- ضبط صدا به روش تکرار..... ۱۵
- ۴-۲- طراحی روش های مناسب ارزیابی سیستم های VC..... ۲۰



|    |                                  |
|----|----------------------------------|
| ۲۰ | ..... ۲-۴-۱- معیار فاصله طیفی    |
| ۲۱ | ..... ۲-۴-۲- تست ABX تعمیم یافته |
| ۲۲ | ..... ۲-۴-۳- تست MOS             |

### فصل ۳.

#### طراحی یک مدل آنالیز/سنتز گفتار با کیفیت بالا

|    |   |
|----|---|
| ۲۴ | ..... ۳-۱- پیش گفتار  |
| ۲۸ | ..... ۳-۲- مدل مکولای/کواتری  |
| ۳۱ | ..... ۳-۳- تغییر و اصلاح زمانی                                      |
| ۳۲ | ..... ۳-۴- تغییر و اصلاح مقیاس فرکانس و پیچ                         |
| ۳۲ | ..... ۳-۵- معرفی مدل هارمونیکی تعمیم یافته                          |
| ۳۶ | ..... ۳-۶- محاسبه سیگنال نرمالیزه شده                               |
| ۳۷ | ..... ۳-۷- تعیین نامزدهای نقاط قابل اطمینان                         |
| ۳۸ | ..... ۳-۸- انتخاب یک دوره پیچ از سیگنال                             |
| ۳۸ | ..... ۳-۸-۱- تخمین نهایی دوره پیچ                                   |
| ۴۰ | ..... ۳-۸-۲- محاسبه سیگنال نرمالیزه شده                             |
| ۴۰ | ..... ۳-۸-۳- الگوریتم تعیین نقاط قابل اطمینان                       |
| ۴۱ | ..... ۳-۸-۴- تخمین نهایی مقدار پیچ و همچنین نواحی واک دار از بی واک |
| ۴۳ | ..... ۳-۸-۵- نتایج  |
| ۴۴ | ..... ۳-۹- آنالیز یک دوره پیچ از سیگنال                             |
| ۴۵ | ..... ۳-۱۰- درون یابی پوش طیف                                       |
| ۴۵ | ..... ۳-۱۱- سنتز سیگنال گفتار                                       |
| ۴۶ | ..... ۳-۱۲- ارزیابی مدل تعمیم یافته هارمونیکی                       |

### فصل ۴.

#### طراحی مجموعه توابع تبدیل اطلاعات طیفی وابسته به متن

|    |                      |
|----|----------------------|
| ۴۹ | ..... ۴-۱- پیش گفتار |
|----|----------------------|

|    |   |
|----|---|
| ۵۱ | ۲-۴- بررسی روشهای مطرح در تبدیل گفتار.....                                    |
| ۵۱ | ۴-۲-۱- تبدیل گفتار به روش VQ.....   |
| ۵۲ | ۴-۲-۱-۱- مرحله یادگیری.....   |
| ۵۳ | ۴-۲-۱-۲- مرحله تبدیل گفتار.....   |
| ۵۳ | ۴-۲-۲- تبدیل گفتار به روش LMR.....  |
| ۵۳ | ۴-۲-۲-۱- مرحله یادگیری.....   |
| ۵۴ | ۴-۲-۲-۲- مرحله تبدیل گفتار.....   |
| ۵۵ | ۴-۲-۳- تبدیل گفتار به روش GMM.....  |
| ۵۶ | ۴-۲-۳-۱- مرحله یادگیری.....   |
| ۵۸ | ۴-۲-۳-۲- مرحله تبدیل گفتار.....   |
|    | ۴-۳- طراحی روش مناسب جهت همترازی زمانی گفتار دو گوینده در مرحله آموزش و       |
| ۵۸ | ارزیابی.....  |
| ۶۰ | ۴-۳-۱- بررسی روش های همترازی زمانی داده های گفتاری آموزشی.....                |
| ۶۱ | ۴-۴- بهبود کیفیت سیستم تبدیل گفتار با استفاده از مدل GMM(2).....              |
| ۶۲ | ۴-۴-۱- روش آموزش.....   |
| ۶۴ | ۴-۴-۲- مرحله تبدیل.....   |
| ۶۷ | ۴-۴-۳- نتایج شبیه سازی.....   |
| ۶۹ | ۴-۴-۶- جمع بندی و نتیجه گیری.....   |
| ۷۰ | ۴-۵-۵- بهبود کیفیت سیستم تبدیل گفتار با استفاده از مدل GMM و GHM.....         |
| ۷۰ | ۴-۵-۱- مدل تعمیم یافته هارمونیک (GHM).....                                    |
| ۷۲ | ۴-۵-۲- نتیجه شبیه سازی.....   |
|    | ۴-۶-۶- بهبود کیفیت سیستم تبدیل گفتار با استفاده از مدل GMM و GHM و استفاده از |
| ۷۳ | واحدهای گفتاری نیمه هجا.....  |
| ۷۴ | ۴-۶-۱- استفاده از واحد گفتاری نیمه هجا در سیستم تبدیل گفتار.....              |
| ۷۸ | ۴-۶-۲- سیستم تبدیل گفتار مبتنی بر GMM و وابسته به متن.....                    |
| ۷۸ | ۴-۶-۳- نتیجه شبیه سازی.....   |
| ۸۰ | ۴-۷- جمع بندی و نتیجه گیری.....   |

## فصل ۵.

### سیستم تبدیل گفتار بر پایه TSVQ

|     |  |
|-----|--|
| ۸۴  | ۱-۵- پیش گفتار.....                            |
| ۸۵  | ۲-۵- تبدیل گفتار به روش کتاب کد.....           |
| ۸۸  | ۳-۵- ارائه روش جدید TSVQ برای تبدیل گفتار..... |
| ۹۰  | ۴-۵- مرحله آموزش.....                          |
| ۹۰  | ۱-۴-۵- روش تصفیه مجموعه های آموزشی.....        |
| ۹۱  | ۵-۵- مرحله تبدیل.....                          |
| ۹۱  | ۱-۵-۵- ساختار شبکه.....                        |
| ۹۱  | ۲-۵-۵- محاسبه هزینه هر گره.....                |
| ۹۴  | ۳-۵-۵- تعریف شاخه هر گره.....                  |
| ۹۵  | ۴-۵-۵- تعریف تابع هزینه.....                   |
| ۹۶  | ۵-۵-۵- الگوریتم جستجو.....                     |
| ۹۶  | ۶-۵-۵- تبدیل و سنتز.....                       |
| ۹۸  | ۶-۵- پیاده سازی سیستم تبدیل گفتار.....         |
| ۹۹  | ۷-۵- نتایج شبیه سازی.....                      |
| ۱۰۱ | ۸-۵- جمع بندی و نتیجه گیری.....                |

## فصل ۶. تایج نهایی، نوآوری ها و کارهای آینده

|     |  |
|-----|--|
| ۱۰۴ | ۱-۶- پیش گفتار.....                    |
| ۱۰۴ | ۲-۶- نتایج.....                        |
| ۱۰۷ | ۳-۶- نوآوری های رساله.....             |
| ۱۰۸ | ۴-۶- پیشنهاداتی برای کارهای آینده..... |
| ۱۱۱ | مراجع.....                             |

## فهرست شکل ها

شماره صفحه

- شکل ۱-۱ بلوک دیاگرام مرحله آموزش سیستم تبدیل گفتار ..... ۷
- شکل ۲-۱ بلوک دیاگرام مرحله تبدیل سیستم تبدیل گفتار ..... ۷
- شکل ۱-۳ مثالی از مسیر فرکانسی مدل سینوسی. (a) مثالی از نحوه مرگ و میر پارامترها (b) مسیرمنتجه از مسیرهای فرکانس سینوسی در تعیین نواحی واکنش از بی واک [۳۹] ..... ۲۹
- شکل ۲-۳ بلوک دیاگرام سیستم آنالیز کننده و سنتز کننده مدل سینوسی [۴۶] ..... ۳۰
- شکل ۳-۳ بلوک دیاگرام سیستم آنالیزر/سنتز سائزر ..... ۳۳
- شکل ۴-۳ سیگنال نرمالیزه شده و نحوه تعیین پارامترهای لازم برای استخراج نقاط قابل اطمینان که در آن  $Z_i$  نقطه گذر از صفر سیگنال بوده و  $V_i$  و  $P_i$  به ترتیب نزدیکترین دره و قله به  $Z_i$  از میان نقاط روی سیگنال است. .... ۳۸
- شکل ۵-۳ نشان دهنده نحوه محاسبه سیگنال نرمالیزه ..... ۳۹
- شکل ۶-۳ نحوه محاسبه نقاط قابل اطمینان ..... ۳۹
- شکل ۷-۳ نمایش سیگنال گفتار به همراه سیگنال نرمالیزه شده و علائم پیچ مربوط به آن ..... ۴۱
- شکل ۸-۳ نحوه تعیین نقاط قابل اطمینان و محاسبه قله های پوش طیف سیگنال در مضارب فرکانس پیچ ۴۴
- شکل ۹-۳ نمایش نمونه سیگنال اصلی و سیگنال سنتز شده به همراه طیف آن ..... ۴۶
- شکل ۱-۴ بلوک دیاگرام سیستم کلاسیک تبدیل گفتار شامل استخراج ویژگی، تابع تبدیل و سنتز کننده گفتار ..... ۴۹
- شکل ۲-۴ بلوک دیاگرام تبدیل گوینده مبدأ به گوینده مقصد با روش VQ ..... ۵۳
- شکل ۳-۴ نمایش نحوه پیاده سازی تابع تبدیل توسط مجموعه ای از مخلوط های وزن دار شده ..... ۵۵
- شکل ۴-۴ نحوه یادگیری برای روش GMM ..... ۵۶
- شکل ۵-۴ بلوک دیاگرام مورد استفاده در سیستم تبدیل گفتار برای روش GMM ..... ۵۸
- شکل ۶-۴ قیود مکانی مختلف مورد استفاده برای همترازی زمانی در روش DTW ..... ۶۱
- شکل ۷-۴ بلوک دیاگرام سیستم تبدیل گفتار به روش GMM(2) ..... ۶۵
- شکل ۸-۴ بلوک دیاگرام سیستم طبقه بندی واحدهای گفتاری ..... ۶۸
- شکل ۹-۴ اسپکتروگرام مربوط به دنبال های واکه اتصالی در: الف) مرز واکه ب) هسته واکه ..... ۷۶
- شکل ۱-۵ بلوک دیاگرام مرحله یادگیری سیستم تبدیل گفتار با استفاده از تبدیل کتاب کد ..... ۸۶
- شکل ۲-۵ بلوک دیاگرام مرحله تبدیل سیستم تبدیل گفتار با استفاده از تبدیل کتاب کد ..... ۸۶

- شکل ۳-۵ فرایند همترازی زمانی دو عبارت مشابه از دو گوینده مختلف. ۸۸.....
- شکل ۴-۵ انتخاب ۲ نمایه قاب در فرایند همترازی زمانی دو عبارت مشابه از دو گوینده مختلف. ۸۹.....
- شکل ۵-۵ نمایش یک ساختار شبکه ای با L مرحله و N حالت به همراه تعدادی گره و شاخه های وارد شده به هر گره و در نهایت انتخاب یک مسیر به عنوان مسیر بهینه با کمترین هزینه ۹۲.....
- شکل ۶-۵ الگوریتم تعیین هزینه برای هر گره از شبکه. ۹۳.....
- شکل ۷-۵ الگوریتم تعیین شاخه برای هر گره از شبکه. ۹۵.....
- شکل ۸-۵ بلوک دیاگرام سیستم تبدیل گفتار جدید. ۹۸.....

## فهرست جداول

شماره صفحه

- جدول ۱-۲** لیست ۸۱ عبارت شامل تمامی نیمه هجا های CV و VC با هسته /a/ , /@/ , /e/ در زبان فارسی مورد استفاده در تهیه پایگاه آموزشی سیستم تبدیل گفتار ..... ۱۷
- جدول ۲-۲** لیست ۸۲ عبارت شامل تمامی نیمه هجا های CV و VC با هسته /u/ , /o/ , /i/ در زبان فارسی مورد استفاده در تهیه پایگاه آموزشی سیستم تبدیل گفتار ..... ۱۸
- جدول ۳-۲** لیست ۲۶ عبارت شامل تمامی نیمه هجا های CV و VC با هسته /w/ در زبان فارسی مورد استفاده در تهیه پایگاه آموزشی سیستم تبدیل گفتار ..... ۱۹
- جدول ۴-۲** نتایج حاصل از تست ABX ..... ۲۱
- جدول ۱-۳** نتایج پیاده سازی الگوریتم جدید ..... ۴۴
- جدول ۱-۴** نتایج آزمون کمی از اعوجاج متوسط (برحسب dB) بین بردار پارامترهای طیفی متناظر دو گوینده قبل و بعد از تبدیل ..... ۶۹
- جدول ۲-۴** نتایج آزمون شنیداری MOS برای روشهای مختلف تبدیل گفتار ..... ۷۰
- جدول ۳-۴** مقایسه نتایج آزمون کمی از اعوجاج متوسط (برحسب dB) بین بردار پارامترهای طیفی متناظر دو گوینده قبل و بعد از تبدیل ..... ۷۳
- جدول ۴-۴** مقایسه نتایج آزمون شنیداری MOS برای روشهای مختلف تبدیل گفتار ..... ۷۳
- جدول ۵-۴** نمایش ۱۴ گروه نیمه هجا مورد استفاده در سیستم تبدیل گفتار ..... ۷۵
- جدول ۶-۴** مقایسه نتایج آزمون کمی از اعوجاج متوسط (برحسب dB) بین بردار پارامترهای طیفی متناظر دو گوینده قبل و بعد از تبدیل ..... ۷۹
- جدول ۷-۴** مقایسه نتایج آزمون شنیداری MOS برای روشهای مختلف تبدیل گفتار ..... ۷۹
- جدول ۸-۴** مقایسه متوسط کاهش اعوجاج برای روش های مختلف (برحسب dB) بین بردار پارامترهای طیفی متناظر دو گوینده قبل و بعد از تبدیل ..... ۸۰
- جدول ۹-۴** مقایسه متوسط نتایج آزمون شنیداری MOS برای روشهای مختلف تبدیل گفتار ..... ۸۰
- جدول ۱-۵** مقایسه نتایج آزمون کمی از اعوجاج متوسط (برحسب dB) بین بردار پارامترهای طیفی متناظر دو گوینده قبل و بعد از تبدیل ..... ۱۰۱
- جدول ۲-۵** مقایسه نتایج آزمون شنیداری MOS برای روشهای مختلف تبدیل گفتار ..... ۱۰۱

# فصل ۱

معرفی و تشریح کاربرد سیستم های تبدیل گفتار و

اهداف رساله

### ۱-۱- پیش گفتار

گفتار، یکی از ساده ترین راه های برقراری ارتباط میان مردم می باشد. از زمان آغاز تحقیقات بر روی روش های ایجاد ارتباط میان انسان و ماشین، گفتار به عنوان یکی از مطلوب ترین راه های تعامل با کامپیوتر بوده است. در نتیجه بازشناسی گفتار و مبدل متن به گفتار به عنوان موضوعات مهم در ایجاد ارتباط موثر میان ماشین و انسان مورد مطالعه قرار گرفته است.

به منظور طبیعی تر شدن محاوره میان انسان و ماشین، بایستی تمامی جنبه های موثر در کیفیت گفتار ماشینی در نظر گرفته شوند. لذا گفتار حاصله علاوه بر داشتن مفهوم و تمامی اطلاعات، بایستی اطلاعاتی در مورد احساسات و هویت گوینده را منتقل سازد.

هویت گوینده یا به عبارت دیگر، صدای فرد، به عنوان یک ویژگی کلیدی در محاورات شفاهی می باشد. هویت گوینده امکان تمایز میان گویندگان را در یک کنفرانس تلفنی، برنامه رادیویی و ... برای ما فراهم می سازد. از طرفی هنگام استفاده از آن در سیستم های محاوره کامپیوتری اغلب افراد خواستار محاوره با کامپیوتر با صدای یک فرد خاص هستند. تا جایی که گاهی این سیستم ها بدلیل مورد پذیرش قرار نگرفتن صدای مورد استفاده برای کامپیوتر مورد رغبت دیگران قرار نمی گیرد. در سالهای اخیر از فن آوری تبدیل گفتار به منظور روشی برای کنترل هویت صوتی<sup>۱</sup> صدای طبیعی و یا گفتار سنتز شده استفاده می شود.

در این رساله به مطالعه و طراحی سیستم تبدیل گفتار با کیفیت بالا برای زبان فارسی می پردازیم. در بخش مقدمه، مفهوم تبدیل گفتار تعریف می شود. در فصل ۱-۲ کاربردهای این تکنولوژی مورد بررسی قرار خواهد گرفت. در فصل ۱-۳ اجزای اصلی سیستم بررسی می گردد و در ادامه اهداف مدنظر رساله تشریح می گردد. در بخش آخر نیز مروری بر فصول رساله صورت می گیرد.

<sup>۱</sup>-Voice Identity



## ۱-۲- تعریف سیستم های تبدیل گفتار

سیستم های تبدیل گفتار، صدای یک گوینده (گوینده مبدا) را به گونه ای تغییر می دهند که گویی یک گوینده دیگر (گوینده هدف) آن را بیان کرده است. بنابراین با فرض داشتن دو گوینده، هدف از بکارگیری یک سیستم تبدیل گفتار عبارت است از: تعیین یک تابع تبدیل به قسمی که گفتار گوینده مبدا چنان به نظر برسد که توسط گوینده مقصد گفته شده باشد. سیگنال گفتار حاوی انواع مختلفی از اطلاعات از قبیل: محتوای زبانی<sup>۱</sup>، هویت گوینده و نویز محیطی<sup>۲</sup> است. زمینه های متعددی از فن آوری گفتار بر روی هر کدام از این انواع اطلاعات متمرکز می باشند. تمرکز سیستم های تبدیل گفتار بر روی هویت گوینده است.

## ۱-۳- کاربردهای تبدیل گفتار

VC<sup>۳</sup> اساساً به عنوان روشی برای ایجاد گویندگان جدید در سیستم تبدیل متن به گفتار مورد استفاده قرار گرفته است. کاربردهای دیگر سیستم VC در حوزه های دیگری از قبیل ترجمه خودکار گفتار به گفتار، آموزش، مدد رسانی پزشکی و سرگرمی مشاهده شده است.

### ۱-۳-۱- کاربرد در سیستم های TTS<sup>۴</sup>:

کیفیت سنتز کننده متن به گفتار با بهره گیری از مجموعه گفتاری بزرگ و همچنین فن آوری انتخاب واحد<sup>۵</sup> افزایش می یابد. این دسته از سیستم ها غالباً با نام TTS بر پایه گفتار<sup>۶</sup>، قادر به تولید گفتار سنتز شده با انتخاب مناسب ترین دنباله واحدهای آکوستیکی از یک پایگاه گفتاری<sup>۷</sup> وابسته به گوینده و سپس بهره گیری از یک استراتژی هموارسازی<sup>۸</sup> در مکان اتصال واحدهای انتخاب شده

<sup>۱</sup>-Linguistic Content

<sup>۲</sup>-Environmental Noise

<sup>۳</sup>-Voice Conversion

<sup>۴</sup>-Text To Speech

<sup>۵</sup>-Unit Selection

<sup>۶</sup>-Corpus-Based TTS

<sup>۷</sup>-Database

<sup>۸</sup>-Smoothing Strategy

می باشد. بنابراین به منظور سنتز دسته های دیگر مانند گفتار احساسی<sup>۱</sup>، شعر<sup>۲</sup> و یا گفتار با صدای گویندگان متعدد بایستی نمونه های گفتار نماینده از قبل ضبط شده باشند. همچنین کیفیت بالای گفتار سنتز شده نیازمند در اختیار داشتن یک مجموعه گفتاری با حجم زیاد می باشد. از طرفی ضبط گفتار و پردازش اطلاعات برای یک TTS مستلزم صرف هزینه و زمان زیادی می باشد. VC به عنوان یک راه حل، وسیله ای سریع و ارزان برای تولید صداهای جدید برای یک TTS می باشد. با استفاده از VC امکان خواندن نامه های الکترونیکی و SMS<sup>۳</sup> با صدای فرستنده بدست می آید. همچنین امکان استفاده از صدای خود و یا سایر دوستان به شخصیت های بازی رایانه ای و یا برای تخصیص صداهای مختلف به کاربردهای مختلف کامپیوتر فراهم می گردد. همچنین VC می تواند با تغییر و اصلاح ویژگی های عروضی<sup>۴</sup> گفتار روی یک جمله طبیعی، آن را به یک گفتار احساسی تبدیل نماید.

### ۱-۳-۲- ترجمه خودکار گفتار به گفتار<sup>۵</sup>:

در کاربرد ترجمه برخط تلفنی، به این فناوری برای مواردی که نیازمند قابلیت تشخیص هویت گوینده در گفتار تبدیل شده توسط شنوندگان هستیم، بسیار ضروری است. برای مثال، در یک تماس تلفنی با بیش از دو نفر، توانایی شناسایی گویندگان مختلف توسط صدایشان بسیار مهم می باشد.

### ۱-۳-۳- آموزش<sup>۶</sup>:

هنگام یادگیری یک زبان خارجی، تلفظ صحیح جملات و ادای واج هایی که در زبان مادری وجود ندارد یکی از مشکل ترین موارد برای دانش آموزان است [۳۶] و [۳۷]. استفاده از VC می تواند برای یادگیری زبان موثر واقع شود، خصوصا در تلفظ تمرینات، هنگامی که دانش آموز به تلفظ زبان خارجی با صدای خودش گوش خواهد داد.

<sup>۱</sup>- Emotional

<sup>۲</sup>- Expressive

<sup>۳</sup>- Short Message Service

<sup>۴</sup>- Prosodic Modification

<sup>۵</sup>- Automatic Speech to Speech Translation

<sup>۶</sup>- Education

### ۱-۳-۴- یاری رسانی پزشکی<sup>۱</sup>:

حوزه دیگر کاربرد VC در بازسازی صدای افرادی است که صدای آنها دچار عیب شده است [۲۷]. در این صورت از سیستم تبدیل گفتار برای بهبود قابلیت فهم گفتار غیر عادی تلفظ شده بوسیله گوینده ای که دستگاه صوتی او دچار مشکل شده است می توان استفاده کرد.

کاربرد دیگر سیستم VC در بازگرداندن قدرت شنوایی برای افرادی است که دچار اختلال شنوایی در بازه های مختلف فرکانسی می باشند. در این صورت به کمک فن آوری VC می توان با تغییر بازه فرکانسی قدرت بازشناسی فرد معلول را افزایش داد.

### ۱-۳-۵- سرگرمی<sup>۲</sup>:

یکی از کاربردهای پر طرفدار VC در حوزه سرگرمی ، استفاده از نرم افزارهایی است که امکان خواندن یک متن با صدای دیگر افراد را فراهم می آورد. کاربرد دیگر این سیستم در صداگذاری بر روی فیلم ها و یا تغییر متن در محتوای فیلم است [۳۰]. در این کاربرد با بکارگیری تنها چند صداگذار، قادر به تولید صدای هنرپیشگان با زبان های مختلف می باشیم. کاربرد دیگر صداگذاری، بازگرداندن صدای بازیگرانی است که کیفیت صدای آنها بدلیل کهولت سن از دست رفته است. تلفیق سیستم VC با فن آوری پویا نمایی<sup>۳</sup> بعدی<sup>۴</sup> چهره ، امکان بوجود آوردن شخصیت های مجازی با ویژگی های مطلوب صوتی را برای کاربردهای گوناگون مانند بازی های ویدیویی فراهم می آورد. بجز موارد نامبرده، تبدیل گفتار را می توان در بسیاری از کاربردهای دیگر مورد استفاده قرار داد. به عنوان مثال در کدینگ گفتار با پهنای خیلی کم<sup>۴</sup>، گفتار ارسال شده بدون اطلاعات مربوط به هویت گوینده بوده و این اطلاعات در مرحله رمزگشایی در مقصد، به گفتار اضافه می گردد.

<sup>۱</sup>- Medical Aids

<sup>۲</sup>- Entertainment

<sup>۳</sup>- Facial Animation

<sup>۴</sup>- Very Low Bandwidth Speech Encoding

علاوه بر این، VC با در اختیار گذاردن اطلاعات سطح بالا در خصوص هویت گوینده، موجب افزایش کارایی فن آوری های دیگر در حوزه گفتار مانند بازشناسی گوینده<sup>۱</sup> و بازشناسی گفتار<sup>۲</sup> می گردد.

#### ۱-۴- اجزای سیستم تبدیل گفتار

هر سیستم تبدیل گفتار دارای دو بخش مجزا می باشد که عبارتند از: مرحله آموزش توابع تبدیل مناسب برای تغییر اصلاح ویژگی های طیفی و عروسی و همچنین مرحله تبدیل گفتار گوینده.

در مراحل آموزش و تبدیل نیازمند بخش های زیر می باشیم:

۱- تهیه پایگاه گفتاری مناسب برای فرآیند تبدیل گفتار که در آن می توان از انواع مختلف دادگان از قبیل کلمات، هجاها<sup>۳</sup>، نیمه هجا و یا آواهای موجود در یک زبان استفاده کرد.

۲- بهینه سازی و توسعه یک روش آنالیز/سنتز مناسب جهت تولید گفتار با کیفیت بالا

۳- استفاده از یک روش همترازی زمانی<sup>۴</sup> که اطلاعات طیفی و عروسی متناظر از پایگاه گفتاری دو گوینده را با حداکثر تطبیق زمانی همتراز نماید.

۴- تکنیک های طراحی تبدیل های مناسب اطلاعات طیفی و پروردیک

در مرحله آموزش، نمونه های گفتار و اطلاعات وابسته به آن برای هر دو گوینده مبدا و مقصد مورد تحلیل قرار گرفته و ویژگی های متناظر با یکدیگر همتراز می شوند و از این اطلاعات برای آموزش توابع تبدیل استفاده می گردد. بنابراین، برای هر زوج جدید مبدا و مقصد نیازمند اجرای یک آموزش جدید روی مجموعه گفتاری می باشیم. بلوک دیاگرام مرحله آموزش سیستم تبدیل گفتار در شکل زیر نشان داده شده است.

<sup>۱</sup>- Speaker Recognition

<sup>۲</sup>- Speech Recognition

<sup>۳</sup>- Syllables

<sup>۴</sup>- Time Alignment