

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه آزاد اسلامی  
واحد تهران مرکز

**موضوع:**

**پردازش زبان فارسی**

## فهرست مطالب

فهرست شکل ها.....	۵
فهرست جداول.....	۵
۱ بازشناسی گفتار گسسته و پیوسته فارسی.....	۶
۱-۱ مقدمه.....	۶
۲-۱ بازشناسی گفتار گسسته فارسی.....	۶
۱-۲-۱ دایره کلمات و دادگان.....	۶
۲-۲-۱ سیستم‌های مورد استفاده.....	۷
۱-۲-۲-۱ مدل‌های مارکوف پنهان با چگالی مشاهدات پیوسته (CDHMMs).....	۷
۱-۲-۲-۱ توپولوژی مدل‌ها.....	۸
۲-۲-۱ استخراج ویژگی‌های طیفی.....	۸
۱-۲-۲-۱ آموزش مدل‌ها.....	۹
۱-۲-۲-۱ بازشناسی کلمات جدا.....	۱۰
۲-۲-۱ شبکه‌های عصبی با تأخیر زمانی.....	۱۱
۱-۲-۲-۱ دادگان به‌کار رفته.....	۱۱
۲-۲-۱ پیاده‌سازی.....	۱۱
۳-۲-۱ مشکلات در بازشناسی و اصلاحات مربوطه.....	۱۲
۱-۳-۲-۱ سرعت بازشناسی.....	۱۲
۲-۳-۱ مقابله با اثرات شرایط مختلف محیطی.....	۱۴
۴-۲-۱ واسط کاربر.....	۱۵
۳-۱ بازشناسی گفتار پیوسته فارسی.....	۱۷
۱-۳-۱ دادگان و دایره لغات.....	۱۷
۱-۳-۱ فارس دات.....	۱۷
۲-۳-۱ دادگان اخبار.....	۱۸
۲-۳-۱ سیستم‌های به‌کار گرفته شده.....	۱۹
۱-۳-۲-۱ مدل‌های مارکوف پنهان با چگالی مشاهدات پیوسته (CDHMMs).....	۱۹
۱-۳-۲-۱ استخراج ویژگی‌ها.....	۲۰
۲-۳-۱ مدل‌های مورد استفاده.....	۲۰
۳-۳-۱ آموزش سیستم.....	۲۰
۴-۳-۱ بازشناسی (دکودینگ) گفتار پیوسته.....	۲۲
۵-۳-۱ استفاده از چگالی مشاهدات مخلوط.....	۲۳
۶-۳-۱ هم‌ردیف‌سازی زمانی [۳][۱۴].....	۲۳
۷-۳-۱ مدل‌سازی وابسته به متن.....	۲۴

۲۵	..... ۱-۳-۲-۱ ۸ مدل‌سازی هجا
۲۵	..... ۱-۳-۲-۲ مدل‌های ترکیبی HMM و شبکه عصبی
۲۷	..... ۴ پیاده‌سازی‌ها و بررسی نتایج
۳۰	..... ۲ سنتز گفتار فارسی
۳۰	..... ۱-۲ مقدمه
۳۱	..... ۲-۲ نرم افزار نهایی
۳۱	..... ۲-۲-۱ لایه کاربری نرم افزار
۳۱	..... ۲-۲-۲ لایه ارتباطی ماژولهای نرم افزار
۳۳	..... ۳-۲ ماژول تبدیل متن به واج (TTP)
۳۳	..... ۱-۳-۲ مشکلات موجود در TTP فارسی
۳۳	..... ۲-۱-۳-۲ عدم فاصله گذاری صحیح بین کلمات
۳۳	..... ۲-۱-۳-۲ نوشته نشدن حرکات
۳۳	..... ۲-۱-۳-۲ تشخیص کسره اضافه
۳۴	..... ۲-۳-۲ بلوکهای ماژول TTP
۳۴	..... ۲-۳-۲-۱ بلوک تقطیع جمله به واژه‌ها
۳۵	..... ۲-۳-۲-۲ بلوک استخراج رشته واج معادل واژه
۳۵	..... ۲-۳-۲-۱-۲ استخراج رشته واج متناظر واژه به کمک واژه‌نامه
۳۵	..... ۲-۳-۲-۲-۲ تحلیل واژه‌های مشتق
۳۶	..... ۲-۳-۲-۲-۳ کلمات ناشناس
۳۶	..... ۲-۳-۲-۲-۳ بلوک تشخیص کسره اضافه
۳۶	..... ۲-۳-۳-۲ محصولات جنبی
۳۶	..... ۲-۳-۳-۱ واژه‌نامه فونتیک
۳۶	..... ۲-۳-۳-۲ بانک جملات زبان فارسی
۳۷	..... ۴-۲ ماژول نوای گفتار
۳۸	..... ۲-۴-۱ میزان تأثیر نوای گفتار در زبان
۳۸	..... ۲-۴-۲ تحقیقات انجام شده تا کنون
۳۹	..... ۲-۴-۳ روشهای مدلسازی نوای گفتار
۳۹	..... ۲-۴-۳-۱ روشهای قاعده‌مدار
۴۰	..... ۲-۴-۳-۲ روشهای داده‌مدار
۴۰	..... ۲-۴-۳-۳ روشهای تلفیقی
۴۱	..... ۲-۴-۴ پیاده‌سازی نوای گفتار در گروه سنتز
۴۱	..... ۲-۴-۴-۱ پیاده‌سازی تکیه
۴۳	..... ۲-۴-۴-۲ پیاده‌سازی آهنگ
۴۴	..... ۲-۴-۴-۳ یک نمونه عملی از بکارگیری نوای گفتار
۴۴	..... ۲-۵ ماژول گفتارساز MBE

۴۶	۱-۵-۲ کلیات روش بهم چسباندن واحدهای گفتار .....
۴۶	۱-۱-۵-۲ انتخاب نوع واحدهای ذخیره شده .....
۴۶	۲-۱-۵-۲ انتخاب روش پردازش سیگنال .....
۴۶	۱-۲-۱-۵-۲ روش LPC .....
۴۷	۲-۲-۱-۵-۲ روش PSOLA .....
۴۷	۱-۲-۲-۱-۵-۲ روش TD-PSOLA .....
۴۷	۲-۲-۲-۱-۵-۲ روش LP-PSOLA .....
۴۷	۳-۲-۲-۱-۵-۲ روش FD-PSOLA .....
۴۸	۲-۵-۲ گفتار ساز فارسی با روش MBR-PSOLA .....
۵۰	۱-۲-۵-۲ روش تغییر گام صحبت در گفتار ساز MBR-PSOLA .....
۵۱	۲-۲-۵-۲ روش تغییر طول و درونیایی بین سیلابها .....
۵۱	۳-۲-۵-۲ روش تغییر انرژی .....
۵۱	۴-۲-۵-۲ اعمال قواعد ثابت آهنگین کردن گفتار و تغییر ساختار سیلاب در کلمات .....
۵۲	۵-۲-۵-۲ تغییر و اصلاح الگوریتم سنتز جملات و سیلابها .....
۵۲	۶-۲-۵-۲ اصلاح و بهینه سازی الگوریتم آنالیز دوفونیهای ذخیره شده .....
۵۳	۷-۲-۵-۲ اصلاح و بهینه سازی الگوریتم سنتز سیلابها و درونیایی خطی بین آنها .....
۵۴	۸-۲-۵-۲ تست واحدهای ذخیره شده و پیشنهاد اصلاح و در صورت لزوم افزایش واحدها .....
۵۴	۹-۲-۵-۲ ادغام برنامه های آنالیز دو واجهای CV و VC و بهبود آنها .....
۵۵	۱۰-۲-۵-۲ برنامه ادغام فایل های پارامترها .....
۵۵	۳-۵-۲ نتیجه گیری .....
۵۶	۳ تصدیق هویت گوینده .....
۵۶	۱-۳ مقدمه و هدف .....
۵۷	۲-۳ مرور منابع علمی .....
۶۱	۳-۳ خصوصیات خط تلفن و مکالمات تلفنی .....
۶۲	۴-۳ تشخیص گفتار از سکوت .....
۶۲	۵-۳ استخراج ویژگی .....
۶۳	۶-۳ مدل نمودن ارقام و گویندگان .....
۶۳	۷-۳ ارزیابی روشهای بازشناسی ارقام و تصدیق هویت گوینده .....
۶۷	۸-۳ تشخیص و تصحیح خطای بازشناسی ارقام کد شناسائی شخصی .....
۶۷	۹-۳ دادگان FARSDIGITS1 .....
۶۸	۱۰-۳ بازشناسی کد شناسائی .....
۶۸	۱-۱۰-۳ بازشناسی ارقام کد شناسائی شخصی بصورت مجزا توسط شبکه عصبی پیشگو .....
۶۸	۱-۱-۱۰-۳ مدل شبکه عصبی پیشگو .....
۶۹	۲-۱-۱۰-۳ الگوریتم بازشناسی کلمات .....

- ۳-۱۰-۱-۳ الگوریتم آموزشی مدل کلمات ..... ۶۹
- ۳-۱۰-۱-۴ استخراج ویژگی و آموزش مدل های ارقام ..... ۶۹
- ۳-۱۰-۱-۵ نتایج آزمایشات ..... ۷۰
- ۳-۱۰-۲-۱ شناسایی کد شناسائی شخصی بصورت ارقام مجزا توسط مدل مخفی مارکوف ... ۷۲
- ۳-۱۰-۲-۱-۱ استخراج ویژگی ..... ۷۲
- ۳-۱۰-۲-۲ آموزش مدل های پنهان مارکف ..... ۷۲
- ۳-۱۰-۲-۳ نتایج آزمایشات ..... ۷۳
- ۳-۱۰-۳-۱ شناسایی کد شناسائی شخصی بصورت ارقام متصل توسط مدل مخفی مارکوف. ۷۴
- ۳-۱۰-۳-۱-۱ پیش پردازش و استخراج ویژگی ..... ۷۵
- ۳-۱۰-۳-۲ آموزش مدل های پنهان مارکوف برای ارقام متصل ..... ۷۵
- ۳-۱۰-۳-۳ آزمایشات و تحلیل نتایج ..... ۷۵
- ۳-۱۱-۱۱-۱ تصدیق هویت گوینده ..... ۷۶
- ۳-۱۱-۱۱-۱-۱ تصدیق هویت توسط تلفیق شبکه عصبی درخت برآمدگی و الگوریتم ژنتیکی ..... ۷۷
- ۳-۱۱-۱۱-۱-۱-۱ شبکه درختی برآمدگی ..... ۷۷
- ۳-۱۱-۱۱-۲ تصدیق هویت گوینده توسط تلفیق درخت برآمدگی و الگوریتم ژنتیکی ..... ۷۹
- ۳-۱۱-۲-۱ تصدیق هویت گوینده توسط سیستم هیبرید متشکل از مدل پنهان مارکف و مدل مخلوط گاوسی ..... ۸۱
- ۳-۱۱-۲-۱-۱ پیش پردازش و استخراج ویژگی ..... ۸۲
- ۳-۱۱-۲-۲ آموزش مدل های گویندگان ..... ۸۲
- ۳-۱۱-۲-۳ نحوه ساختن سیستم هیبرید ..... ۸۳
- ۳-۱۱-۲-۴ معیار تصویر وزن دهی شده ..... ۸۳
- ۳-۱۱-۲-۵ آزمایشات ..... ۸۴
- ۳-۱۱-۳ مقایسه چند روش نرمالیزاسیون امتیازات در سطح گویش و در سطح فریم برای افزایش کارایی تصدیق هویت گوینده بر روی خط تلفن ..... ۸۶
- ۳-۱۱-۳-۱ روش های نرمالیزاسیون امتیازات [۶۴] ..... ۸۶
- ۳-۱۱-۳-۲ روش های نرمالیزاسیون امتیازات در سطح فریم [۶۷] ..... ۹۰
- ۳-۱۱-۳-۳ وزن دهی امتیازات مدل [۶۷] ..... ۹۰
- ۳-۱۱-۳-۴ استخراج ویژگی ..... ۹۱
- ۳-۱۱-۳-۵ آموزش مدل های مخلوط گاوسی ..... ۹۱
- ۳-۱۱-۳-۶ آزمایشات ..... ۹۲
- ۳-۱۲ نتیجه گیری ..... ۹۳
- مراجع ..... ۹۶
- پیوست الف - جداول نشانه های بخش سنتز گفتار ..... ۱۰۱
- پیوست ب - مقالات ..... ۱۰۴

## فهرست شکل ها

- شکل ۱-۱ پنجره اصلی محیط گرافیکی..... ۱۵
- شکل ۱-۲ توپولوژی استفاده شده برای مدل سازی واج ها (مدل چپ - راست Bakis)..... ۲۰
- شکل ۲-۱ شمای کلی پروژه گفتار ساز فارسی..... ۳۰
- شکل ۲-۲ پنجره رابط کاربر نرم افزار..... ۳۱
- شکل ۲-۳ شمای کلی ماژول تبدیل متن به واج..... ۳۴
- شکل ۲-۴ تفکیک مراحل اعمال نوای گفتار..... ۴۵
- شکل ۲-۵ روش بدست آوردن فرمول برونیابی..... ۵۳
- شکل ۳-۱ مقادیر آستانه تصمیم گیری EER بر حسب میانگین منهای واریانس فواصل برون  
گوینده ای نظیر آنها..... ۶۵
- شکل ۳-۲ مدل پنهان مارکوف برای ارقام متصل..... ۷۵

## فهرست جداول

- جدول ۱-۳ بعضی از کارهای انجام شده در زمینه تصدیق هویت گوینده بصورت وابسته به متن  
..... ۵۹
- جدول ۲-۳ بعضی از کارهای انجام شده در زمینه تصدیق هویت گوینده بصورت مستقل از متن  
..... ۵۹
- جدول ۳-۳ صحت بازشناسی ارقام متصل..... ۷۶
- جدول نشانه های قراردادی واج نویسی..... ۱۰۱
- جدول نشانه های انواع صرفی..... ۱۰۲
- جدول نشانه های نقشهای نحوی..... ۱۰۳

## ۱ بازشناسی گفتار گسسته و پیوسته فارسی

### ۱-۱ مقدمه

هدف در بخش بازشناسی گفتار گسسته و پیوسته فارسی، دستیابی به سیستم‌های بازشناسی گفتار فارسی با توانایی بازشناسی گفتار گسسته در چارچوب دایره کلمات تعیین شده و نرخ بازشناسی قابل قبول، بر روی دادگان تعریف شده و نیز بازشناسی گفتار پیوسته فارسی با دایره کلمات مشخص و در چارچوب دادگان اصلاح شده پیوسته و با نرخ بازشناسی مورد نظر می‌باشد. در هر یک از این دو راستا، اقدامات صورت گرفته شامل مطالعه، بررسی و پیاده‌سازی روش‌های مطرح در بازشناسی جهت دستیابی به نتایج مطروحه بوده است. تهیه دادگان‌های گفتاری گسسته و پیوسته با مشخصات از پیش تعیین شده، پیاده‌سازی روش‌های مبتنی بر HMM‌های با چگالی مشاهدات پیوسته و نیز شبکه‌های عصبی، ارزیابی و بهبود روش‌های پیاده‌سازی شده و احیاناً اصلاح یا پیشنهاداتی جهت اصلاح این روش‌ها می‌باشد.

### ۱-۲ بازشناسی گفتار گسسته فارسی

بازشناسی گفتار گسسته از جمله بخش‌های مهم در بحث بازشناسی گفتار می‌باشد که در کاربردهای مختلف از جمله فرمان و کنترل موارد استفاده فراوانی دارد.

#### ۱-۲-۱ دایره کلمات و دادگان

دایره کلمات از جمله عوامل مهم در تأمین بازشناسی با کیفیت مناسب گفتار گسسته می‌باشد. هدف در این طرح پژوهشی دستیابی به بازشناسی در محدوده کلمات صد کلمه بوده است. با این همه، از آنجاکه پیش از آغاز این طرح، یک دادگان گفتاری از کلمات گسسته (ارقام فارسی) با تعداد ده کلمه و بصورت ناوابسته به گوینده موجود بوده است، مرحله ابتدائی کار بر روی این دادگان به انجام رسید. لازم به اشاره است که این دادگان از مجموعه‌ای از دانشجویان در محدوده سنی ۲۰ تا ۳۰ سال از هر دو جنسیت ضبط گردیده و مشتمل بر دو بخش آموزشی و آزمایشی بوده است. شرایط ضبط، محیط با نویز زمینه کم بوده است.



دادگان دیگری برای دستیابی به اهداف این طرح پژوهشی طراحی و ضبط گردید. در طراحی این دادگان تلاش شد کلمات مورد استفاده، ضمن پوشش یک محدوده صدکلمه‌ای، از ویژگی‌های خاصی برخوردار باشند. به همین جهت، مجموعه کلمات شامل ۳۸ کلمه اصلی تشکیل‌دهنده کلیه اعداد تا شش رقمی در زبان فارسی (صفر تا نوزده، مضارب ده، مضارب صد و مضارب هزار) به همراه اسامی شهرهای مهم ایران می‌باشد. در عین حال ضبط دادگان بصورت مستقیم به داخل کامپیوتر از طریق یک میکروفون دینامیک مناسب (Sony F-VK98) و در محیط اداری (دفتر کار) با نویز محیط معمولی صورت گرفت. در مجموع از ۱۰۲ گوینده (نیمی مرد و نیمی زن)، هر یک از ۱۰۰ کلمه تعریف شده، ۵ بار ضبط گردیدند. البته باتوجه به ضبط گفتار در محیط دانشگاهی و از افرادی که دارای مدرک تحصیلی حداقل دیپلم متوسطه بوده‌اند، دسترسی به افراد با سن بالای ۴۰ سال عملاً مقدور نبوده و لذا سن افراد شرکت‌کننده در تهیه این دادگان محدود و وسیعی را شامل نمی‌شود [۱][۲].

پس از جمع‌آوری دادگان، اقدام به بررسی تک تک حدود ۵۱/۰۰۰ بیان گفتاری موجود گردید و اشکالات مختلف موجود در آنها از قبیل صداهای اضافی ضبط شده، بریدگی کلمات در آغاز یا پایان، برخی لهجه‌ها و گویش‌های نامعمول، صدای دهان در آغاز ادای کلمات و ... که علی‌رغم پیش‌بینی‌های لازم در ضبط برخی کلمات وجود داشتند، مشخص و بیان‌های دارای مشکل حذف گردیدند [۴]. سپس دادگان به دو بخش آموزشی و آزمایشی تقسیم گردید. این دادگان به عنوان دادگان اصلی در این بخش از طرح پژوهشی پردازش گفتار فارسی مورد استفاده واقع گردید. در عین حال در بعضی مراحل، از جمله در بازشناسی به کمک شبکه‌های عصبی نیز یک مجموعه ۱۲ کلمه‌ای از کلمات برای کنترل یک یونیت دندانپزشکی مورد استفاده قرار گرفت که بنابه عللی که بعداً اشاره خواهد گردید مورد استفاده بیشتری نیافت [۲].

## ۱-۲-۲ سیستم‌های مورد استفاده

جهت دستیابی به بازشناسی کلمات گسسته، دو شیوه مورد توجه قرار گرفت که ذیلاً به این دو شیوه می‌پردازیم.

### ۱-۲-۲-۱ مدل‌های مارکوف پنهان با چگالی مشاهدات پیوسته (CDHMMs)

مدل‌های مارکوف پنهان به عنوان یکی از شیوه‌های موفق، امروزه در کاربردهای بازشناسی گفتار پیوسته و گسسته کاربردهای فراوانی یافته‌اند. این مدل‌ها باتوجه به توانایی بالائی که در مدل‌نمودن ویژگی‌های گفتار و بخصوص ویژگی دینامیک گفتار دارند، مورد بررسی و استفاده فراوانی در این زمینه قرار گرفته‌اند. مناسب‌ترین این مدل‌ها جهت اینگونه مدل‌سازی، از نقطه نظر

چگالی احتمالات خروجی، انواع با چگالی مشاهدات پیوسته می‌باشند. این گونه از HMM ها به دلیل دقت بالایی که در مدل‌سازی ارائه می‌دهند مناسب تشخیص داده شده‌اند [۵][۶]. به همین دلیل، در این پیاده‌سازی نیز از اینگونه مدل‌های آماری استفاده گردید.

#### ۱-۲-۱-۱ توپولوژی مدل‌ها

پیاده‌سازی اولیه بر روی دادگان گفتاری ده‌کلمه‌ای براساس CDHMMs دارای ۶ حالت صورت‌گرفته بود. در کاربرد ۱۰۰ کلمه‌ای، باتوجه به اینکه طول برخی کلمات طولانی‌تر بوده، از CDHMMs با ۸ حالت استفاده گردید. این HMM ها از نوع چپ - راست و بدون انتقال جهشی در نظر گرفته شدند [۴]. اشاره به این نکته نیز در اینجا خالی از اهمیت نیست که باتوجه به توانایی HMM ها در مدل‌سازی دینامیک سیگنال گفتاری، مدل‌نمودن سکوت ابتدا و انتهای کلمه نیز برعهده حالت‌های آغازین و پایانی HMM ها گذاشته شد و از آشکارسازی نقاط ابتدا و انتهای گفتار پرهیز گردید.

#### ۱-۲-۲-۱ استخراج ویژگی‌های طیفی

ویژگی‌های طیفی مورد استفاده در این بخش از طرح پژوهشی، ضرائب کپسترال LP انتخاب گردیدند. برای استخراج این ضرائب از سیگنال گفتاری رقمی شده، مراحل زیر به ترتیب به انجام رسید [۱][۴].

- ۱- پیش‌تأکید<sup>۱</sup>. وظیفه این بخش مسطح کردن طیف فرکانسی سیگنال گفتار برای پیشگیری از مشکلات احتمالی ناشی از محدودیت طول لغت دیجیتال در پردازش‌های بعدی می‌باشد.
- ۲- تقسیم به قاب‌ها<sup>۲</sup>. سیگنال گفتار در این بخش به مجموعه‌ای از بلوک‌های متداخل با طول زمانی حدود ۲۰ تا ۲۵ میلی‌ثانیه و فاصله فریم ۱۰ تا ۱۵ میلی‌ثانیه تقسیم می‌گردد. هر قاب سپس مورد پردازش مستقل قرار گرفته و ویژگی‌های سیگنال گفتاری در داخل آن پایدار فرض می‌شود.
- ۳- اعمال پنجره همینگ<sup>۳</sup> که باعث کاهش دامنه نمونه‌های کناری قاب می‌گردد.
- ۴- استخراج ضرائب LPC<sup>۴</sup>.
- ۵- بدست آوردن ضرائب کپسترال<sup>۵</sup> با استفاده از ضرائب LPC.
- ۶- اعمال لیفتر جوانگ<sup>۶</sup> جهت وزن‌دهی ضرائب کپسترال.

<sup>1</sup> Pre-emphasis

<sup>2</sup> Frame Blocking

<sup>3</sup> Hamming Window

<sup>4</sup> Linear Predictive Coding

<sup>5</sup> Cepstral Coefficients

<sup>6</sup> Juang Lifter

۷- بدست آوردن ضرائب دلتا و دلتا - دلتا برای ضرائب کپسترال و انرژی لگاریتمی.  
بردارهای ویژگی بدست آمده از مراحل فوق به عنوان ویژگی‌های طیفی نماینده گفتار در  
مراحل بعدی پردازش مورد استفاده قرار می‌گیرند.

### ۱-۲-۳ آموزش مدل‌ها

برای بدست آوردن یک سیستم بازشناسی گفتار مناسب، لازم است ابتدا مدل‌ها به نحو مطلوبی  
مورد آموزش قرار گیرند. یکی از پرستفاده‌ترین روش‌ها در آموزش مدل‌های HMM، روشی مبتنی  
بر تکنیک درستنمایی بیشینه (ML<sup>V</sup>) است، موسوم به Baum - Welch. این الگوریتم یک روش تکراری  
است و رسیدن آن به یک ماکزیمم محلی برای درستنمایی اثبات شده است [۷]. با استفاده از این  
الگوریتم و در شرایط دنباله‌های مشاهده چندگانه، روابط باز تخمین برای پارامترهای اصلی در  
یک CDHMM به قرار زیر می‌باشند [۱][۶]:

$$\hat{c}_{jk} = \frac{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(j, k)}{\sum_{l=1}^L \sum_{t=1}^T \sum_{m=1}^M \gamma_t^{(l)}(j, m)} \quad (1-1)$$

$$\hat{\mu}_{jk} = \frac{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(j, k) \cdot \mathbf{O}_t^{(l)}}{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(j, k)} \quad (2-1)$$

$$\hat{U}_{jk} = \frac{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(j, k) \cdot (\mathbf{O}_t^{(l)} - \mu_{jk})(\mathbf{O}_t^{(l)} - \mu_{jk})^T}{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(j, k)} \quad (3-1)$$

در این روابط،  $c_{jk}$  وزن مخلوط  $k$  ام از حالت  $j$  ام،  $\mu_{jk}$  بردار میانگین آن و  $U_{jk}$  ماتریس  
کوواریانس آن می‌باشند و روابط برای  $\mu_{jk}$  و  $U_{jk}$  به صورت برداری (یا ماتریسی) نوشته شده‌اند.  
همچنین،  $\gamma(j, k)$  احتمال بودن در مخلوط  $k$  ام از حالت  $j$  ام در لحظه  $t$  و مشاهده  $\mathbf{O}_t$  است.

پیش از اینکه مدل‌های CDHMM بتوانند آموزش ببینند، لازم است مدل‌های اولیه با مقادیر  
مناسبی مقداردهی شوند تا دستیابی به ماکزیمم محلی بتواند مترادف با رسیدن به ماکزیمم کلی تلقی  
شود. از این رو مقداردهی اولیه پارامترها از اهمیت فوق‌العاده‌ای برخوردار است. برای این منظور در  
دو مرحله به شرح زیر عمل می‌شود:

الف: کلیه مقادیر دنباله‌های مشاهده آموزشی موجود برای هر مدل بطور یکنواخت به  
تعداد حالت‌های آن مدل تقسیم شده و بر روی تمامی بردارهای بدست آمده برای هر حالت مقادیر

<sup>7</sup> Maximum Likelihood

بردارهای میانگین و واریانس بدست می‌آیند و به عنوان مقادیر ابتدائی برای پارامترهای آن حالت مورد استفاده قرار می‌گیرند.

ب: با استفاده از مقادیر ابتدائی فوق، تمامی دنباله‌های مشاهده موجود با استفاده از الگوریتم ویتربی با حالات مدل مربوطه هم‌ردیف می‌شوند. سپس مقادیر هم‌ردیف شده بند الف فوق برای بدست آوردن پارامترهای هر حالت مورد استفاده قرار می‌گیرند. این عمل تا رسیدن به یک معیار همگرایی و یارسیدن به تعداد موردنظر از دفعات تکرار ادامه می‌یابد.

پارامترهای بدست آمده از مراحل فوق، سپس به عنوان پارامترهای اولیه در الگوریتم Baum-Welch مورد استفاده قرار می‌گیرند.

درخصوص آموزش مدل‌های دارای چگالی‌های مشاهده مخلوط گوسین<sup>۸</sup>، اگرچه استفاده از روش فوق نیز جایز می‌باشد ولی ترجیح داده می‌شود که در مراحل بعدی نسبت به افزایش تعداد عناصر مخلوط اقدام نمود. بنابراین مراحل ابتدائی آموزش عموماً با چگالی‌های مشاهده تک‌گوسین صورت می‌پذیرد و سپس برای افزایش تعداد عناصر مخلوط از الگوریتمی نظیر روش شکافت مخلوط<sup>۹</sup> که در بحث بازشناسی گفتار پیوسته به تفصیل بیشتری مورد اشاره قرار خواهد گرفت، استفاده می‌شود.

نکته دیگری که در اینجا حائز اهمیت است، استفاده از نمایش درستنمایی بصورت لگاریتمی است که لازمه پیاده‌سازی الگوریتم‌های نظیر Baum-Welch می‌باشد چرا که مقادیر درستنمایی بعد از چند لحظه زمانی به علت ضرب شدن در مقادیر احتمالی بسیار کوچک، به سمت صفر میل خواهد نمود که عمل آموزش را غیرممکن می‌سازد. بهمین جهت لازم است که در پیاده‌سازی از نمایش لگاریتمی احتمالات و درستنمایی استفاده نمود. در این صورت ضرب‌ها هم تبدیل به جمع شده و بنابراین مشکل میل نمودن سریع مقادیر حاصلضرب به سمت صفر نیز از بین خواهد رفت.

#### ۱-۲-۲-۱-۴ بازشناسی کلمات جدا

مرحله بازشناسی کلمات جدا با استفاده از HMMها عموماً به کمک الگوریتم ویتربی<sup>۱۰</sup> صورت می‌گیرد [۵][۶]. این الگوریتم که براساس شیوه برنامه‌نویسی پویا طراحی گردیده از مزایای چندی به شرح زیر برخوردار است:

- یافتن مسیر بهینه حالت براساس اصل بهینگی Bellman .
- کاهش فراوان حجم محاسبات به دلیل استفاده از اصول برنامه‌نویسی پویا.

<sup>8</sup> Mixture-Gaussian Observation Densities

<sup>9</sup> Mixture Splitting

<sup>10</sup> Viterbi Algorithm

- یافتن همزمان احتمال دنباله مشاهدات به شرط وجود مدل  $(P(O|\lambda))$  که در بازشناسی گفتار کاربرد دارد.

- عدم استفاده از جمع (برخلاف الگوریتم جلورونده<sup>۱۱</sup>) و درمقابل استفاده از بیشینه‌سازی که باعث ساده‌شدن پیاده‌سازی آن بصورت لگاریتمی می‌شود.

باتوجه به موارد فوق، الگوریتم ویتربی یکی از روش‌های مناسب و مطلوب جهت انجام بازشناسی گفتار تلقی شده و لذا در این بخش از طرح پژوهشی جاری نیز مورد استفاده قرار گرفته است.

### ۱-۲-۲-۲ شبکه‌های عصبی با تأخیر زمانی<sup>۱۲</sup>

یکی دیگر از روش‌های مطرح در بازشناسی گفتار، استفاده از شبکه‌های عصبی با تأخیر زمانی (TDNN) می‌باشد. این نوع شبکه به عنوان گونه‌ای تغییر یافته از MLP<sup>۱۳</sup> جهت کاربرد در بازشناسی گفتار مطرح گردیده است. از آنجائی که شبکه‌های عصبی توانائی بالائی در طبقه‌بندی الگوهای استاتیک دارند، جهت اعمال آنها در این زمینه، نیاز به افزودن ویژگی توانائی برخورد با دینامیک الگوهای گفتاری به آنها می‌باشد. در TDNN، اطلاعات چند فریم گفتاری، به همین منظور، بطور همزمان به سیستم اعمال می‌شود. در عمل ملاحظه گردیده است که توانائی نسبتاً قابل قبولی را می‌توان در این زمینه از این گونه شبکه‌ها انتظار داشت [۸].

### ۱-۲-۲-۱ دادگان به کار رفته

برای این مرحله از پیاده‌سازی، از دادگانی که با استفاده از دوازده کلمه کترلی اشاره‌شده قبلی تشکیل گردیده بود استفاده گردید. این دادگان با استفاده از گفتار ۷۰ گوینده که به تساوی از مردان و زنان تشکیل شده بودند و با سه بار تکرار هر کلمه تشکیل گردیده بود. علاوه بر این یک بخش وابسته به گوینده نیز در این دادگان پیش‌بینی شده بود تا بتوان توانائی سیستم بازشناسی را روی اینگونه کاربرد نیز آزمود. [۱]

### ۱-۲-۲-۱ پیاده‌سازی

در مرحله پیاده‌سازی این سیستم نیز همانند سیستم مبنی بر CDHMM، نیاز به مراحل استخراج ویژگی‌های گفتاری، آموزش سیستم بازشناسی و بالاخره شبیه‌سازی و بازشناسی می‌باشد. موارد فوق بر روی یک شبکه عصبی از نوع TDNN پیاده‌سازی گردیده و شبکه با دو مجموعه دادگان،

<sup>11</sup> Forward Algorithm

<sup>12</sup> Time Delay Neural Networks

<sup>13</sup> Multi-Layer Perceptron

یکی دادگان مبتنی بر مجموعه ارقام اشاره شده در بخش ۱-۲-۱ و دیگری مجموعه اشاره شده در بخش ۱-۲-۲-۱-۲-۱ آزمایش گردید [۲]. اگرچه نتایج بدست آمده از این مرحله نسبتاً مناسب می‌باشند، ولی کیفیت بالاتر سیستم CDHMM در بازشناسی گفتار گسسته کاملاً آشکار بوده است. بررسی دقیق تر نشان می‌دهد که شبکه TDNN، علیرغم تأخیر زمانی ایجاد شده، قادر به دخالت دادن مناسب و مطلوب دینامیک گفتار در طبقه‌بندی نمی‌باشد. یکی از روش‌های پیشنهاد شده استفاده از نوعی پیچش زمانی به عنوان پردازش اولیه جهت انطباق مناسب زمانی سیگنال گفتار با الگوهای استفاده شده برای آموزش شبکه می‌باشد. از آنجا که روش‌های نظیر پیچش زمانی پویا (DTW<sup>14</sup>) خود به عنوان الگوریتم‌های بازشناسی گفتار مستقیماً مورد استفاده می‌باشند، تصمیم گرفته شد تحقیق بیشتر در این زمینه متوقف شده و پیاده‌سازی تنها براساس CDHMMs ادامه یابد.

### ۱-۲-۳ مشکلات در بازشناسی و اصلاحات مربوطه

سیستم بازشناسی ایجاد شده براساس CDHMMs، علیرغم توانایی‌های خوبی که از خود نشان می‌دهد، از دو مشکل اساسی رنج می‌برد: محدودیت سرعت و کاهش کیفیت در شرایط مختلف محیطی (نظیر وجود نویز زمینه). لذا دنباله کار در این بخش بر روی رفع این دو مشکل و یافتن راه‌حل‌هایی برای آنها متمرکز گردید.

#### ۱-۲-۳-۱ سرعت بازشناسی

مسئله سرعت در بازشناسی در مراحل مختلف کار قابل طرح می‌باشد. پیاده‌سازی بهینه الگوریتم‌ها موجب می‌گردد که با سرعت بالاتری به انجام برسند. چه در مرحله آموزش و چه در مرحله بازشناسی، این فاکتور دارای اهمیت زیادی می‌باشد. داشتن سرعت بالا در آموزش، درعین حال دارای اهمیت سرعت در بازشناسی نمی‌باشد چرا که آموزش عموماً در شرایط off-line صورت می‌گیرد درحالی‌که در کاربردهای واقعی، بازشناسی اغلب on-line می‌باشد. به همین جهت اهمیت سرعت در بازشناسی بیش از آموزش می‌باشد. البته باید توجه نمود که بهر صورت کندبودن آموزش خود باعث تأخیر در پیاده‌سازی موارد آزمایشی سیستم می‌گردد.

علاوه بر موارد فوق باید توجه نمود که چه در آموزش و چه در بازشناسی، از آنجا که نیاز به استخراج ویژگی‌های گفتاری می‌باشد، سرعت استخراج ویژگی‌ها دارای اهمیت می‌باشد. با این همه، در این خصوص نمی‌توان بهبودی چندانی، جز از طریق بهینه‌سازی برنامه‌نویسی و یا برخی تغییرات جزئی بدست آورد [۱]. در خصوص الگوریتم آموزش نیز باتوجه به اینکه در این طرح پژوهشی،

<sup>14</sup> Dynamic Time Warping

آموزش بصورت off-line صورت می‌گرفته است، افزایش سرعت اجرای الگوریتم چندان مورد توجه قرار نگرفته است.

الگوریتم بازشناسی، باتوجه به چند مورد، نیاز به افزایش سرعت دارد. اول اینکه این الگوریتم می‌تواند به صورت on-line مورد استفاده واقع شود. علاوه بر این، باتوجه به استفاده از HMM های باچگالی پیوسته، عمده زمان صرف محاسبه تابع چگالی احتمال مخلوط پیوسته (گوسین‌ها) می‌گردد که شامل محاسبه مقادیر نمائی و ضرب‌های برداری می‌باشد. باتوجه به زمان‌گیر بودن این محاسبات و اینکه باید به‌ازاء هر بردار ورودی این کار تکرار گردد، زمان صرف‌شده کلی به‌ازاء هر دنباله بردارهای ورودی (بیان گفتاری) قابل توجه می‌باشد. از آنجا که در بازشناسی با دایره کلمات ۱۰۰ کلمه، تعداد مدل‌هایی که باید مورد بررسی قرار گیرند، زیاد است (۱۰۰ عدد)، این محاسبات باید ۱۰۰ بار تکرار گردد تا مدل مناسب بتواند بررسی و یافت شود. بنابراین زمان زیادی باید صرف شود تا بازشناسی بتواند صورت گیرد.

شیوه‌های چندی برای بهبود سرعت در الگوریتم بازشناسی به‌کار گرفته شده‌اند. آنچه در این طرح پژوهشی برای این منظور مورد توجه قرار گرفته از دو جنبه قابل طرح است. جنبه اول بهبود پیاده‌سازی الگوریتم جهت دستیابی به سرعت‌های بازشناسی قابل قبول می‌باشد. دومین جنبه از کار، توجه به روش‌های جستجوی بهبودیافته جهت کاهش فضای جستجو در هنگام کار با تعداد کلمات بالا می‌باشد.

در پیاده‌سازی الگوریتم، توجه به ماتریس انتقال می‌تواند از جهت افزایش سرعت راهگشا باشد. توجه به غیرارگودیک<sup>۱۵</sup> بودن HMM و نیز محدود نمودن اختصاص فریم‌ها به حالت‌ها از نقطه نظر زمانی، باتوجه به محدودیت‌های طبیعی در گفتار، می‌تواند باعث افزایش نسبی سرعت اجرای الگوریتم شوند [۱][۲].

از جنبه کاهش فضای جستجو، می‌توان بجای اقدام به اعمال تک‌تک مدل‌ها (مثلاً ۱۰۰ مدل) به داده ورودی جهت محاسبه درست‌نمایی مربوطه و سپس انتخاب بالاترین درست‌نمایی جهت بدست آوردن کاندیدای مناسب، به شیوه دیگری عمل نمود [۴]. در شیوه اخیر، اقدام به اعمال همزمان الگوریتم و تریبی جهت هم‌ردیف‌سازی و بدست آوردن درست‌نمایی نهائی بین گفتار ورودی و تمامی ۱۰۰ مدل گفتاری موجود می‌گردد. این شیوه، اگرچه در ظاهر امر مانند شیوه قبلی است، با این تفاوت که تمام محاسبات همزمان و بطور موازی صورت می‌گیرند، اما دارای این ویژگی مطلوب است که می‌توان با استفاده از تکنیک جستجوی شعاعی<sup>۱۶</sup>، در مقاطعی از عمل، فضای جستجو را با استفاده از یک شعاع جستجوی محدود که از طریق اعمال یک سطح آستانه نسبی (نسبت به

<sup>15</sup> Non-Ergodic

<sup>16</sup> Beam Search

درستنمائی بیشینه در هر مقطع ایجاد می‌شود، کاهش داد. اینگونه جستجوها در موارد مختلف و با درجات موفقیت قابل قبولی در کاربردهای مختلف و بویژه کاربردهای بازشناسی گفتار مورد استفاده قرار گرفته‌اند [۹][۱۰]. به این ترتیب و با اعمال سطح آستانه مناسب می‌توان ضمن داشتن نرخ خطای قابل قبول در بازشناسی، سرعت اجرای الگوریتم بازشناسی را نیز افزایش داد.

### ۱-۲-۳-۲ مقابله با اثرات شرایط مختلف محیطی

شرایط محیطی مختلف می‌توانند به شدت کیفیت یک سیستم بازشناسی گفتار را تحت تأثیر قرار دهند. این شرایط شامل انواع نویزهای جمع‌شونده با سیگنال<sup>۱۷</sup> ناشی از منابع مختلف آلاینده فضای صوتی می‌باشند. از جمله این موارد می‌توان به انواع نویز ناشی از کار دستگاههای مختلف برقی، موتورهای احتراقی و نظایر آنها، صدای سایر افراد که به عنوان نویز زمینه سیگنال گفتار را تحت تأثیر قرار می‌دهد و یا هرگونه صدای دیگری نظیر صدای باد، ضربات و اصطکاک، جریان آب و نظائر آنها باشد. این منابع می‌توانند از نظر طیف فرکانسی، انرژی سیگنال جمع‌شونده، و پایداری زمانی، کاملاً ویژگی‌های متفاوتی داشته باشند.

منابع دیگری نیز که می‌توانند بر سیگنال گفتار موردنظر تأثیر بگذارند، شامل منابع نویز ضرب‌شونده<sup>۱۸</sup> یا Convolutional است که عموماً مشخصات فرکانسی سیگنال گفتاری را تحت تأثیر قرار می‌دهند. این منابع عموماً شامل کانال‌های انتقال می‌باشند نظیر خط تلفن، میکروفون، ضبط گفتار، کانال رادیویی و امثال آنها.

منابع نویز، تأثیر آنها بر کیفیت بازشناسی و روش‌های معمول در مقابله با آنها در [۲] مورد بررسی قرار گرفتند. یکی از روش‌هایی که در مقابله با نویز کارآئی خوبی از خود نشان داده است "معیار تصویروزن‌دهی شده"<sup>۱۹</sup> (WPM) نام دارد. در این روش، با توجه به اثر نویز سفید بر روی بردارهای کپسترال یک معیار تصویر معرفی می‌شود که به کمک آن می‌توان برای هر بردار درستنمائی را به شکل جدیدی در هر حالت از HMM محاسبه نمود [۱]. در بررسی‌های به عمل آمده مزایای چندی به شرح زیر برای روش WPM بدست آمد:

الف - این روش علاوه بر نویز سفید، برای نویزهای رنگی نیز تا حد زیادی مفید می‌باشد.

ب - دقت بازشناسی در شرایط نویزی با استفاده از این روش افزایش قابل ملاحظه‌ای می‌یابد.

پ - آموزش مدل‌ها برای اعمال این روش دستخوش تغییر نشده و WPM تنها در مرحله بازشناسی قابل اعمال است.

<sup>17</sup> Additive Noise

<sup>18</sup> Multiplicative

<sup>19</sup> Weighted Projection Measure