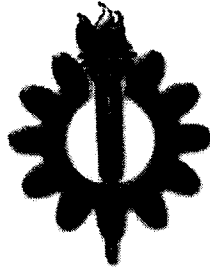


بسم الله الرحمن الرحيم

مكتب  
الادارة العامة  
للمحكمة  
القضاء  
الاولى  
بغداد

٣٩٧٧٩

وزارت اطلاعات و امور خارجه  
جمهوری اسلامی ایران



دانشگاه علم و صنعت ایران

دانشکده مهندسی کامپیوتر

۱۳۸۰ / ۱۲ / ۲۸

016174

موضوع: اعمال روشهای بهبود گفتار به عنوان پیش پردازش، جهت بالا بردن دقت  
بازشناسی گفتار فارسی

ارائه کننده: محسن رحمانی

پایان نامه برای دریافت درجه کارشناسی ارشد  
در رشته مهندسی کامپیوتر

استاد راهنما: دکتر احمد اکبری

دی ماه ۱۳۸۰

۳۹۷۷۹

تقديم

به پدر و مادر عزیزم

## چکیده

کارایی سیستمهای بازشناسی در حضور نویز کاهش می‌یابد. در این پایان‌نامه روشهای مقاوم کردن سیستم بازشناسی در برابر نویز دسته‌بندی شده و از بین آنها روشهای مبتنی بر داده بررسی می‌شود.

روشهای مختلف تفاضل طیف، جبران کپسترال، نگاشت ضرایب و روشهای مبتنی بر ویژگیهای مقاوم تحلیل شده‌اند. یک رابطه جدید برای تسطیح نویز در تفاضل طیف و یک لیفتر جبران کننده برای وزندهی ضرایب کپسترال پیشنهاد شده است. همچنین ایده تخمین ویژگیهای مقاوم از ویژگی‌های نویزی ارائه شده است.

نرخ بازشناسی با استفاده از پیش‌پردازنده‌های پیشنهادی افزایش داشته است. نگاشت ضرایب کپسترال نویزی با شبکه‌های عصبی نتایج قابل قبولی داشته است. بین همه روشها، نگاشت ضرایب تفاضل کپسترال میانگین با استفاده از شبکه عصبی بهترین نتیجه را دارد.

بر خود لازم میدانم از استاد راهنمای گرامی خود، جناب آقای دکتر احمد اکبری که راهنمایی‌ها و کمکهای ارزشمند ایشان نقش بسزایی در پیشبرد پروژه داشته است، تشکر و قدردانی کنم.

همچنین از دوستان گرامیم آقایان، مهندس بابک ناصرشریف و مهندس ستار هاشمی که از رهنمودها و کمکهای آنها در این پروژه بهره بسیار بردم، صمیمانه تشکر می‌کنم.

و نیز تشکر می‌کنم از تمامی کسانی که در ضبط پایگاه صدا و دیگر مراحل انجام پروژه یاریگر من بوده‌اند.

صفحه	عنوان
	<b>مقدمه</b>
۱	
۳	<b>فصل ۱: سیستمهای بازشناسی گفتار</b>
۴	۱-۱ پارامترهای بازشناسی گفتار
۴	۱-۱-۱ وابسته یا مستقل از گوینده
۴	۲-۱-۱ گفتار مجزا // متصل / پیوسته
۵	۳-۱-۱ اندازه کتاب لغت
۵	۴-۱-۱ محدودیتهای زبانی
۶	۵-۱-۱ گفتار مکالمه‌ای
۶	۶-۱-۱ محیط
۶	۲-۱ اجزای یک سیستم بازشناسی
۶	۱-۲-۱ نمونه برداری از سیگنال صوتی
۷	۲-۲-۱ استخراج ویژگی از سیگنال گفتار
۷	۱-۲-۲-۱ قاب‌بندی
۸	۲-۲-۲-۱ پیش‌تاکید
۸	۳-۲-۲-۱ پنجره‌گذاری
۹	۴-۲-۲-۱ بانک فیلتر
۹	۵-۲-۲-۱ آنالیز پیشگویی خطی
۱۰	۶-۲-۲-۱ آنالیز کپسترال
۱۱	۷-۲-۲-۱ ضرایب دلتا کپسترال
۱۲	۸-۲-۲-۱ استفاده از مقیاس MEL در آنالیز کپسترال
۱۳	۹-۲-۲-۱ ضرایب انرژی و مشتقات آن
۱۴	۳-۲-۱ تطبیق الگو
۱۵	۱-۳-۲-۱ مدل مخفی مارکف
۱۸	۴-۲-۱ پردازش زبان
۱۸	۳-۱ سیستم بازشناسی پایه
۱۹	۴-۱ نتیجه‌گیری
۲۱	<b>فصل ۲: تاثیر نویز بر سیستم بازشناسی</b>
۲۲	۱-۲ تاثیر نویز بر سیگنال گفتار
۲۲	۱-۱-۲ مدلی برای نویز
۲۴	۲-۱-۲ توزیع بردارهای ویژگی گفتار نویزی
۲۵	۲-۲ کاهش کارایی سیستمها بازشناسی در حضور نویز

۲۷	۱-۲-۲ اثر نویز بر کلاسه‌بندی
۲۸	۳-۲ روشهای مقاومت در برابر نویز
۳۱	۴-۲ نتیجه‌گیری
۳۲	<b>فصل ۳: ویژگی‌های مقاوم</b>
۳۳	۱-۳ تفاضل میانگین در حوزه کپسترال
۳۶	۲-۳ ضرایب کپسترال حاصل از پیشگویی خطی مبتنی بر ادراک انسان
۳۷	۳-۳ ویژگیهای RASTA-PLP
۳۸	۱-۳-۳ تحلیل طیف نسبی
۴۰	۴-۳ اعمال لیفتر
۴۳	۵-۳ نتیجه‌گیری
۴۵	<b>فصل ۴: روشهای بهبود گفتار</b>
۴۷	۱-۴ روشهای تخمین نویز
۴۷	۱-۱-۴ استفاده از معیار انرژی
۴۸	۲-۱-۴ استفاده از نرخ عبور از صفر و معیار انرژی
۵۰	۳-۱-۴ شاخه‌بندی انرژی
۵۱	۲-۴ تفاضل طیف
۵۱	۱-۲-۴ تفاضل طیف استاندارد
۵۳	۲-۲-۴ پیاده‌سازی تفاضل طیف
۵۵	۳-۲-۴ پارامترهای تفاضل طیف
۵۷	۳-۴ نگاشت ضرایب کپسترال
۵۸	۱-۳-۴ جبران کپسترال وابسته به نویز
۵۹	۲-۳-۴ نگاشت ضرایب کپسترال با استفاده از لیفتر
۶۰	۴-۴ شبکه عصبی
۶۳	۱-۴-۴ استفاده از اطلاعات مسیر
۶۴	۲-۴-۴ استفاده از ویژگیهای مقاوم‌تر گفتار نویزی برای تخمین
۶۵	۳-۴-۴ استفاده از ویژگی‌های مقاوم برای تخمین ویژگی‌های تخریب شده
۶۶	۵-۴ نتیجه‌گیری
۶۷	<b>فصل ۵: مقایسه روشها و ارائه چند پیش‌پردازنده</b>
۶۸	۱-۵ مقایسه روشها
۷۲	۲-۵ بازشناسی گفتار متصل
۷۲	۳-۵ طراحی پیش‌پردازنده‌هایی برای سیستم بازشناسی
۷۳	۱-۳-۵ استفاده از پایگاه داده استریو
۷۴	۲-۳-۵ عدم استفاده از پایگاه داده استریو
۷۵	۴-۵ نتیجه‌گیری

۷۶	فصل ششم: نتیجه گیری و ادامه کار
۷۷	۳-۶ خلاصه کارهای انجام شده
۷۹	۲-۶ پیشنهاد ادامه کار
۸۰	پیوست
۸۱	الف-۱ پایگاه داده گفتار
۸۲	الف-۲ نويز
۸۲	الف-۳ آزمایشات انجام شده
۸۳	الف-۳-۱ ویژگی‌های مقاوم
۸۴	الف-۳-۲ تفاضل طیف
۸۶	الف-۳-۳ شبکه‌های عصبی
۸۸	الف-۳-۴ جبران کپسترال وابسته به نويز و ليفتر جبران کننده
۸۸	الف-۳-۵ تخمین نويز
۸۹	الف-۳-۶ قطعه‌بندی گفتار
۸۹	الف-۳-۷ پیش‌پردازنده‌ها
۹۲	مراجع
۹۶	واژه‌نامه



## فهرست اشکال

صفحه	عنوان
۱۱	شکل ۱-۱: محاسبه ضرایب کپستروم حقیقی با استفاده از DTFT
۱۶	شکل ۲-۱: مثالی از مدل مخفی مارکف سه حالتی
۱۹	شکل ۳-۱: مدل مخفی مارکف، حالتها و نحوه انتقال بین حالتها در سیستم پایه
۱۹	شکل ۴-۱: سیستم بازشناسی پایه
۲۳	شکل ۱-۲: مدلی برای نویز وارد بر گفتار
۲۵	شکل ۲-۲: تخمین توزیع گفتار نویزی با معادلات ۱۱-۲ و ۱۲-۲
۲۶	شکل ۳-۲: تاثیر نویز بر درصد بازشناسی
۲۷	شکل ۴-۲: تاثیر نویز بر میزان بازشناسی سیستم پایه
۲۷	شکل ۴-۲: کلاسه‌بند الگو: مرز تصمیم انتخاب شده خطا را حداقل می‌کند
۲۸	شکل ۵-۲: اثر نویز بر خطای تصمیم‌گیری. خطایی که به دلیل انتخاب نادرست مرز تصمیم بدست آمده است با خطای مرز تصمیم بهینه جمع می‌شود
۲۹	شکل ۶-۲: مقایسه روشهای جبران داده و جبران کلاسه‌بند
۳۰	شکل ۷-۲: مقایسه روشهای مقاومت در برابر نویز. الف: سیستم پایه، ب: بهبود مدل ج: ویژگیهای مقاوم، د: تخمین گفتار تمیز
۳۶	شکل ۱-۳: تحلیل PLP
۳۸	شکل ۲-۳: روش RASTA-PLP
۳۹	شکل ۳-۳: بلاک دیاگرام عملیات Rasta
۳۹	شکل ۴-۳: فیلتر RASTA. الف: پاسخ ضربه، ب: پاسخ فرکانسی
۴۱	شکل ۳-۳: لیفترهای بکار رفته در سیستمهای بازشناسی. الف: لیفتر خطی ب: لیفتر سینوسی ج: لیفتر نمایی
۴۱	شکل ۳-۶: واریانس اختلاف ضرایب کپسترال نویزی و تمیز
۴۱	شکل ۳-۷: معکوس واریانسها و منحنی برازش شده
۴۲	شکل ۳-۸: لیفترهای پیشنهادی. الف: لیفتر از معکوس واریانسها ب: حذف ضرایب
۴۸	شکل ۱-۴: استفاده از مقادیر آستانه انرژی در مرزبندی کلمات
۵۰	شکل ۲-۴: هیستوگرام انرژی برای بیان ده بار عبارت یک
۵۱	شکل ۳-۴: مقایسه روشهای تخمین نویز
۵۳	شکل ۴-۴: انواع مختلف تفاضل طیف
۵۴	شکل ۵-۴: مراحل پیاده‌سازی الگوریتم تفاضل طیف
۵۵	شکل ۴-۶: مقایسه تاثیر افزایش $\alpha$ در سیگنال به نویزهای مختلف
۵۶	شکل ۴-۷: تاثیر تغییرات مقدار $\beta$ در رابطه ۴-۱۲
۵۶	شکل ۴-۸: تاثیر تغییرات مقدار $\beta$ در رابطه ۴-۱۳

- شکل ۴-۹: نتایج بازشناسی برای مقادیر مختلف  $\beta_1$  و  $\beta_2$  در رابطه ۴-۱۴
- شکل ۴-۱۰: نگاشت در حوزه زمان
- شکل ۴-۱۱: بلوک دیاگرام نگاشت در حوزه‌های تبدیل یافته
- شکل ۴-۱۲: شبکه عصبی مورد استفاده برای تخمین ضرایب کپستروم
- شکل ۴-۱۳: الف: ضرایب کپسترال اول برای ۲۵ قاب متوالی. ب: ضرایب کپسترال ۱ تا ۱۲ متعلق به یک قاب.
- شکل ۴-۱۳: نحوه وزن دهی ضرایب کپسترال
- شکل ۴-۱۴: بلوک دیاگرام تخمین ضرایب تخریب شده با ضرایب نسبتا سالم
- شکل ۵-۱: دسته‌بندی روشهای ارائه شده
- شکل ۵-۲: مقایسه روشهای بهبود گفتاری را که از پایگاه داده استریو استفاده می‌کنند
- شکل ۵-۳: مقایسه ویژگی‌های مقاومی که از پایگاه داده استریو استفاده می‌کنند
- شکل ۵-۴: مقایسه ویژگی‌های مقاومی که از پایگاه داده استریو استفاده نمی‌کنند
- شکل ۵-۵: مقایسه نماینده دسته‌های مختلف
- شکل ۵-۶: مقایسه روشها. مقادیر میله‌های نمودار نشان‌دهنده میانگین نتایج بازشناسی در سیگنال به نویزهای مختلف می‌باشد.
- شکل ۵-۶: مقایسه روشها. مقادیر میله‌های نمودار نشان‌دهنده نتایج بازشناسی در سیگنال به نویز 10db می‌باشد.
- شکل ۵-۷: پیش‌پردازش پیشنهادی برای وقتی که پایگاه داده‌ای از ضبط همزمان نویزی و تمیز در دسترس باشد
- شکل ۵-۸: پیش‌پردازش بدون استفاده از پایگاه داده استریو
- شکل ۵-۹: استفاده از تفاضل طیف برای بهبود گفتار در پیش‌پردازش
- شکل ۵-۱۰: مقایسه نتایج برای گویندگان زن و مرد

## فهرست جداول

صفحه	عنوان
۱۸	جدول ۱-۱: نتایج بازشناسی برای سیستم پایه
۳۶	جدول ۱-۳: مقایسه درصد بازشناسی استفاده از CMS و MFCC
۴۲	جدول ۲-۳: نتیجه آزمایشات با اعمال لیفترهای مختلف
۵۱	جدول ۱-۴: مقایسه روشهای تخمین نویز
۵۵	جدول ۲-۴: بازشناسی با تفاضل طیف برای مقادیر مختلف $\alpha$
۵۹	جدول ۳-۴: نتایج استفاده از روش SDCN در مقایسه با سیستم پایه
۵۹	جدول ۴-۴: نتایج استفاده از لیفتر جبران کننده
۶۱	جدول ۵-۴: درصد بازشناسی. استفاده از شبکه عصبی برای نگاشت در حوزه کپسترال
۶۳	جدول ۶-۴: درصد بازشناسی. استفاده از شبکه عصبی برای نگاشت در حوزه کپسترال
۶۴	جدول ۷-۴: نتایج استفاده از اطلاعات مسیر در نگاشت ضرایب کپسترال
۶۵	جدول ۸-۴: استفاده از ویژگیهای مختلف برای آموزش شبکههای عصبی
۶۵	جدول ۹-۴: استفاده از ویژگیهای مطمئن برای تخمین ویژگیهای نامطمئن
۶۸	جدول ۱-۵: روشهای آزمایش شده
۷۲	جدول ۲-۵: خطای مرزبندی برای سیگنال تمیز
۷۲	جدول ۳-۵: خطای مرزبندی برای سیگنال نویزی با و بدون تفاضل طیف
۷۳	جدول ۴-۵: مقایسه پیش‌پردازش‌هایی که از پایگاه داده استریو استفاده می‌کنند.
۷۵	جدول ۵-۵: مقایسه دو پیش‌پردازش که از پایگاه داده استریو استفاده نمی‌کنند
۸۱	جدول الف-۱: گویندگان پایگاه داده
۸۳	جدول الف-۲: چگالی طیف مربوط به نویزهای پایگاه داده NOISEX 92
۸۳	جدول الف-۳: مقایسه درصد بازشناسی استفاده از ویژگیهای مختلف، نویزهای مختلف و دو گوینده
۸۴	جدول الف-۴: تفاضل طیف با فراتخمین‌های مختلف
۸۴	جدول الف-۵: تفاضل طیف با فراتخمین‌ها و تسطیح نویزهای مختلف
۸۵	جدول الف-۶: تفاضل طیف با تغییر پارامترهای $\alpha$ ، $\beta_1$ و $\beta_2$
۸۶	جدول الف-۷: نتایج استفاده از شبکههای عصبی
۸۸	جدول الف-۸: نتایج بازشناسی وقتی که از SDCN و لیفتر استفاده شده‌است.
۸۸	جدول الف-۹: مقایسه روشهای تخمین طیف
۸۹	جدول الف-۱۰: خطای تقطیع خودکار
۸۹	جدول الف-۱۱: نتایج برای پیش‌پردازش‌هایی که از پایگاه داده استریو استفاده می‌کنند.
۹۰	جدول الف-۱۲: نتایج برای پیش‌پردازش‌هایی که از پایگاه داده استریو استفاده نمی‌کنند.
۹۰	جدول الف-۱۳: نتایج بازشناسی برای ۳۰ گوینده و برای پیش‌پردازش FE(SS)

## مقدمه

هدف نهایی سیستمهای بازشناسی گفتار فراهم آوردن قابلیت درک گفتار انسان بوسیله ماشین است [۱]. از آنجا که گفتار یکی از مهمترین ابزار ارتباطی انسان با محیط اطرافش محسوب می‌شود، ایجاد توانایی در برقراری ارتباط در ماشین‌ها، تحول بزرگی در استفاده از ماشین می‌باشد.

در یک سیستم بازشناسی گفتار، پارامترهایی مختلفی تعیین کننده درجه توانایی سیستم بازشناسی هستند. این پارامترها عبارتند از: وابسته و مستقل بودن از گوینده، بازشناسی کلمات مجزا و گفتار پیوسته، اندازه کتاب لغت، محدودیتهای زبانی، گفتار مکالمه‌ای و شرایط محیطی که بازشناسی در آن انجام می‌گیرد. در این پایان‌نامه روشهای مقاوم کردن سیستم بازشناسی در محیطهای نویزی بررسی می‌شود.

سیستمهای بازشناسی گفتار اساساً کلاسه‌بند<sup>۱</sup>هایی هستند که قطعات گفتار را به عنوان اعضای کلاسه‌ها طبقه‌بندی می‌کنند. سیستم بازشناسی توزیع بردارهای ویژگی را که از هر واحد گفتاری حاصل می‌شود، با استفاده از مجموعه‌ای از گفتار آموزشی فرا می‌گیرد. به هنگام بازشناسی هر قطعه گفتار به عنوان یک واحد گفتاری که توزیع آن به توزیعی از بردارهای ویژگی نزدیکتر باشد، کلاسه‌بندی می‌شود.

هرگاه گفتاری که باید تشخیص داده شود، با نویز تخریب گردد کارایی سیستمهای بازشناسی به شدت کاهش می‌یابد. اثر تخریبی سیگنال گفتار آن است که توزیع بردارهای ویژگی گفتار تخریب شده، شبیه توزیعی که سیستم بازشناسی با آن آموزش دیده است، نمی‌باشد. این عدم تطابق باعث کاهش کارایی سیستمهای بازشناسی در شرایط نویزی می‌شود.

روشهای مختلفی جهت مقابله با اثر نویز بر سیستمهای بازشناسی ارائه شده است. روشهای مقابله با نویز را می‌توان به سه دسته تقسیم کرد. دسته اول روشهایی هستند که از ویژگیهای مقاوم به نویز استفاده می‌کنند. در این روشها ویژگی‌هایی از گفتار استخراج می‌شود که در مقابل نویز حساسیت کمتری داشته باشند. دسته دوم روشهایی هستند که بر پایه تخمین گفتار تمیز عمل می‌کنند. در این روشها گفتاری که باید تشخیص داده شود با استفاده از الگوریتمهای بهبود گفتار، بهبود می‌یابد. و دسته سوم، روشهای مبتنی بر اصلاح مدل آکوستیکی سیستم بازشناسی هستند. در این روشها سعی می‌شود سیستم بازشناسی به گونه‌ای اصلاح شود که نسبت به نویز مقاوم باشد. در این پایان‌نامه دو دسته اول بررسی شده‌اند. اهمیت روشهای این دو دسته آن است که در آنها

نیازی به تغییر دادن شیوه تطبیق الگو نیست و فرایند مقاوم‌سازی فقط در مرحله پیش‌پردازنده<sup>۱</sup> سیستم بازشناسی انجام می‌شود.

فصل اول این پایان‌نامه سیستم‌های بازشناسی گفتار را به طور کلی بررسی می‌کند و یک سیستم بازشناسی مبتنی بر مدل مخفی مارکف را به عنوان سیستم پایه معرفی می‌کند. فصل دوم به بررسی مدل نویز، تاثیر آن بر سیگنال گفتار و نتیجه این تاثیر بر کارایی سیستم بازشناسی می‌پردازد. در فصل دوم همچنین روش‌های مقاوم کردن سیستم بازشناسی در برابر نویز دسته‌بندی شده و مشخصات هر دسته توضیح داده شده است. فصل سوم این پایان‌نامه درباره ویژگی‌های مقاوم در برابر نویز بحث می‌کند. در این فصل چند ویژگی مقاوم در برابر نویز را معرفی شده و نتیجه بکارگیری بعضی از آنها گزارش شده است. در فصل چهارم، چند روش مختلف بهبود گفتار معرفی شده و در آزمایشات مختلف، نتیجه اعمال آنها بر سیگنال گفتار نویز بررسی شده است. فصل پنجم پایان‌نامه روش‌های ارائه در فصول قبل را باهم مقایسه کرده و چند پیش‌پردازنده برای سیستم بازشناسی گفتار پایه پیشنهاد می‌دهد. فصل ششم به نتیجه‌گیری و تحقیقات آتی اختصاص دارد.

## فصل اول

### سیستم‌های بازشناسی گفتار

بازشناسی گفتار با توجه به کاربردهای وسیع آن در ارتباطات، تبادل اطلاعات میان انسان و ماشین و بکارگیری ماشین در ترجمه مکالمات از یک زبان به زبان دیگر، مورد توجه قرار گرفته است. در این فصل مسائل بازشناسی گفتار همچون پیوسته یا گسسته بودن، وابسته یا مستقل از گوینده بودن، اندازه کتاب لغت، محدودیتهای زبان و محیط و نوع بیان گفتار بررسی شده‌اند. همچنین اجزای یک سیستم بازشناسی و مدل مخفی مارکف به عنوان جزئی از این سیستم بازشناسی مورد بررسی قرار گرفته‌اند و در انتهای فصل یک سیستم بازشناسی پایه معرفی شده است.

### ۱-۱ پارامترهای بازشناسی گفتار

پارامترهایی مختلفی در یک سیستم بازشناسی گفتار موثر هستند که تعیین کننده درجه پیچیدگی سیستم می‌باشند. این پارامترها عبارتند از: وابسته و مستقل بودن از گوینده، بازشناسی کلمات مجزا و گفتار پیوسته، اندازه کتاب لغت، محدودیتهای زبانی، گفتار مکالمه‌ای و شرایط محیطی که بازشناسی در آن انجام می‌گیرد. در این بخش این پارامترها به اختصار مورد بررسی قرار می‌گیرند.

#### ۱-۱-۱ وابسته یا مستقل از گوینده

یک سیستم وابسته به گوینده فقط برای استفاده یک گوینده طراحی می‌شود، درحالی‌که یک سیستم مستقل از گوینده برای استفاده بوسیله هر گوینده‌ای طراحی می‌گردد. بطور معمول سیستمهای وابسته به گوینده دقیق‌تر از یک سیستم مستقل از گوینده هستند و نتایج بهتری را ارائه می‌دهند. ضعف عمده سیستم وابسته به گوینده این است که هر بار نیاز به بازشناسی گفتار گوینده جدیدی باشد، لازم است سیستم مجدداً برای این گوینده آموزش ببیند. سیستم میانه سیستمهای وابسته و مستقل از گوینده، سیستم چند گوینده است که برای تعداد ثابت و کم گوینده‌ها بکار می‌رود.

#### ۱-۱-۲ گفتار مجزا / متصل / پیوسته

در بازشناسی کلمات مجزا<sup>۱</sup>، هر کلمه بصورت جداگانه و واضح بیان می‌شود و سیستم بازشناسی با هر کلمه بطور مستقل سروکار دارد. در بازشناسی کلمات متصل<sup>۲</sup>، دنباله‌ای از کلمات برای بازشناسی مورد توجه قرار می‌گیرند، ولی کلمات جمله باید بطور مجزا و با فواصل زمانی سکوت از هم جدا شوند. در بازشناسی گفتار

<sup>۱</sup> - Isolated word recognition

<sup>۲</sup> - Connected word recognition