

# بهینه سازی پایگاه داده واژگان برای ساخت یک سیستم گفتار متن فارسی

ایمان راسخ<sup>۱\*</sup>، رضا جاویدان<sup>۲</sup>

<sup>۱</sup>دانشگاه آزاد اسلامی - واحد صفاشهر، <sup>۲</sup>دانشگاه آزاد اسلامی - واحد بیضا فارس

reza.javidan@gmail.com<sup>۱</sup>، iman.rasekh@gmail.com<sup>۲</sup>

چکیده: سیستم گفتار متن یکی از زمینه های روز در مهندسی پزشکی است که می تواند بعنوان ابزاری جهت کمک به افراد ناتوان بخصوص افراد نابینا در خواندن متون از طریق شنیدن بجای دیدن بکار رود. روشهای گوناگونی در زمینه ترکیب گفتار در جهان توسعه یافته است. ترکیب اتصالی گفتار (concatenative speech synthesis) یکی از روشهای متداول در این زمینه است. با این حال یکی از مهمترین مشکلات در ترکیب اتصالی متن فارسی، حجم بالا و ناکارآمد بودن سیستم پایگاه داده ها جهت ذخیره واژکها و اجزای تشکیل دهنده کلمات می باشد. در این مقاله به منظور بهینه سازی پایگاه داده واژگان از نظر حجم و کارایی، از یک واحد صوتی به نام دو صدایی به عنوان واحد بنیادی ذخیره سازی داده ها استفاده گردیده و در نهایت طبقه طراحی و آموزش پایگاه داده ها مورد بررسی قرار گرفته است. نتیجه روی داده های شبیه سازی شده بیانگر موثر بودن روش ارائه شده در این مقاله می باشد.

واژه های کلیدی: سنتز اتصالی گفتار، پایگاه داده واژگان، واژک، نرمالسازی، دوصدایی، سه صدایی

## ۱- مقدمه

طبیعی، تعیین ساختار متن، آنالیز تکواژشناختی (morphophonemics)، نرمال سازی متن، تعیین ساختار صرفی و نحوی و تحلیل معنایی، روشهای شناسایی الگو، روشهای محاسبات نرم، هوش مصنوعی، آمار، الگوریتم های پردازش سیگنالهای دیجیتال، تحلیل فرآیندهای تصادفی و مباحث مربوط به پردازش گفتار دیجیتال و مانند آن اشاره کرد [۱]، [۶].

تبدیل متن به صوت را می توان به دو قست اصلی تبدیل متن به واج (Phoneme) و تبدیل واج به صوت تقسیم کرد. خروجی قسمت اول کلیه اطلاعات کمی و کیفی لازم ساخت صوت می باشد. در مرحله دوم بایستی این اطلاعات بصورت صوتی قابل فهم و با کیفیت مطلوب ارائه گردند [۲]. در مرحله اول تبدیل متن به واج با استفاده از روالهای حرف به صدا انجام می گیرد که این روالها با استفاده از شبکه های عصبی پیاده سازی می شوند. مهمترین فناوری برای تبدیل متن به واج عبارتند از ترکیب فرمنت (Formant Synthesis) و ترکیب اتصالی، که هر یک از این روشها مزایا و معایب مخصوص به خود را دارند. در روش ترکیب اتصالی گفتار ذخیره شده طبیعی بصورت تکه تکه در کنار هم قرار می گیرند تا تولید یک گفتار خروجی کنند.

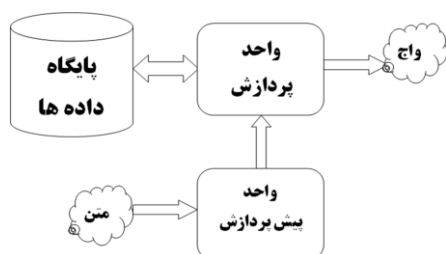
تبدیل متن به گفتار یکی از مباحث مهم در طراحی و ساخت سیستم های چند رسانه ای در فناوری اطلاعات و ارتباطات می باشد که کاربردهای گسترده ای نیز در پزشکی دارد. از این سیستم می توان به افراد دارای معلولیت کمک کرد تا مطالب نوشته شده را بجای مطالعه کردن بشنوند. پردازش گفتار در مهندسی پزشکی با سایر رشته های مهندسی دارای تفاوت هایی است. در سایر رشته ها معمولاً به سیگنال گفتار به صورت یک سیگنال عادی که حاوی اطلاعاتی است که باید به هر صورت ممکن از آن استخراج شود، نگاه می شود در حالی که در مهندسی پزشکی، محققان در پی دنبال کردن مسئله و مدلسازی آن به صورتی هستند که تا حد ممکن با اصول عملکرد جهاز صوتی و مبانی زیستی تولید گفتار در انسان هماهنگی داشته باشد و سعی می کنند از روش های استخراج ویژگی و تولید مدل هایی استفاده کنند که به روش های زیستی انسانی نزدیک تر باشند.

تبدیل متن به گفتار طیف گسترده ای از علوم را شامل میشود که از آن جمله میتوان به زبانشناسی، پردازش زبان

معماری کلی برای سیستم TTS (Text to speech Synthesizer) مورد استفاده در این مقاله ارائه شده است. در بخش سه روشقطعه قطعه کردن متن جهت قراردادن در پایگاه داده ها بررسی شده است. در بخش چهار ساختمان پایگاه داده های مورد استفاده تشریح شده و در بخش پنج ضمن معرفی دوصدایی و نیز نحوه استخراج دوصدایی طراحی پایگاه داده دوصدایی نیز مورد بررسی قرار گرفته و در نهایت نتیجه گیری در بخش شت ارائه شده است.

## ۲- معماری تبدیل متن فارسی به واج

بطور معمول معماری سیستم ترکیب متن به شامل سه بخش اساسی می باشد. شکل ۱ ساختمان کلی این سیستم را نشان می دهد [۴].



شکل ۱: سیستم ترکیب متن فارسی

در واحد پیش پردازش پردازشهای کلی بر روی متن ورودی انجام می گیرد. که به آن *نرمال سازی* می گویند. در طی این پردازش اعداد به حروف (و یا واجها) تبدیل می شوند برای مثال نشانه "۲" به "دو" تبدیل می شود. علائم به حروف (و یا واجها) تبدیل می شوند برای مثال نشانه ":" به "دو نقطه" تبدیل می گردد. تشدید بصورت تکرار حرف یکبار با واک نامشخص و یکبار بدون واک پیاده سازی می شود؛ مثلاً تشدید در *بنّا* بصورت *بننا* اعمال می گردد. تنوین، بدین شکل پیاده سازی می شود که تنوین و الف ماقبل آن حذف شده و یک *"نون"* ماقبل مفتوح اضافه می گردد. و او مکسوره در موارد خاص چون تشخیص *"خوا"* و جداسازی کلماتی چون *"خواهر"* از *"خواص"* جهت تشخیص واجها اعمال می شود [۳].

واحد پردازش خروجی مرحله پیش پردازش را گرفته و با الگوریتمهای بسیار ساده تبدیل به واجها در خروجی می کند. در واقع در این مرحله دو نوع پردازش انجام می گیرد یکی پردازش بر مبنا قوانین برای طبقه بندی هجاهای حافظه و دیگری جهت ساخت واج [۲]. در نهایت هم پایگاه داده ها برای ذخیره واجهای ایجاد شده به کار می رود. قبل از قرار دادن هر متنی در

مهمترین مساله در ترکیب اتصالی صدا ساخت پایگاه داده ها برای آن می باشد. روشهای گوناگونی برای ساخت این پایگاه داده ها وجود دارند. ساده ترین روش برای ساخت پایگاه داده های تولید گفتار توسط کامپیوتر روش ذخیره گفتار (Speech Storing) است که در این روش گفتار را در یک پایگاه داده ذخیره کرده و سپس به وسیله فرمان مربوط آن را عیناً در خروجی ظاهر می کنند. اگر مجموعه تمامی پیغامهای مورد نیاز محدود، کم و تغییر ناپذیر باشد ممکن است این روش مفید واقع شود [۹]. یک روش بهتر برای این کار این است که به جای تبدیل لغت به رشته آوایی، از یک پایگاه داده ها املاء - تلفظ (Orthographic-Pronunciation Database) استفاده شود و برای تمام لغات تلفظ آنها را در این پایگاه داده ها نگهداری شود. می توان از چنین روشی استفاده کرد ولی هزینه حافظه آن بسیار بالا است. روش دیگر این است که فقط کلماتی را که نمی توان توسط قوانین حرف به صدا واج آنها را بدست آورد در یک پایگاه داده ها استثنائات املاء - تلفظ (Exceptions Orthographic-Pronunciation Database) قرار داده و برای بقیه کلمات از قوانین تبدیل حرف به صدا استفاده شود. این روش سه مزیت دارد اول اینکه لغات بیگانه ای مانند بوئینگ و توپولوف که از سایر زبانها وارد زبان فارسی شده است را پوشش می دهد. دوم اینکه مشکل تاثیر ساختار داخلی بر تلفظ لغات را نیز حل می کند و خطاهای قواعد حرف به صدا نیز کاهش می یابد ولی هنوز دو مشکل اساسی وجود دارد اول اینکه لغات جدیدی مرتباً به زبان اضافه می شوند که لازم است به همراه تلفظشان به پایگاه داده ها اضافه شوند. دوم اینکه با اضافه کردن پیشوندها و پسوندها یی به لغت، لغات جدیدی بوجود می آید. بنابراین اگر تمام لغات در پایگاه داده ها نوشته شوند به طور غیر ضروری حجم پایگاه داده ها افزایش پیدا می کند [۷].

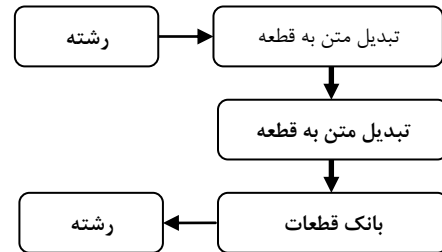
برای حل این مشکل بهتر است متن را به واحدهایی کوچکتر از لغت تجزیه کرد. یکی از این واحدها که زیر مجموعه لغت است واژک (Morpheme) کوچکترین واحد ترکیبی در زبان است. به پایگاه داده متشکل از واژکها پایگاه داده واژگان می گویند. در این مقاله از واحد صوتی دوصدایی (Diphone) برای طراحی پایگاه داده واژگان استفاده شده است. دوصدایی شامل دو نیم-واژک (Half-phoneme) مجاور می باشد در این روش از دوصدایی ها به عنوان عنصر پایه در پایگاه داده استفاده می کنیم

ساختار ارائه مقاله به صورت ذیل می باشد: در بخش دوم یک

پایگاه داده ها باید ابتدا آن را قطعه قطعه کرد

### ۳- قطعه قطعه کردن متن

در اغلب پروژه های سنتز صحبت، از سیستم تبدیل متن به صحبت با دیاگرام شکل ۲ استفاده می شود.



شکل ۲- بلوک دیاگرام سیستم

گرچه سنتز صحبت در زبان فارسی با توجه به ویژگی های آن نسبت به زبان های دیگر ساده تر است، اما تبدیل متن به دو صدایی یا واج بسیار مشکل می باشد. زیرا در فارسی واکه های " \_ "، " \_ "، " \_ " و در متن های معمولی کمتر دیده می شوند. همچنین سکون و علامت تشدید در متن وجود ندارد. برای حل این مسأله راه حل های متعددی ارائه شده است که هر کدام دارای مزایا و معایبی است. یکی از این راه حل ها استفاده از بانکی است که مشابه کلمات متن را به همراه اعراب و حرکات لازم در بر داشته باشد. با استفاده از این روش می توان هجاها را از هم جدا نموده و از روی هجاها دو صدایی های مناسب استخراج شوند.

مثلاً چنانچه داشته باشیم "مُحَمَّد"، رشته واجگانی به این صورت داریم: "CVCVCVCVC". در این رشته از انتها هجاها را جدا می کنیم. برای جدا کردن هجا باید در نظر داشته باشیم که واکه مرکز هجاست و نمی تواند در اول هجا قرار گیرد و قبل از واکه تنها یک همخوان می تواند در هجا وجود داشته باشد. بنابراین هجای "CVC" از انتها جدا می شود و به همین ترتیب الگوریتم می تواند تکرار شود تا همه هجاها به دست آیند. مرحله بعد قرار دادن این قطعات در پایگاه داده هاست.

### ۴- طراحی و بهینه سازی پایگاه داده واژگان

به پایگاه داده متشکل از واژکها پایگاه داده واژگان می گویند. واژکها از یک سطح پایدار ساده برخوردارند و کوچکترین واحد ترکیبی زبان هستند و همچنین پایه ساختمانی لغات را تشکیل می دهند. آنها دارای خاصیت پایداری در یک زبان هستند. به عبارت دیگر خیلی کم اتفاق می افتد که واژک جدیدی به زبان

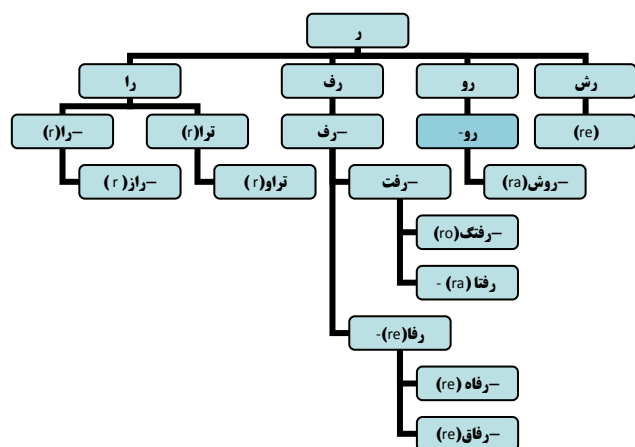
اضافه شود و این به این معناست که پایگاه داده واژک اغلب احتیاجی به بروز رسانی ندارد [۱۶]، [۳]. اگر لغت جدیدی به زبان اضافه شود و آن لغت ترکیبی از واژکهای درون فرهنگ واژک باشد دیگر احتیاجی به اضافه کردن آن به فرهنگ نیست. لغات بسیاری از زبانهای جهان را می توان با استفاده از قواعد لغت به واژک همان زبان تجزیه کرد و فقط تلفظ واژکهای آن زبان را نگهداری نمود. مهمترین مسأله در طراحی پایگاه داده واژگان انتخاب واحد صوتی مناسب برای آن می باشد.

پایگاه داده ها باید در ابتدا آموزش داده شود. یعنی داده های مورد نظر (واژگان زبان و معادل صوتی آنها) باید در پایگاه داده ها قرار داده شوند. آموزش پایگاه داده ها در چهار مرحله انجام می گیرد: ابتدا رشته ورودی به واژگان معادل با استفاده از روالهای تجزیه واژک تبدیل می شود؛ در مرحله دوم لغاتی که پوشش واژکی معادل ندارند (در پایگاه داده ها معادلی برای آنها وجود ندارد) تجزیه می شوند این مرحله به معنی جداسازی کلمه از وند (پیشوند یا پسوند) می باشد؛ در مرحله بعد وندها و ریشه هایی از لغت را که در پایگاه داده واژک وجود دارند و سیستم آنها را می شناسد جداسازی شده و معادل گفتاری آنها درج می شود و نهایتاً اگر در آخر ریشه ای باقی ماند که واژک معادل آن در پایگاه داده واژک موجود نبود. روالهای حرف به صدا فراخوانی می شود و این ریشه باقی مانده توسط آنها به گفتار معادلش تبدیل می شود. به منظور یاد گیری سیستم این ریشه ناشناخته را به پایگاه داده ها واژگان اضافه کرد.

به طور مثال فرض کنید روالهای تجزیه لغت به واژک در مورد لغت " غیر مسؤولانه " به شکست انجامیده است. حال این لغت توسط روالهای وند عریانی (Bared Affix Procedures) واژکهای آشنای "غیر" و "انه" از آن جدا می شوند و فقط "مسؤول" می ماند که در فرهنگ واژک وجود ندارد. سپس "مسؤول" توسط روالهایی که از قواعد حرف به صدا استفاده میکنند به گفتار معادل تبدیل می شود ولی گفتار معادل "غیر" و "انه" از درون فرهنگ واژک بدست می آید [۷].

آموزش اتوماتیک قوانین جهت نگاشت حرف به واج توسط قانون کوهونن که بنام " گسترش پویای بافت " معروف است انجام می گیرد. در این روش قوانینی که عمومی تر هستند در یک درخت تصمیم قرار می گیرند. این درخت تصمیم، بر اساس گسترش بافت در اطراف هر حرف، فقط تا سه حرف بافت گسترش پیدا کند، الگوریتم گسترش پویای بافت (Context Dynamically Expanding) یا DEX سری قوانین وابسته به متن را توسط مثالهایی از رشته های مبدأ و مقصد، ایجاد می

متوقف شویم یعنی اینکه متناظر با گسترش بافت که بایستی انجام پذیرد، هیچ شاخه متناظری در درخت وجود نداشته باشد، شاخه ای که احتمال وقوعش بیشتر است را در نظر می گیریم [۸].



شکل ۳: درخت چند سطحی مربوط به حرف "ر"

## ۵- سیستمهای TTS مبتنی بر پایگاه داده دوصدایی

در این مقاله از واحد صوتی دوصدایی (Diphone) برای طراحی پایگاه داده واژگان استفاده شده است. دوصدایی شامل دو نیم-واژک (Half-phoneme) مجاور می باشد در این روش از دوصدایی ها به عنوان عنصر پایه در پایگاه داده استفاده می کنیم.

### ۵-۱ استخراج دوصدایی از صحبت فارسی

می توان محاسبه کرد که تعداد دوصدایی ها در زبان فارسی ۱۰۰۲ عدد است [۲]. اگر بازه زمانی هر دو صدایی به طور متوسط ۱۵۰ میلی ثانیه باشد، کل فضای لازم جهت ذخیره دو صدایی ها برابر مقدار زیر خواهد بود:

$$H = 1002 * 0.15 * B * S \quad (1)$$

و چنانچه دو صدایی ها با کیفیت CD (S 44100 B = 2) = ذخیره شوند. برابر ۱۳ MB خواهد بود.

دو صدایی ها را نمی توان به صورت جداگانه ضبط نمود و باید آنها را از صحبت طبیعی استخراج نمود. برای این منظور ابتدا باید صحبتی تهیه شود که حاوی تمامی دو صدایی ها باشد و در هنگام ضبط نیز باید سعی کرد تا محیط ضبط یکنواخت و عاری از نویز باشد. همانگونه که گفته شد، سیگنال صحبت از آواهای بی واک و واک دار تشکیل شده است. آواهای واک دار شبه متناوبند و آواهای بی واک به شکل نویز هستند. برای واج های واک دار محل شروع و خاتمه دو صدایی مهم است. محل

کند که این قوانین قادر هستند متن را به واکه متناظر آن، تبدیل نمایند. در اینجا فقط خود حرف مورد توجه نیست بلکه اطراف آن نیز در مورد توجه می باشد و اگر حروف کناری تاثیری بر روی تلفظ این حرف می گذارند، این حالات نیز در یک درخت تصمیم ذخیره می شود. در این روش چون تک تک حروف کلمه واج نگاری می شود نمی توان نوع کلمه را تشخیص داد بنابراین نمی توان بر روی آن تکیه قرار داد.

هدف اصلی DEX تولید قوانین تبدیل یک سمبل واحد به یک رشته از سمبل ها است. جداسازی واژکها از وند در این الگوریتم به دو روش انجام می گیرد گسترش متقارن و گسترش نامتقارن؛ در روش گسترش متقارن از دو طرف حرف ابتدا از سمت چپ یک حرف پیش رویم و سپس یک حرف از سمت راست همان حرف مورد نظر پیش می رویم اما در روش گسترش نامتقارن از سمت چپ دو برابر و از سمت راست یک برابر گسترش داده می شود. این روش معمولاً بهتر جواب می دهد و دارای درصد کارایی بیشتری است. گسترش به سمت چپ در زبان انگلیسی متناظر با گسترش سمت راست در زبان فارسی می باشد.

جدول ۱ نمایش دو نوع گسترش حول یک حرف از کلمه

گسترش به راست	گسترش متقارن	سطح بافت
t	t	۰
ti	ti	۱
Tio	atio	۲
atio	atio	۳
ation	iatio	۴

گسترش متقارن و نامتقارن برای حرف "t" در کلمه abbreviation در جدول ۱- نشان داده شده است. در نظر داشته باشید که گسترش به سمت راست در زبان انگلیسی متناظر با گسترش سمت چپ در زبان فارسی می باشد [۱۳].

در شکل ۳ درخت چند سطحی مربوط به حرف (ر) که از کلمات "راز، رفتار، رفتگر، رفاقت، رفاه، تراوت، پرورش و روش" استخراج شده است ارائه گردیده است.

برای استفاده از چین درختی، جستجو از بالای درخت شروع میشود و هرگاه در هر گره به تداخل قوانین برخورد کردیم بر طبق گسترش بافت کلمه، از شاخه های درخت به سمت پایین حرکت می نمایم تا به جایی برسیم که فقط یک قانون واحد وجود داشته باشد و تداخلی نداشته باشیم. اگر در حین پایین آمدن از گره ها، قبل از رسیدن به برگها، در گره های میانی

لغات ورودی توسط روالهای تجزیه واژگ به دوصدایی های تشکیل دهنده تجزیه می شوند بردار ورودی (شامل دوصدایی ها) وارد یک شبکه عصبی می شود شبکه عصبی یک دسته خاص که نشان دهنده بردار ورودی می باشد ایجاد می کند در مرحله بعد هر دسته توسط مبدل به یک رشته دودویی از ۰ و ۱ تبدیل می شود این رشته باینری در نهایت به سیگنالهای صحبت تبدیل می شود و در پایگاه داده ها ذخیره می شود.

### ۳-۵ - معماری سیستم TTS مبتنی بر پایگاه داده دوصدایی ها

در این مرحله باید رشته دوصدایی ورودی را به صدا تبدیل کنیم هر دو صدایی با دو کرکتر نشان داده می شود مراحل تبدیل دوصدایی به صوت در این سیستم به ترتیب زیر است:

۱. ابتدا هر لغت به دوصدایی های تشکیل دهنده آن شکسته می شود.

۲. هر دوصدایی وارد پروسه آموزش می شود تا الگوی خروجی مناسبی تشکیل گردد.

۳. این الگو با الگوهای ذخیره شده در پایگاه داده های دوصدایی ها مقایسه می شود و با توجه به معیار فاصله اقلیدسی (distance metric Euclidean) با دوصدایی های موجود در پایگاه داده ها طبقه بندی می شود.

۴. دوصدایی تشخیص داده شده (با استفاده از فاصله اقلیدسی) به صدای معادل تبدیل می شود [۱۰].

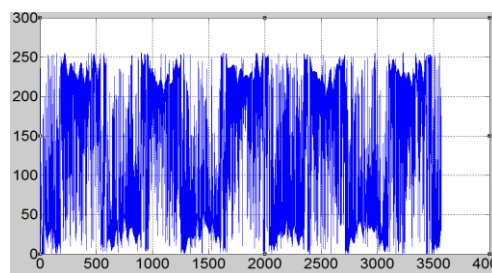
### ۶- جمع بندی و تست

برای شبیه سازی سیستم جهت تشخیص هجا از نرم افزار متلب و زبان VB.Net انجام گرفت. در این روش پردازش کلمه به کلمه یا بر طبق فواصل بین حروف انجام گرفت. در این طراحی سیستم حروف بصورت سریال وارد بلوک می گردند و پس از تشخیص کلمات با توجه به موقعیت فضاهای خالی پردازش روی کلمات انجام گرفته و واجها در خروجی قرار می گیرند. سپس برای تکمیل مراحل و ایجاد صوت با استفاده از روش سنتز اتصالی مرحله تبدیل واج به صوت نیز انجام گرفت.

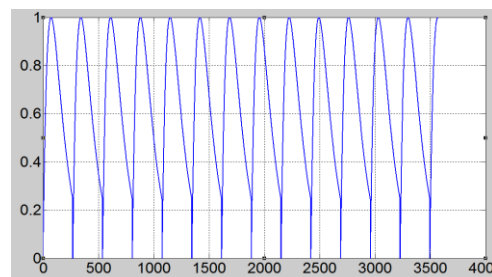
جهت تست سیستم از یک پایگاه داده لغات فارسی شامل ۱۸۰ لغت که واجهای مربوط به هر واژگ به آن افزوده شده بود استفاده گردی. و سیستم با کلیه پایگاه داده تست گردید. از کل مجموعه ۱۶۰ هجا استخراج گردید. نتایج تست سیستم با آموزشهای مختلف در جدول ۲ ذکر گردیده است. نمودار کارایی سیستم نیز در شکل ۵ رسم گردیده است، همانطور که مشخص

شروع دو صدایی چنانچه یک آوای واک دار باشد، از محل شروع ناحیه کمینه سیگنال شبه پریودیک و محل خاتمه دو صدایی نیز محل شروع ناحیه کمینه در نقطه صفر خواهد بود. برای آواهای بی واک که شکل موج آنها نویزی است. همچنین در تقطیع دو صدایی باید دقت نمود محل شروع دو صدایی در ناحیه درنگ (و تقریباً وسط ناحیه درنگ (یک واج و محل خاتمه آن نیز در ناحیه درنگ واج مجاور باشد.

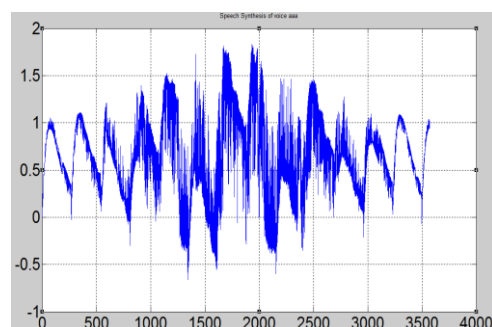
شکل-۴ سنتز دصدایی "Bi" را با استفاده از متلب ۷ نشان می دهد.



(الف)



(ب)



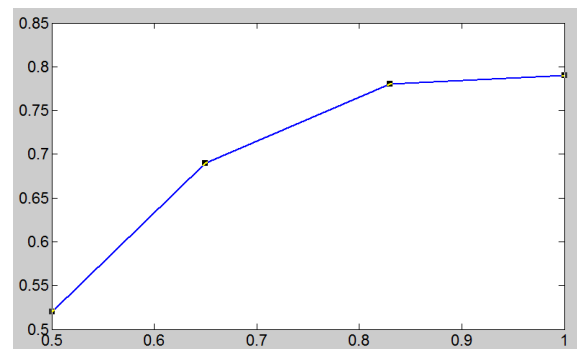
(ج)

شکل ۴ - الف - شکل موج اولیه دو صدایی "بی" ب- پالس منبع تحریک مورد استفاده برای سنتز دو صدایی "بی" ج- سنتز دوصدایی "بی"

### ۵-۲ - آموزش پایگاه داده های دوصدایی

هدف از این مرحله آموزش پایگاه داده ها (پر کردن فیلدهای پایگاه داده ها) با سیگنالهای صوتی می باشد. ابتدا

است افزایش درصد آموزش منجر به افزایش کارایی سیستم می گردد. برای تست سیستم از سه پایگاه داده استفاده شد، پایگاه داده اول شامل ۹۰ کلمه (۵۰٪ لغات) پایگاه داده دوم شامل (۵۰٪ لغات) ۱۵۰ کلمه و پایگاه داده سوم شامل ۱۸۰ کلمه (۱۰۰٪ لغات) بود. سنتز انجام شد و نتایج به شکا زیر بود. همانطور که مشخص است افزایش درصد آموزش منجر به افزایش کارایی سیستم می گردد. اما افزایش در محدوده بالا با افزایش کمتری در کارایی سیستم همراه می باشد که می تواند به دلیل ازدیاد هجاهای با نوشتار یکسان ولی تلفظ متفاوت باشد



شکل ۵. منحنی افزایش کارایی با بالا رفتن آموزش

## ۷- نتیجه گیری

سنتز صحبت فارسی می تواند از روش اتصال قطعات صحبت با کیفیتی مناسب انجام پذیرد. انتخاب قطعه باید به گونه ای باشد که بتواند سیگنال صحبت مشابه سیگنال طبیعی تولید کند. یکی از قطعات مناسب دو صدایی می باشد. ابتدا باید واژگانی به سیستم آموزش داده شود و سپس سیستم با مقایسه ورودیهای متنی با واژگان موجود در پایگاه داده ها خروجی مطلوب را ایجاد می کند. در این مقاله به منظور بهینه سازی پایگاه داده واژگان از لحاظ حجم و کارایی از واحدهای کوچکتری به نام دوصدایی استفاده شده است و طریقه طراحی و آموزش پایگاه داده آموزش داده شده است. استفاده از این واحد منجر به کارایی بهتری نسبت به سیستمهای قبلی شده است و کارایی تا ۸۳.۳٪ افزایش یافته است.

## مراجع:

- [۲] مجید نم نبات، محمد مهدی همایون پور، "تبدیل حرف به صدا در زبان فارسی به کمک شبکه های عصبی پرسپترون چند لایه ای" نشریه مهندسی برق و کامپیوتر ایران، ۱۳۸۵
- [۳] آرمین سجادی، محمد مهدی همایونپور، "مدلسازی دانش تکواژ شناختی زبان فارسی به کمک گرامرهای پیوندی"، نامه ای از پژوهشکده، شماره ۱ پیاپی ۵، ۱۳۸۵.
- [۴] (دبیدیان، ۱۳۷۸) دبیدیان ک، "مطالعه و پیاده سازی مدار تولید صدا فارسی"، دانشگاه شهید بهشتی، پایان نامه
- [۵] (طاهری، ۱۳۸۱) طاهری ر، "شبه سازی سنتز کننده KLATT به زبان VHDL" دانشگاه شهید بهشتی، پایان نامه کارشناسی ارشد الکترونیک
- [6] H. R. Abutalebi and M. Bijankhan, "Implementation of a Text-to-Speech System for Farsi Language," Sixth International Conference on Spoken language Processing (ISCA), 2000.
- [7] F. Hendessi, A. Ghayoori, and T. A. Gulliver, "A Speech Synthesizer for Persian Text Using a Neural Network with a Smooth Ergodic HMM," ACM transactions on Asian Language Information Processing (TALIP), 2005.
- [8] W. Barkhoda, F. Daneshfar, and B. ZahirAzami, "Design and Implementation of a Persian TTS System Based on Allophones Using Neural Network," SCEE'08, Zanjan, Iran, 2008 (In Persian).
- [9]
- [10] F. Daneshfar, W. Barkhoda, and B. ZahirAzami, "Implementation of a Text-to-Speech System for Persian Language," ICDT'09, Colmar, France, July 009.
- [11] W. M. Thackston, Sorani Persian: "A Reference Grammar with Selected Reading, Harvard: Iranian Studies at Harvard University", 2006.
- [12] S. Baban, Phonology and Syllabication in Persian Language, Persian Academy Press, first edition, Arbil, 2005 (In Persian).
- [13] R. J. Deller, et al., Discrete time processing of speech signals, John Wiley and Sons, 2000.
- [14] M. N. Rao, S. Thomas, T. Nagarajan, and H. A. Murthy, "Text-to-Speech Synthesis using syllable-like units," National Conference on Communication, ndia, 2005.
- [15] H. Sak, A Corpuse-Based Concatenative Speech Synthesis System for Turkish, M.Sc. Thesis, Bogazici University, 2004
- [۱۶] Mario , Malcangi , David ,Frontini , " Language-independent, neural network-based, text-to-phones conversion " Volume 73 , Issue 1-3 (December 2009) Pages 87-96 ISSN:0925-2312

[۱] محمد مهدی همایون پور، سید مصطفی موسوی، "تولید پارامترهای

سنتز گفتار فارسی با استفاده از مدلهای مخفی مارکف و درخت

تصمیم گیری"، نشریه علمی پژوهشی انجمن کامپیوتر ایران، مجلد ۲،

شماره ۱ و ۳ (الف)، صفحات ۱۹-۳۰، بهار و پائیز ۱۳۸۳.