١١٢٢٧٨

دانشکده مهندسی برق و کامپیوتر
گروه مهندسی برق – مخابرات

پایان‌نامه
برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق - مخابرات

عنوان

تبدیل متن به گفتار با استفاده از تبدیل ویولت

استاد راهنما
دکتر محمد علی طینتی

استاد مشاور
دکتر بهزاد مظفری تازه کند

پژوهشگر
علی قلی پور

اسفند ۱۳۸۷

۱۱۲۲۷۸

دانشکده مهندسی برق و کامپیوتر
گروه مهندسی برق – مخابرات

پایان‌نامه
برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق - مخابرات

عنوان

تبدیل متن به گفتار با استفاده از تبدیل ویولت

استاد راهنما
دکتر محمد علی طینتی

استاد مشاور
دکتر بهزاد مظفری تازه کند

پژوهشگر
علی قلی پور

اسفند ۱۳۸۷

این پایان‌نامه طی قرارداد شمارهٔ ۵۰۰/۲۰۶۰۹/ت مورخ ۸۶/۱۲/۲۸ از طرف مرکز تحقیقات مخابرات ایران مورد حمایت مالی قرار گرفته است که بدین‌وسیله کمال قدردانی از حمایت مادی و معنوی مرکز تحقیقات ایران می‌نمائیم.

تقدیر و تشکر

در ابتدا بر خود لازم می دانم از همهٔ عزیزانی که در تهیهٔ این پایان نامه، حقیر را یاری نمودند تشکر نمایم:

جناب آقای دکتر محمد علی طینتی استاد راهنمای بنده در این پایان نامه،

جناب آقای دکتر بهزاد مظفری تازه کند، استاد مشاور اینجانب که در تمام این مدت ناظر و راهنمای من بودند،

و بالاخره پدر و مادر مهربانم که در این مدت لطف بی دریغشان همواره شامل حال بنده بود.

علی قلی پور

نام: علی	**نام خانوادگی دانشجو:** قلی پور
عنوان پایان‌نامه: تبدیل متن به گفتار با استفاده از تبدیل ویولت	
استاد راهنما: دکتر محمد علی طینتی	**استاد مشاور:** دکتر بهزاد مظفری تازه کند
مقطع تحصیلی: کارشناسی ارشد	**رشته:** مهندسی برق **گرایش:** مخابرات- سیستم
دانشگاه: تبریز	**دانشکده:** مهندسی برق وکامپیوتر
تاریخ فارغ‌التحصیلی: ۸۷/۱۲/۰۳	**تعداد صفحات:** ۱۱۵
واژه‌نامه: تبدیل متن به گفتار - تبدیل ویولت - سنتز صوت - آنالیز صوت – یافتن گام- پارامترهای LPC	

چکیده: در این پایان نامه هدف بررسی روش‌های مختلف تبدیل متن به گفتار و مدل سازی این تبدیل برای زبان فارسی می‌باشد. روند تبدیل متن به گفتار یک روند بسیار طولانی می‌باشد که انجام کامل آن در طی یک پایان‌نامه مقدور نمی‌باشد هدف ما از این پایان‌نامه مدل سازی این تبدیل برای تعداد محدودی از کلمات در زبان فارسی است همچنین تکیه بیشتر بر روی سنتز صوت می‌باشد. در این پایان نامه بررسی تبدیل ویولت در بهبود کیفیت گفتار سنتز شده و همچنین کاربرد این تبدیل در تسریع این روند مورد بررسی قرار خواهد گرفت.

روند تبدیل متن به گفتار به صورت زیر خلاصه می‌شود:

- شناسائی کاراکترهای تایپ شده
- شناسائی کلمه تایپ شده به همراه تلفظ مربوطه
- در نهایت ایجاد گفتار با استفاده از بانک اطلاعاتی ذخیره شده

تمام این بخش‌ها در محیط MATLAB انجام گرفته است. از پارامترهای بسیار مهم این تبدیل قدرت درک توسط شنونده است که نتایج مربوط به تست در پایان نامه آورده شده است که نتایج قابل قبولی می‌باشند. یکی از مسائلی که قدرت درک را به شدت پائین می‌آورد ناپیوستگی بین قطعات از قبل ضبط شده می‌باشد در این پایان نامه سعی بر از بین بردن این ناپیوستگی‌ها شده است تا قدرت درک افزایش پیدا کند. از دیگر پارامترهای مهم این تبدیل سرعت دسترسی به بانک اطلاعتی و به صورت کل سرعت تبدیل متن به گفتار است، با توجه به نحوه برنامه نویسی این کار تقریباً به صورت آنلاین انجام می‌گیرد. از دیگر پارامترهای مهم طبیعی به نظر رسیدن صوت سنتز شده است که این کار هم با استفاده از روش‌های تغییر همپوشانی بین فریم‌های مربوط به بخش‌های صدا دار و الگوهای در دست قابل حصول است. این پایان نامه یک پایه مناسب برای کسانی که می‌خواهند در این زمینه کار کنند می‌باشد.

فهرست مطالب

فصل اول: پیش‌گفتار ۱

۱-۱- مقدمه ۲

۱-۲- تبدیل متن به گفتار ۲

۱-۲-۱- مخابرات بین انسان و ماشین ۲

۱-۲-۲- سیستم‌های خوانا ۵

۱-۲-۳- سیستم‌های مترجم ۵

۱-۲-۴- کاربردهای دیگر ۶

۱-۳- تاریخچه ۶

۱-۴- نگاهی گذرا به پایان نامه ۷

فصل دوم: بررسی منابع ۹

۲-۱- مقدمه ۱۰

۲-۲- نگاهی به تبدیل متن به گفتار ۱۰

۲-۲-۱- ورودی ۱۱

۲-۲-۲- نرمالیزه کننده متن ۱۲

۲-۲-۳- دیکشنری ۱۲

۲-۲-۴- تبدیل گرافم به واج ۱۲

۲-۲-۵- قوانین مربوط به زیر و بمی صدا ۱۳

۲-۲-۶- قوانین مربوط به فونتیک ۱۳

۲-۲-۷- بلوک تولید کننده صدا ۱۴

۲-۳- ابزارهای استفاده شده در تبدیل متن به گفتار ۱۴

۲-۳-۱- تبدیل متن به گفتار با استفاده از شبکه‌های عصبی ۱۴

۲-۳-۲- تبدیل متن به گفتار با استفاده از تبدیل فوریه سریع ۱۶

۲-۳-۳- تبدیل متن به گفتار با استفاده از تبدیل ویولت ۱۷

۲-۳-۴- مدل ترکیبی از سیگنال و نویز برای تبدیل متن به گفتار ۲۰

۲-۴- الگوریتم‌های ارائه شده در سنتز صوت ۲۱
۲-۴-۱- روش LPC ۲۲
۲-۴-۲- روش PSOLA ۲۲
۲-۴-۲-۱- روش TD-PSOLA ۲۳
۲-۴-۲-۲- روش FD-PSOLA ۲۳
۲-۴-۲-۳- روش LP-PSOLA ۲۴
۲-۴-۳- روش ESNOLA ۲۴
۲-۵- تبدیل متن به گفتار در زبان فارسی ۲۴
۲-۶- پارامترهای ویولت در سنتز صوت ۲۶

فصل سوم: مواد و روش‌ها ۳۱
۳-۱- مقدمه ۳۲
۳-۲- پردازش گفتار ۳۲
۳-۲-۱- نحوه تولید صدا ۳۳
۳-۲-۲- فریم‌های زمانی ۳۷
۳-۲-۳- ضرائب LPC ۴۰
۳-۲-۴- مدل سازی صوت ۴۱
۳-۳- سنتز صوت ۴۵
۳-۳-۱- پارامترهای سنتز صوت ۴۶
۳-۳-۲- سنتز کننده فورمانت ۴۷
۳-۳-۲-۱- تولید شکل موج از روی مقادیر فورمانت ۴۸
۳-۳-۲-۲- تولید فورمانت با استفاده از قواعد ۵۰
۳-۳-۳- سنتز صوت به روش به هم چسباندن قطعات ۵۱
۳-۳-۳-۱- انتخاب بخش‌ها ۵۲
۳-۳-۳-۱-۱- واج ۵۳
۳-۳-۳-۱-۲- دوواج ۵۴

۳-۳-۳-۱-۳- سه‌واج ۵۵
۳-۳-۳-۱-۴- کلمات ۵۵
۳-۳-۳-۲- اصلاح نوای گفتار ۵۶
۳-۳-۳-۲-۱- SOLA ۵۶
۳-۳-۳-۲-۲- PSOLA ۵۷
۳-۳-۳-۳- مشکلات PSOLA ۵۹
۳-۴- تبدیل ویولت ۶۰

فصل چهارم: نتایج و شبیه سازی‌ها (سنتز به وسیله واج‌ها) ۶۶
۴-۱- مقدمه ۶۷
۴-۲- تبدیل متن به واج ۶۸
۴-۲-۱- شکستن جمله به کلمه ۷۰
۴-۲-۲- دیکشنری ۷۰
۴-۲-۲-۱- ایجاد تلفظ ۷۱
۴-۲-۲-۲- بلوک تشخیص کسره اضافی ۷۳
۴-۳- بلوک ایجاد نوای گفتار ۷۴
۴-۳-۱- نحوه ایجاد تکیه در کلمه ۷۶
۴-۴- بلوک سنتز صوت ۷۷
۴-۴-۱- تولید بانک اطلاعاتی ۷۸
۴-۴-۲- بلوک دیاگرام سنتز صوت ۷۸
۴-۴-۲-۱- آشکار کردن گام و استفاده از الگوریتم TD- PSOLA ۸۱
۴-۴-۲-۲- کاهش حجم بانک اطلاعاتی با استفاده از ضرائب LPC ۸۴

فصل پنجم: نتایج و شبیه‌سازی‌ها (سنتز به وسیله دوواج‌ها) ۸۶
۵-۱- مقدمه ۸۷
۵-۲- نحوه ایجاد بانک اطلاعاتی ۸۷
۵-۳- بلوک دیاگرام سنتز گفتار به وسیله دوواج ۸۹

۵-۳-۱- یافتن انرژی دوواج‌ها ۹۰
۵-۳-۲- اتصال دوواج‌ها به هم ۹۰
۵-۳-۳- اصلاح نهایی جهت افزودن نوای گفتار به سیگنال سنتز شده ۹۲
۵-۳-۳-۱- یافتن تبدیل ویولت سیگنال سنتز شده ۹۳
۵-۳-۳-۲- یافتن انرژی ضرائب ویولت در سطح پنجم ۹۴
۵-۴- مقایسه بین سنتز به وسیله دوواج‌ها و واج‌ها ۹۵
۵-۵- نتیجه‌گیری ۹۷
۵-۶- پیشنهادات ۹۸
منابع ۱۰۰

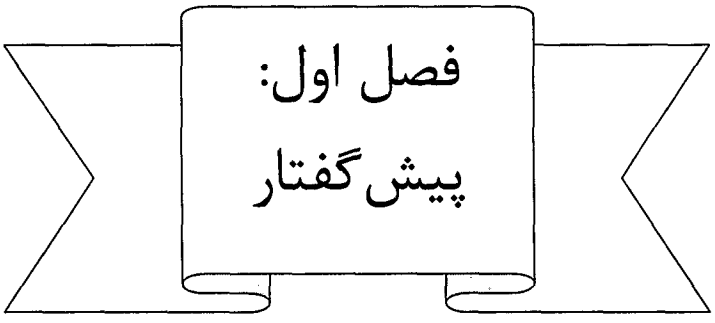
فهرست جدول ها

جدول ۲-۱ : جدول مربوط به عملکرد بخش سنتز کننده متن به گفتار با استفاده از هجاها..... ۲۶
جدول ۳-۱ : جدول مربوط به بخش‌های به کار برده شده در سنتز زنجیره‌ای..... ۵۲
جدول ۴-۱ : جدول مربوط به دیکشنری و چند کلمه نمونه..... ۷۱
جدول ۴-۲ : جدول مربوط به گروه بندی اصوات زبان فارسی..... ۷۲
جدول ۴-۳ : جدول مربوط به دو نمونه از الگوهای موجود و اعمال شده در کلمات..... ۷۵

فهرست شکل‌ها

شکل ۱-۱ : بلوک دیاگرام اولیه تبدیل متن به گفتار ۲
شکل ۱-۲ : دایره مکالمه صوتی(۱)..... ۳
شکل ۱-۳ : دایره مکالمه صوتی(۲)..... ۴
شکل ۱-۴ : بلوک دیاگرام سیستم مترجم..... ۶
شکل ۱-۵ : گذر تاریخی تبدیل متن به گفتار..... ۶
شکل ۲-۱ : بلوک دیاگرام کامل تبدیل متن به گفتار..... ۱۱
شکل ۲-۲ : بلوک دیاگرام تبدیل متن به گفتار با استفاده از شبکه‌های عصبی ۱۶
شکل ۲-۳ : بلوک دیاگرام تبدیل متن به گفتار با استفاده از تبدیل ویولت ۱۹
شکل ۲-۴ : بلوک دیاگرام تبدیل متن به واج ۲۵
شکل ۲-۵ : بلوک دیاگرام مربوط به بخش سنتز کننده ۲۵
شکل ۲-۶ : بلوک دیاگرام کانالِ مدلِ تأثیر صوت ۲۷
شکل ۲-۷ : مدل STV برای اداهای صدادار ۳۰
شکل ۲-۸ : کانال coarticulation حنجره ۳۰
شکل ۳-۱ : آناتومی انسان و نحوه تولید صدا ۳۵
شکل ۳-۲ : نمایش حروف صدا دار و بی صدا در زبان فارسی ۳۶
شکل ۳-۳ : نمایش فورمانت و هارمونیک‌های تشکیل دهنده یک حرف صدادار..... ۳۷
شکل ۳-۴ : نمایش حوزه زمان و حوزه فرکانس پنجره‌های همینگ و مستطیلی..... ۳۹
شکل ۳-۵ : مدل تولید ضرائب LPC با درجه P ۴۰
شکل ۳-۶ : نمایش یک سیگنال و تخمین سیگنال با استفاده از ضرائب LPC به همراه خطای تخمین..... ۴۱
شکل ۳-۷ : مدل ریاضی ارائه شده توسط Fant برای تولید صدا در انسان..... ۴۲
شکل ۳-۸ : مدل چند لوله‌ای برای Vocal tract..... ۴۳
شکل ۳-۹ : مدل تولید انواع صداها اعم از پریودیکی، نویزی و ضربه‌ای..... ۴۴
شکل ۳-۱۰ : نمودار کیفیت بدست آمده در مقابل درصد جملات تولید شده به وسیله هر روش..... ۴۶
شکل ۳-۱۱ : بلوک دیاگرام سنتز صوت به وسیله روش قاعده مدار در سنتز به وسیله فورمانت..... ۴۸
شکل ۳-۱۲ : مدل سنتز کننده klatt متشکل از مدل‌های سری و موازی..... ۴۹
شکل ۳-۱۳ : نمایش حالت گذر بین دو حرف صدادار در سنتز کننده فورمانت..... ۵۱
شکل ۳-۱۴ : سنتز کلمه "ترازو"به وسیله تک واج‌ها..... ۵۴
شکل ۳-۱۵ : سنتز کلمه "دامداری"به وسیله دو واج‌ها..... ۵۵
شکل ۳-۱۶ : نمایش نحوه استخراج دوواج‌ها از یک کلمه نمونه..... ۵۵
شکل ۳-۱۷ : نحوه استفاده از الگوریتم SOLA و اشکال این الگوریتم..... ۵۷
شکل ۳-۱۸: نمایش نحوه تغییر فرکانس و دوره گام..... ۵۸
شکل ۳-۱۹ : آنالیز ویولت با استفاده از الگوریتم Multiresolution..... ۶۳
شکل ۳-۲۰ : سنتز ویولت با استفاده از الگوریتم Multiresolution..... ۶۳
شکل ۳-۲۱ : نمایش درخت تشکیل دهنده ویولت بسته‌ای به همراه ضرائب موجود در یک گره خاص..... ۶۴
شکل ۳-۲۲ : نمایش یک سیگنال صدادار به همراه ضرائب تبدیل ویولت در سطح پنجم..... ۶۵
شکل ۴-۱ : بلوک دیاگرام اولیه تبدیل متن به گفتار در زبان فارسی..... ۶۷

شکل ۴-۲ : بلوک دیاگرام کامل تبدیل متن به گفتار با استفاده از تک‌واج‌ها..... ۶۸
شکل ۴-۳ : بلوک دیاگرام تبدیل جمله فارسی به رشته واج متناظر..... ۶۹
شکل ۴-۴ : نمایش نحوه کد کردن اصوات فارسی برای کدهای ۵ رقمی ۷۲
شکل ۴-۵ : بلوک دیاگرام تبدیل جمله فارسی به رشته واج متناظر، با در نظر گرفتن کسره ربط دهنده..... ۷۳
شکل ۴-۶ : نمایش الگوی نمونه برای تولید نوای گفتار به همراه اعمال این الگو به یک حرف صدادار..... ۷۶
شکل ۴-۷ : بلوک دیاگرام اولیه سنتز صوت..... ۷۸
شکل ۴-۸ : نمایش نحوه فراخوانی سیگنال‌های ضبط شده برای سنتز یک کلمه خاص..... ۷۹
شکل ۴-۹ : بلوک دیاگرام ثانویه سنتز صوت جهت بالا بردن کیفیت..... ۸۰
شکل ۴-۱۰ : نمایش پنجره های متناظر مربوط به حروف تشکیل دهنده یک کلمه نمونه..... ۸۰
شکل ۴-۱۱ : بلوک دیاگرام آشکار کردن یک سیکل از حرف صداردار با استفاده از الگوریتم AMDF..... ۸۱
شکل ۴-۱۲ : شکل حرف صدادار شناسائی شده به همراه تابع خود همبستگی ۸۱
شکل ۴-۱۳ : شکل یک سیکل هر حرف صدادار آشکار شده..... ۸۲
شکل ۴-۱۴ : بلوک دیاگرام الگوریتم TD-PSOAL..... ۸۳
شکل ۴-۱۵ : سیگنال بدست آمده از الگوریتم TD-PSOLA..... ۸۳
شکل ۴-۱۶ : نمایش یک سیگنال نمونه به همراه تخمین سیگنال به وسیله ضرائب LPC..... ۸۴
شکل ۵-۱ : نمایش یک کلمه در زبان فارسی و استخراج دوواج‌ها..... ۸۷
شکل ۵-۲ : نحوه کاهش بانک اطلاعاتی در دوواج‌ها و نمایش تفاوت دو سیگنال بدست آمده..... ۸۸
شکل ۵-۳ : بلوک دیاگرام تبدیل کلمه به گفتار به وسیله دوواج‌ها..... ۸۹
شکل ۵-۴ : نمایش یک دوواج نمونه به همراه انرژی بدست آمده از این دوواج..... ۹۰
شکل ۵-۵ : بلوک دیاگرام شناسائی فریم‌های صدادار از فریم‌های بی صدا و جدا کردن بخش صدادار از بخش بی صدا..... ۹۰
شکل ۵-۶ : سنتز یک کلمه به وسیله دوواج‌ها و نمایش اغتشاش موجود در محل اتصال..... ۹۱
شکل ۵-۷ : نمایش یک کلمه سنتز شده به وسیله دوواج‌ها بدون وجود اغتشاش..... ۹۲
شکل ۵-۸ : نمایش بلوک دیاگرام اصلاح نهایی برای اضافه کردن نوای گفتار در حوزه ویولت..... ۹۳
شکل ۵-۹ : نمایش سیگنال سنتز شده به همراه ضرائب ویولت..... ۹۳
شکل ۵-۱۰ : نمایش ضرائب ویولت پنجره شده به همراه انرژی موجود در هر فریم..... ۹۴
شکل ۵-۱۱ : نمایش سیگنال سنتز شده بعد از حذف اغتشاش به وسیله تبدیل ویولت ۹۵


فصل اول:
پیش‌گفتار

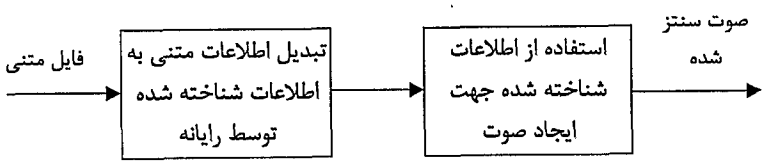
۱-۱- مقدمه

تبدیل متن به گفتار[۱] اطلاعات مربوط به زبان شناسی ذخیره شده به صورت متن را به حالت صوت تبدیل می‌کند. این تبدیل به صورت بسیار گسترده در وسائل گویا برای افراد نابینا استفاده می‌گردد. از این تبدیل برای خواندن پست‌های الکترونیکی، خواندن انواع متن‌ها اعم از اسکن شده، تایپ شده استفاده می‌شود. در این بخش هدف این است که به تعدادی از کاربردهای مهم اشاره گردد تا اهمیت این تبدیل روشن شود.

۱-۲- تبدیل متن به گفتار

هر تبدیل متن به گفتار از دو بخش عمده تشکیل شده است: یکی تبدیل اطلاعات مربوط به یک فایل متنی به اطلاعات شناخته شده توسط رایانه، دومی استفاده از این اطلاعات جهت ایجاد صوت می‌باشد که هر کدام از این بخش‌ها از اجزاء کوچکتری تشکیل شده است که به صورت مفصل مورد بررسی قرار خواهد گرفت.

برای انجام این تبدیل بلوک دیاگرام کلی شکل (۱-۱) را در نظر می‌گیریم.



شکل ۱-۱: بلوک دیاگرام اولیه تبدیل متن به گفتار

در بخش‌های بعدی به تبدیل متن به گفتار به صورت کامل خواهیم پرداخت و به بلوک‌های تشکیل دهنده هر بخش به صورت کامل اشاره خواهیم کرد.

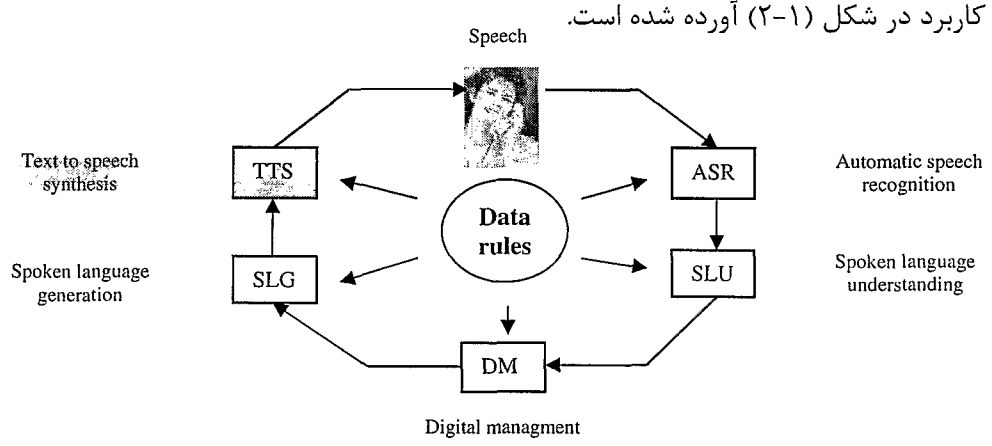
۱-۲-۱- مخابرات بین انسان و ماشین

یکی از کاربردهای مهم تبدیل متن به گفتار، دایره مکالمه صوتی می‌باشد.

---

[۱] Text To Speech

این کاربرد در شکل (۱-۲) آورده شده است.



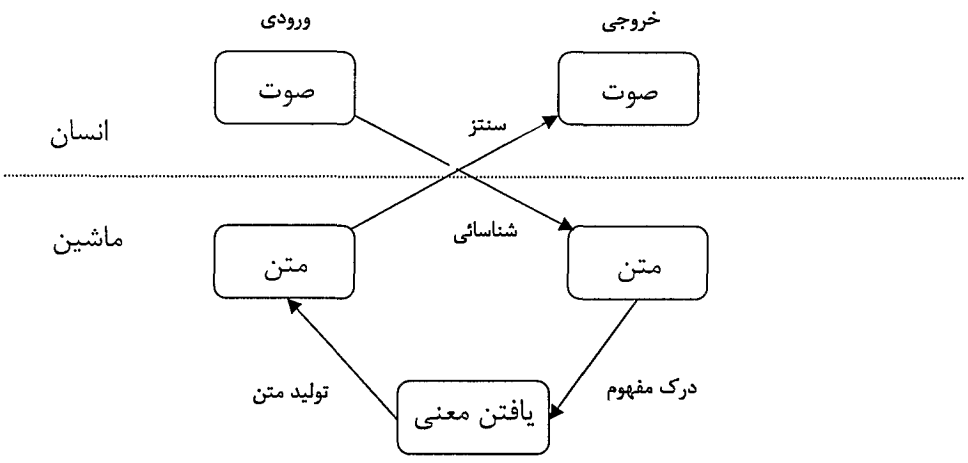
شکل۱-۲ : دایره مکالمه صوتی

از این بلوک دیاگرام بیشتر برای کیوسک‌های اطلاعاتی و همچنین به منظور کاربردهای مربوط به مکالمات دو طرفه بین انسان و ماشین استفاده می‌گردد. مثلاً یک صحبتی در پشت تلفن انجام می‌گیرد. در ابتدا باید این صحبت توسط رایانه شناسائی گردد. این کار توسط بلوک ASR[1] که بلوک شناسائی صوت است صورت می‌پذیرد. این بلوک بر مبنای مدل‌های آماری و مدل مخفی مارکوف[2] عمل می‌کند. بعد از این بخش اطلاعات شناسائی شده باید درک گردد. این کار توسط بلوک SLU[3] که کار آن درک صحبت می‌باشد انجام می‌شود. منظور این است که مفهوم صحبت بیان شده چیست و تصمیم گیری بر مبنای مفهوم صحبت انجام گرفته صورت می‌پذیرد. تا اینجا صحبت شخص پشت کیوسک توسط بلوک‌های مختلف درک شده است. حالا باید برای این صحبت بیان شده یک جواب مناسب اتخاذ گردد. لذا باید یک مدیریتی بین بلوک‌ها انجام گیرد. این کار توسط بلوک DM[4] انجام می‌گیرد. در بخش بعد با توجه به صحبت درک شده باید یک جواب مناسب برای مخاطب انتخاب شود. این کار توسط بلوک SLG [5] و با توجه به الگوهای ذخیره شده انجام می‌شود. مثلاً یک سری اطلاعات به صورت خام ذخیره شده است که با توجه به صحبت درک شده یکی از این الگوها انتخاب می‌شود. این اطلاعات مانند اطلاعات خروجی بلوک SLU نیست. این اطلاعات را می‌توانیم مانند یک

[1] Automatic Speech Recognition
[2] Hidden Markov Modeling
[3] Spoken Language Understanding
[4] Digital Management
[5] Spoken Language Generation

متن در نظر بگیریم. در آخر نیز این اطلاعات باید به صورت صوت تبدیل گردد. این کار توسط بلوک TTS[1] انجام می‌شود. این بلوک باید قادر باشد که یک صدای مفهوم با کیفیت بالا تولید کند که شنونده قادر به درک آن باشد. علاوه بر این، گفتار تولید شده حالت طبیعی داشته باشد، یعنی جمله از نظر زیر و بمی صدا مانند سؤالی بودن، امری بودن، خبری بودن قابل شناسائی باشد.

این نوع مخابرات را می‌توانیم به شکل ساده‌تر نیز در نظر بگیریم که به صورت ساده از دو بخش سنتز و آنالیز تشکیل شده است که برای درک بهتر شکل (۱-۳) را در نظر می‌گیریم.



شکل ۱-۳ : دایره مکالمه صوتی

همانطور که از روی شکل (۱-۳) نیز مشخص است در ابتدا یک صحبتی توسط شخصی بیان می‌گردد. سپس این صحبت به متن تبدیل می‌شود، یا به عبارت بهتر به پارامترهای قابل شناسائی توسط رایانه تبدیل می‌گردد. در مرحله بعد به وسیله این پارامترها مفهوم صحبت بیان شده درک می‌گردد و با توجه به صحبت درک شده، متن مورد نظر از بین الگوهای موجود یافت می‌شود و این متن بعد از سنتز به صوت تبدیل می‌گردد. از این نوع مخابرات می‌توان به عنوان مخابرات بین انسان و ماشین یاد کرد.

[1] Text To Speech

۱-۲-۲- سیستم‌های خوانا[1]

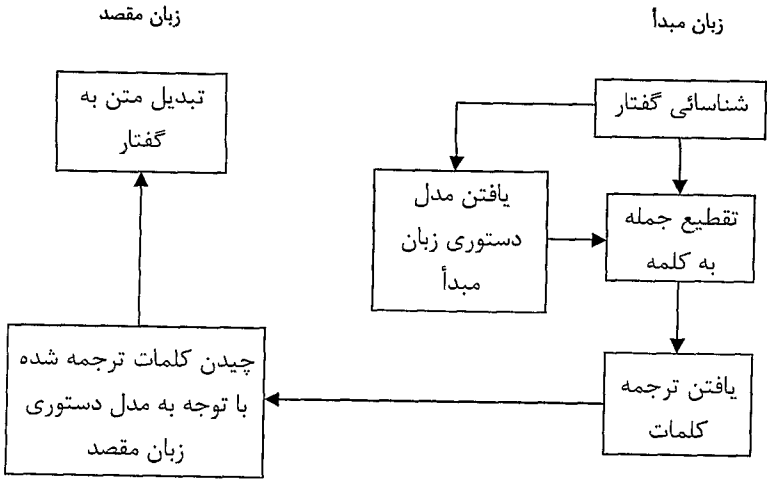
در بسیاری از اداره‌ها از این سیستم‌ها برای خواندن متن‌ها و کارهایی که باید در طی یک روز انجام گیرد استفاده می‌شود. به این صورت که یک سری از کارهای روزمره که باید انجام گیرد توسط یک گوینده ضبط می‌شود. این اطلاعات ضبط شده در زمان‌های مشخص و برنامه ریزی شده خوانده می‌شود. این سیستم‌ها دارای کیفیت خوبی بوده و قدرت درک[2] بالا در این سیستم‌ها وجود دارد. دلیل این موضوع این است که قطعات ضبط شده بزرگ بوده و ناپیوستگی بین قطعات کم بوده که این مسئله سبب افزایش کیفیت صوت سنتز شده خواهد شد.

یکی دیگر از کاربردهای سیستم‌های خوانا در ساعت گویا است، به این صورت که وقتی شماره مورد نظر گرفته می‌شود، در لحظه برقراری ارتباط با سیستم گویا، زمان ثبت می‌شود. با توجه به زمان برقراری ارتباط، اعدادِ از قبل ضبط شده در کنار هم قرار می‌گیرند و ساعتِ مورد نظر به صورت صوت بیان می‌شود.

۱-۲-۳- سیستم‌های مترجم

از دیگر کاربردهای تبدیل متن به گفتار می‌توان به سیستم‌های مترجم اشاره کرد که در ترجمهٔ گفتار بیان شده از یک زبان به زبان دیگر استفاده می‌شود. به این صورت که در ابتدا صحبتی بیان می‌شود، این صحبت شناسائی می‌گردد، جمله به کلمات تقسیم بندی می‌شوند، به طوریکه باید جایگاه دستوری کلمه نیز مشخص شود تا ترجمه به خوبی انجام گیرد، سپس معنی هر کلمه از بانک اطلاعاتی یافت می‌شود و با توجه به مدل زبان و همچنین مدل دستوری زبان مقصد، کلمات جدید در کنار هم قرار می‌گیرند و جمله جدید ایجاد شده توسط بلوک تبدیل متن به گفتار خوانده می‌شود. برای توضیح بیشتر شکل (۱-۴) را در نظر می‌گیریم.

[1] Read aload systems
[2] Inteliigibility



شکل۱-۴ : بلوک دیاگرام سیستم مترجم

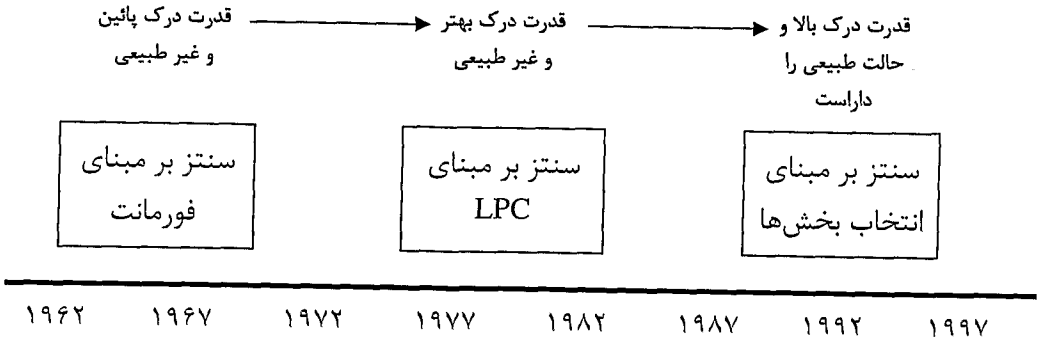
۱-۲-۴- کاربردهای دیگر

از دیگر کاربردها می‌توان به سیستم‌های تبدیل متن به گفتار که بر روی سیستم‌های عامل مانند ویندوز XP و VISTA وجود دارند، اشاره کرد. این کاربرد جهت استفاده افراد نابینا از این سیستم عامل‌ها می‌باشد.

از دیگر کاربردها می‌توان به استفاده از این تبدیل در ربات‌های سخن گو اشاره کرد، این کاربرد نیز مانند کیوسک‌های اطلاعاتی است، فقط شکل موضوع تغییر پیدا کرده است.

۱-۳- تاریخچه

در این بخش بیشتر هدف در این است که به گذر تاریخی تبدیل متن به گفتار اشاره کنیم. برای درک بهتر شکل (۱-۵) را در نظر بگیرید.



شکل۱-۵ : گذر تاریخی تبدیل متن به گفتار

مسئله تبدیل متن به گفتار به سال ۱۹۶۰ بر می‌گردد. در این سال استفاده از روش‌های سنتز قاعده مدار[1] که بر مبنای مدل‌های ریاضی استوار بوده انجام می‌گرفت. در آن زمان برای سنتز صوت، باید از روش‌ها و مدل‌های پیچیده ریاضی استفاده می‌شد. در این نوع سنتز علاوه برکیفیت پائین، اصوات سنتز شده، از نظر طبیعی بودن در سطح بسیار پائینی قرار داشتند. اما با پیشرفت روش‌های آنالیز صوت مانند روش LPC[2]، سنتز صوت نیز بر مبنای همین روش انجام می‌گرفت. استفاده از این روش موجب افزایش کیفیت صوت سنتز شده می‌شد، اما همچنان از نقطه نظرِ طبیعی به نظر رسیدن صوت، در سطح پائینی قرار داشت. اما از سال ۱۹۸۷ به بعد با ابداع روش‌های داده مدار[3] کیفیت صوت سنتز شده به شدت افزایش پیدا کرده و همچنین از نقطه نظرِ طبیعی به نظر رسیدن صوت، کیفیت سیگنال سنتز شده افزایش چشمگیری داشته است. در روش‌های داده مدار دیگر از روابط پیچیده ریاضی خبری نیست و سنتز بر مبنای چیدن قطعات از قبل ضبط شده صورت می‌پذیرد. در این نوع سنتز، چون قطعات صوت از قبل به وسیله گوینده‌هایی ضبط شده است، انتظار سنتز سیگنال با کیفیت خوب دور از ذهن نیست. در حال حاضر محققان به دنبال این مطلب هستند که به صوت سنتز شده احساس را نیز اضافه کنند به این مفهوم که از روی صوت سنتز شده ما متوجه شویم که گوینده چه احساسی دارد مثلاً اینکه یک ربات سخن‌گو از گفته ما ناراحت شده است یا اینکه عصبانی شده است.

۱-۴- نگاهی گذرا به پایان نامه

در این پایان‌نامه ما به دنبال این مطلب هستیم که صوت سنتز شده دارای کیفیت مطلوب باشد، برای بررسی این موضوع از روش تستِ سیگنال سنتز شده میان شنوندگان مختلف استفاده شده است.

به دلیل اینکه تبدیل متن به گفتار یک روند طولانی دارد، تکیه ما بیشتر بر بخش سنتز است،

[1] Rule-based synthesis
[2] Linear prediction coding
[3] Data driven-based