





سازمان پژوهش و فناوری

به نام خدا

## منشور اخلاق پژوهش

بیاییم از خداوند سبحان و اعمقاً به این که عالم محضر خداست و همواره ناظر بر اعمال انسان و به منظور پاس داشتن مقام بلند دانش و پژوهش و نظر به اهمیت جایگاه دانشگاه در اعلائی فرهنگ و تمدن بشری، ما دانشجویان و اعضاء هیات علمی واحد های دانشگاه آزاد اسلامی متعهد می گردیم اصول زیر را در انجام فعالیت های پژوهشی مد نظر قرار داده و از آن تخلفی نکنیم:

- ۱- اصل برائت: التزام به برائت جویی از هرگونه رفتار غیر حرفه ای و اعلام موضع نسبت به کسانی که حوزه علم و پژوهش را به شائبه های غیر علمی می آلائند.
- ۲- اصل رعایت انصاف و امانت: تعهد به اجتناب از هرگونه جانب داری غیر علمی و حفاظت از اموال، تجهیزات و منابع در اختیار.
- ۳- اصل ترویج: تعهد به رواج دانش و ارائه نتایج تحقیقات و انتقال آن به همکاران علمی و دانشجویان به غیر از مواردی که منع قانونی دارد.
- ۴- اصل احترام: تعهد به رعایت حریم ها و حرمت ها در انجام تحقیقات و رعایت جانب تقد و خودداری از هرگونه حرمت شکنی.
- ۵- اصل رعایت حقوق: التزام به رعایت کامل حقوق پژوهشگران و پژوهشگران (انسان، حیوان و نبات) و سایر صاحبان حق.
- ۶- اصل رازداری: تعهد به صیانت از اسرار و اطلاعات محرمانه افراد، سازمان ها و کشور و کلیه افراد و نهادهای مرتبط با تحقیق.
- ۷- اصل حقیقت جویی: تلاش در راستای پی جویی حقیقت و وفاداری به آن و دوری از هرگونه پنهان سازی حقیقت.
- ۸- اصل مالکیت مادی و معنوی: تعهد به رعایت کامل حقوق مادی و معنوی دانشگاه و کلیه همکاران پژوهش.
- ۹- اصل منافع ملی: تعهد به رعایت مصالح ملی و در نظر داشتن پیشبرد و توسعه کشور در کلیه مراحل پژوهش.



واحد علوم و تحقیقات کرمانشاه

## تعهد نامه اصالت رساله یا پایان نامه

اینجانب حمید رضا نیرومند دانش آموخته مقطع کارشناسی ارشد ناپیوسته در رشته مهندسی فناوری اطلاعات که در تاریخ ۱۳۹۲/۱۰/۳۰ از پایان نامه خود تحت عنوان " بررسی روش‌های تبدیل متن به گفتار و ارائه راهکاری عملی جهت کاهش خستگی ذهنی ناشی از شنیدن گفتار ماشینی یکنواخت " با کسب نمره‌ی و درجه‌ی دفاع کرده‌ام بدین وسیله متعهد می‌شوم:

۱- این پایان نامه حاصل تحقیق و پژوهش انجام شده توسط اینجانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران (اعم از پایان نامه، کتاب، مقاله و...) استفاده کرده‌ام، مطابق ضوابط و رویه‌ی موجود، نام منبع مورد استفاده و سایر مشخصات آن را در فهرست مربوطه ذکر و درج کرده‌ام.

۲- این پایان نامه قبلاً برای دریافت هیچ مدرک تحصیلی (همسطح، پایین‌تر یا بالاتر) در سایر دانشگاه‌ها و مؤسسات آموزش عالی ارائه نشده است.

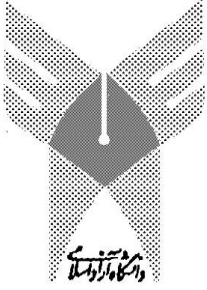
۳- چنانچه بعد از فراغت از تحصیل، قصد استفاده و هرگونه بهره‌برداری اعم از چاپ کتاب، ثبت اختراع و... از این پایان‌نامه داشته باشم، از حوزه معاونت پژوهشی واحد مجوزهای مربوطه را اخذ نمایم.

۴- چنانچه در هر مقطع زمانی خلاف موارد فوق ثابت شود، عواقب ناشی از آن را می‌پذیرم و واحد دانشگاهی مجاز است با اینجانب مطابق ضوابط و مقررات رفتار کرده و در صورت ابطال مدرک تحصیلی‌ام هیچگونه ادعایی نخواهم داشت ./ت

نام و نام خانوادگی:

حمید رضا نیرومند

تاریخ و امضا:



دانشگاه آزاد اسلامی  
واحد علوم و تحقیقات کرمانشاه  
دانشکده فنی و مهندسی

پایان نامه برای دریافت درجه کارشناسی ارشد (M.Sc)

رشته: مهندسی فناوری اطلاعات

گرایش: شبکه های کامپیوتری

عنوان:

بررسی روش های تبدیل متن به گفتار و ارائه راهکاری عملی جهت  
کاهش خستگی ذهنی ناشی از شنیدن گفتار ماشینی یکنواخت

استاد راهنما:

جناب آقای دکتر فرهاد مردوخی

استاد مشاور:

جناب آقای دکتر محمود احمدی

نگارش:

حمید رضا نیرومند

زمستان ۱۳۹۲

## سپاسگزاری:

با سپاس از خداوند کریم که قدرت خواندن و نوشتن و درک به حقیر عنایت فرمود.

دروود فراوان به روح پر فتوح پدر بزرگووارم، تقدیر و تشکر فراوان از مادر عزیزم که با دعای خیرشان همیشه بدرقه کننده راه و یاورم بودند و همچنین سپاس بیکران بر همدلی و همراهی تمامی دوستانی که این بنده کمترین را در این عرصه کمک نمودند.

با تقدیر و تشکر شایسته از استادان فرهیخته و فرزانه، حکیم دلسوز جناب آقای دکتر فرهاد مردوخی و جناب آقای دکتر محمود احمدی که با نکته‌های ناب و گفته‌های بلند، همواره راهنما و راه‌گشای نگارنده در اتمام و اکمال پایان‌نامه بوده است.

با تشکر از همه معلمان و اساتید عزیزم که با شمع وجودشان، روشنی‌بخش تمامی مراحل زندگی‌ام بودند و هستند.

از خداوند تبارک و تعالی توفیق و عنایات برای همه عزیزان را خواستارم.

تقدیم به :

قطب عالم امکان، امام عصر، مهدی موعود (عجل الله تعالی فرجه الشریف)

و عاشقان علم آموزی

## فهرست مطالب

چکیده:	۵
فصل اول	۱
۱-۱- مقدمه	۲
۲-۱- تعریف تبدیل متن به گفتار و اهداف آن	۳
۳-۱- کاربردهای سیستم تبدیل متن به گفتار و کاربران عمده آن	۴
۴-۱- بیان مسأله	۶
۵-۱- اهمیت و ضرورت پژوهش	۷
۶-۱- فرضیه‌های پژوهش	۸
۷-۱- اهداف پژوهش	۸
۱-۷-۱- اهداف علمی:	۹
۲-۷-۱- اهداف کاربردی:	۹
۸-۱- تعریف متغیرها	۹
۱-۸-۱- تعریف تبدیل متن به گفتار	۹
۲-۸-۱- تعریف لحن در گفتار	۹
۳-۸-۱- تعریف صدای یکنواخت	۹
۹-۱- راهکار پیشنهادی	۹
۱۰-۱- ساختار پایان‌نامه	۱۰
فصل دوم	۱۱
۲-۲- تأثیر «لحن» در روان‌شناسی گفتار:	۱۲
۳-۲- لحن در TTS	۱۳
۴-۲- کارهای انجام شده بر روی زیر و بمی و نوای فارسی	۱۳
۲-۵- خلاصه فصل	۱۸

فصل ۳:	۲۰
۳-۱- مقدمه	۲۱
۳-۲- توصیف راهکار	۲۱
۳-۲-۱- ایده اول: کاهش و افزایش صدا به صورت سینوسی	۲۱
۳-۲-۲- ایده دوم: کاهش در سکوت، افزایش در ۵ ثانیه	۲۲
۳-۲-۳- ایده سوم: کاهش در سکوت، افزایش تا سکوت بعدی	۲۲
۳-۳- یافتن سکوت در کاربرد تبدیل متن به گفتار	۲۳
۳-۴- یافتن سکوت در کاربرد صدای آنالوگ ضبط شده	۲۴
۳-۴-۱- کمترین میدان در فواصل زمانی مشخص	۲۴
۳-۴-۲- اولین لحظه‌ای که میدان از مقدار مشخصی کمتر شود	۲۴
۳-۵- یافتن سکوت در کاربرد پخش زنده صدا	۲۴
۳-۶- افزایش به نرمی در کاربرد تبدیل متن به صدا	۲۵
۳-۷- افزایش به نرمی در کاربرد صدای آنالوگ ضبط شده	۲۵
۳-۸- افزایش به نرمی در کاربرد پخش زنده صدا	۲۵
۳-۹- خلاصه فصل	۲۶
فصل ۴:	۲۷
۴-۱- مقدمه	۲۸
۴-۲- نرم افزار تبدیل متن به گفتار پارس خوان	۲۸
۴-۳- خلاصه‌ای از الگوریتم تبدیل متن به گفتار در پارس خوان	۲۹
۴-۴- متن انتخابی برای تبدیل به گفتار	۳۰
۴-۵- روش شبیه‌سازی ایده‌ی این پژوهش بر روی خروجی پارس خوان	۳۱
۴-۶- روش اول اندازه‌گیری: استفاده از دستگاه EEG	۳۲
۴-۶-۱- تأثیرات صدا بر روی مغز انسان و احساس تأثیرات با دستگاه EEG	۳۲
۴-۶-۲- نوار مغز چیست؟	۳۷



- ۳۸ ..... ۳-۶-۴-توصیف روال انجام آزمایش
- ۴۰ ..... ۷-۴- نتیجه اندازه‌گیری
- ۴۰ ..... ۸-۴- روش دوم اندازه‌گیری: نظرسنجی
- ۴۲ ..... ۹-۴- نتیجه اندازه‌گیری به روش نظرسنجی
- ۴۴ ..... ۱۰-۴- خلاصه فصل
- ۴۵ ..... فصل ۵:
- ۴۶ ..... ۵-۱- نتیجه‌گیری
- ۴۷ ..... ۲-۵- پیشنهادات
- ۴۷ ..... ۱-۲-۵- تخمین سکوت بعد
- ۴۷ ..... ۲-۲-۵- آثار القائات قبل از شنیدن یک صدا بر روی شنوندگان آن صدا

## فهرست اشکال

- شکل ۱ - ارتباط بخش‌های مختلف یک سیستم تبدیل متن به گفتار..... ۳
- شکل ۲ - انواع گراف مغز..... ۳۴
- شکل ۳ - ضربان دو گوشی..... ۳۵
- شکل ۴ - ایده‌ی دوم کاهش در سکوت و افزایش به نرمی..... ۲۲
- شکل ۵ - ایده‌ی سوم کاهش حجم صدا در سکوت و افزایش به نرمی..... ۲۳
- شکل ۶ - بزرگ‌نمایی ایده‌ی سوم..... ۲۳
- شکل ۷ - نمای نرم افزار پارس‌خوان..... ۲۸
- شکل ۸ - نمای نرم افزار Corel Video Studio..... ۳۲
- شکل ۹ - یکی از سوژه‌های آزمایش EEG..... ۳۸
- شکل ۱۰ - بخشی از گراف EEG مربوط به کلیپ اول..... ۳۹
- شکل ۱۱ - بخشی از گراف EEG مربوط به کلیپ دوم..... ۴۰
- شکل ۱۲ - نمایی از صفحه نظرسنجی آنلاین..... ۴۱

## چکیده:

امروزه یکی از مهم‌ترین و پرتعدادترین مباحث در زمینه علوم کامپیوتر، مبحث «تبدیل متن به گفتار» است. اما علاوه بر مشکلاتی که بر سر راه تبدیل متن به گفتار وجود دارد، یکی از مشکلات نیز بعد از انجام عملیات تبدیل متن به صدا ظاهر می‌شود و آن، خستگی ذهنی مخاطبان بر اثر شنیدن صدای تبدیل شده از روی متن یا همان «گفتار ماشینی» است. به ویژه اگر این متن کمی طولانی باشد. بنابراین، هدف از این پژوهش، بررسی تأثیر تنوع حجم صدا در رفع خستگی ذهنی ناشی از شنیدن صدای تبدیل شده از روی متن به ویژه هنگام شنیدن صدای مربوط به متون طولانی می‌باشد.

ما در این پژوهش، ایده «کاهش حجم صدا در سکوت و افزایش حجم صدا به نرمی» را مطرح کرده‌ایم و ایده‌مان را از طریق بررسی گراف به دست آمده از طریق دستگاه EEG<sup>1</sup> که نتیجه آزمایش صدای تولید شده با این راهکار بر روی مخاطبان مختلف است و همینطور تحلیل نتایج نظرسنجی انجام شده از قریب به ۸۰ مخاطب مختلف پس از شنیدن صدای تولید شده با این راهکار اثبات نموده‌ایم.

نتایج این پژوهش نشان داد که تنوع در حجم صدا هر چند ممکن است از طرف کاربر تا حدودی نامحسوس باشد اما تا حد زیادی تحمل کاربر را در هنگام شنیدن صدای یکنواخت طولانی افزایش می‌دهد.

کلیدواژه‌ها: تنوع مصنوعی حجم صدا، کاهش خستگی ذهنی ناشی از گفتار ماشینی، تبدیل متن به گفتار، بهبود صدای یکنواخت ماشینی، مشکل یکنواختی صدا

---

<sup>1</sup> Electroencephalography

فصل اول

کلیات پژوهش

## ۱-۱- مقدمه

با پیشرفت تکنولوژی، سیستم‌های تبدیل متن به گفتار هر روز بیشتر و بیشتر مورد استقبال قرار می‌گیرند، این سیستم‌ها به منظور قرائت متون برای استفاده در رایانه‌ها طراحی شده‌اند و هر روز بیشتر به استفاده از آن در دیگر سیستم‌های هوشمند توجه می‌شود. پیش‌بینی می‌شود به زودی در اکثر وسایل، قطعات هوشمندی تعبیه شود که از سیستم پردازش صوت و همچنین تبدیل متن به گفتار برای برقراری ارتباط با انسان استفاده می‌کنند. برخی از کاربردهای سیستم‌های تبدیل متن به گفتار<sup>۱</sup> عبارتند از:

- کاربرد در بازی‌ها
- کاربرد در تلفن‌های گویا (اعلام وضعیت مسیر، آب و هوا، موجودی حساب بانکی و غیره)
- خواندن و صحبت کردن برای معلولین
- آموزش‌دهنده گویا
- و ده‌ها کاربرد دیگر

برای تولید گفتار در **TTS** بایستی از روش‌ها و الگوریتم‌هایی جهت خواندن متون استفاده شود چرا که ذخیره‌سازی تمامی کلمات یک زبان (با توجه به تعداد بسیار زیاد کلمات و رشد همیشگی آن) غیرممکن و در راستای تولید گفتار طبیعی بی‌فایده خواهد بود. «زیر و بمی» یا تغییرات «فرکانس گام»<sup>۲</sup>، «دیرش»<sup>۳</sup>، «شدت»<sup>۴</sup> و نیز «درنگ»<sup>۵</sup> چهار عنصر نوایی گفتار هستند که معمولاً در سطوح مختلف اعم از هجا، واژه یا جمله اثر خود را نشان می‌دهند و در سیستم‌های **TTS** بایستی گنجانده شوند.

اساساً تمامی سیستم‌های تبدیل متن به گفتار دارای دو بخش اصلی هستند که عبارتند از [1]:

۱- استخراج اطلاعات آوایی و سایر اطلاعات زبانی مانند تکیه و نوا از متن ورودی. (مبدل متن به دنباله آوایی)

۲- تبدیل این اطلاعات آوایی به شکل موج گفتار. (سنتز گفتار)

<sup>1</sup> Speech Synthesis / Text-to-Speech (TTS)

<sup>2</sup> Text To Speech

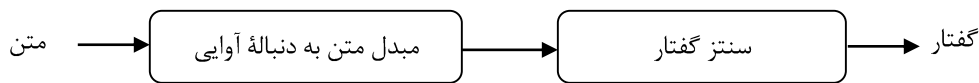
<sup>3</sup> Pitch

<sup>4</sup> Duration

<sup>5</sup> Intensity

<sup>6</sup> Pause

که نحوه ارتباط این دو بخش و تولید صدا از روی یک متن ورودی به صورت کلی در شکل زیر نشان داده شده است.



شکل 1- شکل کلی ارتباط بخش‌های مختلف یک سیستم تبدیل متن به گفتار

دو رویکرد رایج برای تولید گفتار وجود دارد، اولی روش سنتز مبتنی بر قاعده است که در آن پارامترهای مشخصه گفتار در هر بازه زمانی توسط مجموعه‌ای از قواعد تولید می‌شوند و بعدی روش اتصال قطعات گفتار که در آن واحدهای از پیش ذخیره شده‌ی صوتی برای تولید عبارتی دلخواه در کنار هم چیده می‌شوند. نمونه مشهور سنتز کننده‌های مبتنی بر قاعده، Klatt و مدل تجاری تر آن DECTalk می‌باشد.

## ۱-۲- تعریف تبدیل متن به گفتار و اهداف آن

تولید گفتار به معنای تولید مصنوعی گفتار انسان است. یک سیستم تبدیل متن به گفتار یک سیستم مبتنی بر سخت‌افزارهای دیجیتال یا رایانه است که باید توانایی خواندن متن‌های یک یا چندین زبان را داشته باشد. این متون ممکن است از طریق صفحه کلید و یا به صورت یک فایل متنی و یا پس از شناسایی توسط یک سیستم شناسایی نوری نویسه‌ها<sup>۱</sup> دریافت شوند. در حقیقت تبدیل متن به گفتار تلاش برای تقلید توانایی‌های انسان در خواندن متون است. [2]

گفتار می‌تواند از اتصال قطعات گفتاری ضبط شده که در یک دادگان ذخیره شده‌اند و یا از نتیجه اعمال یک سیگنال تحریک مناسب به فیلتری که پارامترهای آن از واحدهای گفتار طبیعی به دست آمده‌اند تولید شود. سیستم‌های تولید گفتار از جهت نوع واحد گفتاری ذخیره شده و یا روش بکار رفته در استخراج پارامترهای گفتار و نحوه استفاده مجدد از آن‌ها برای تولید گفتار متفاوت هستند.

واحدهای گفتاری واج یا دوآوایی قادر به تولید محدوده وسیعی از گفتار هستند. لیکن دسترسی به کیفیت بالایی از طبیعی بودن گفتار تولیدشده مشکل است. در کاربردهای خاص، ذخیره مستقیم کلیه کلمات و حتی جملات و استفاده مجدد از آن‌ها کیفیت بالایی را در تولید گفتار باعث می‌شود. اما این روش با محدودیت تعمیم‌پذیری در تولید متونی با محتوایی متفاوت با آنچه که کلمات و جملات آن را ضبط نموده‌ایم مواجه است.

<sup>1</sup> Optical Character Recognition (OCR)

کیفیت خروجی یک سیستم تبدیل متن به گفتار معمولاً با میزان شباهت آن با صدای انسان و میزان قابلیت فهم آن سنجیده می‌شود.

در نگاه اول شاید ساخت یک سیستم تبدیل متن به گفتار زیاد مشکل به نظر نرسد، اما در واقع این گونه نیست. در حقیقت هدف اصلی یک سیستم تبدیل متن به گفتار، تولید گفتاری هر چه شبیه‌تر به گفتار انسان است که در آن علاوه بر استفاده از دانش طبیعی، از هوش مصنوعی نیز در مواقع نقص دانش استفاده می‌شود. به نظر می‌رسد که انسان دارای یک نوع دانش و توانایی ناخودآگاه برای خواندن متون زبان مادری خود است. قابلیت‌هایی که به انسان این امکان را می‌دهد که حتی کلمات ناشناخته را نیز بیان کند. نیاز به وجود هر دو عامل هوش و دانش در کنار یکدیگر برای تولید گفتار مناسب باعث می‌شود که ساخت و توسعه سیستم‌های تبدیل متن به گفتار چندان هم ساده نباشد.

### ۳-۱- کاربردهای سیستم تبدیل متن به گفتار و کاربران عمده آن

زمینه‌های کاربردی فراوانی برای سیستم‌های تبدیل متن به گفتار وجود دارد. در واقع هرکجا که نیاز به عاملی است که وظیفه آن انتقال اطلاعات به تعدادی از مخاطبان با خواندن متن است، سیستم‌های تبدیل متن به گفتار بسیار مفید خواهند بود. استفاده از تبدیل متن به گفتار حیطه گسترده‌ای از نیازهای ما به خواندن متون، از خواندن کتب و روزنامه‌ها گرفته تا انتقال هم زمان اخبار و وقایع را شامل می‌شود. البته حتی اگر سیستم‌های تبدیل متن به گفتار قابلیت‌های خوبی در زمینه طبیعی بودن بیان و خطای بسیار ناچیز در تولید گفتاری عبارات نوشتاری داشته باشند، در مواردی چون اجرای برنامه‌های رادیویی که می‌بایست ضمن برقراری ارتباط گفتاری، ارتباط عاطفی مناسبی نیز با مخاطبین برقرار گردد، سیستم‌های تبدیل متن به گفتار فعلی هنوز به این قابلیت نرسیده‌اند و شاید بتوان گفت که این گونه جایگزینی انسان با ماشینی که فاقد روح و عاطفه و احساس است، در این گونه مواقع کمتر مورد توجه است. تعدادی از کاربردهای سیستم‌های تبدیل متن به گفتار عبارت‌اند از:

- کمک به نابینایان: یکی از اهداف طراحی سیستم‌های تبدیل متن به گفتار، کمک به افرادی بوده که از معلولیت‌های کم بینایی و نابینایی رنج می‌برند. با توجه به تعداد قابل توجه این افراد از بیماران، ارائه وسیله‌ای که بتواند آن‌ها را در استفاده از کتاب‌ها، روزنامه‌ها و سایر متون یاری نماید، سخت مورد توجه محققان بوده است. یک سیستم تبدیل متن به گفتار هم می‌تواند در راستای شیوه‌های قبل برای تهیه نوارهای کاست یا فایل‌های صوتی مورد استفاده قرار گیرد و هم این که یک فرد نابینا با داشتن چنین سیستمی خواهد توانست متنی را به عنوان ورودی به سیستم داده، به شنیدن گفتار تولیدی پردازد و بدین صورت از صحت متنی که تهیه کرده است، اطلاع یابد و یا به تصحیح آن پردازد و همچنین می‌تواند صوت نهایی به دست آمده را ذخیره کرده و به دیگران ارسال کند.

- اهداف آموزشی: تولید گفتار اگر با کیفیت مناسب انجام شود، می تواند در تهیه برنامه های آموزشی نیز استفاده شود. به عنوان مثال، می توان با پیاده سازی یک دادگان بزرگ از یک زبان خارجی به آموزش زبان پرداخت. بدین گونه که کاربر، متن مورد نظر خود را وارد نموده و منتظر بماند تا سیستم، تلفظ دقیق آن را بخواند، نظیر شرکت میکروسافت است که کاربر پس از وارد کردن **talkit** چنین سیستمی، سیستم متن، می تواند گفتار متناظر متن را به تلفظ انگلیسی یا تلفظ اسپانیولی بشنود. علاوه بر این، مبدل های متن به گفتار می توانند در آموزش دیگر علوم نیز استفاده شوند.
- استفاده جهت کاهش حجم اطلاعات: برای توضیح این قسمت به ذکر یک مثال را به مقصدی ارسال کنیم «من به خانه آمدم» می پردازیم. فرض کنید بخواهیم جمله و برای سادگی فرض می کنیم، تنها مفهوم جمله برای گیرنده کافی باشد و از اطلاعات مربوط به نوای گفتار و این که گوینده کیست، صرف نظر کنیم و همچنین فرض می کنیم، هر واج نیاز به یک بایت برای ارسال داشته باشد. از آن جا که جمله فوق از پانزده واج تشکیل شده، پس با پانزده بایت می توانیم مفهوم آن را به گیرنده ارسال کنیم، البته در صورتی که از قالب متن برای ارسال استفاده کنیم. حال فرض کنیم همین جمله بخواهد به صورت صوتی ارسال شود، اگر گفتار با فرکانس نمونه برداری ۸ کیلوهرتز نمونه برداری و هر نمونه با یک بایت ضبط شود و حدوداً دو ثانیه نیز بیان جمله طول بکشد، حجم اطلاعات حاصل از نمونه برداری گفتار فوق معادل ۱۶۰۰۰ نمونه خواهد بود، که در مقایسه با ارسال اطلاعات به صورت متن، ۱۰۶۶ برابر بیشتر است. در ضمن می دانیم که در کاربردهای امروزی استفاده از فرکانس نمونه برداری ۱۶۰۰۰ نمونه در ثانیه و ۱۶ بیت به ازای هر نمونه مرسوم تر است، پس دیده می شود که سیستم های تبدیل متن به گفتار در صورتی که از کیفیت مناسب برخوردار باشند، می توانند گزینه بسیار مناسبی جهت کاهش حجم اطلاعات ارسالی باشند.
- محاوره انسان و کامپیوتر: در صورتی که بازشناسی گفتار و تولید گفتار (تبدیل گفتار به متن و تبدیل متن به گفتار) به گونه ای مناسبی با هم ترکیب شوند، سیستم ترکیبی قادر خواهد بود تا به عنوان یک وسیله ای محاوره با انسان به کار رود، به عنوان مثال، بخش بازشناسی، سؤال شفاهی کاربر را به متن تبدیل نموده و آن گاه سیستم از طریق یک دادگان اطلاعاتی، پاسخی برای آن پرسش می یابد و سپس بخش سنتز پاسخ را برای کاربر قرائت می کند.
- ادغام با سیستم تشخیص متن دست نویسی یا تایپی: سیستم های تشخیص متن دست نویسی یا تایپی که پیشرفت های شایانی در زمینه آن انجام شده است نیز می توانند با سنتزکننده ها ترکیب شوند و امکان قرائت از روی صفحه کاغذ یا کتاب فراهم گردد و دیگر به وارد کردن مطالب مورد نظر در فایل های متنی نیاز نباشد.



- استفاده در ترجمه‌ی خودکار: در این گونه سیستم‌ها ابتدا یک جزء بازشناسی گفتار، گفتار تولید شده توسط گوینده را به معادل متنی آن تبدیل می‌کند، سپس جزء دیگر معادل کلمه گفته شده در زبان مبدأ را از طریق یک مترجم متنی به زبان مقصد تبدیل می‌کند. آن گاه سیستم سنتز گفتار، گفتار زبان مقصد را تولید می‌کند. کاربرد چنین سیستمی در کنفرانس‌ها و مجامع بین المللی، مسافرت به کشورهای خارجی و یا ارتباط از طریق صوت و بر روی اینترنت<sup>۱</sup> می‌باشد.
  - تبدیل متن به پیغام‌های شفاهی در اماکن عمومی مثل فرودگاه‌ها: در اماکن عمومی نظیر فرودگاه ه، ایستگاه‌های راه آهن و ترمینال‌های اتوبوسرانی نیز می‌توان از سیستم سنتز گفتار استفاده کرد. به عنوان مثال در یک فرودگاه، ساعات حرکت، شماره پرواز، مقصد و ساعت پرواز هواپیما و اطلاعاتی از این قبیل هر از چند گاهی به گفتار تبدیل شده و به اطلاع مسافران رسانیده می‌شود.
  - کمک به افراد گنگ برای بیان گفتار: این کاربرد از جمله کاربردهای بسیار مفیدی است که به افرادی که دارای ناتوانایی‌های گفتاری هستند کمک می‌کند که بهترین شیوه انتقال منظور خود را انتخاب نمایند.
  - سرویس‌های مدیریت ارتباط با مشتری<sup>۲</sup>: برای مؤسسات و سازمان‌هایی که مدام با مشتریان خود در ارتباط هستند و نظر آن‌ها و همچنین جلب آرای آن‌ها برای مدیران این سازمان‌ها مهم است، بسیار مفید خواهد بود که این ارتباط به صورت تمام وقت و شبانه روزی برقرار باشد و در این گونه موارد است که یک ماشین پاسخ گو با گفتاری با درجه قابل فهم بودن و خوشایند بودن بالا می‌تواند بسیار برای مدیران و مشتریان مناسب باشد.
- برای تمام زمینه‌هایی که در بالا به عنوان کاربردهای سیستم‌های تبدیل متن به گفتار برشمردیم، محصولات تجاری در بازار موجود می‌باشد. اشتیاق کاربران به استفاده از چنین سیستم‌هایی به قدری زیاد بوده که حتی سیستم‌های با کیفیت اندک نیز از اقبال و توجه عمومی فراوانی برخوردار شده‌اند.

#### ۴-۱- بیان مسأله

تبدیل متن به گفتار با دو قالب اساسی زبان یعنی قالب نوشتاری و قالب گفتاری آن در ارتباط است. در قالب نوشتاری با پیچیدگی‌های فراوان در انواع تحلیل‌های صرفی، نحوی، معنایی به منظور استخراج انواع اطلاعات مورد نیاز از متن ورودی رو به رو هستیم. بسیاری از این تحلیل‌ها در کاربردهای دیگری چون ترجمه ماشینی،

<sup>1</sup> VoIP

<sup>2</sup> CRM

خلاصه‌سازی و طبقه‌بندی متون، سیستم‌های پرسش و پاسخ، بازیابی اطلاعات، جستجو در پایگاه‌های داده متنی و مانند آن نیز مورد نیاز است. [3]

علاوه بر این در تبدیل متن به گفتار با قالب گفتاری زبان نیز سر و کار داریم و لذا با هدف دستیابی به گفتاری نزدیک به گفتار طبیعی، با مشکلات عدیده‌ای در تولید تلفظ و نوای مناسب برای گفتار بازسازی شده رو به رو می‌باشیم. این مشکلات همواره مورد توجه محققان در سراسر دنیا قرار داشته است و مقالات فراوانی همه ساله در بسیاری از کنفرانس‌های علمی معتبر ارائه می‌گردد. با وجود تحقیقات فراوان و سرمایه‌گذاری‌های بسیار در این زمینه، پیچیدگی ذاتی مسائل در این حوزه و گستردگی علوم و فناوری‌های مرتبط با آن سبب شده است که راه زیادی تا تولید محصولات ایده آل تبدیل متن به گفتار باقی باشد.

در این میان، یکی دیگر از چالش‌ها در تولید نرم افزارهای مبدل متن به صدا، این است که صدای تولیدی به طور خسته کننده‌ای، یکنواخت است. به عبارت دیگر حجم صدای تولید شده، از ابتدا تا انتهای متن، در یک سطح خواهد بود و این امر در متون طولانی (به طور مثال، اگر از نرم افزار خواسته شود که یک مقاله یا یک داستان را بخواند) بسیار خسته کننده خواهد بود.

ما فکر می‌کنیم یکی از دلایلی که باعث می‌شود ذهن مخاطب از شنیدن صدای طبیعی انسان خسته نشود اما از شنیدن صدای ماشینی احساس خستگی ذهنی کند، این است که هنگامی که یک انسان یک مقاله یا داستان را می‌خواند، به طور هوشمندانه حجم صدای خود را کم و زیاد می‌کند و یا گاهی عواملی مانند جا به جا شدن شخص و یا چرخش سر او، باعث این کم و زیاد شدن حجم صدا می‌شود و بنابراین تنوع در حجم صدا، مانع خستگی ذهنی مخاطب می‌شود.

در این تحقیق، بر آنیم تا بر اساس همین پدیده طبیعی، یک راه حل مصنوعی برای کاهش اثر خستگی در صدای تولید شده توسط الگوریتم‌ها و نرم افزارهای تبدیل متن به صدا ارائه دهیم.

## ۵-۱- اهمیت و ضرورت پژوهش

گفتار، ابزار اولیه ارتباط بین انسان‌هاست. انسان‌ها به کمک گفتار خود می‌توانند مفاهیم متفاوتی را به مخاطبان خود انتقال دهند. گفتار علاوه بر این که ابزاری مناسب برای انتقال دانسته‌ها و بیان نیازها است، بهترین روش انتقال مفاهیم ذهنی و احساسات درونی نیز به حساب می‌آید. در واقع برتری بارز گفتار بر نوشتار متناظر با آن این است که گفتار اطلاعات جانبی بیشتری را به شنونده انتقال می‌دهد که بعضاً هدف اصلی بیان جملات نیز انتقال همین اطلاعات می‌باشد. یکی دیگر از ویژگی‌های بارز گفتار سهولت انتقال آن به مخاطب است. به این ترتیب که در مواقعی مناسب‌تر است ما به جای انتقال تصویری یا نوشتاری اطلاعات آن‌ها را به صورت

گفتاری منتقل نماییم تا بازه وسیع‌تری از مخاطبان را در برگیرد. برای مثال در اماکن عمومی که نصب تابلوهای اطلاعاتی به شکل گسترده امکان‌پذیر نمی‌باشد، انتقال اطلاعات از طریق بلندگوها یک گزینه کم هزینه است. از طرف دیگر این نوع انتقال پیام موجب می‌شود که مخاطب زودتر در جریان قرار گیرد. [4]

قابلیت‌های فوق‌الذکر باعث جلب توجه و علاقه بسیاری به سمت ایجاد سیستمی جهت تولید گفتار شده است. تمام این مطالب در کنار کاربرد وسیع سیستم‌های تولید گفتار در عرصه زندگی انسان، باعث رشد روز افزون زمینه‌ای از مباحث پردازش‌های هوشمند تحت عنوان تبدیل متن به گفتار گردیده است.

از طرفی، بر اساس آنچه در مطالب آینده خواهد آمد، انسان از صدای ناهنجار دور می‌شود و به صداهای خوشایند نزدیک و این موضوع، تنها چیزی نیست که صداهای ناهنجار به آن آسیب می‌رساند! بر اساس تحقیقات، بیشتر صداهای فروشگاه‌های خرده‌فروشی، نامناسب، تصادفی و حتی خصومت‌آمیز محسوب می‌شوند و اثر باور نکردنی روی فروش دارند. برخی فروشگاه‌ها ۳۰ درصد از فروش خود را با خارج شدن سریع مشتری‌ها از مغازه یا حتی برگشتن آن‌ها از مقابل در فروشگاه، از دست می‌دهند. همه این‌ها تأثیر صدا در زندگی انسان و جذب شدن و یا نشدن او به یک موضوع خاص را می‌رساند. حال، چطور می‌توان انتظار داشت کاربران به سمت استفاده از مبدل‌های متن به گفتار جذب شوند در حالی که با شنیدن صدای تولید شده توسط آن‌ها احساس خستگی می‌کنند؟

در این تحقیق ما بر آنیم که تا حد ممکن از این اثر منفی بکاهیم تا در نتیجه‌ی آن، رغبت کاربران برای استفاده از نرم افزارهای تبدیل متن به گفتار افزایش یابد.

#### ۱-۶- فرضیه‌های پژوهش

با توجه به سوال اصلی این پژوهش و به منظور یافتن جواب مناسب، فرضیه‌های اصلی زیر در ابتدا مورد نظر بوده‌اند:

- صدا بر روی مغز انسان تأثیر می‌گذارد. [5]
- با بهبود کیفیت صدای تولیدی توسط مبدل‌های متن به گفتار، کاربران رغبت بیشتری برای استفاده از چنین نرم افزارهایی خواهند داشت.

#### ۱-۷- اهداف پژوهش

هر چند نتیجه این پژوهش می‌تواند در زمینه‌های مختلفی به کار گرفته شود اما به طور خاص، دو هدف ذیل مورد توجه بوده است:

### ۱-۷-۱- اهداف علمی:

بررسی تأثیر صدای یکنواخت تولیدی توسط مبدل‌های متن به گفتار بر روی مغز انسان و کاهش اثرات منفی آن مد نظر است.

### ۱-۷-۲- اهداف کاربردی:

ارائه یک راهکار عملی برای کاهش خستگی ذهنی ناشی از شنیدن متن تبدیل شده به گفتار توسط مبدل‌ها به ویژه در مورد متون طولانی، هدف اصلی تحقیق است.

### ۱-۸- تعریف متغیرها

#### ۱-۸-۱- تعریف تبدیل متن به گفتار

تولید گفتار به معنای تولید مصنوعی گفتار انسان است. یک سیستم تبدیل متن به گفتار یک سیستم مبتنی بر سخت‌افزارهای دیجیتال یا رایانه است که باید توانایی خواندن متن‌های یک یا چندین زبان را داشته باشد. این متون ممکن است از طریق صفحه کلید و یا به صورت یک فایل متنی و یا پس از شناسایی توسط یک سیستم شناسایی نوری نویسه‌ها<sup>۱</sup> دریافت شوند. در حقیقت تبدیل متن به گفتار تلاش برای تقلید توانایی‌های انسان در خواندن متون است. [6]

#### ۱-۸-۲- تعریف لحن در گفتار

لحن، همان جنبه موسیقایی از صدای انسان است که به بمی، حجم، سرعت و تأکید اشاره دارد. [7]

#### ۱-۸-۳- تعریف صدای یکنواخت

صدای یکنواخت به صدایی گفته می‌شود که از ابتدا تا انتها در یک سطح باقی می‌ماند و هیچ تنوعی در فشار در آن وجود ندارد. [7]

### ۱-۹- راهکار پیشنهادی

راهکار کلی مطرح شده در این پژوهش، «ایجاد تنوع مصنوعی در حجم صدا» است که این «تنوع» ممکن است به روش‌های مختلفی مانند «کاهش و افزایش سینوسی حجم صدا» مطرح شود. اما چیزی که ما پس از آزمایشات مختلف بدان رسیده‌ایم، «کاهش حجم صدا در سکوت و افزایش حجم صدا به نرمی تا سکوت بعدی» است که توضیحات بیشتر را به فصل سوم موکول می‌کنیم.

<sup>1</sup> Optical Character Recognition (OCR)

## ۱-۱۰ - ساختار پایان‌نامه

به طور کلی پایان‌نامه را در پنج فصل مختلف ارائه کرده‌ایم:

- فصل اول: کلیات پژوهش

این فصل حاوی مفاهیم و متغیرهای مورد نیاز برای درک فصول بعدی و نگاهی کلی به روال طی شده در پژوهش است.

- فصل دوم: پیش‌زمینه و راهکارهای مرتبط

در این فصل نگاهی داشته‌ایم به تلاش‌های محققین ایرانی و خارجی در زمینه تبدیل متن به گفتار و تلاش‌های صورت گرفته برای بهبود کیفیت خروجی نرم افزارهای مبدل.

- فصل سوم: راهکار پیشنهادی

در این فصل به مشکل اصلی مد نظر در این پژوهش یعنی خستگی ذهنی بر اثر شنیدن متون یکنواخت ماشینی اشاره داشته‌ایم و راهکارمان را برای کاهش این خستگی در کاربردهای مختلف به تفصیل بیان نموده‌ایم.

- فصل چهارم: شبیه‌سازی و بیان نتایج

با توجه به سابقه ارائه یک نرم افزار مبدل متن به گفتار توسط نگارنده، در این فصل به شبیه‌سازی ایده با نرم‌افزار مبدل متن به گفتار «پارس‌خوان» و بررسی نتیجه راهکار بر روی قریب به ۸۰ مخاطب مختلف پرداخته‌ایم.

- فصل پنجم: نتیجه‌گیری و بیان پیشنهادات

نهایتاً در فصل آخر یک نتیجه‌گیری کلی از آزمایشات و بررسی‌های انجام شده داشته‌ایم و همچنین پیشنهادات ارزشمندی برای برنامه‌نویسان و پژوهشگران در این حیطه داشته‌ایم.

## فصل دوم

### پیش‌زمینه و راهکارهای مرتبط

## ۱-۲- مقدمه:

آنچه از بررسی ده‌ها مقاله فارسی و انگلیسی بر می‌آید این است که بسیاری از محققان در زمینه تبدیل متن به گفتار پی به این موضوع برده‌اند که صدای تبدیل شده با الگوریتم‌های تبدیل متن به گفتار، برای مخاطب، کمی نامأنوس و خسته کننده است. اما تمامی آن‌ها به اتفاق، به این نتیجه رسیده‌اند که اگر به صدای تبدیل شده، لحن نیز اضافه گردد، این خستگی ذهنی کاهش می‌یابد. این در حالی است که اولاً به ازعان بسیاری از محققان، پیاده‌سازی لحن در TTS بسیار پیچیده است و ثانیاً ما در سخنرانی‌هایی که یک انسان به طور طبیعی صحبت می‌کند نیز شاهد هستیم که با اینکه او جملاتش دارای لحن است اما اگر به طور یکنواخت صحبت کند، مخاطبان خیلی سریع خسته خواهند شد. با توجه به تعریف صدای یکنواخت، می‌توان به این نتیجه رسید که حتی اگر لحن نیز در یک مبدل متن به گفتار پیاده‌سازی شود، باز هم نیاز به تغییر حجم صدا لازم به نظر می‌رسد.

در این فصل ما به کارهای انجام شده در زمینه پیاده‌سازی لحن در تبدیل متن به گفتار خواهیم پرداخت.

## ۲-۲- تأثیر «لحن» در روان‌شناسی گفتار:

لحن، همان جنبه موسیقایی از صدای انسان است که به بمی، حجم، سرعت و تأکید اشاره دارد. مخاطبان به طور غریزی به لحن صدا واکنش نشان می‌دهند (گاهی واکنش مثبت و گاهی منفی).

تحقیقاتی که توسط دپارتمان روان‌شناسی دانشگاه پیتسبرگ در آمریکا<sup>۱</sup> انجام شده است نشان می‌دهد که مردم به طور غریزی بر اساس لحن صدای صحبت کننده در مورد او قضاوت می‌کنند. به خصوص، انسان‌ها فکر می‌کنند کسانی که صداهای عمیق‌تری دارند برای کارهای مدیریتی بهتر هستند و برعکس، صدای نازک و تیز در دیگران حس ناتوانی در مدیریت ایجاد می‌کند.

صدای یکنواخت به صدایی گفته می‌شود که از ابتدا تا انتها در یک سطح باقی می‌ماند و هیچ تنوعی در فشار در آن وجود ندارد. روان‌شناسان سخنرانی معتقدند: با تغییر دادن تن صدا، زندگی و انرژی به پیغامی که قصد

<sup>1</sup> Department of Psychology, University of Pittsburgh, USA

انتقال آن را دارید اضافه می‌شود. احساساتی مانند شور و اشتیاق و طنز و هیجان می‌تواند از لحن صدای گوینده به مخاطب منتقل شود، بنابراین می‌توان انتظار داشت که صدایی که دارای لحن درستی نیست، روی مخاطب تأثیرات منفی مختلفی خواهد گذاشت و این، همان چیزی است که باعث می‌شود خیلی از کاربران به استفاده از نرم افزارهای تبدیل متن به صدا رغبت نداشته باشند. [8]

### ۲-۳- لحن در TTS

بخش مهمی از اطلاعات صوتی موجود در گفتار به نوای گفتار بر می‌گردد و نوای گفتار علاوه بر اینکه حاوی اطلاعات مهم گفتاری اطراف جهت تأکیدات و نوع جمله از جهت سوالی، خبری بودن است، در طبیعی بودن گفتار نیز تأثیر به‌سزایی دارد. لذا در این قسمت به فعالیت‌های صورت گرفته برای مدل‌سازی نوای گفتار فارسی شامل زیر و بمی، دیرش و شدت نیز امکان تغییر لحن جمله و نوع خبری سوالی بودن آن پرداخت شده است.

یکی از روش‌های استفاده شده در سیستم‌های مبدل متن به گفتار برای تولید لحن، روش تیلت<sup>۱</sup> می‌باشد. در این روش سعی می‌شود قسمت‌های مهم منحنی گام را که به آن‌ها رویداد می‌گویند به کمک پارامترهای تیلت مدل‌سازی کنند. منظور از قسمت‌های مهم منحنی گام بخش‌هایی است که از نظر زبان‌شناسی و درک شنیداری دارای اهمیت است و تغییرات این قسمت‌ها باعث تغییر در معنا و کیفیت گفتار می‌شود. محل این رویدادها به کمک فایل برچسب<sup>۲</sup> مشخص می‌شوند و هر رویداد به کمک پنج پارامتر تیلت مدل‌سازی می‌شود. تخمین این پارامترها و همچنین دیرش واج‌ها و شدت به کمک درخت CART<sup>۳</sup> صورت می‌گیرد. به منظور تولید منحنی گام، از روش فوجی ساکی، تیلت و منحنی‌های قطعه استفاده می‌شود. روش فوجی ساکی برای منحنی گام، دو جزء دستورات تکیه و عبارت را فرض می‌کند که هرکدام دارای پارامترهای خاص خود هستند. پارامترهای دستورات تکیه، برای هجاهای تکیه‌بر و پارامترهای دستورات عبارت، برای اولین هجای عبارت‌های نوایی گفتار تخمین زده می‌شود و با استفاده از این پارامترها، منحنی گام با به کارگیری فرمول فوجی ساکی، تولید می‌شود. به منظور تخمین پارامترها، روش‌های مارس، شبکه عصبی و ماشین پشتیبان بردار به کار گرفته شده‌اند که نتایج آزمایش‌ها نشان می‌دهد که روش مارس قادر به تخمین کلیه پارامترهای فوجی ساکی نمی‌باشد.

### ۲-۴- کارهای انجام شده بر روی زیر و بمی و نوای فارسی [2]

<sup>1</sup> Tilt

<sup>2</sup> Tag File

<sup>3</sup> Classification And Regression Tree (CART)



از لحاظ سنتی، نوا به مطالعه قوانین مربوط به گرامر وزن و قافیه و آهنگ در شعر، اشاره دارد. نوا با آن الگوهایی سر و کار دارد که در سنت‌های ادبی خاص توسعه یافته‌اند تا امکان سرودن اشعار در آن سبک‌ها را فراهم نمایند. در همین راستا نویسندگانی چند به نوشتن در مورد نوا و عروض از گذشته‌های دور پرداخته‌اند. صرف نظر از برخی کارهای سنتی، و تا آن‌جا که مربوط به ایده‌های مدرن در نوا می‌شود، زبان‌شناسان و دست‌نویسانی چند بر روی ویژگی‌های زیر و بمی و آهنگ در فارسی مطالعه نموده‌اند ( [9]، [10]، [11]، [12]، [13]، [14]، [15]، [16]، [17] و ...). بسیاری از فعالیت‌ها در این زمینه تنها شامل مطالعات پایه‌ای بر روی الگوهای تکیه و واژگانی و بعضی از جنبه‌های عمومی آهنگ فارسی می‌باشد.

یکی از اولین افرادی که تحقیقات سودمندی را در مورد بررسی واج‌های فارسی و الگوهای تأکید لغوی انجام داده آقای خانلری بوده است. او برخی مفاهیم سطحی آهنگ و ویژگی‌های زبر زنجیری زبان فارسی را معرفی نموده است. از بین بررسی‌های انجام شده توسط افراد فوق، کارهای کمی را می‌توان یافت که در حد مناسبی نقش‌های آهنگ را با تمرکز بر معانی مختلفی که هر الگوی اصلی آهنگ به شنونده منتقل می‌کند بررسی کرده باشند. یکی از اولین کارهای انجام شده در این زمینه در پایان‌نامه‌ی آقای وحیدیان کامیار تحت عنوان «رویدادهای زبر زنجیری» در سال ۱۳۷۲ انجام گرفته است [18]. آقای کامیار در ابتدا، الگوهای تأکید و واژگانی فارسی از جمله محل قرارگیری تکیه در دسته‌های مختلف گرامری همچون اسم، صفت، ضمیر، حروف اضافه و غیره را مورد بررسی قرار داده است. او توجه ویژه‌ای به تغییرات تأکیدهای واژگانی در انواع جملات نموده و به بررسی پدیده‌های مختلفی از جمله تضعیف و تقویت تأکید، حذف تأکید، جا به جایی و تغییر مکان تکیه می‌پردازد.

کامیار در بررسی خود، به تأثیرهای احتمالی الگوی تأکید لغات عربی بر روی کلمات فارسی و همچنین تفاوت الگوهای تکیه و واژگانی در لهجه‌های مختلف زبان فارسی پرداخته است. او ابتدا به بررسی الگوهای تکیه در اشعار فارسی پرداخته و سپس توجه اصلی خود را به چگونگی استفاده از تکیه برای تأکید روی یک کلمه و یا گروهی از کلمات در یک جمله هنگامی که منجر به معانی مختلف شود معطوف می‌سازد. یعنی بررسی این که چگونه یک گویش بسته به نوع منظور و مقصود گوینده، به گونه‌های مختلفی خوانده می‌شود. بررسی‌های او درباره آهنگ، گونه‌های مختلف جملات شامل جملات خبری، سؤالی، امری و غیره را شامل می‌شود. کامیار سعی می‌کند که انواع مختلف آهنگ را به انواع مقصودهای گوینده ارتباط دهد. یکی از مهم‌ترین کارهای آقای کامیار در مطالعاتی که وی صورت داده است، شناسایی محل تکیه اصلی جمله و یا گروه آهنگی یا به عبارتی هجایی در درون جمله یا گروه آهنگی که قوی‌ترین تکیه را دریافت می‌نماید، می‌باشد. بنا بر این کار او علاوه بر تحقیق در زمینه آهنگ زبان فارسی، می‌تواند به عنوان کاری نه چندان عمیق در خصوص شناسایی

قوانین تکیه هسته نیز مرتبط می‌باشد که بیان‌کننده تغییرات محل هسته و تغییرات احتمالی ناشی از آن در معنای جمله است. کامیار به بعضی تغییرات در الگوی تکیه واژگانی تک تک کلمات در جمله مانند خنثی شدن تکیه واژگانی بعد از هسته در یک گروه آهنگی و نیز به بعضی از تعاملات تکیه و گرامر در زبان فارسی اشاره می‌کند.

با پیدایش نظریه‌هایی در زمینه واج‌شناسی آهنگ که قلمرو آن تنها به واج‌شناسی فراواژگانی و نمایش آوایی آن مربوط می‌شود، یکی از این نظریه‌ها که توجه زیادی را به خود معطوف داشته نظریه واج‌شناسی لایه‌ای<sup>۱</sup> است. بعضی از تحقیقات اخیر در زمینه آهنگ زبان فارسی در قالب این نظریه و خصوصاً در چارچوب رهیافت پیرهامبرت صورت گرفته است که از آن جمله می‌توان به تحقیقات آقایان محرم اسلامی [19]، [16] و بهزاد ماهجانی [17] اشاره کرد.

اسلامی با این باور تحقیق خود را انجام داده است که زیر و بمی و آهنگ دارای یک نظام واج‌شناختی است و با شناسایی عناصر بنیادین آهنگ زبان فارسی از یک طرف و ترکیب نظام‌مند آن‌ها از طرف دیگر، به ترتیب، الگوهای آهنگ و معنای فراواژگانی هر کدام از آن‌ها به دست می‌آید. در چارچوب نظریه به کار رفته در تحقیقات اسلامی، آهنگ به معنای کاربرد مشخصه‌های آوایی زبرزنجیری است که از آن برای انتقال معنای فراواژگانی (معنا در جمله) به شیوه‌ای نظام‌مند به لحاظ زبانی استفاده می‌شود. او در تحقیقات خود علاوه بر هدف تبیین نظام آهنگ و شناسایی واحدهای بنیادین آن، قصد داشته است که نتایج تحقیق خود را در تولید الگوی آهنگ جملات در تبدیل متن به گفتار فارسی به کار برد. او در تحقیق خود پس از معرفی عناصر بنیادین آهنگ به همراه معنای آن‌ها، بیان می‌دارد که چگونه از رهگذر ترکیب عناصر بنیادین، الگوهای بالقوه آهنگ شکل می‌گیرند و این که هر کدام از آن‌ها نماینده چه بافتی هستند. از مطالب مهم مطرح شده در این تحقیق تفاوت بین تکیه واژگانی و انتزاعی، ثابت و قابل پیش‌بینی بودن آن و تکیه‌ی زیر و بمی و عینی، متغیر و غیر قابل پیش‌بینی بودن است.

از دیدگاه اسلامی [16] جملات زبان دارای ساخت سازه‌ای هستند و سازه‌ها به شکل سلسله مراتبی در قالب واحدهای بزرگ‌تری به نام گروه‌ها سازماندهی می‌شوند. از این رو کلمات، سازه‌های بلافصل جمله نیستند بلکه آن‌ها سازه‌های پایانی جمله به حساب می‌آیند. با استفاده از آزمایش‌هایی مانند پیش‌آیند سازی، جایگزینی ضمیر و غیره مرز گروه‌های نحوی مشخص می‌شود. هسته تنها عنصر اجباری گروه است و به تنهایی می‌تواند نقش گروه کامل را بازی کند. هسته‌ها به عنوان مقولات نحوی از ویژگی‌های بیشترین گسترش برخوردارند. صورت گسترش یافته هسته، دارای ساخت سازه‌ای درون‌گروهی است که در آن، هسته گروه وابسته‌هایی را

<sup>1</sup> Autosegmental-Metrical (AM) Phonology

قبل یا بعد از خود می‌پذیرد. وابسته‌ها دارای رابطه‌ای نظام‌مند با هسته هستند و با توجه به رابطه ساختاری که با هسته برقرار می‌کنند، نوع وابستگی آن‌ها مانند آویزه و یا متمم بودن مشخص می‌شود. گروه‌های اسمی، فعلی، حرف اضافه‌ای، صفتی و گروه قیدی، گروه‌های نحوی اصلی به حساب می‌آیند. پیش بینی محل تکیه زیر و بمی در گروه‌های نحوی بر اساس اصلی به نام اصل هسته‌گزینی صورت می‌گیرد. اصل هسته‌گزینی ناظر بر این واقعیت است که الگوی برجستگی در تولید بی‌نشان واحدهای نحوی به گونه‌ای است که در آن تکیه گروه روی دورترین وابسته به هسته واحد نحوی قرار می‌گیرد. در نتیجه با پیدا کردن واحدهای نحوی به کمک بیشترین گسترش هر گروه و به کمک اصل هسته‌گزینی می‌توان محل تکیه‌ی زیر و بمی را در گفتار بی‌نشان مشخص کرد [16]. اسلامی در تحقیقات خود به بعضی از نظرات محققین قبلی ایراداتی را وارد نموده است که از آن جمله می‌توان به موارد زیر اشاره کرد:

- تمام واحدهای واژگانی دارای تکیه واژگانی هستند و به همین دلیل می‌توانند در بافت‌های خاص، محل ظهور تکیه‌ی زیر و بمی باشند و لذا ادعای وجود واحدهای بدان بی تکیه (مانند حروف اضافه) در زبان را که بدان معتقد است، می‌توان رد نمود.
- عدم تمایز بین تکیه‌ی واژگانی و تکیه‌ی زیر و بمی باعث شده است که بسیاری از محققینی که تکیه را مورد بررسی قرار داده‌اند، فرقی بین تکیه انتزاعی (تکیه‌ی واژگانی) و تکیه عینی (تکیه‌ی زیر و بمی) قائل نبوده و لذا بحث‌هایی را پیرامون ضعیف شدن تکیه، حذف تکیه، انتقال تکیه، افزوده شدن تکیه در گفتار فارسی مطرح نمایند که در صحت بسیاری از این بحث‌ها جای تردید است [16].
- تمایز بین دو مفهوم انتزاعی و عینی برای تکیه موجب شده است که برخی از پژوهشگران چون «سامعی» [20] در واحدهای نحوی به دنبال یک ویژگی واژگانی باشند (به عنوان مثال «می‌رفت» یک واحد نحوی است و نه یک واحد واژگانی). اسلامی بحث روی تغییرات مداوم در جایگاه تکیه در فعل را موجه نمی‌داند و معتقد است که تغییرات در الگوی برجستگی صورت‌های تصریف شده فعل مربوط به تکیه‌ی زیر و بمی است که جایگاه ثابتی دارد.
- تمام واحدهای واژگانی دارای تکیه واژگانی هستند و به همین دلیل می‌توانند در بافت‌های خاص محل ظهور تکیه‌ی زیر و بمی باشند و بدین ترتیب ادعای وجود واژه‌های بی تکیه (مانند حروف اضافه) در زبان فارسی (وحیدیان [14] و...) نمی‌تواند صحیح باشد.
- پیرو آزمایشات تجربی در ساخت‌های ندایی که در آن‌ها معمولاً اسم حالت ندایی به خود می‌گیرد، محل برجستگی تغییر نمی‌کند و تنها تفاوت بین اسم در حالت ندایی و غیر ندایی این است که هجاهای بی تکیه در حالت ندایی از نظر سطح زیر و بمی در یک سطح بالاتری تولید می‌شوند. ذکر

این نکته اعتقاد به تغییر جایگاه تکیه در ساخت‌های ندایی که قبلاً توسط ( [14] و [21] ) بیان شده بود را متزلزل می‌سازد [19].

اسلامی مشکل اصلی اکثر تحقیقات پیشین را نداشتن چارچوب نظری مشخص بیان نموده، گرچه معتقد است که اکثر آن‌ها در قالب نظریه تحقیق، به نیکی و در حد کمال مطلب را ادا نموده‌اند. ضمن این که در بعضی از تحقیقات گذشته موضوع تکیه و دیگر مسائل زیرزنجیری تنها بخش کوچکی از تحقیقات را تشکیل می‌دهد است و لذا نمی‌توان انتظار داشت که در چنین شرایطی حق مطلب به تمام معنا بیان شده باشد. اسلامی در [19]، تحقیق توحیدی (۱۹۷۴) را به عنوان کامل‌ترین اثر در چارچوب رهیافت بریتانیایی می‌داند و در جدولی نواخت‌های هسته‌ای فارسی پیشنهادی توحیدی را با نواخت‌های هسته‌ای خود مقایسه می‌نماید.

دسته دیگری از کارهای انجام شده بر روی زیر و بمی و نوا در زبان فارسی، در پایان نامه‌ها و پروژه‌های انجام شده توسط دانشگاه‌های فنی و مراکز تحقیقاتی به عنوان بخشی از فعالیت تحقیقاتی عمومی مطالعه در مورد روش‌های شناسایی و بازسازی گفتار بوده است. یکی از این موارد مطالعاتی «ساختار نوایی در زبان فارسی، تحقیق و پیاده‌سازی با استفاده از ترکیب‌کننده گفتاری<sup>۱</sup>» توسط آقای شیخ زاده در سال ۱۹۹۹ است [22]. در این پژوهش در مورد ویژگی‌های نوایی زبان فارسی تحقیقاتی صورت گرفته و برخی تلاش‌ها برای تعریف یک سری قوانین برای تعیین تکیه انجام یافته است و تعدادی قوانین آهنگ برای اهداف بازسازی گفتار فارسی پیشنهاد شده‌اند. این تحقیق بیش از هر چیز بر تغییرات گام دلالت دارد، اما از دیدگاه تحقیق در زمینه آهنگ، تنها چند قانون تجربی برای آهنگ جملات فارسی ارائه شده است. بر این اساس، نگارندگان این تحقیق، چندین قانون اساسی را بر اساس شنیدن و مشاهده و طبقه‌بندی جملات فارسی استخراج و سپس جملات فارسی را به گروه‌های اصلی شامل خبری، سئوالی ساده، سئوالی با کلمات پرسشی، تعجبی و امری طبقه‌بندی کرده‌اند. آن‌ها با استفاده از تحلیل گرامری، کلمه حاوی تکیه اصلی (هسته) را شناسایی کرده و سپس یک منحنی زیر و بمی اصلی را برای شبیه‌سازی تغییرات فرکانس زیر و بمی طبیعی که مطابق با الگوی پیشنهادی آهنگ باشد به کار برده‌اند.

یکی از تحقیقات انجام شده دیگر پیاده‌سازی یک سیستم تبدیل متن به گفتار برای زبان فارسی توسط آقای ابوطالبی در سال ۱۳۷۹ است [23]. این سیستم، یک تولید کننده به روش هم‌گذاری است که توسط الحاق هجاهای فارسی با استفاده از یک روش موسوم به TD-PSOLA<sup>۲</sup> عمل می‌نماید. برای دستیابی به یک بازسازی کننده کارا، محققان بررسی‌های زیادی بر روی تغییرات زیر و بمی در جملات فارسی انجام داده‌اند و ارائه

<sup>1</sup> Speech Synthesizer

<sup>2</sup> Time Domain Pitch Synchronous Overlap and Add

برخی قوانین برای مدل کردن این تغییرات، با همین هدف بوده است. روش کار بدین صورت است که ابتدا بر مبنای محل هجاهای تکیه بر، یک منحنی زیر و بمی اولیه برای هر کلمه به دست آورده می‌شود. به صورت مشابهی منحنی زیر و بمی برای دسته‌های مختلف گروه‌های نوایی و نوع جملات به دست آمده و منحنی نهایی زیر و بمی به دست می‌آید. یک جنبه بسیار مهم در این کار آن است که علاوه بر تکیه کلمه و الگوی زیر و بمی مربوط به نوع جمله، یک سطح نوایی میانی به نام گروه‌بندی نوایی وجود دارد. این نوع گروه‌بندی‌ها توسط بعضی از نشان‌گرهای مرزی تعیین می‌شوند (در حقیقت آن‌ها ابتدا و انتهای گروه‌ها را مشخص می‌نمایند). این نشان‌گرها شامل برخی از کلمات دستوری (مثل حروف اضافه، حروف عطف و مانند آن) یا برخی از علامت‌های نشان‌گذاری (مانند ویرگول و نقطه-ویرگول و غیره) هستند.

کار دیگری که می‌توان بدان اشاره کرد با نام «تحلیل ساختار گفتار فارسی با استفاده از اطلاعات نوایی سیگنال گفتار» است که توسط آقای فرشاد الماس گنج در سال ۱۳۷۷ انجام شده است. در این تحقیق وی سعی در استفاده از ساختار نوایی کلمات، در تعیین یک معیار برای شناسایی کلمات فارسی، داشته است [24].

از دیگر کارهایی که در این زمینه انجام شده، پایان‌نامه‌ی دوره کارشناسی ارشد آقای علی دفتریان [25] تحت عنوان «تقطیع جمله به اجزای آن و اعمال نوای گفتار فارسی» است. در این پروژه الگوریتمی جهت ویرایش هم‌زمان مقادیر گام و دیرش واحدهای هجایی در زبان فارسی توسط روش بازسازی هم‌نواختی-نوفه‌ای (HNM) طراحی و پیاده‌سازی شده است که نتایج به دست آمده در شرایط مختلف و ضمن بازسازی واحدهای گفتاری متعدد، کاملاً رضایت‌بخش و بدون ایجاد حالت‌هایی چون ایجاد ناپیوستگی در مرز بخش‌های بازسازی شده و ایجاد نویز<sup>۱</sup> و یافت نامطلوب کیفیت در گفتار بازسازی شده بوده است.

## ۲-۵- خلاصه فصل

می‌توان به جرأت ادعا کرد که هیچ پژوهشگری نیست که آگاه از مصنوعی بودن و خسته‌کننده بودن صدای تولیدی از نرم افزارهای مبدل متن به گفتار نباشد، اما چیزی که از بررسی پایان‌نامه‌ها، کتب، مقالات و نظرات آن‌ها مشخص است این است که آن‌ها «لحن» و شبیه بودن گفتار مصنوعی به گفتار طبیعی و انسانی را تنها راه رفع این خستگی ذهنی می‌دانند. بنابراین در ابتدا بخش عظیمی از پژوهش‌ها به بهبود الگوریتم‌های تبدیل متن به گفتار معطوف شده و بخش دیگر به بهبود «لحن».

اما ما در این پژوهش معتقدیم حتی اگر پس از زحمات فراوان، نهایت کیفیت الگوریتم‌های تبدیل و لحن را به دست آوریم و خروجی، کاملاً شبیه به گفتار طبیعی باشد، ما مشکل مطرح شده در این پژوهش را در بین

<sup>1</sup> Buzziness

گفتار طبیعی نیز شاهد هستیم! یعنی یک سخنران که تنوع چندانی در حجم صدای خود ندارد، موجب خستگی ذهنی مخاطبان می‌شود. راهکار بیان شده در این پژوهش هر چند در تبدیل متن به گفتار بسیار مفیدتر خواهد بود اما در گفتار طبیعی نیز مفید فایده خواهد بود.

## فصل ۳:

# راهکار پیشنهادی

### ۳-۱- مقدمه

در این بخش به بررسی راهکار خود برای کاهش خستگی ذهنی ناشی از شنیدن صدای یکنواخت به ویژه صدای تولید شده از روی متون طولانی توسط مبدل‌های متن به گفتار می‌پردازیم. پس از توصیف راهکار، با توجه به اینکه راهکار ما دو بخش خواهد داشت: یکی «کاهش در سکوت» و دیگری «افزایش به نرمی»، در قسمت اول از این فصل به بررسی راهکارهای یافتن «سکوت» در سه کاربرد مهم این پژوهش؛ یعنی:

- کاربرد در تبدیل متن به گفتار

- کاربرد در صدای آنالوگ ضبط شده‌ی یکنواخت

- کاربرد در پخش زنده‌ی صدای یکنواخت

خواهیم پرداخت و سپس در مورد «افزایش به نرمی» در این کاربردها صحبت خواهیم کرد. در بخش بعد نیز در مورد نتایج به دست آمده از آزمایشات انجام شده روی نمونه‌های مختلف بحث خواهیم کرد.

### ۳-۲- توصیف راهکار

#### ۳-۲-۱- ایده اول: کاهش و افزایش صدا به صورت سینوسی

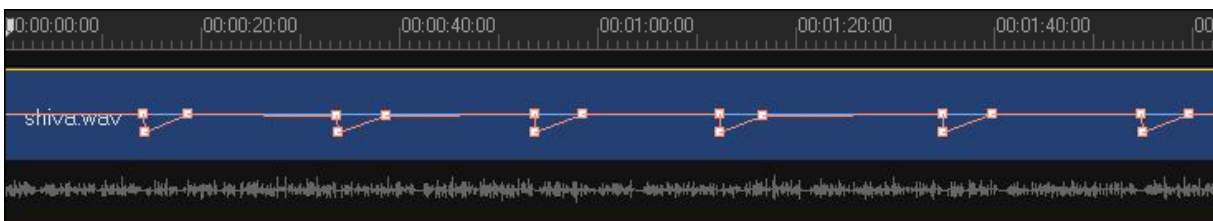
همانطور که پیش از این توضیح داده شد، در راهکار ارائه شده در این تحقیق، ما به دنبال ایجاد تنوع مصنوعی بر روی گفتار یکنواخت هستیم. برای انجام این کار، اولین ایده‌ای که به ذهنمان رسید این بود که حجم صدا را بدون توجه به کلمات و جملات بیان شده، به صورت سینوسی کم و زیاد کنیم. به این معنی که صدا در مدت زمان مشخصی (مثلاً ۱۰ ثانیه) به مرور، ۲۰ درصد کاهش یابد و سپس در همان مدت زمان به حالت ۱۰۰ درصد برگردد. اما پس از انجام آزمایشات مختلف، متوجه شدیم که در صورت انجام چنین کاری، با توجه به اینکه کاربر در حال توجه به صدا است، هر تغییری در حجم صدا برای او مشخص خواهد بود. بنابراین شنونده کاملاً متوجه کاهش و افزایش صدا خواهد شد و همین موضوع باعث می‌شود حواس کاربر از توجه به متن، به افزایش و کاهش حجم صدا معطوف شده و این مشکل به نامفهوم بودن طبیعی صداهای تبدیل شده از متن به گفتار اضافه شود و در نتیجه شنونده به سختی متوجه مفهوم متن خواهد شد.

بنابراین، این ایده از نگاه ما رد شد اما بهبود آن را در دستور کار قرار دادیم.



### ۲-۲-۳- ایده دوم: کاهش در سکوت، افزایش در ۵ ثانیه

در مرحله دوم، ایده جدیدی را به کار گرفتیم: با توجه به حافظه موقت شنونده، صدا را در سکوت به اندازه ۲۰ درصد کاهش دهیم و در عرض ۵ ثانیه به حالت اول برگردانیم. منظور از سکوت در تبدیل متن به گفتار می‌تواند لحظه رسیدن به یک نقطه و یا ویرگول و یا هر پایان‌دهنده یا ایجاد کننده‌ی مکث دیگری باشد و در یک سخنرانی یکنواخت، سکوت می‌تواند یک لحظه وقفه (هر چند بسیار کوتاه) در صحبت‌های سخنران باشد. در این صورت شنونده در مدتی که سکوت رخ داده، حجم صدای قبلی را تا حد زیادی فراموش خواهد کرد و از لحظه‌ی آغاز مجدد صدا، بدون اینکه حواس او پرت شود، صدا با حجم کمتری پخش خواهد شد و به مرور و در عرض ۵ ثانیه صدا به حالت ۱۰۰ درصد خواهد رسید. تصویر زیر، پیاده‌سازی این ایده را بر روی یک فایل صوتی نمونه نشان می‌دهد:

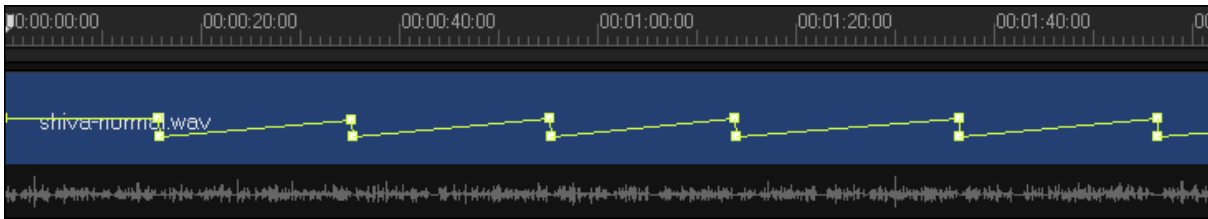


شکل ۲ - ایده‌ی دوم: کاهش در سکوت و افزایش به نرمی

این تغییرات در حجم می‌تواند به صورت متناوب، در هر ۲۰ تا ۶۰ ثانیه یک بار اتفاق بیفتد. البته اگر در این فاصله یک سکوت یافت شود، در غیر این صورت یک دوره ۶۰ ثانیه‌ای دیگر به دنبال سکوت گشته می‌شود. نتیجه‌ی این ایده هر چند بهتر از ایده‌ی اول بود اما با توجه به کوتاه بودن مدت اوج‌گیری صدا (۵ ثانیه) مجدداً این کاهش و افزایش محسوس بود. بنابراین این امکان وجود داشت که حواس شنونده متوجه تغییرات حجم صدا شود.

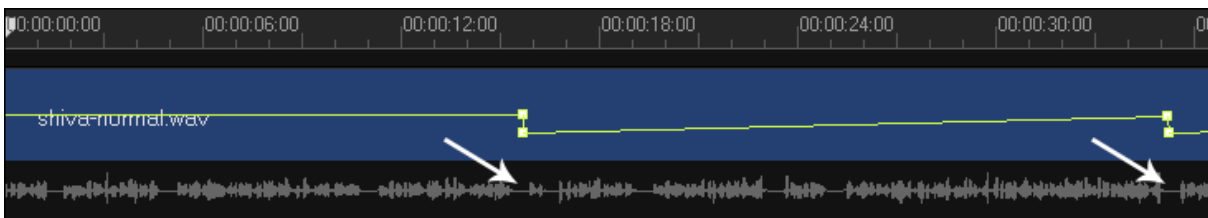
### ۳-۲-۳- ایده سوم: کاهش در سکوت، افزایش تا سکوت بعدی

در مرحله سوم، ایده‌ی دوم کمی تغییر کرد: همانند مرحله دوم، حجم صدا در سکوتی که در فاصله زمانی ۲۰ تا ۶۰ ثانیه قرار دارد به اندازه ۲۰ درصد کاهش خواهد یافت اما اوج‌گیری صدا با طول زمانی بیشتری اتفاق خواهد افتاد. به طور دقیق، صدا تا سکوت بعدی و به مرور افزایش خواهد یافت. تصویر زیر، پیاده‌سازی این ایده را بر روی یک فایل صوتی نمونه نشان می‌دهد:



شکل ۳ - ایده‌ی سوم: کاهش حجم صدا در سکوت و افزایش به نرمی

در تصویر زیر که بزرگنمایی شده‌ی تصویر بالا است، کاهش حجم صدا در سکوت، بهتر مشخص است:



شکل ۴ - بزرگ‌نمایی ایده‌ی سوم

پس از انجام آزمایش بر روی چند نمونه محدود به این نتیجه رسیدیم که این همان چیزی است که مد نظر این تحقیق است؛ یعنی تنوع در صدا به طوری که مخاطب در عین حال که از این تنوع راضی بوده و احساس آرامش می‌کند، حواس او متوجه این تنوع در حجم صدا نباشد.

### ۳-۳- یافتن سکوت در کاربرد تبدیل متن به گفتار

یافتن سکوت در نرم افزارهای تبدیل متن به گفتار ساده‌تر از یافتن سکوت در فایل سخنرانی یکنواخت خواهد بود. در تبدیل متن به گفتار می‌توان مدت زمان تولید و یا پخش فایل صوتی مربوط به یک صامت و مصوت را به طور میانگین محاسبه کرد، سپس یک مدت زمان حداقلی برای یافتن سکوت بعدی در نظر گرفت (مثلاً ۲۰ ثانیه) و آنرا را تقسیم بر مدت زمان به دست آمده (یعنی میانگین مدت زمان تولید صامت و مصوت) کرد تا تعداد صامت و مصوت‌هایی که باید پخش شود و سپس جستجو برای سکوت بعدی شروع شود، به دست آید. به طور مثال اگر طول فایل صوتی مربوط به هر صامت و مصوت را یک بیست و پنجم ثانیه بگیریم، باید ۵۰۰ ترکیب صامت و مصوت بگذرد و سپس شروع به جستجوی سکوت (نقطه، ویرگول، علامت تعجب و غیره) کنیم.

البته شرط استفاده از این ایده این است که طراح نرم افزار مبدل متن به گفتار برای نقاط و دیگر علامات ایجاد کننده‌ی مکث، مدتی سکوت در نظر بگیرد. این موضوع، نکته بسیار مهمی در کاهش خستگی ناشی از شنیدن صدای ماشینی است. آزمایشات ما نشان می‌دهد که کاربران از اینکه در حین شنیدن صدای ماشینی، شاهد مدتی سکوت هستند، بسیار راضی‌اند چرا که باعث آرامش ذهن آن‌ها در همین مدت کوتاه می‌شود. به ویژه

هنگامی که نرم افزار به یک پاراگراف جدید (و یا در اصطلاح برنامه نویسی، به یک «n جدید») می رسد، مدت این سکوت باید کمی بیشتر باشد.

بنابراین پیشنهاد ما این است که در صورتی که مبدل، بدون در نظر گرفتن مکث برای علامات ایجاد کننده مکث طراحی شده است، ابتدا فکری به حال این نقص ننماید و سپس الگوریتم اصلی مطرح شده در این پژوهش را به کار گیرد.

### ۳-۴- یافتن سکوت در کاربرد صدای آنالوگ ضبط شده

یافتن سکوت در فایل های ضبط شده از حالت آنالوگ یکی از پیچیده ترین مباحث در علم کامپیوتر است و نیاز به محاسبات ریاضی بسیار پیچیده ای دارد، چرا که ابتدا باید تعریف مشخصی از «سکوت» ارائه شود و این تعریف باید نویزها و صدای پس زمینه را مد نظر قرار دهد. [26] اما با توجه به اینکه در این تحقیق به دنبال ارائه راهکارهای ساده تر و عملی هستیم، دو راهکار ساده تر را معرفی می کنیم:

#### ۳-۴-۱- کمترین میدان در فواصل زمانی مشخص

یکی از راهکارها این است که کمترین میدان<sup>۱</sup> در یک مدت زمان مشخص به عنوان سکوت در نظر گرفته شود. به طور مثال در هر ۲۰ تا ۶۰ ثانیه، میدان صدا تحلیل شود و لحظه ای که کمترین میدان را دارد به عنوان «سکوت» در نظر گرفته شود.

برای اطمینان می توان برای این افت میدان، شرط طول زمانی نیز قائل شد. به عبارت دیگر اگر این میدان برای مدت زمان مشخصی ادامه داشت، سکوت در نظر گرفته شود در غیر این صورت، خیر. مثلاً اگر این افت میدان، حداقل ۲ ثانیه ادامه داشت، یعنی سکوت رخ داده است.

#### ۳-۴-۲- اولین لحظه ای که میدان از مقدار مشخصی کمتر شود

راهکار دیگر می تواند این باشد که در فواصل زمانی مشخص، به محض اینکه میدان از مقدار مشخصی (فرضاً 20dbA) کمتر شد و برای مدت مشخصی (فرضاً ۲ ثانیه) ادامه داشت، آن لحظه، سکوت در نظر گرفته شود.

### ۳-۵- یافتن سکوت در کاربرد پخش زنده صدا

---

<sup>1</sup> Amplitude

در صورتی که نتیجه این تحقیق بخواهد بر روی پخش زنده صدا اعمال شود؛ از جمله در مکالمات تلفنی و یا پخش زنده سخنرانی‌ها، با توجه به اینکه میدان صدا در آینده در دسترس نخواهد بود، راهکار دوم یعنی اولین افت میدان می‌تواند به کار گرفته شود.

اسکرپت‌های مختلفی برای یافتن سکوت در فایل‌های صوتی طراحی شده است که می‌توان به Sox<sup>1</sup> به عنوان یکی از آن‌ها اشاره کرد. این اسکرپت منبع‌باز این امکان را فراهم می‌کند که سکوت در یک فایل صوتی به راحتی تشخیص داده شود.

### ۳-۶- افزایش به نرمی در کاربرد تبدیل متن به صدا

موضوع دیگر در پیاده‌سازی این ایده، افزایش تدریجی صدا در مدت زمانی است که به سکوت بعد برسیم. این کار می‌تواند با شمارش تعداد ترکیب صامت و مصوت‌ها بین سکوت فعلی و سکوت بعدی و تقسیم کردن عدد ۲۰ (درصد صدایی که در سکوت فعلی کاهش یافته) به آن تعداد شمارش شده، و افزایش حجم صدا در هر ترکیب، پیاده‌سازی شود؛ برای مثال اگر بین سکوت فعلی تا سکوت بعد، ۵۲۰ ترکیب صامت و مصوت قرار داشته باشد، نتیجه‌ی تقسیم ۲۰ بر ۵۲۰، یعنی ۰.۰۳ درصد باید در هر ترکیب صامت و مصوت، حجم صدا افزایش یابد.

هر چند این افزایش صدا به صورت پله‌ای است و این را به ذهن می‌آورد که ممکن است کاربر متوجه این افزایش پله‌ای شود اما باید گفت که این افزایش آنقدر نرم و سریع اتفاق می‌افتد که نمی‌توان «پله‌ای» بودن آن را احساس کرد. مانند خطوط مورب در یک مانیتور که هر چند به صورت پیکسلی و پله‌ای رنگ‌آمیزی صورت گرفته، اما آنقدر پله‌ها کوچک هستند که چشم انسان تقریباً پله‌ای بودن خط را احساس نخواهد کرد.

### ۳-۷- افزایش به نرمی در کاربرد صدای آنالوگ ضبط شده

در این کاربرد، با دانستن سکوت بعدی و محاسبه مدت زمان بین سکوت فعلی و سکوت بعدی، می‌توان به صورت دوره‌ای در مدت زمان مشخصی به مرور صدا را افزایش داد. برای مثال اگر بین سکوت فعلی و سکوت بعدی، ۲۲ ثانیه فاصله باشد، می‌توان با تقسیم ۲۰ (یعنی درصد کاهش صدا) به ۲۲ (یعنی طول گام)، عدد ۰.۹ را به دست آورد و در هر ۰.۹ ثانیه یک درصد به حجم صدا افزود.

### ۳-۸- افزایش به نرمی در کاربرد پخش زنده صدا

<sup>1</sup> <http://sox.sourceforge.net/>

در این کاربرد، با توجه به اینکه محل سکوت بعدی مشخص نیست، نمی‌توان از راهکاری که در کاربرد صدای آنالوگ ضبط شده معرفی شده است استفاده نمود. شاید بتوان با «پیش‌بینی سکوت بعدی» راهکاری مانند راهکار قبل ارائه کرد که البته این موضوع، خود می‌تواند یک تحقیق جداگانه باشد، بنابراین آن را به فرصتی دیگر و یا به محققینی دیگر واگذار می‌کنیم. اما به عنوان یک راهکار ابتدایی می‌توان از سکوت بعدی صرف نظر کرد و در عوض، از ایده ۳.۲.۲ (یعنی ایده دوم: کاهش در سکوت، افزایش در ۵ ثانیه) استفاده نمود با این تفاوت که زمان ثابت ۵ ثانیه را تا لحظه شروع به کار برای یافتن سکوت بعدی به تأخیر انداخت؛ به طور مثال، پس از تشخیص یک سکوت و کاهش حجم صدا به میزان ۲۰ درصد، می‌توان حجم صدا را طی مدت زمان ۲۰ ثانیه به حالت ۱۰۰ درصد رسانید و سپس شروع به یافتن سکوت بعدی کرد.

### ۳-۹- خلاصه فصل

دو چالش اصلی در استفاده از ایده‌ی مطرح شده در این پژوهش، عبارتند از: «یافتن سکوت بعدی» و «افزایش به نرمی». در این فصل پاسخی برای این دو چالش در کاربردهای مختلف یعنی: «تبدیل متن به گفتار»، «صدای آنالوگ ضبط شده» و «پخش زنده‌ی صدا» ارائه نمودیم.

## فصل ۴:

# شبیه سازی و ارزیابی نتایج

## ۴-۱- مقدمه

در این فصل طرح کلی پژوهش، جامعه آماری، نمونه و روش نمونه‌گیری مورد بحث قرار گرفته و در ادامه ابزارهای اندازه‌گیری معرفی می‌شود، سپس روش گردآوری اطلاعات و روش تجزیه و تحلیل ارائه می‌شود. هدف اصلی این بود که بررسی شود آیا الگوریتم پیشنهاد شده در این تحقیق به راستی باعث خستگی ذهنی کمتر و آرامش ذهنی بیشتری در شنوندگان خواهد شد؟

## ۴-۲- نرم افزار تبدیل متن به گفتار پارس خوان

در این پژوهش، از نرم افزار «پارس خوان» به عنوان مبدل متن به گفتار استفاده شده است. دلیل اصلی انتخاب این نرم افزار، این است که نگارنده‌ی این پژوهش، خود، برنامه‌نویس این نرم افزار بوده است. بنابراین به روش کار و کدهای نرم افزار احاطه کامل داشته‌ایم.



شکل ۵ - نمای نرم افزار پارس خوان

### ۳-۴- خلاصه‌ای از الگوریتم تبدیل متن به گفتار در پارس‌خوان<sup>۱</sup>

- ۱- متن از کاربر دریافت می‌شود.
- ۲- متن از نظر نوع نگارش، یکدست می‌شود. به طور مثال، «ی» (ی عربی همراه با دو نقطه در زیر آن) و یا «ک» عربی به حالت فارسی تبدیل می‌شود، اعداد با نگارش‌های مختلف مثل ۱ و 1 به یک حالت تبدیل می‌شوند.
- ۳- متن با معیار «فاصله» شکسته می‌شود به کلمات.
- ۴- کلمات یک به یک بررسی می‌شوند.
- ۵- اگر کلمه در دیکشنری پارس‌خوان موجود بود و تلفظ آن درج شده بود، تلفظ با معیار «فاصله» شکسته می‌شود به «بخش» و سپس فایل صوتی هر بخش، پخش می‌شود. مثال: کلمه «آفتابگردان» در دیکشنری به صورت «@ا ف^ ت^ @ب^ گ^ ر^ د^ @ن^» ثبت شده است. تک تک حروف الفبا با حرکات مختلف، نیز ضبط شده‌اند. به طور مثال: ف a ف e ف o ف @ ف i ف u ف^ . در کلمه آفتابگردان، فایل مربوط به ف^ پخش می‌شود.
- ۶- اگر کلمه در دیکشنری موجود نبود، بررسی‌های مختلف بر روی آن شروع می‌شود.
  - a. آیا ابتدای آن «می» یا «ب» یا «نمی» یا «نا» یا «ن» است و آیا ادامه کلمه در دیکشنری موجود است؟ اگر چنین شرایطی بود، ابتدا فایل صوتی پیشوند و سپس تلفظ کلمه پخش می‌شود. مثال: نگو < ن a + گو
  - b. آیا انتهای کلمه «م»، «ت»، «ش»، «یم»، «یید»، «یند» و پسوندهای دیگر زبان فارسی است و باقیمانده کلمه در دیکشنری موجود است؟ اگر چنین شرایطی برقرار بود، ابتدا تلفظ کلمه استخراج و پخش می‌شود و سپس تلفظ پسوند.
- ۷- اگر در کلمه حروف انگلیسی به کار رفته بود، پردازش‌هایی صورت می‌گیرد تا اعداد، حروف فارسی و انگلیسی مختلط در آن، جدا شود. به طور مثال برنامه باید بتواند B2الف را تلفظ کند.
- ۸- اگر کلمه کاملاً انگلیسی بود، توسط Text To Speech تعبیه شده در ویندوز، تلفظ می‌شود.
- ۹- اگر کلمه، عدد بود، توسط الگوریتم تبدیل عدد به حروف، آن عدد به حروف تبدیل می‌شود و تلفظ می‌شود.
- ۱۰- پس از بررسی‌های مختلف، اگر نهایتاً کلمه‌ای در متن بود که از تمام این الگوریتم‌ها گذشت و تلفظ دقیق آن یافت نشد، کلمه شکسته می‌شود به «حروف» و تک تک حروف با صدای ساکن تلفظ می‌شود.

<sup>۱</sup>منقول از مستندات پارس‌خوان



#### ۴-۴- متن انتخابی برای تبدیل به گفتار

برای شبیه‌سازی ایده، ابتدا یک متن نسبتاً طولانی برای تبدیل به گفتار انتخاب شد که فایل صوتی به دست آمده از نرم افزار پارس‌خوان برای این متن، ۳ دقیقه و ۴۵ ثانیه شد.

در انتخاب متن، به این موضوع توجه شد که متن، حاوی یک داستان جذاب باشد تا کاربر را بیشتر مجذوب نماید و از توجه به حواشی اجتناب کند.

متن انتخابی همراه با علامت‌گذاری دستی برای افزایش کیفیت خروجی، به صورت زیر بود:

#### در انتظار سرمایه

روزی شیوانا از نزدیک مزرعه‌ای می‌گذشت. مرد میانسالی را دید که کنار حوضچه نشسته و غمگین و افسرده به آن خیره شده است. شیوانا کنار مرد نشست و علت افسردگی‌اش را جویا شد. مرد گفت: "این زمین را از پدرم به ارث گرفته‌ام. از جوانی آرزو داشتم در این جا ماهی پرورش بدهم. همه چیز آماده است. فقط نیازمند سرمایه‌ای بودم که این حوضچه را لایروبی و تمیز کنم و فضای سرریسته‌ی مناسبی برای پرورش و نگهداری ماهی ایجاد کنم. این آرزو را از همان ایام جوانی داشتم و الان بیش از ده سال است که هنوز چنین سرمایه‌ای نصیبم نشده است.

بچه‌هایم در فقر و دست‌تنگی بزرگ می‌شوند و آرزوی من برای رسیدن به سرمایه‌ی لازم برای آماده‌سازی این حوض بزرگ هر روز کم رنگ‌تر و محال‌تر می‌شود. ای کاش خالق هستی همراه این حوض بزرگ به من سرمایه‌ای هم می‌داد تا بتوانم از آن، ثروت مورد نیاز خانواده‌ام را بیرون بکشم."

شیوانا نگاهی به اطراف انداخت و سپس حوضچه‌ی سنگی کوچکی را در فاصله‌ی دور از حوض بزرگ نشان داد و گفت: "چرا از آن‌جا شروع نمی‌کنی؟ هم کوچک و قابل نگهداری است و هم می‌تواند دستگرمی خوبی برای شروع کار باشد."

مرد میانسال نگاهی ناامیدانه به شیوانا انداخت و گفت: "من می‌خواستم با این حوض بزرگ شروع کنم تا به یک باره به ثروت عظیمی برسم و شما آن حوضچه‌ی کوچک سنگی را به من پیشنهاد می‌کنید. آن را که همان ده سال پیش، خودم به تنهایی می‌توانستم راه بیندازم."

شیوانا سری تکان داد و گفت: "من اگر جای تو بودم به جای دست روی دست گذاشتن و حسرت خوردن لااقل با آن حوضچه‌ی کوچک آرزوی بزرگم را تمرین می‌کردم تا کمرنگ نشود و از یادم نرود!"

مرد میانسال آهی کشید و نظر شیوانا را پذیرفت و به سوی حوضچه‌ی کوچک رفت تا خودش را سرگرم کند.

چند ماه بعد به شیوانا خبر دادند که مردی با یک گاری پر از خرچنگ خوراکی نزدیک مدرسه ایستاده و می‌گوید همه‌ی این‌ها را به رایگان برای مدرسه هدیه آورده است و می‌خواهد شیوانا را ببیند.

شیوانا نزد مرد رفت و دید او همان مرد میانسالی است که آرزوی پرورش ماهی را داشت. او را به داخل مدرسه آورد و جویای حالش شد. مرد میانسال گفت: "شما گفتید که اگر جای من بودید اول از حوضچه‌ی سنگی شروع می‌کردید. من هم تصمیم گرفتم چنین کنم. وقتی به سراغ حوضچه‌ی سنگی رفتم متوجه شدم که آبی که حوضچه را پر می‌کند از چشمه‌ای زیر زمینی و متفاوت می‌آید و املاح آن برای پرورش ماهی اصلاً مناسب نیست اما برای پرورش میگو عالی است. به همین دلیل بلافاصله همان حوضچه‌ی کوچک را راه انداختم و در عرض چند ماه به ثروت زیادی رسیدم. ای کاش همان ده سال پیش همین کار را می‌کردم و این قدر به خود و خانواده‌ام سختی نمی‌دادم."

شیوانا تبسمی کرد و گفت: "حال می‌خواهی چه کنی؟" مرد گفت: "ثروت حاصل از این حوضچه‌ی سنگی و پرورش میگو تمام خانواده‌ی مرا کفایت می‌کند. می‌خواهم از این به بعد در راحتی و آسایش به پرورش میگو در حوضچه‌ی سنگی کوچک بپردازم." شیوانا تبسمی کرد و گفت: "من اگر جای تو بودم با سرمایه‌ای که اکنون به دست آورده‌ام به سراغ حوض بزرگ می‌رفتم و در آن پرورش ماهی را هم شروع می‌کردم. مردم این دهکده و دهکده‌های اطراف به ماهی نیاز دارند و حوض بزرگ تو می‌تواند بسیاری را از گرسنگی نجات دهد."

#### ۴-۵- روش شبیه‌سازی ایده‌ی این پژوهش بر روی خروجی پارس‌خوان

برای انجام آزمایشات نیاز داشتیم که ایده‌ی مطرح شده در این پژوهش را بر روی یک فایل صوتی خروجی گرفته شده از یک مبدل متن به گفتار اعمال کنیم و سپس نتیجه را تجزیه و تحلیل نماییم. برای شبیه‌سازی از نرم افزار Corel Video Studio جهت یافتن سکوت‌ها و کاهش حجم صدا و افزایش صدا به نرمی، استفاده شد.

این نرم افزار جهت انجام تدوین صدا و تصویر کاربرد دارد.



شکل ۶ - نمای نرم افزار Corel Video Studio

#### ۴-۶- روش اول اندازه گیری: استفاده از دستگاه EEG

##### ۴-۶-۱- تأثیرات صدا بر روی مغز انسان و احساس تأثیرات با دستگاه EEG

انسان به طور دائم در معرض صداهای مختلف قرار دارد. تأثیر بعضی از این صداها، گذراست ولی بعضی دیگر، عمیقاً ما را تحت تأثیر قرار می دهند. معمولاً صدای بلند، تأثیری عمیق تر بر روی انسان می گذارد. برای بررسی تأثیرات صدا بر روی مغز انسان باید با روش کار مغز آشنا شویم [8]:

مغز انسان از میلیاردها سلول به نام «نورون» تشکیل شده است که (مانند دیگر سلولهای بدن) از الکتریسیته برای ارتباط با یکدیگر استفاده می کنند. همانطور که می شود حدس زد، وقتی این میلیونها نورون به یک باره با هم سیگنال ارسال می کنند، یک فعالیت الکتریکی عظیمی را در مغز شاهد خواهیم بود و این فعالیت می تواند توسط تجهیزات پزشکی مانند EEG که میزان الکتریسیته در بخشهای مختلف جمجمه را اندازه می گیرد، اندازه گیری شود.

وقتی فعالیت الکتریکی مغز، توسط EEG به شکل گراف در آید، به آن «الگوی امواج مغز»<sup>۱</sup> گفته می شود. کلمه «موج» به خاطر شکل سیکل مانند آن (دقیقاً مشابه شکل موج در طبیعت) استفاده شده است.

<sup>1</sup> brainwave pattern

الگوهای امواج مغز به چهار دسته تقسیم می‌شوند:

**Beta** (بین ۱۴ تا ۳۰ هرتز):

حالت‌های تمرکز، تحریک، هوشیاری، ادراک سطوح بالاتر از ۳۰ هرتز، مربوط هستند به نگرانی، مریضی، احساس دوری، جنگ یا پرواز

**Alpha** (بین ۸ تا ۱۳.۹ هرتز)

حالت‌های آرامش، یادگیری فوق‌العاده، تمرکز آرام، از خود بی‌خود شدن سبک، افزایش تولید سروتونین، پیش‌خواب، خواب‌آلودگی بعد از بیداری، عبادت، شروع دسترسی به ذهن ناخودآگاه

**Theta** (بین ۴ تا ۷.۹ هرتز)

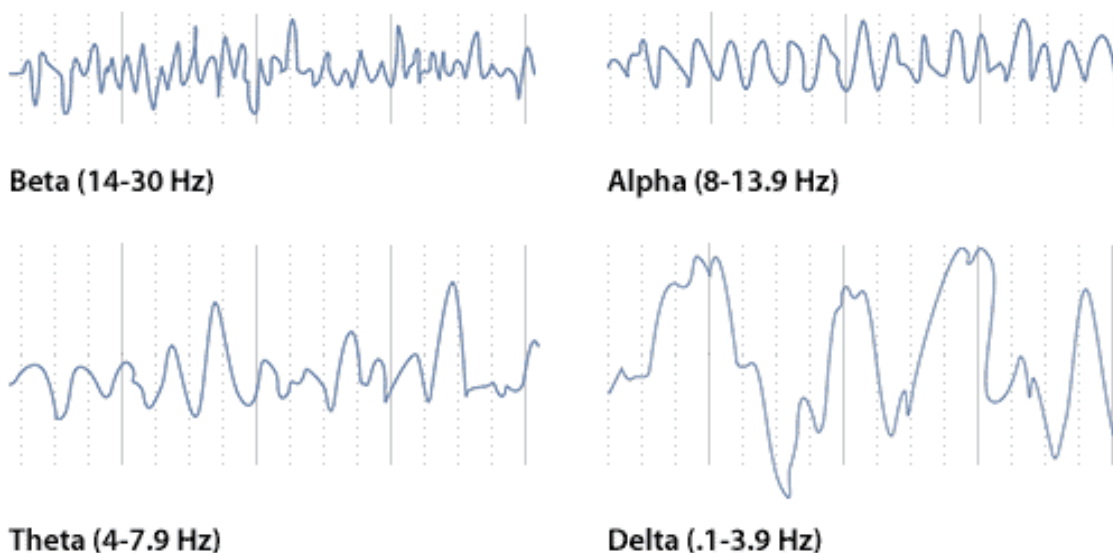
خواب توأم با رؤیا (خواب REM)<sup>۱</sup>، افزایش تولید کاتکول آمین‌ها (که برای یادگیری و حفظیات، حیاتی هستند)، افزایش خلاقیت، تجربیات احساسی، تغییر بالقوه در رفتار، افزایش ماندگاری آموخته‌ها، تصورات خواب‌آور، از خود بی‌خود شدن، مدیتیشن عمیق، دسترسی به ذهن ناخودآگاه

**Delta** (بین ۰.۱ تا ۳.۹ هرتز)

خواب بدون رؤیا، انتشار هورمون رشد انسان، حالت غیرفیزیکی عمیق شبیه به بی‌هوشی، کاهش آگاهی بدن، دسترسی به ذهن ناخودآگاه و «ناخودآگاه جمعی»، بیشترین ورودی به مغز وقتی با هولوسینک القا شده است.

---

<sup>۱</sup> حرکت سریع چشم در خواب (REM) ششمین مرحله از خواب که در آن بدن در حالت خواب عمیق به سر می‌برد ولی واکنش مغز در حالت بیداری است. مرحله‌ای است در خواب که مشخصه آن حرکات سریع چشم است. اکثر رؤیاها در این مرحله از خواب روی می‌دهند. (ویکی‌پدیا)



شکل ۷- انواع گراف مغز - منبع: [8]

اکثر انسان‌ها در بخش عمده‌ای از زندگی‌شان در حالت امواج مغزی بتا زندگی می‌کنند؛ یعنی تحریک شده، هوشیار، متمرکز البته با کمی استرس.

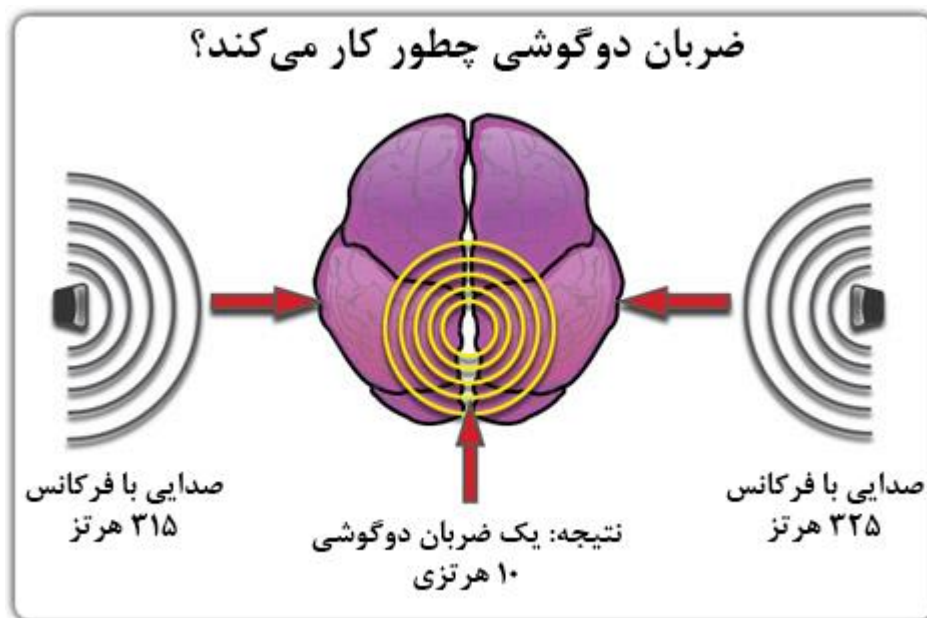
وقتی بسامد امواج مغز از بتا به آلفا کاهش می‌یابد، انسان در حالت ایده‌آل برای یادگیری اطلاعات جدید، انجام کارهای سخت‌تر، یادگیری زبان‌های جدید، تحلیل مسائل و موقعیت‌های پیچیده قرار می‌گیرد و حتی توانایی انجام ورزش‌هایی را که روانشناسان آن‌را «قلمرو» می‌نامند - و آن، حالت تمرکز و کارایی عالی در مسابقه و ورزش کشتی است - خواهد داشت. بخشی از این به خاطر این است که کاهش فعالیت الکتریکی در مغز می‌تواند منجر به کاهش شدید مواد شیمیایی مربوط به احساس خوب مثل اندروفین، نوراپی‌نفرین، دوپامین شود.

بنابراین، وقتی انسان برای مثال، مدیتیشن انجام می‌دهد، روی یک چیز متمرکز می‌شود، حالا چه یک شعله شمع باشد و یا دم و بازدم نفس‌هایش و یا یک مانترا و یا یک دعا. وقتی به این صورت متمرکز می‌شود، الگوی الکتریکی در مغز او آهسته می‌شود و احساس آرامش می‌کند و میدان امواج مغز او در رنج امواج آلفا ثابت می‌شود.

می توان به این نتیجه رسید که لازم نیست انسان یک راهبِ ماهر باشد یا آخر هفته‌ها مدیتیشن انجام دهد تا به حالت آرامشی آلفا دست یابد، در عوض می‌تواند از مفهومی به نام «شگرفی امواج مغز»<sup>۱</sup> استفاده کند تا تأثیرات مشابهی را به دست آورد.

«شگرفی امواج مغز» به هر روشی اطلاق می‌شود که باعث شود بسامدهای امواج مغز انسان به بسامد خاصی کاهش یابد. این مفهوم بر اساس این نظریه است که مغز انسان تمایل دارد که بسامد EEG غالب خود را به سمت «بسامد محرک خارجی غالب»<sup>۲</sup> (مانند موسیقی یا صدا) تغییر دهد. [27]

بسامدهای صوتی که معمولاً در «شگرفی امواج مغز» استفاده می‌شود، «ضربان دو گوش»<sup>۳</sup> نامیده می‌شود. روش کار این صداها به این صورت است که دو صدا که در بسامد نزدیک به هم هستند یک بسامد ضربه‌ای به اندازه تفاوت دو بسامد ایجاد می‌کنند. برای مثال، یک صدا با بسامد ۴۹۵ هرتز و یک صدا با بسامد ۵۰۵ هرتز یک ضربه‌ی ۱۰ هرتزی ایجاد می‌کنند که تقریباً در رنج امواج مغزی آلفا قرار دارد.



شکل ۱ - ضربان دوگوشی - منبع: [8]

صدا تمام مدت روی ترشح هورمون‌های بدن انسان، همچنین نفس کشیدن، ضربان قلب و امواج مغزی اثر می‌گذارد. [28] دانشمندان معتقدند صدا به چهار صورت اساسی روی انسان اثر می‌گذارد:

<sup>1</sup> brainwave entrainment  
<sup>2</sup> dominant external stimulus  
<sup>3</sup> binaural beats

۱- به صورت فیزیولوژیکی: به طور مثال، با شنیدن صدای زنگ ساعت بلند یک افزایش کورتیزول<sup>۱</sup>، همان هورمون ستیز / گریز، در انسان ایجاد می‌شود. فقط صداهای ناهنجار نیستند که این اثر را دارند. به طور مثال، بسامد صدای امواج دریا تقریباً ۱۲ چرخه در دقیقه است. برای خیلی‌ها این صدا، آرامش‌بخش است، و جالب است که ۱۲ چرخه در دقیقه تقریباً فرکانس نفس کشیدن انسان در زمان خوابیدن است. یعنی انسان احساس می‌کند که در حال استراحت است و استرسی ندارد، و یا در تعطیلات به سر می‌برد.

۲- به صورت روانی: موسیقی قوی‌ترین نوع از صدا است که روی احساسات انسان اثر می‌گذارد. به طور مثال برخی موسیقی‌ها انسان را غمگین می‌کنند و برخی شاد. البته موسیقی تنها صدایی نیست که روی احساسات انسان اثر می‌گذارد. صداهای طبیعی هم همین اثر را دارند. برای مثال آواز پرندگان برای بیشتر مردم اطمینان‌بخش است. چرا که در طول صدها هزار سال، انسان یاد گرفته است که زمانی که پرندگان می‌خوانند، همه چیز امن است و زمانی که نمی‌خوانند باید نگران شد.

۳- به صورت ادراکی: همه ما متوجه صحبت‌های دو نفر که به طور هم‌زمان صحبت می‌کنند نمی‌شویم. انسان برای پردازش ورودی صدا، پهنای باند خیلی کمی دارد برای همین صدایی مثل سر و صدا در دفتر کار راندمان او را به شدت کاهش می‌دهد.

۴- به صورت رفتاری: با توجه به آن همه تأثیراتی که در بالا بیان شد، اگر رفتار انسان تغییر نکند، چیز عجیبی خواهد بود. برای مثال می‌توان به راحتی حدس زد که راننده‌ای که در ماشین خود آهنگ تکنو گوش می‌کند، بعید است که با سرعت عادی ۴۰ کیلومتر در ساعت رانندگی کند!

در مجموع، انسان از صدای ناهنجار دور می‌شود و به صداهای خوشایند نزدیک و این تنها چیزی نیست که صداهای ناهنجار به آن آسیب می‌رساند! بر اساس تحقیقات، بیشتر صداهای فروشگاه‌های خرده‌فروشی، نامناسب، تصادفی و حتی خصومت‌آمیز محسوب می‌شوند و اثر باور نکردنی روی فروش دارند. برخی فروشگاه‌ها ۳۰ درصد از فروش خود را با خارج شدن سریع مشتری‌ها از مغازه یا حتی برگشتن آن‌ها از مقابل در فروشگاه، از دست می‌دهند.

این نکات و گراف‌ها می‌تواند معیار خوبی برای این تحقیق به حساب آید. با بررسی گراف EEG افراد مورد آزمایش در حین شنیدن صدای تولید شده با الگوریتم ما، می‌توان به این نتیجه رسید که مغز فرد در چه حالتی قرار گرفته است

---

<sup>1</sup> cortisol

## ۲-۶-۴- نوار مغز چیست؟

نوار مغز یا ثبت فعالیت الکترونیکی مغز انسان و بررسی الگوی امواج مغزی می‌تواند برای ارزیابی کارکرد مغز در بعضی از بیماری‌ها مورد استفاده قرار گیرد. الکترودهای فلزی بر روی نقاط متفاوت پوست سر قرار گرفته و امواج مغزی را دریافت می‌کنند. یک ماشین پلی‌گراف (چند ثبتی) کار ثبت امواج را بر روی کاغذ انجام می‌دهد و البته در بعضی موارد اطلاعات به یک کامپیوتر منتقل شده و بر روی صفحه نمایشگر نشان داده می‌شود. به طور متوسط انجام یک نوار مغز حدود ۴۵ دقیقه طول می‌کشد (بین ۳۰ تا ۹۰ دقیقه). امروزه انواعی از دستگاه‌های ثبت نوار مغز تولید شده است که بیمار به راحتی می‌تواند به اطراف حرکت کرده و فعالیت‌های عادی روزانه خود را انجام دهد در حالی که دستگاه مشغول ضبط امواج مغز وی است. [29]

نوار مغز برای چه مقاصدی درخواست می‌شود؟

نوار مغز برای ارزیابی بیماران مبتلا به تشنج، گیجی و کاهش سطح هوشیاری، ضربه سر و دیگر شرایطی که ممکن است، به علت اختلال در کارکرد مغز ایجاد شوند، درخواست می‌گردد. همچنین این وسیله برای کمک به تشخیص بعضی از بیماری‌های مغزی که باعث بدتر شدن کارکرد روانی فرد (مثل بیماری فراموشی یا نسیان) و یا اختلال کارکرد مغزی او می‌شوند مفید می‌باشد. همچنین گاهی در بیماری‌های کلیوی و یا کبدی شدید نیز که باعث اختلال کارکرد مغزی (آنسفالوپاتی) می‌شوند، می‌توان از این وسیله استفاده کرد. گاهی نیز برای تأیید تشخیص شرایطی مثل مرگ مغزی، از نوار مغز استفاده می‌شود. محققان اعصاب و روان نیز گاهی از نوار مغز برای اتخاذ تصمیمات خود کمک می‌گیرند.

آمادگی و شرایط لازم برای انجام نوار مغز

در بسیاری از موارد برای انجام نوار مغز هیچ‌گونه آمادگی نیاز نمی‌باشد ولی گاهی برای انجام EEG (نوار مغز) پزشک توصیه می‌کند که فرد خود را محدود کرده و در شرایط کم‌خوابی برای انجام تست مراجعه کند. این حالت که بیشتر در مبتلایان به بیماری صرع و گاهی در سایر اختلالات توصیه می‌شود، باعث می‌گردد که اختلال بطور واضح‌تری در نوار مغزی خود را نشان دهد.

نوار مغز یا الکتروانسفالوگرام چگونه صورت می‌گیرد؟

فرد در حالی که نشسته و یا بر روی تخت خوابیده است، تکنسین الکترودهای کوچکی را به سر او وصل می‌کند. گرچه برای انجام این تست نیازی نیست سر تراشیده شود ولی باید سر کاملاً تمیز بوده و محلی که الکتروده قرار می‌گیرد باید از هرگونه چربی اضافه پاک باشد. ممکن است از ژل مخصوص در محل اتصال الکتروده استفاده شود تا امواج مغزی به راحتی بتوانند توسط آنها دریافت شوند.



پس از دریافت امواج توسط الکترودها آنها به ماشین پلی گراف فرستاده شده و بر روی کاغذ ثبت می شوند. در هنگام انجام نوار مغز تکنسین ممکن است از بیمار بخواهد چشم‌های خود را باز و یا بسته کند و یا اینکه به سرعت نفس بکشد و یا تنفس عمیق انجام دهد. گاهی فرد در مقابل نور درخشان و چشمک زن قرار می‌گیرد. در موارد خاص نیز انجام نوار مغز حتی پس از به خواب رفتن بیمار ادامه می‌یابد.

آیا انجام نوار مغز برای انسان خطر دارد؟

این آزمایش، کاملاً ایمن بوده و فقط گاهی در افراد مبتلا به صرع در اثر تحریک مغز (مثلاً در اثر نور درخشان و چشمک زن و یا تنفس‌های عمیق و سریع) می‌تواند باعث تشنج شود.

### ۳-۶-۴- توصیف روال انجام آزمایش

در اولین گام با استفاده از دستگاه<sup>۱</sup> EEG به ثبت واکنش‌های مغز انسان در هنگام شنیدن یک متن تبدیل شده به گفتار، در دو حالت مختلف؛ یک بار در حالت معمولی و بدون اعمال ایده‌ی مطرح شده در این پژوهش و بار دوم پس از اعمال ایده‌مان پرداختیم.



شکل ۹- یکی از سوژه‌های آزمایش EEG

در این آزمایش یک نوجوان ۸ ساله انتخاب شد تا مطمئن باشیم که حواس او به طور کامل به داستان متوجه است. استفاده از یک بزرگسال می‌توانست این ریسک را به همراه داشته باشد که درگیری‌های ذهنی وی باعث تغییر گراف EEG گردد.

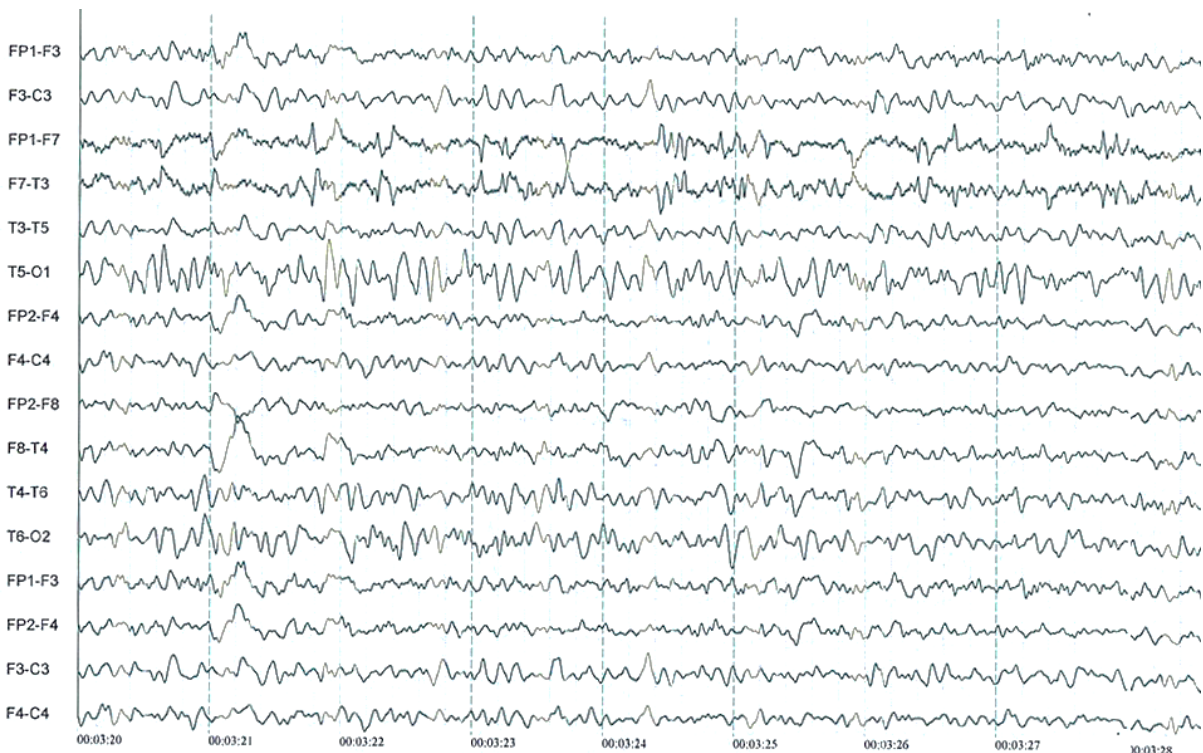
<sup>1</sup> ElectroEncephaloGram

ابتدا صدای اول (بدون اعمال ایده) توسط هدفون‌هایی که به لپ‌تاپ متصل بود، برای سوژه پخش شد. دلیل پخش صدای معمولی در نوبت اول، این بود که اجازه دهیم ذهن او بر اثر شنیدن صدای طولانی کمی خسته شود تا در کلیپ صوتی دوم، تأثیر ایده را بهتر مشاهده کنیم.

بلافاصله بعد از اتمام کلیپ اول، کلیپ دوم که ایده‌ی این پژوهش بر روی آن اعمال شده بود، پخش شد. در پایان، ۶۳ صفحه نوار مغز به دست آمد که ۳۱ صفحه مربوط به کلیپ اول، ۱ صفحه مربوط به وقفه‌ای که تا پخش کلیپ دوم ایجاد شد و ۳۱ صفحه مربوط به کلیپ دوم بود.

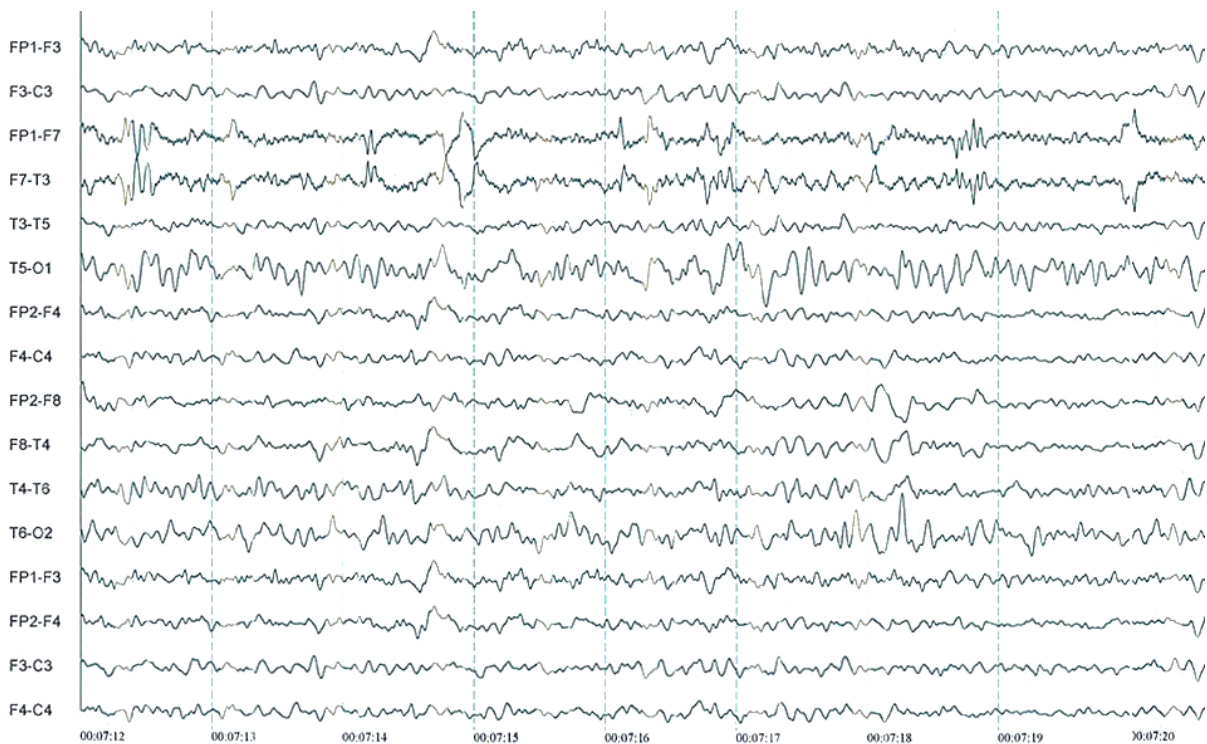
ما ۴۵ ثانیه‌ی آخر از هر دو کلیپ را جدا کرده و در نرم افزار فتوشاپ بر روی یکدیگر قرار دادیم.

یک نمونه‌ی ۸ ثانیه‌ای از کلیپ اول:



شکل ۱۰ - بخشی از گراف EEG مربوط به کلیپ اول

یک نمونه‌ی ۸ ثانیه‌ای از کلیپ دوم:



شکل ۱۱ - بخشی از گراف EEG مربوط به کلیپ دوم

#### ۴-۷- نتیجه اندازه‌گیری

نوار مغزهای گرفته شده به سه متخصص مغز و اعصاب نشان داده شد تا شاید مهر تأییدی بر آرامش بیشتر در هنگام شنیدن کلیپ دوم زده شود، اما هر سه بیان داشتند که امکان تشخیص چنین تغییر حسی اندکی در نوار مغزهای EEG وجود ندارد و نیاز به دستگاه‌های بسیار حساس‌تر و دقیق‌تر است که در دسترس همگان نیست.

بنابراین تصمیم گرفتیم به روش دیگری به بررسی ایده‌مان بپردازیم.

#### ۴-۸- روش دوم اندازه‌گیری: نظرسنجی


در این مرحله، یک نظرسنجی تحت وب طراحی کردیم که در آن کلیپ‌ها برای شرکت کنندگان در نظرسنجی پخش می‌شد و سپس از افراد خواسته می‌شد که کلیپی که بهتر تشخیص می‌دهند را انتخاب کنند و دلیل انتخاب خود را توضیح دهند.

با سلام،  
 نظرسنجی زیر، بخشی از پایان‌نامه اینجانب (حمید رضا نیرومند) می‌باشد.  
 همانطور که می‌دانید یکی از مباحث مهم در دنیای امروز، بحث تبدیل متن به گفتار (Text To Speech) است، یکی از بزرگ‌ترین مشکلات بر سر این راه این است که صدای تولید شده توسط ماشین، صدایی یکنواخت و در متون بلند، خسته‌کننده است. هدف از این تحقیق این است که با به کارگیری روش‌های مختلف، تأثیرات منفی یکنواخت بودن صدا کاهش یابد. نتیجه‌ی این تحقیق می‌تواند باعث بهبود کیفیت خروجی نرم‌افزار تبدیل متن به صدا باشد.


**خواهشمند است با دقت به دو کلیپ زیر گوش کنید و نهایتاً انتخاب نمایید که به نظر شما کدام کلیپ از نظر روانی، کمتر باعث رنجش شما شد؟ (در مجموع ترجیح می‌دهید خروجی یک نرم‌افزار تبدیل متن به صدا، کدام یک از این کلیپ‌ها باشد؟)**  
 توجه:

- در صورت امکان برای شنیدن از هدفون استفاده نمایید.
- لطفاً کلیپ‌ها را تا انتها گوش کنید چون یکی از موارد مهم برای ما خستگی ذهنی شما بر اثر شنیدن متون طولانی است. پس نباید با شنیدن یک دقیقه اول تصمیم بگیرید. (ما می‌توانیم ثبت کنیم که شما کلیپ‌ها را تا انتها شنیده‌اید یا خیر. نظرسنجی افرادی که کلیپ‌ها را تا انتها گوش نکرده باشند برای ما معتبر نیست)
- کلیپ‌ها را حتماً با بخش‌کننده‌ی مرورگر بشنوید. (نباید آن‌ها را دانلود و سپس گوش کنید چون جایگاه کلیپ‌ها به صورت رندوم است)
- ممکن است مرورگر شما برای مشاهده بخش‌کننده نیاز به آپدیت شدن داشته باشد. (ترجیحاً از آخرین نسخه‌ی مرورگر گوگل کروم استفاده کنید)
- لازم نیست به کلیپ‌ها دقت کنید که تفاوت‌ها را کشف کنید. خیلی عادی به آن‌ها گوش کنید. فقط اگر در یکی از کلیپ‌ها احساس کردید مغز شما آرمش یا استراحت بیشتری داشت، آنرا انتخاب کنید.

**کلیپ اول:**



**کلیپ دوم:**



انتخاب شما؟  
 کلیپ اول  
 کلیپ دوم

لطفاً دلیل انتخاب خود را به طور کامل بیان نمایید (اجباری):

آیا از هدفون استفاده کردید؟  
 بله  
 خیر

شکل ۱۲ - نمایی از صفحه نظرسنجی آنلاین

برای رفع هر گونه ابهام، در نظرسنجی، در مورد الگوریتم خود توضیحی ندادیم. یعنی گفته نشد که کلیپ دوم چه بهبودی نسبت به اولی داشته. حتی توضیح ندادیم که کدام کلیپ از نگاه ما بهتر است و حتی برای اینکه کاربر در ذهن خود تصور نکند که لابد کلیپ دوم بهبود یافته است و آنرا انتخاب کند، اینطور القا کردیم که جایگاه کلیپ‌ها رندوم است اما در عمل کلیپ معمولی را ابتدا پخش کردیم و کلیپ بهبود یافته را در جایگاه دوم.

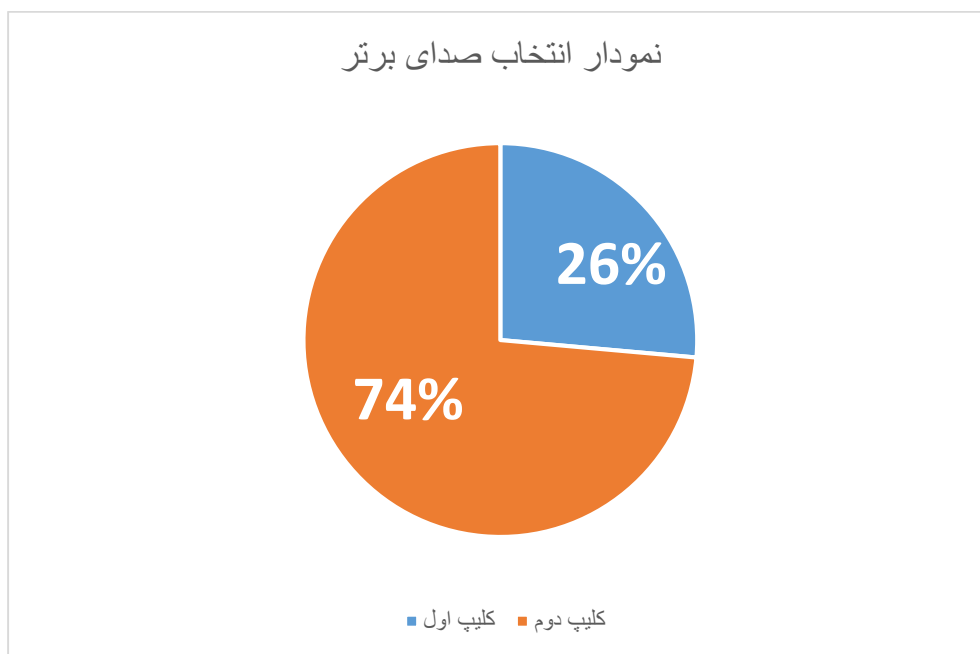
از کاربران خواسته شد که در صورت امکان از هدفون برای شنیدن کلیپ‌ها استفاده کنند.

برای اینکه از تقلب و پاسخ‌های غیرمعتبر جلوگیری شود، مدت زمان حضور کاربر در صفحه اندازه‌گیری شد و پاسخ‌هایی که مدت حضور آن‌ها کمتر از مجموع زمان دو کلیپ بود، حذف گردید. همچنین، مشخصات سیستم و آی‌پی هر کاربر نیز ثبت شد و مواردی که احتمال داده شد که یک کاربر به صورت تکراری و با نام‌های مختلف در نظرسنجی شرکت کرده حذف گردید.

مواردی که کاملاً مشخص بود که معتبر نیست؛ مانند کاربری که به خاطر سرعت کم اینترنت و بافر کردن صدا، دلیل خود را قطع و وصلی کمتر ذکر کرده بود، حذف گردید.

#### ۹-۴- نتیجه اندازه‌گیری به روش نظرسنجی

از مجموع ۷۷ نظر، ۵۳ نظر معتبر شناخته شد و تحلیل بر روی آن‌ها با استفاده از نرم افزار اکسل<sup>۱</sup> شروع شد. بر این اساس، درصد افرادی که کلیپ اول یعنی کلیپ بدون اعمال ایده را انتخاب کرده‌اند نسبت به کسانی که کلیپ دوم یعنی کلیپ پس از اعمال ایده را انتخاب کرده‌اند، در نمودار زیر مشخص است:



همانطور که مشخص است، به طور شگفت‌آوری، ۷۴ درصد شرکت‌کنندگان، کلیپ دوم را ترجیح داده‌اند. برخی از دلایل که در نظرات ثبت شده است، کاملاً مشخص می‌کند که ایده این پژوهش به خوبی توانسته باعث بهبود صدای تبدیل شده از متن شود؛ به طور مثال:

- مهدی، دانشجو، ۱۹ ساله گفته است:

<sup>1</sup> Microsoft Excel

- کلیپ دوم مکث دارد و باعث میشه گوش رو اذیت ندهد. حرف زدن گوینده یکنواخت نیست و همیشه تمرکز کرد. و قشنگ میشه حرف گوینده رو فهمید.
- حسن، دانشجو، ۲۰ ساله:
- تا جایی که من به این دو کلیپ گوش کردم. کلیپ اول خیلی خسته کننده بود چون نمیشد درست فهمید که چه میگوید ولی صدای دومی یکم پایین تر بود و قابل تحمل
- سید حسین، کارمند، ۳۶ ساله:
- با سلام و خسته نباشید
- کلیپ دوم از نظر اینجانب بهتر و آرامش در صحبت بیشتر بود و طبیعی به نظر می رسید هر چند که تفاوت زیادی دیده نشد وان تفاوت کم بود ولی خوب کلمات بهتر ادا می شد در کلیپ دوم با تشکر
- فاطمه، دانشجو، ۲۴ ساله:
- با سلام و خسته نباشید بابت ا ارائه کلیپ ها. من کلیپ دوم را پسندیدم چون متن ادبی بود باید بین جملات مکسی صورت میگرفت تا زیبایی متن مشخص میشد. این مکث که در کلیپ دوم وجود داشت به زیبایی متن اضافه میکرد.
- محمد رضا، کارمند، ۳۲ ساله نظر جالبی دارد که نکات جالبی را مشخص می کند:
- با سلام. من هر دو را گوش دادم ولی کلیپ اول خسته کننده بود انگار فقط قراره این متن رو تند تند بخونه و تمام بشه ولی کلیپ دوم با آرامش خاصی و مکث لازم زمانی که جمله تمام میشود وجود دارد که عجله نکردن روی تمام کردن متن این نوع بیان را شیوا تر و متناسب با نوع ادبی متن قرار داده است.
- سید بهزاد، دانشجو، ۲۴ ساله معایب کلیپ اول را اینطور بیان می کند:
- ۱- به نظر من در کلیپ اول سرعت گفتگو بالا بود
- ۲- انعکاس صدا زیاد بود و برخی از جمله ها قابل فهم نبود
- ۳- هنوز جمله قبلی تمام نشده بود جمله ی بعدی شروع می شد
- مریم، دانش آموز، ۱۷ ساله:
- کلیپ دوم آرامشش حد اقل بیشتر بود .
- مینا، دانشجو، ۲۲ ساله گفته است:
- به نظر من کلیپ دوم واضح تر بود و با شنیدن اون ذهن خسته نمیشه و اعصاب کمتر آشفته میشه!

- وحید، دانشجو، ۲۸ ساله، همان نظری را دارد که ما در مورد بسیاری از مبدل‌های متن به گفتار داریم:
  - در کلیپ اول احساس کردم کسی دنبال این بنده خدا می‌کرد و تندتر می‌خواند؛ ولی در کلیپ دوم تقریباً این احساس کمتر بود.
  - نگارنده: این نظر و نظرات مشابه زمانی جالب می‌شود که بدانیم ما در کلیپ دوم، سرعت تلفظ را هرگز تغییر ندادیم!

در مقابل، کسانی که کلیپ اول را انتخاب کرده‌اند، نظراتی شبیه به این نظرات داشته‌اند:

- امید، فوق دیپلم، ۱۸ ساله گفته است:
  - اولین و مهم‌ترین به نظر من باید گویا بودن صدا باشد تا کاربر بفهمد چه می‌گوید! در کلیپ دوم اینگونه نیست! کلیپ اول گویا تر است به نظر من حیاتی‌ترین موضوع گویا بودن صدا است...
- عباس، ۳۰ ساله:
  - درک آن راحت تر است و قابل فهم تر است
- محمد، کارمند، ۴۳ ساله ظاهراً متوجه کم و زیاد شدن صدا شده است و از موضوع راضی نبوده است:
  - ۱- تن صدای کلیپ دوم کم و زیاد میشد. ۲- کلیپ دوم بازگشت صدا داشت.
- فهیمه، ۲۳ ساله:
  - به نظر من کلیپ اول واضح تر بود کمتر باید توجه می‌کردی مطالب راحت تر شنیده میشد.

#### ۱۰-۴- خلاصه فصل

در این فصل با دو ابزاری که در اختیار داشتیم؛ یعنی استفاده از دستگاه EEG و نظرسنجی آنلاین، به بررسی ایده‌مان پرداختیم. هر چند به خاطر عدم وجود امکانات و متخصصین مورد نیاز، امکان اثبات ایده‌مان از طریق روش اول را نداشتیم اما به خوبی واقف بودیم که «نظرسنجی» یکی از پایه‌های تصمیم‌گیری در علوم مختلف بوده است. بنابراین با اعمال ایده‌مان بر روی یک صدای تبدیل شده از روی متن، و پخش آن در کنار صدای خام برای حدود ۸۰ نفر از کاربران با جنسیت و سنین مختلف، اقدام به برگزاری یک نظرسنجی نمودیم و در نهایت با تحلیل آمار، این روش توانست به خوبی ایده‌مان را اثبات نماید. ۷۴ درصد از شرکت کنندگان با ما موافق بودند و فایل صوتی بهبود یافته را قابل تحمل‌تر دانستند.

## فصل ۵:

# نتیجه‌گیری و پیشنهادات



## ۱-۵- نتیجه‌گیری

گفتار از ابزارهای اولیه‌ی انسان‌ها برای برقراری ارتباط به شمار می‌رود. تبدیل متن به گفتار و یا سنتز گفتار به معنای تولید خودکار شکل موج سیگنال گفتار است و در طول دهه‌های اخیر مورد توجه فراوانی قرار گرفته است. به طوری که سیستم‌های بسیاری در این زمینه تولید شده است. به همان میزان، ضرورت تولید یک صدای قابل قبول و قابل تحمل نیز افزایش یافته است. تجربه نشان می‌دهد که کاربران هر چند در ابتدا از کار با نرم افزارهای تبدیل متن به گفتار استقبال می‌کنند و تا حدودی ذوق‌زده می‌شوند اما به خاطر عدم شباهت گفتار مصنوعی به گفتار طبیعی و از آن مهم‌تر، خستگی ذهنی آن‌ها در هنگام شنیدن گفتار مصنوعی به ویژه در متون طولانی مدت، به مرور رغبت خود را برای استفاده از چنین نرم افزارهایی از دست داده و ترجیح می‌دهند زحمت و هزینه‌های خواندن متون با چشم و یا ضبط صدای طبیعی را قبول نمایند.

پژوهش حاضر تلاشی بود در جهت کاهش نواقص صدای ماشینی و در نتیجه، افزایش رغبت کاربران برای استفاده از نرم افزارهای مبدل متن به گفتار.

ایده‌ی مطرح شده در پژوهش، از تحلیل رفتار روزانه انسان‌ها شکل گرفت. به طور مثال، انسان‌ها پس از شنیدن طولانی مدت صدای یکنواخت مجری اخبار، ترجیح می‌دهند کمی حجم صدای تلویزیون را کاهش دهند و سپس ناخواسته پس از مدتی حجم صدا را افزایش می‌دهند و این روال چندین بار ممکن است تکرار شود. ما قصد داشتیم در این پژوهش ثابت کنیم که این رفتار ناخودآگاه می‌تواند در نرم افزارهای پخش صدا و یا نرم افزارهای تبدیل متن به صدا به صورت آگاهانه انجام شود و نیازی نباشد که کاربر خود اقدام به کاهش و افزایش صدا نماید.

بر اساس آنچه در فصول گذشته آمد، می‌توان به این نتیجه رسید که ایده‌ی مطرح شده در این پژوهش، یعنی کاهش حجم صدا در سکوت و افزایش حجم صدا به نرمی، می‌تواند تا حد زیادی از خستگی ذهنی کاربر بر اثر شنیدن طولانی مدت صدای یکنواخت به ویژه صدای یکنواخت تولید شده در نرم افزارهای تبدیل متن به گفتار، بکاهد.

این ایده پژوهش می‌تواند در کاربردهای بسیاری مورد استفاده قرار گیرد. می‌توان روزی را تصور کرد که یک نرم افزار مبدل متن به گفتار بتواند بدون ایجاد خستگی ذهنی اقدام به خواندن مقالات و داستان‌های طولانی برای مخاطبان کند، از تأثیر منفی صدای یکنواخت مجری اخبار در رادیو و تلویزیون کاسته شود، در نرم افزارهای پخش صدا دکمه‌ای تعبیه گردد که کاربران به محض مواجه شدن با یک سخنرانی یکنواخت، با فعال کردن آن دکمه، صدای سخنران را قابل تحمل‌تر کنند و ده‌ها کاربرد دیگر.

هر چند می‌شد با در اختیار داشتن دستگاه‌های نوار مغز دقیق‌تر، این ایده را به روشی قابل دفاع‌تر اثبات نمود اما نتایج نظرسنجی بر روی بیش از ۷۰ نفر نشان داد که قریب به سه چهارم افراد شرکت کننده صدای بهبود یافته با الگوریتم مطرح شده در این پژوهش را به عنوان صدای قابل تحمل‌تر نسبت به حالت عادی (بدون بهبود) انتخاب کردند و این نشان از تأثیر مثبت الگوریتم ما بر روی صداها و نتیجه‌بخش بودن پژوهش دارد.

## ۲-۵- پیشنهادات

انتظار می‌رود طراحان نرم افزارهای تبدیل متن به گفتار با جدی گرفتن مشکل مطرح شده در این پژوهش، بر اساس ایده‌های مطرح شده، اقدام به بهبود کارایی نرم افزار خود کنند. همچنین طراحان نرم افزارهای پخش کننده صدا و ویدئو نیز در این راه بکوشند.

و اما دانشجویان و پژوهشگران می‌توانند به تحقیق بر روی ابهامات اشاره شده در این پژوهش پردازند.

### ۱-۲-۵- تخمین سکوت بعد

یکی از موضوعاتی که در حین انجام این پژوهش بدان نیاز بود و می‌تواند یک موضوع بسیار جالب برای پژوهش‌های بعدی باشد، تخمین سکوت بعدی در یک مکالمه زنده است. در صورتی که بتوان الگوریتمی برای این ابهام کشف کرد، می‌توان از آن حتی در نوبت‌دهی در الگوریتم‌هایی مانند SDM<sup>۱</sup> در شبکه‌های سوئیچینگ بسته‌ای استفاده کرد. به طور مثال می‌توان با تشخیص سکوت، به محض سکوت، نوبت استفاده از کانال را هر چه سریع‌تر به مشترک دیگر واگذار کرد. علاوه بر این، نتیجه آن پژوهش می‌تواند در ایده‌ی مطرح شده در پژوهش حاضر نیز مفید فایده باشد.

### ۲-۲-۵- آثار القائات قبل از شنیدن یک صدا بر روی شنوندگان آن صدا

یکی از نکات جالب در این پژوهش، این بود که بسیاری از افرادی که در نظرسنجی، هر دو کلیپ را شنیده بودند، بر اثر القائاتی که قبل از شنیدن کلیپ‌ها و با خواندن متن موجود در بالای کلیپ‌ها در صفحه نظرسنجی،

<sup>1</sup> Statisticstical Division Multiplexing

در ذهنشان ایجاد شده بود، در کلیپ‌ها چیزهایی احساس کردند که ما واقعاً آن تغییرات را اعمال نکرده بودیم! به طور مثال: یکی از افراد احساس کرده بود که کلیپ دوم مکث بیشتری نسبت به کلیپ اول دارد و یا یکی دیگر احساس کرده بود که در کلیپ دوم سرعت خواندن متن کمتر است در حالی که تنها تفاوت کلیپ دوم، کاهش صدا در سکوت و افزایش آن به نرمی بود. بنابراین، این موضوع که القائات قبل از نظرسنجی‌ها چه تأثیری در نتیجه نظرسنجی دارد، می‌تواند یک موضوع بسیار جالب برای پژوهش‌های بعدی باشد.

## فهرست منابع فارسی

- [۱] م. آریانا، مستندات نرم افزار آریانا، ۱۳۸۹.
- [۲] م. م. همایونپور، پژوهشنامه تبدیل متن به گفتار، شورای عالی انقلاب فرهنگی، دبیرخانه شورای عالی اطلاع رسانی، ۱۳۹۱.
- [۳] د. ا. اکبری، "استخراج ابعاد نیازمندی‌های تبدیل متن به گفتار در ایجاد پیکره‌های متناظر زبان فارسی،" شورای عالی اطلاع رسانی، ۱۳۸۸.
- [۴] م. م. همایونپور، ارائه داده‌ها، دستورالعمل و نرم افزارهای ارزیابی عملکرد سیستم های تبدیل متن به گفتار فارسی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، ۱۳۸۶.
- [۶] "مستندات صفحه‌خوان شیوا،" پژوهشکده پردازش هوشمند علام، ۱۳۸۶.
- [۱۱] ح. فؤادی، "آهنگ زبان فارسی،" مجله مهر، p. 968, 1312.
- [۱۳] س. سپینتا، "بررسی فونتیکی خصوصیات واج‌های زبان فارسی،" پایان‌نامه دکتری، دانشگاه تهران، ۱۳۵۱.
- [۱۴] ت. و. کامیار، "تکیه در زبان فارسی،" پایان‌نامه کارشناسی ارشد، گروه زبان‌شناسی دانشگاه تهران، ۱۳۴۷.
- [۱۵] ع. حق‌شناس، "آواشناسی،" ۱۳۷۱.
- [۱۶] م. اسلامی، کتاب «نحو و واج‌شناسی (یک سطح مشترک)»، ۱۳۸۲.
- [۱۸] ت. و. کامیار، "رویدادهای زیر زنجیری،" ۱۳۷۲.
- [۱۹] م. اسلامی، "نظام آهنگ زبان فارسی،" مجله زبان‌شناسی، جلد ۲، شماره ۱۷، ۱۳۸۱.
- [۲۰] سامعی، فرهنگ نظام و فرهنگ معین، ۱۳۶۸.
- [۲۴] فرشاد الماس گنج، سید محمد رض اهاشمی گلپایگانی، کارو لوکس، سید علی سید صالحی، محمود بی‌جن خان، "افزایش دقت بازشناسی واج‌های زبان فارسی با استفاده از خصوصیات نوایی گفتار در سطح واج‌ها،" مجموعه مقالات چهارمین کنفرانس بین‌المللی سالانه انجمن کامپیوتر ایران، تهران، ۱۳۷۷.
- [۲۵] ع. دفتریان، "تقطیع جمله به اجزای آن و اعمال نوای گفتار فارسی،" پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی امیرکبیر، دانشکده برق، ۱۳۷۹.
- [۲۹] ش. ملایی، "دستگاه EEG،" ۱۳۹۲. [درون خطی]. Available: <http://youurl.ir/5i3>.

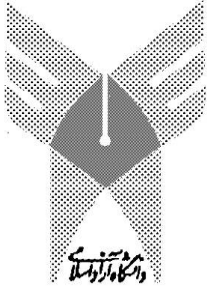
- [5] Yuxuan Wang, Kun Han, and DeLiang Wang, Fellow,, "Exploring Monaural Features for Classification-Based Speech Segregation," *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, 2012, *IN PRESS*, 2012.
- [7] C. Farmer, "Communication skills training - Voice tone," 2011. [Online]. Available: <http://www.corporatecoachgroup.co.uk/blogDetail.asp?blogid=169>.
- [8] B. Greenfield, "How You Can Use Sound And Music To Change Your Brain Waves With Laser Accuracy And Achieve Huge Focus And Performance Gains," [Online]. Available: <http://www.bengreenfieldfitness.com/2012/05/>.
- [9] T. T. K. Masuko, "Speech synthesis from HMMs using dynamic features," *Proc. of ICASSP*, p. 389–392, 1996.
- [10] M. J. a. H. Paper, "A reference grammar of Modern Persian," in *University of Michigan*.
- [12] L. Yarmohammadi, *Contrastive study of Modern English and Modern Persian*, Ph.D. thesis, Indiana University, 1964.
- [17] B. Mahjani, "An Instrumental Study of Prosodic Features and Intonation in Modern Farsi (Persian)," *Department of Theoretical and Applied Linguistics, School of Philosophy, Psychology, Linguistics and Social Sciences*, 2003.
- [21] Y.-a. Samareh, *A Course In Colloquial Farsi*, Tehran University, 1977.
- [22] Sheikhzadeh, "Farsi Language Prosodic structure, Research and implementation using a Speech Synthesizer," *EE Dept., AmirKabir Univ. of Tech*, 1999.
- [23] H. R. Abutalebi, "Implementation of a Text-To-Speech System for Farsi Language," Research Center of Intelligent Signal Processing (RCISP), Tehran, Iran, 2000.
- [26] S. Gillbee, 2008. [Online]. Available: <http://stackoverflow.com/questions/19353/detecting-audio-silence-in-wav-files-using-c-sharp>.
- [27] D. Jurafsky, "Speech Recognition and Synthesis," 2007. [Online]. Available: <http://www-nlp.stanford.edu/courses/Isa352/>.
- [28] J. Treasure, "The 4 ways sound affects us," in *TED*, California, 2009.

**Abstract:**

One of the most attractive research subjects nowadays is “Text To Speech”. People’s interest in “Siri” on Apple iPhone showed that people love talking and listening to machine! But the biggest problem in this case is that the voice produced by a machine is monotone and boring. There are dozens of articles about prosody in TTS but the main problem is that even if we consider this approach has been implemented as good as a real human voice, but we still have monotone voice problem even in between humans! Many audiences claim the speakers about their monotone voice.

This research is something beyond the prosody of voice. It tries to suggest a solution to decrease the mental fatigue while listening to a monotone sound.

In one sentence it suggests developers to create an artificial diversity on sounds.



Islamic Azad University  
Kermanshah Science and Research Branch  
Faculty of Engineering

Thesis for receiving «M.Sc» degree on  
Information Technology Engineering  
Computer Network

Subject:

**Suggesting a method for solving monotone  
voice problem in TTS**

Thesis Advisor:

**F. Mardukhi Ph.D.**

Consulting Advisor:

**M. Ahmadi Ph.D.**

By:

**Hamid Reza Niroomand**

Winter 2013