

## طبقه‌بندی ویژگی‌های گفتار مقاوم به نوفه در سامانه تعیین هویت گوینده

محدثه میربیگی<sup>۳،۱</sup>، امین‌اله مه‌آبادی<sup>۳،۱\*</sup>، اکبر رنجبر<sup>۲،۲</sup>

۱. گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه شاهد

۲. گروه مهندسی الکترونیک، دانشکده فنی مهندسی، دانشگاه شاهد

۳. مرکز تحقیقات آکوستیک، دانشگاه شاهد

دریافت: ۱۳۹۹/۰۶/۲۰، پذیرش: ۱۴۰۰/۰۴/۳۱

### چکیده

تشخیص خودکار هویت گوینده کاربردهای وسیعی در سامانه‌های صنعتی و امنیتی دارد و وابسته به ویژگی علامت گفتار است. کاربرد ماتریس ویژگی در شناسایی بی‌درنگ گوینده بسیار مهم و وجود نوفه محیطی و پردازشی منجر به تغییر مشخصات ویژگی‌ها و تولید خطا در تعیین هویت است. افزایش دقت در تشخیص هویت به فرایند حذف نوفه برای تعیین صحیح ویژگی‌های انرژی، آنتروپی انرژی، نرخ عبور از صفر، مرکز ثقل طیفی، گسترش طیفی، آنتروپی طیفی، شار طیفی، و رل آف طیفی از علامت گوینده نیاز دارد. در طراحی الگوریتم‌های بی‌درنگ و قابل اعتماد، فرایندهای مهم استخراج صحیح گفتار، شناسایی میزان حساسیت و سنجش میزان مقاومت مؤلفه‌های علامت برای حذف نوفه و بهبود کیفیت گفتار در بهبود علامت به نوفه نقش اساسی دارند. هدف اصلی این مقاله ارائه روش طبقه‌بندی ویژگی‌های علامت گفتار جهت طراحی الگوریتم‌های بی‌درنگ تعیین هویت گوینده و مقاوم به نوفه با سنجش میزان مقاومت آن است. روش پیشنهادی حذف نوفه از ماسک دودویی با ویژگی مقاوم مشخص بهره می‌برد و نتایج تجربی آزمایش‌ها روی داده‌ها، بهبود علامت به نوفه ۲ الی ۳ دسی‌بل را نشان می‌دهد. ارزیابی ماتریس ویژگی در سامانه تشخیص هویت از ضریب کپسترال بسامدمدل و ضریب پیشگویی خطی و ضریب کپستروم تشکیل شده که با روش فاصله‌یابی اقلیدسی در مجموعه داده‌های استاندارد ارزیابی شده است. روش پیشنهادی با وجود داده‌های نوفه‌ای توانسته قدرت تشخیص تعیین هویت گوینده را به دقت بالای ۸۰ درصد و سرعت بی‌درنگ افزایش دهد.

**کلیدواژه‌ها:** طبقه‌بندی ویژگی‌های گفتار، تعیین هویت گوینده، مقاوم به نوفه، علامت گفتار، نوفه.

### ۱. مقدمه

گوینده در حضور نوفه برای طراحی الگوریتم‌های مقاوم به نوفه بسیار مهم است. طبقه‌بندی<sup>۲</sup> و استخراج این ویژگی‌ها، امر مهمی در طراحی الگوریتم‌های استخراج گفتار مقاوم به نوفه محسوب می‌شود. از این الگوریتم‌ها در طراحی دستگاه‌های ارتباطی و سمعک‌ها استفاده می‌شود تا شنوندگان ضمن حذف صداهای مزاحم بتوانند بهتر صداهای بهبودیافته را بشنوند [۲]. از دیگر کاربردهای حذف نوفه در طراحی سامانه تعیین هویت<sup>۳</sup> است که همواره در صدد یافتن ویژگی‌های خاص برای هر شخص است [۳]. سامانه رایانه‌ای که هویت هر فرد را براساس مشخصات صوتی او تشخیص می‌دهد، سامانه شناسایی

گفتار مهم‌ترین پروتکل ارتباطی در زندگی روزمره بشر است که در حوزه پردازش علامت‌های گفتار قرار دارد و دارای ویژگی‌های مهم است. نرم‌افزارهای یاخته‌ای و ماهواره‌ای ارتباطی در بین مردم محبوبیت بسیار دارند که هر روز در جستجوی نرم‌افزارهای پیشرفته‌تر هستند. محققان در چهار ویژگی پیچیدگی، کیفیت، نرخ‌بیت و تأخیر رمزگذاری گفتار در حال تحقیق هستند [۱]. در شرایط عادی گفتار توسط نوفه<sup>۱</sup> تنکشی (تضعیف) می‌شود و کیفیت علامت اصلی را ندارد. لذا بررسی ویژگی‌های

<sup>۲</sup> Classification

<sup>۳</sup> Authentication

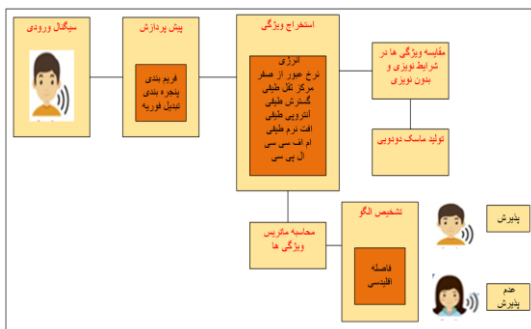
\* نویسنده پاسخگو: mahabadi@shahed.ac.ir

<sup>۱</sup> Noise

صفر<sup>۶</sup>، مرکز ثقل طیفی<sup>۷</sup>، گسترش طیفی<sup>۸</sup>، آنتروپی طیفی<sup>۹</sup>، شار طیفی<sup>۱۰</sup> و افت نرم طیفی<sup>۱۱</sup> از علامت گوینده نیازمند است. در طراحی الگوریتم‌های خودکار بی‌درنگ<sup>۱۲</sup> و قابل اعتماد و مقاوم، استخراج گفتار، شناسایی میزان حساسیت و سنجش میزان مقاومت مؤلفه‌های علامت برای حذف نوفه و بهبود کیفیت گفتار نقش اساسی در بهبود علامت به نوفه دارد. در این مقاله تمامی ویژگی‌های حوزه زمان و بسامد مورد بررسی قرار می‌گیرند تا میزان مقاومت‌شان در حضور نوفه‌های مختلف تولیدی از قطار، هواپیما، خیابان، ماشین، ازدحام و همه‌همه سنجش شوند تا از علامت خروجی مقاوم به نوفه در سامانه تشخیص هویت استفاده شود.

### ۱-۲. روش پیشنهادی و ارزیابی

بلوک دیاگرام روش پیشنهادی تعیین هویت گوینده مقاوم به نوفه، در شکل ۱ نمایش داده شده است.



شکل ۱ بلوک دیاگرام روش پیشنهادی طبقه‌بندی ویژگی‌های گفتار.

این روش دارای گام‌های پیش‌پردازش، استخراج ویژگی، بررسی ویژگی‌ها در حضور و غیاب نوفه، محاسبات ماتریس ویژگی، یافتن ماسک دودویی<sup>۱۳</sup> برای بهبود علامت، تشخیص الگو، و تعیین هویت گوینده است. ابتدا علامت ورودی پنجره‌بندی و ویژگی‌های انرژی، نرخ عبور از صفر،

گوینده خودکار نام دارد. چالش‌های مرتبط با پردازش گفتار عبارت از فهم انسان از گفتار در مقایسه با سامانه تشخیص هویت، زبان بدن، نوفه، زبان گفتار در مقابل زبان نوشتار، گفتار به هم پیوسته، تغییرپذیری مجرا (کانال)، تغییرپذیری گوینده، فهم، الگوهای گفتار، جنسیت گوینده، ساختمان تارهای صوتی، سرعت گفتار، گویش‌های اجتماعی و منطقه‌ای و ابهام است [۴]. از مهم‌ترین چالش‌های پردازش گفتار، تنکش (تضعیف) و حذف نوفه است. حذف نوفه در یک سامانه تعیین هویت بسیار مهم و مقاومت در برابر نوفه برای سامانه قابل اعتماد<sup>۱</sup> لازم است [۵-۷]. در این مقاله بررسی علامت گفتار، نقش نوفه و میزان تنکش و حذف آن در فرآیند تعیین هویت مورد بررسی قرار می‌گیرد تا با حذف نوفه بهترین تصمیم‌سازی نسبت به ویژگی‌های گفتار اخذ شود [۸].

### ۱-۱. استخراج ویژگی گفتار

صداهاى مختلف از طریق ارتعاش اجسام در هوا مخابره می‌شوند و در بسامد بین ۲۰ هرتز تا ۲۰ کیلوهرتز برای انسان قابل شنیدن هستند. گفتار نوعی موج صوتی است که جهت برخورداری از وضوح شنوایی بیش‌تر در نرم‌افزارهای مختلف مورد تحلیل و بررسی قرار می‌گیرد. پردازش گفتار در حوزه رمزگذاری گفتار مربوط به ارتباطات صوتی انسان به انسان، ترکیب گفتار مرتبط با ارتباطات ماشین با انسان و تشخیص گفتار در رابطه با ارتباط انسان با ماشین است [۱]. تشخیص خودکار هویت گوینده کاربردهای وسیعی در صنعت بشری و امنیت افراد دارد و نیازمند استخراج ویژگی علامت گفتار است. کاربرد ماتریس ویژگی در شناسایی بی‌درنگ گوینده بسیار مهم است.

وجود نوفه محیطی<sup>۲</sup> و پردازشی<sup>۳</sup> منجر به تغییر مشخصات ویژگی‌ها و تولید خطا در تعیین هویت می‌شود. افزایش دقت در تشخیص هویت به فرآیند حذف نوفه برای تعیین صحیح ویژگی‌ها به انرژی<sup>۴</sup>، آنتروپی انرژی<sup>۵</sup>، نرخ عبور از

<sup>5</sup> Entropy energy

<sup>6</sup> Zero crossing rate

<sup>7</sup> Spectral centroid

<sup>8</sup> Spectral spread

<sup>9</sup> Spectral entropy

<sup>10</sup> Spectral flux

<sup>11</sup> Spectral roll off

<sup>12</sup> Real time

<sup>13</sup> Binary mask

<sup>1</sup> Robust

<sup>2</sup> Environmental

<sup>3</sup> Processing

<sup>4</sup> Energy

شده است [۱۳]. در پژوهش دیگری نیز الگویی برای استخراج ویژگی‌های مقاوم آرایه گردیده است [۱۴]. روش پیشنهادی حذف نوفه پژوهش حاضر از ماسک دودویی با ویژگی مقاوم مشخص بهره می‌برد که نتایج تجربی آزمایش‌ها روی داده‌ها، بهبود علامت به نوفه ۲ الی ۳ دسی‌بل را نشان می‌دهد. این روش طبقه‌بندی ویژگی‌های علامت گفتار جهت طراحی الگوریتم‌های بی‌درنگ تعیین هویت گوینده و مقاوم به نوفه با سنجش میزان مقاومت آرایه داده است. ارزیابی ماتریس ویژگی آن در سامانه تشخیص هویت از ضریب کپسترال بسامدمل و ضریب پیشگویی خطی و ضریب کپستروم تشکیل شده که با روش فاصله‌یابی اقلیدسی ارزیابی شده است. روش پیشنهادی با وجود داده‌های نوفه‌ای و بدون پیش پردازش‌های سنتی توانسته قدرت تشخیص تعیین هویت گوینده را به دقت بالای ۸۰ درصد و سرعت بی‌درنگ افزایش دهد.

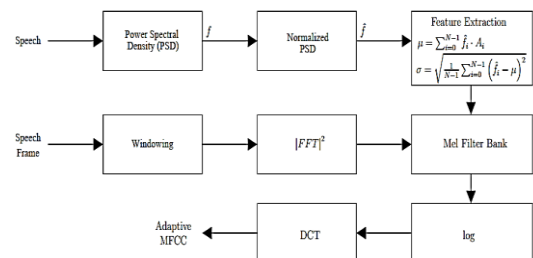
### ۳-۱. نوآوری و ساختار مقاله

تفکر اصلی روش، بیان ایده مبتنی بر آرایه طبقه‌بندی ویژگی‌های مهم برای هر علامت ورودی، لزوم داشتن ابزار لازم برای مقاوم‌سازی سامانه به ازای شرایط علامت در تهدیدات مختلف است. ممکن است در یک سامانه و در یک برهه زمانی، میزان تهدید نوفه در یک علامت در آن برهه زمانی نسبت به همان علامت در برهه زمانی دیگر متفاوت باشد. لذا یک الگوریتم به ازای علامت‌های ورودی مختلف در شرایط متفاوت جواب نادرست می‌دهد. به دلیل لزوم استخراج دانش لازم از بردار ویژگی برای تصمیم‌گیری پردازش هر علامت، سامانه باید با توجه به تهدیدات مهم در آن شرایط و برهه زمانی مقاوم باشد و با ارزیابی ویژگی‌ها و مقاوم‌سازی همان تهدید، تصمیم صحیح بگیرد. با این توجه، دیدگاه مقاوم‌سازی هر الگوریتم در شرایط متفاوت می‌تواند تغییر کند. این نظر روی سامانه تعیین هویت گوینده برای برآورد میزان مقاوم‌سازی در برابر نوفه جهت تشخیص بهتر، آزمون شده و نتایج خوبی آرایه داده است.

ساختار مقاله به این شرح است که در ابتدا در بخش ۲ به شرح جزئیات روش پیشنهادی می‌پردازد. در بخش ۳ نتایج آزمایش‌های تجربی روی داده‌های مختلف و

آنتروپی انرژي، مرکز ثقل طیفی، گسترش طیفی، آنتروپی طیفی، شار طیفی، رل آف طیفی ضرایب کپسترال و پیشگویی خطی استخراج می‌شود. سپس این ویژگی‌ها در علامت‌های دارای نوفه و بدون نوفه با هم مقایسه می‌شوند. در نهایت از ویژگی که تغییر کم‌تری در برابر نوفه داشته باشد برای ماسک دودویی استفاده می‌شود. به منظور تشخیص هویت گوینده از سه ویژگی ضریب کپسترال بسامد مل<sup>۱</sup>، ضریب پیشگویی خطی و ضریب کپستروم استفاده شده است.

در طراحی الگوریتم خودکار بی‌درنگ قابل اعتماد، فرآیندهای مهم استخراج صحیح گفتار، شناسایی میزان حساسیت و سنجش میزان مقاومت مؤلفه‌های علامت برای حذف نوفه و بهبود کیفیت گفتار وجود دارد که در بهبود علامت به نوفه نقش اساسی دارند. طبقه‌بندی ویژگی‌های علامت گفتار آرایه شده برای طراحی الگوریتم بی‌درنگ پیشنهادی باعث مقاومت آن به نوفه در تعیین هویت گوینده است. در گذشته از یک بانک صافی (فیلتر) تطبیقی برای یافتن این ویژگی استفاده شده تا بتواند به تشخیص هویت گوینده پردازد که شکل ۲ نمایانگر این روش است [۹].



شکل ۲ بلوک دیاگرام روش تطبیقی یافتن ضریب مل [۱۶].

در تحقیقی طبقه‌بندی مبتنی بر ماشین بردار پشتیبان آرایه شده که به طبقه‌بندی گفتار از موسیقی پرداخته است [۱۰]. در پژوهش دیگر گفتار با استفاده از استخراج ویژگی مورد بررسی قرار گرفته است [۱۱]. یک روش استخراج ویژگی مقاوم برای تشخیص هویت گوینده آرایه شده است که می‌تواند تمامی اطلاعات گفتار را شامل شود [۱۲]. برای شناسایی زبان نیز از استخراج ویژگی استفاده

<sup>۱</sup> MEL

نمی‌گذارد [۶]. برای اینکه علامت ایستان<sup>۴</sup> باشد و مشخصه‌های علامت صحبت تقریباً ثابت باقی بماند، علامت گفتار به قاب‌های ۲۰ تا ۳۰ میلی‌ثانیه تقسیم شده، ویژگی‌ها از هر قاب استخراج می‌شوند. قاب‌ها معمولاً به‌نحوی انتخاب می‌شوند که با یکدیگر هم‌پوشانی داشته باشند. میزان هم‌پوشانی معمولاً ۱۰ تا ۱۵ میلی‌ثانیه انتخاب می‌گردد. هر چه طول قاب انتخابی کوچک‌تر باشد، تعداد بردارهای ویژگی بیش‌تر و در نتیجه هزینه محاسبات بالاتر می‌رود. با افزایش طول قاب‌ها، تعداد بردارهای ویژگی<sup>۵</sup> کم‌تر و در نتیجه حجم محاسبات کم می‌شود ولی اطلاعات کم‌تری از علامت گفتار در اختیار قرار می‌گیرد.

## ۲-۱-۲. پنجره‌گذاری

در این مرحله هر قاب به‌طور جداگانه در یک پنجره ضرب می‌شود تا اثر ناپیوستگی علامت در ابتدا و انتهای هر قاب کم شود [۶]. انتخاب پنجره بسیار مهم است زیرا حاشیه‌های یک قاب در کم یا زیاد شدن علامت خطا مؤثرند. لذا باید از پنجره‌ای استفاده شود که حاشیه‌های قاب را به‌طور یکنواخت باریک کند. اگر پنجره با  $W(n)$  نمایش داده شوند، اعمال پنجره مطابق رابطه (۲) خواهد بود. پنجره‌های همینگ<sup>۶</sup> و هنینگ<sup>۷</sup> دو نمونه از پنجره‌هایی هستند که معمولاً در این قبیل کاربردها استفاده می‌شوند.

$$\bar{X}(n) = X_k(n)W(n) \quad 0 \leq n \leq N-1 \quad (2)$$

$$\text{Hamming} : W(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N-1}, 0 \leq n \leq N-1$$

$$\text{Hanning} : W(n) = \frac{1}{2} \left( 1 - \cos \frac{2\pi n}{N-1} \right), 0 \leq n \leq N-1$$

## ۲-۲. استخراج ویژگی

پس از قطعه‌بندی<sup>۸</sup> اجزای سازنده گفتار باید ویژگی‌هایی را از علامت صحبت استخراج کرد که در مرحله تطبیق الگو استفاده شوند. علامت گفتار ویژگی‌های زیادی دارد که عموماً با طیف لحظه‌ای علامت گفتار یا شکل مجرای

نوفه‌های متفاوت بیان و در بخش ۴ بحث می‌گردد. نهایتاً در بخش ۵ به نتیجه‌گیری و کارهای آینده می‌پردازد.

## ۲. مواد و روش‌ها

گام‌های روش پیشنهادی در شکل ۱ بیان شد. ابتدا علامت ورودی قاب‌بندی<sup>۱</sup>، پنجره‌بندی و سپس ویژگی‌های اساسی انرژی، نرخ عبور از صفر، آنتروپی انرژی، مرکز ثقل طیفی، گسترش طیفی، آنتروپی طیفی، شار طیفی، رل آف طیفی، ضرایب کپسترال و پیشگویی خطی استخراج می‌شود. این ویژگی‌ها در ارزیابی علامت‌های نوفه‌ای و بدون نوفه با هم مقایسه می‌شوند. در نهایت از ویژگی که تغییر کم‌تری در برابر نوفه داشته باشد برای ماسک دودویی جهت حفظ بخش مفید علامت استفاده می‌شود. به منظور تشخیص هویت از سه ویژگی ضریب کپسترال بسامد مل و ضریب پیشگویی خطی و ضریب کپستروم<sup>۲</sup> برای شناسایی گوینده استفاده شده است. ماسک دودویی نیز براساس ویژگی مورد نظر، بخش مورد نظر علامت را حذف و بخش دیگر را حفظ می‌کند. یکی از ماسک‌های متداول ماسک دودویی است که برای تخمین این هدف در رابطه (۱) بیان می‌شود. توان علامت گفتار و نوفه در واحد دسی‌بل  $S_{ab}(t,f)$  و  $N_{ab}(t,f)$  می‌باشد و  $LC$  حد آستانه‌ای را برای الگوی ماسک در نظر می‌گیرد [۵].

$$IBM(t, f) = \begin{cases} 1 & f S_{ab}(t, f) > N_{ab}(t, f) + LC \\ 0 & otherwise \end{cases} \quad (1)$$

## ۲-۱-۲. مراحل پیش‌پردازش

به‌منظور استخراج بردارهای ویژگی باید پردازش‌هایی روی علامت انجام شود [۲]. این پردازش عبارت از قاب‌بندی، پنجره‌گذاری<sup>۳</sup> و تبدیل فوریه زمان کوتاه است که در این بخش شرح داده می‌شوند.

## ۲-۱-۲. قاب‌بندی

به دلیل تغییر مشخصه‌های علامت گفتار در طول زمان و به عبارتی ناپیوستگی بودن آن‌ها، استخراج ویژگی از یک حوزه‌ی زمانی نسبتاً بزرگ، اطلاعات معتبری را در اختیار

<sup>4</sup> Stationary signal

<sup>5</sup> Feature vectors

<sup>6</sup> Hamming

<sup>7</sup> Hanning

<sup>8</sup> Segmentation

<sup>1</sup> Framing

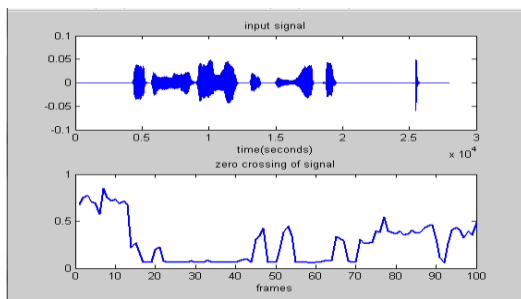
<sup>2</sup> Cepstrum coefficient

<sup>3</sup> Windowing

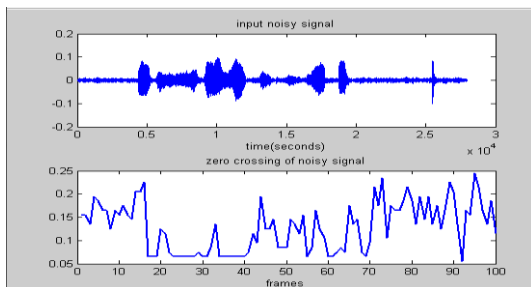
علامت در طول زمان یک قاب است. به عبارت دیگر تعداد دفعاتی که مقدار علامت از مثبت به منفی یا از منفی به مثبت تغییر می‌کند را نرخ عبور از صفر گویند [۶]. رابطه (۵) نرخ عبور از صفر را بیان می‌کند. تابع  $\text{sgn}[x_i(n)]$  تابع علامت است که با رابطه (۶) تعریف می‌شود. شکل ۴ نرخ عبور از صفر علامت بدون حضور نوفه را نشان می‌دهد. نرخ عبور از صفر در قسمت‌های نوفه‌ای بیش‌تر می‌شود که در شکل ۵ مشخص است.

$$Z(i) = \frac{1}{W_L} \sum_{n=1}^{W_L} |\text{sgn}[x_i(n)] - \text{sgn}[x_i(n-1)]| \quad (5)$$

$$\text{sgn}[x_i(n)] = \begin{cases} 1 & x_i(n) \geq 0 \\ -1 & x_i(n) < 0 \end{cases} \quad (6)$$



شکل ۴ نرخ عبور از صفر علامت گفتار بدون حضور نوفه.



شکل ۵ نرخ عبور از صفر علامت گفتار در حضور نوفه.

### ۲-۲-۳. آنتروپی انرژی

آنتروپی انرژی، معیاری برای اندازه‌گیری تغییرات ناگهانی انرژی علامت صوتی طبق رابطه (۷) است. برای محاسبه آن لازم است که هر قاب کوتاه‌مدت به  $K$  زیرقاب‌هایی با طول مشخص تقسیم شود. برای هر زیرقاب انرژی محاسبه و بر انرژی کل قاب کوتاه‌مدت تقسیم می‌شود. عملگر تقسیم یک عمل استاندارد است که میانگین مقادیر انرژی زیرقاب را به عنوان یک توالی احتمالی در نظر می‌گیرد [۶]. رابطه (۸) آنتروپی انرژی را نشان می‌دهد و

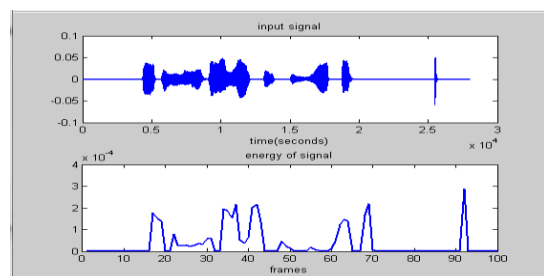
گفتار مرتبط هستند. پردازش همه ویژگی برای کاربردهای بازشناسی گفتار، عملی نخواهد بود لذا تبدیلاتی روی علامت گفتار انجام می‌شود تا بتوان ویژگی‌های مهم را استخراج کرد. استخراج ویژگی به دو دلیل بهبود شباهت و کاهش داده‌ها انجام می‌گیرد. این کار اولاً سبب تمرکز روی اطلاعات موجود در علامت می‌شود و منجر به بهبود میزان شباهت و عدم‌شباهت میان دسته‌های مختلف می‌گردد. ثانیاً حجم داده‌ها را به نحو قابل ملاحظه‌ای کاهش داده و بار محاسبات را کاهش می‌دهد [۶]. هشت ویژگی مهم حوزه زمان و بسامد یعنی سه ویژگی در حوزه زمان و پنج ویژگی در حوزه بسامد از قبیل انرژی، انرژی، نرخ عبور از صفر، مرکز ثقل طیفی، گسترش طیفی، آنتروپی طیفی، شار طیفی، و رل آف طیفی مورد بررسی قرار می‌گیرند. سپس رویکرد مهم حذف نوفه توسط روش پیشنهادی ارائه می‌شود.

### ۲-۲-۱. انرژی

اگر  $x_i(n)$  که  $n = 1, \dots, W_L$  توالی نمونه‌های  $i$ -امین قاب علامت و  $W_L$  طول قاب باشد، انرژی علامت از رابطه (۳) محاسبه می‌شود [۶]. انرژی را با تقسیم بر طول قاب می‌توان بهنجار (نرمال) کرد تا به طول قاب وابسته نباشد لذا رابطه (۳) به رابطه (۴) تبدیل می‌شود. رابطه (۴) توان علامت نامیده می‌شود و شکل ۳ انرژی علامت را نشان می‌دهد.

$$E(i) = \sum_{n=1}^{W_L} (x_i(n))^2 \quad (3)$$

$$E(i) = \frac{1}{W_L} \sum_{n=1}^{W_L} |x_i(n)|^2 \quad (4)$$



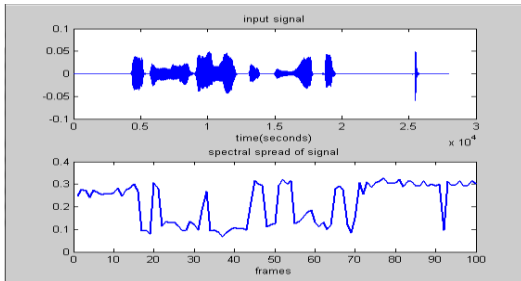
شکل ۳ انرژی علامت گفتار.

### ۲-۲-۲. نرخ عبور از صفر

نرخ عبور از صفر قاب‌های علامت صوتی برابر با نرخ تغییر

ویژگی آن را نشان می‌دهد که این ویژگی بیان‌کننده چگونگی توزیع اسپکترا از مرکز طیفی علامت است.

$$S_i = \sqrt{\frac{\sum_{k=1}^{W_{iL}} (k - C_i)^2 X_i(k)}{\sum_{k=1}^{W_{iL}} X_i(k)}} \quad (11)$$



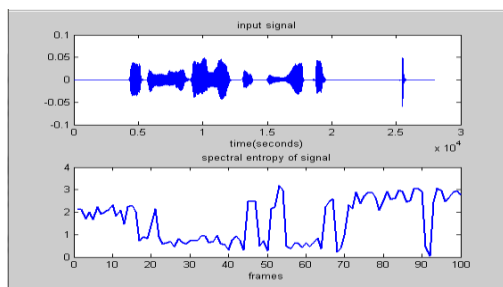
شکل ۸ گسترش طیفی علامت گفتار.

### ۶-۲-۲. آنتروپی طیفی

آنتروپی طیفی مانند آنتروپی انرژی است ولی محاسبات آن در حوزه بسامد انجام می‌شود. در ابتدا اسپکترا قاب‌های کوتاه‌مدت به  $L$  زیرقسمت تقسیم می‌شوند. انرژی برای  $E_f$  مربوط  $f$ -آمین زیرقسمت که  $f = 0, \dots, L-1$  است برای بهنجارسازی (نرمال‌سازی)<sup>۱</sup> به انرژی طیفی تقسیم می‌شود که رابطه (۱۲) آن را بیان می‌کند [۶]. آنتروپی انرژی با رابطه (۱۳) محاسبه می‌شود و شکل ۹ آنتروپی طیفی علامت گفتار را نشان می‌دهد.

$$n_f = \frac{E_f}{\sum_{f=0}^{L-1} E_f} \quad (12)$$

$$H = -\sum_{f=0}^{L-1} n_f \cdot \log_2 n_f \quad (13)$$



شکل ۹ آنتروپی طیفی علامت گفتار.

### ۷-۲-۲. شار طیفی

شار طیفی، اندازه‌گیری تغییرات طیفی بین دو قاب پشت

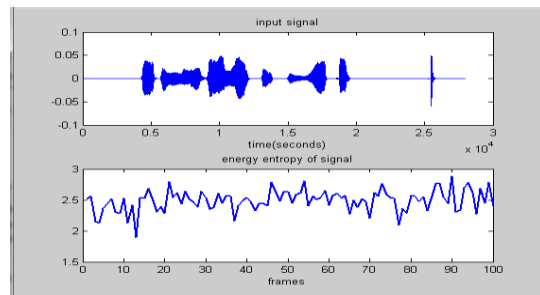
<sup>1</sup> Normalization

$K = 1 \dots z$  در نظر گرفته می‌شود. در مرحله آخر آنتروپی هر  $e_j$  با رابطه (۹) محاسبه می‌شود. مقادیر آنتروپی، زمانی کم می‌شود که تغییرات انرژی به سرعت کاهش یابد. شکل ۶ آنتروپی انرژی علامت را نشان می‌دهد.

$$e_j = \frac{E_{subFrame_j}}{E_{subFrame_i}} \quad (7)$$

$$E_{subFrame_i} = \sum_{k=1}^K E_{subFrame_k} \quad (8)$$

$$H(i) = -\sum_{j=1}^K e_j \cdot \log_2(e_j) \quad (9)$$

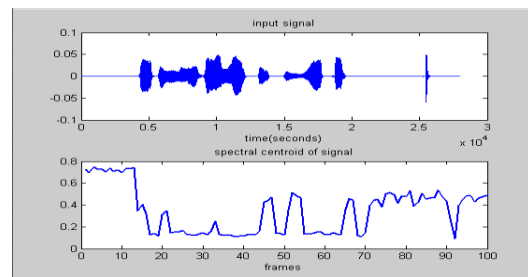


شکل ۶ آنتروپی انرژی علامت گفتار.

### ۴-۲-۲. مرکز ثقل طیفی

مرکز ثقل طیفی معیاری برای اندازه‌گیری شکل طیفی علامت است. مرکز ثقل طیفی مرکز ثقل اسپکترا علامت می‌باشد. رابطه (۱۰) مقدار مرکز ثقل قاب  $i$ -ام را محاسبه می‌کند [۶]. شکل ۷ مرکز ثقل طیفی علامت گفتار را نشان می‌دهد.

$$C_i = \frac{\sum_{k=1}^{W_{iL}} k X_i(k)}{\sum_{k=1}^{W_{iL}} X_i(k)} \quad (10)$$



شکل ۷ مرکز ثقل طیفی علامت گفتار.

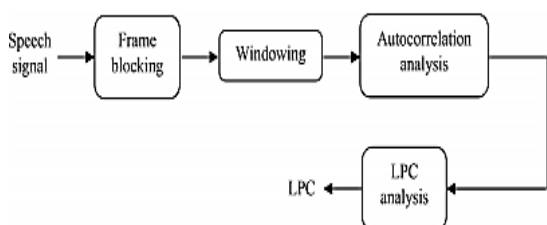
### ۵-۲-۲. گسترش طیفی

گسترش طیفی معیاری برای تعیین موقعیت طیفی علامت است. مرکز دوم لحظه‌ای اسپکترا است و برای محاسبه این ویژگی از رابطه (۱۱) استفاده می‌شود [۶]. شکل ۸

### ۲-۲-۹. ضریب پیش‌گویی خطی

پیشگویی خطی یکی از مشهورترین فن‌های سنتز گفتار است که به دلیل توانایی آن در نمایش کارای شکل موج گفتار برحسب مؤلفه‌های متغیر با زمان، در بسیاری از کاربردهای کدگذاری گفتار به کار می‌رود [۲]. فن‌های پیشگویی خطی از نظر محاسباتی بسیار سریع هستند و در تخمین مؤلفه‌های علامت گفتار همانند دوره اصلی تناوب و فرمت‌ها، ابزاری ضروری محسوب می‌شوند. ایده اصلی روش، مبتنی بر این نظریه است که می‌توان طیف علامت گفتار را با پاسخ بسامدی یک صافی (فیلتر) تمام قطب تقریب زد و ضرایب این صافی را به‌عنوان بردار مشخصه آن مورد استفاده قرار داد. در تحلیل پیشگویی خطی، یک نمونه از علامت گفتار در یک زمان مشخص، با ترکیب خطی تعدادی از نمونه‌های قبلی آن تخمین زده می‌شود. اگر  $s(n)$  نمونه‌ای در زمان  $n$  باشد می‌توان آن را با  $p$  نمونه قبلی‌اش به صورت رابطه (۱۷) تقریب زد. این ضریب با استفاده از پنجره‌بندی و محاسبه خود هم‌بستگی قابل محاسبه و در شکل ۱۲ آمده است.

$$s(n) = \sum_{k=1}^p a_k s(n-k) \quad (17)$$



شکل ۱۲ روش محاسبه ضریب پیش‌گویی خطی [۵].

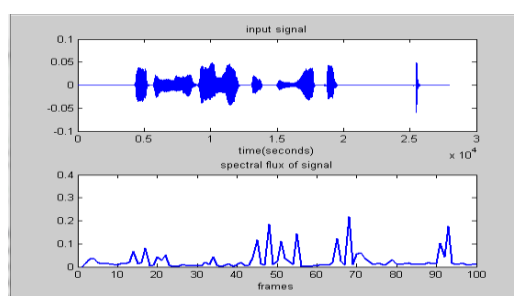
### ۲-۲-۱۰. ضریب کپسترال

ضرایب کپسترال از مشخصه‌هایی است که از علامت گفتار استخراج و در بسیاری کاربردها مورد استفاده قرار می‌گیرند [۲]. این ضرایب نه تنها اطلاعات صافی (فیلتر) مجرای گفتار را در خود دارند بلکه حاوی اطلاعات علامت تحریک نیز هستند. لذا مشخصه‌های مناسبی را برای تشخیص گفتار در اختیار قرار می‌دهند. تحلیل کپسترال به دو دسته کپستروم حقیقی و کپستروم مختلط تقسیم می‌شود. تفاوت کپستروم حقیقی و مختلط این است که تحلیل کپستروم مختلط، حاوی اطلاعات فاز علامت

سر هم است. این محاسبه توسط مربعات تفارق بزرگی اسپکترام دو قاب متوالی انجام می‌شود. رابطه (۱۴) این ویژگی را محاسبه می‌کند که در آن  $EN_i(k)$ ،  $k$ -امین ضریب تبدیل فوریه مربوط به  $i$ -امین قاب می‌باشد و در رابطه (۱۵) آمده است [۶]. شکل ۱۰ شار طیفی علامت گفتار را نشان می‌دهد.

$$Fl_{(i,i-1)} = \sum_{k=1}^{W_{fl}} (EN_i(k) - EN_{i-1}(k))^2 \quad (14)$$

$$EN_i(k) = \frac{X_i(k)}{\sum_{l=1}^{W_{fl}} X_i(l)} \quad (15)$$

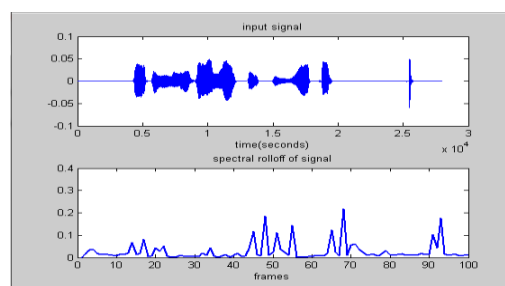


شکل ۱۰ شار طیفی علامت گفتار.

### ۲-۲-۸. افت نرم طیفی

این ویژگی با عنوان توزیع بزرگی اسپکترام زیر بسامد مشخص را تعیین می‌کند.  $m$ -امین ضریب تبدیل فوریه مربوط به  $i$ -امین قاب برای این ویژگی را تعیین می‌کند. رابطه (۱۶) افت نرم طیفی را بیان می‌دارد [۶]. درصد مشخص شده  $C$  توسط کاربر صورت می‌گیرد که این ویژگی به عنوان توصیف‌گر شکل طیفی به کار می‌رود و از آن برای تشخیص بخش‌های صحبت از بخش‌های بدون صحبت استفاده می‌گردد. شکل ۱۱ این ویژگی را نشان می‌دهد.

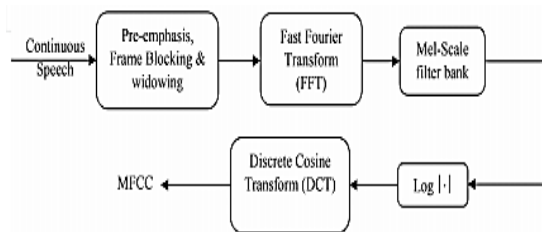
$$\sum_{k=1}^m X_i(k) = C \sum_{k=1}^{W_{fl}} X_i(k) \quad (16)$$



شکل ۱۱ افت نرم طیفی علامت گفتار.

کپستروم را به دست آورد. از این ویژگی برای تشخیص گوینده و سنتز گفتار استفاده می‌شود [۸].

$$mel(f) = 2595 \times \log_{10} \left( 1 + \frac{f}{700} \right) \quad (19)$$



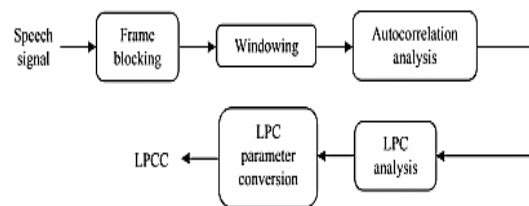
شکل ۱۴ روش محاسبه MFCC سی [۵].

کارهای متفاوتی در پردازش علامت برای حذف نوفه ارایه شده که رشد زیاد الگوریتم‌های پردازش رقومی و فن‌آوری‌های مرتبط با آن، انگیزه زیادی به ارایه نرم‌افزارهای اجرایی در دنیای واقعی داده است. پردازش‌گرهای علامت رقومی از نظر اندازه کوچک ولی توانایی پردازش آن‌ها به صورت نمایی افزایش یافته که در همین زمان مصرف توان این پردازشگرها کاهش داده شده است. لذا از این پردازشگرها در دستگاه‌های کوچک قابل حمل بسیاری مانند سمعک، هدفون و محافظ‌های گوش استفاده می‌شود که گرفتار نوفه برقی<sup>۳</sup> هستند. برای حذف نوفه برقی، دو روش متفاوت منفعل<sup>۴</sup> و فعال<sup>۵</sup> وجود دارد. روش منفعل در حذف نوفه برقی منفعل در سمعک و کاشت حلزون گوش مورد استفاده قرار می‌گیرد. در این روش علامت و نوفه محدود با استفاده از میکروفون ضبط شده‌اند و فن‌های به کار رفته در آن، تفارق طیفی و الگوریتم‌های کم‌ترین مربعات میانگین<sup>۶</sup> و صافی‌های مختلف طیفی هستند. شنونده فقط علامت بدون نوفه را می‌شنود. از فرض‌های مهم این روش آن‌که شنونده از محیط بیرون ایزوله شده است. این فرض در مکان‌هایی که نوفه محدود دارای دامنه زیاد باشد اعتبار ندارد. لذا از روش دوم یعنی حذف نوفه فعال استفاده می‌شود. در پژوهشی کار بهبود کیفیت گفتار با استفاده از شبکه‌های

<sup>2</sup> MFCC  
<sup>3</sup> Electrical noise  
<sup>4</sup> Passive  
<sup>5</sup> Active  
<sup>6</sup> Sum of square error

صحت نیز می‌باشد. ولی از آن‌جا که اطلاعات فاز در شنوایی انسان اهمیت کمی دارد معمولاً از تحلیل کپستروم حقیقی استفاده می‌شود. شکل ۱۳ بیانگر چگونگی محاسبه ضرایب کپسترال است. ضرایب کپسترال با استفاده از ضرایب پیشگویی خطی نیز محاسبه می‌شوند. اگر  $a_m$  ضرایب پیشگویی خطی باشد ضرایب صافی رابطه (۱۸) از مرتبه  $p$  ضرایب ال‌پی‌سی‌سی<sup>۱</sup> به صورت رابطه (۱۸) تعیین می‌شوند.

$$C_m = a_m + \sum_{k=1}^{m-1} \left[ \frac{k}{m} \right] c_k a_{m-k} \quad (18)$$



شکل ۱۳ روش محاسبه ضریب کپسترال [۵].

## ۲-۲-۱۱. ضریب کپسترال بسامد مل

یکی از راه‌های بهبود دقت تشخیص سامانه هنگامی که ضرایب کپسترال به عنوان بردارهای ویژگی به کار می‌روند، استفاده از تبدیل غیرخطی است که حساسیت گوش را نسبت به حوزه‌های مختلف بسامد الگو کند [۲]. مسأله قابل توجه این است که باندهای مختلف بسامد از نظر اطلاعات شنوایی، دارای ارزش‌های مختلفی هستند. خاصیت مهم مقیاس مل ارزش بیش‌تر دادن به اطلاعات حوزه پایین بسامد است. مل واحد ارزیابی صدای درک‌شده یا بسامد آهنگ است. نگاشتی بین مقیاس واقعی هرگز و مقیاس بسامد درک‌شده مل وجود دارد. این نگاشت برای بسامدهای کم‌تر از یک کیلوهرتز خطی و برای بسامدهای بیش‌تر از آن لگاریتمی است [۷]. مقیاس مل با رابطه (۱۹) تعریف می‌شود و روش محاسبه آن در شکل ۱۴ آمده است. برای محاسبه ضرایب مل کپستروم می‌توان لگاریتم انرژی کل در باندهای بحرانی حول بسامدهای مل را به عنوان ورودی تبدیل فوریه گسسته معکوس به کار گرفت و با محاسبه این تبدیل ضرایب مل

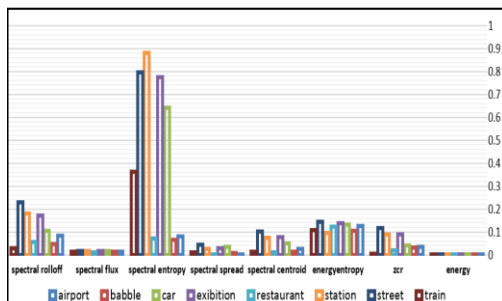
<sup>1</sup> LPCC



مورد آزمایش حاوی پردازنده کُرآی سِوِن<sup>۶</sup> با بسامد ۲/۵ گیگاهرتز و رم ۶ گیگابایت است.

### ۲-۲. آزمایش‌های تجربی

ابتدا هشت ویژگی صوتی انسان در علامت‌های حاوی نوفه و بدون آن استخراج که ویژگی‌ها براساس تغییرات در مقابل نوفه تقسیم‌بندی شده تا ویژگی‌های مقاوم به دست آیند. به این منظور، میانگین هشت ویژگی را در نه محیط مختلف بررسی و تغییرات میانگین مد نظر قرار می‌گیرد. سپس از ویژگی‌های مقاوم به نوفه در ماسک دودویی برای جداسازی استفاده می‌شود. از حد آستانه برای تقسیم‌بندی ویژگی‌ها در برابر نوفه استفاده می‌گردد. میزان تغییرات در حدود ۰/۱ به آن غیرمقاوم، در حدود ۰/۰۱ به آن نیمه‌مقاوم و اگر در حدود ۰/۰۰۱ باشد مقاوم در برابر نوفه گویند. طبقه‌بندی ویژگی‌ها براساس مقاومت در برابر نوفه در نظر گرفته شده است. ویژگی مقاوم، انرژی و شار طیفی است که در طراحی ماسک دودویی استفاده می‌شود. ویژگی‌های نیمه‌مقاوم عبارت از نرخ عبور از صفر، مرکز ثقل طیفی و گسترش طیفی است. ویژگی‌هایی که در برابر نوفه تغییرات زیادی دارند عبارت از آنتروپی طیفی، افت نرم طیفی و آنتروپی انرژی هستند. در آزمایش‌های انجام‌شده ویژگی‌های یک گوینده، در حضور نوفه‌های مختلف مورد بررسی قرار گرفته است. آزمون با داده‌های مختلف در حضور نوفه‌های مختلف فرودگاه، ماشین، قطار، ایستگاه، خیابان، رستوران، نمایشگاه و زمزمه‌های مختلف ارزیابی شده که هشت ویژگی اصلی آن در شکل ۱۵ نشان داده شده است.



شکل ۱۵ میزان تغییرات ویژگی یک گوینده در حضور نوفه‌های متفاوت.

عمیق عصبی انجام گرفته است [۱۵].

روش فعال به روش‌های الکترومکانیکی یا برق پژواک‌شناسی<sup>۱</sup> مرتبط است که حذف اختلالات صوتی برای رسیدن به محیط آرام‌تر را در پی خواهد داشت. در این روش، قاعده اصلی بر تعریف علامت «ضد نوفه»<sup>۲</sup> که دارای دامنه یکسان با گفتار ولی فاز دقیقاً مخالف است که کاربردهای زیادی در طراحی محافظ‌های گوش و هدفون دارد. الگوریتم‌های سنتی حذف نوفه فعال در باندهای بسامدی پایین‌تر بهتر کار می‌کنند. اگر پهنای باند و مرکز بسامد نوفه افزایش یابد کارایی به سرعت کاهش می‌یابد. بیش‌تر منابع نوفه تمایل به داشتن پهنای باند گسترده در طبیعت دارند. گرچه مقدار زیادی از انرژی در بسامدهای پایین متمرکز شده، تمایل به داشتن عنصر مؤثر در بسامد بالا نیز دارند. علاوه بر این که سامانه حذف نوفه فعال با دستگاه‌های ارتباطی و سامانه‌های صوتی ترکیب‌شده لازم و ضروری است که سامانه حذف نوفه وابسته به بسامد نیز برای جلوگیری از تأثیرات مخرب بر علامت وجود داشته باشد [۱۶-۱۷].

### ۳. نتایج

در این بخش محیط و تنظیمات آزمایش‌ها، داده‌های مورد آزمایش، ویژگی‌های مورد استفاده، روش آزمایش‌های تجربی و مشخصات علامت‌های خروجی ارایه می‌شود.

#### ۳-۱. تنظیمات و داده‌ها

به‌منظور طراحی ماسک برای بهبود کیفیت علامت ورودی از مقایسه ویژگی‌های استخراجی در محیط نوفه‌ای و بدون نوفه استفاده شده و با مجموعه داده‌ای آرورا<sup>۳</sup> آزمایش و ارزیابی گردیده است [۱۸]. این مجموعه داده‌ای به‌صورت تک‌مجرا (تک‌کانال) و با ۱۶ بیت پی‌سی‌ام<sup>۴</sup> ضبط شده است. برای شناسایی گوینده از مجموعه داده‌ای ای‌بی‌ال‌اس‌دی‌آر<sup>۵</sup> بهره‌برداری شده است [۱۹]. آزمایش‌های مذکور در محیط متلب ۲۰۱۴ انجام شده است. ماشین

<sup>۱</sup> Electroacoustic

<sup>۲</sup> Anti noise

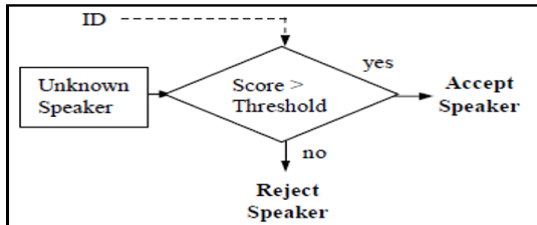
<sup>۳</sup> AURORA

<sup>۴</sup> PCM

<sup>۵</sup> ELSDSR

<sup>۶</sup> Corei7

صورت صفر فرض شده که در شکل ۱۶- ج مشخص است. میزان علامت به نوفه در علامت بهبود یافته با ماسک دودویی برابر  $4/52$  می‌باشد.

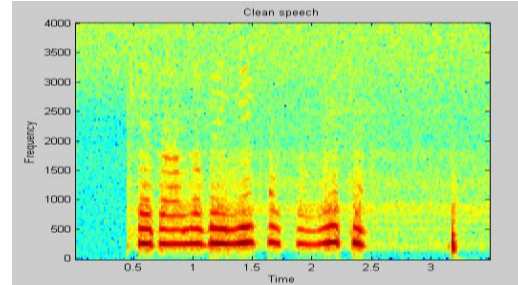


شکل ۱۷ شناسایی گوینده با استفاده از فاصله اقلیدسی.

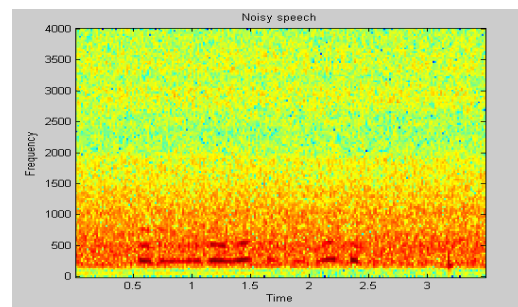
#### ۴. بحث

لازم است هر سامانه برای تصمیم‌گیری صحیح و پردازش درست علامت با وجود تهدیدات مهم در آن شرایط و برهه زمانی مقاوم باشد و با ارزیابی صحیح ویژگی‌ها و مقاوم‌سازی همان تهدیدات تصمیم بگیرد [۲۰]. در شرایط عادی با وجود نوفه، گفتار تنگش (تضعیف) می‌شود و کیفیت علامت اصلی از بین می‌رود لذا حذف نوفه در یک سامانه تعیین هویت بسیار مهم و مقاومت به آن برای طراحی سامانه قابل اعتماد لازم است. انتخاب نامناسب ویژگی‌ها برای تعیین هویت باعث افزایش حجم ماتریس‌های محاسباتی، نفی بی‌درنگی پاسخ و اختلال در پاسخ یکتای تعیین هویت می‌شود. انتخاب مناسب ویژگی‌ها به پاسخ یکتا، کاهش حجم محاسبات و افزایش دقت تعیین هویت کمک می‌کند. در سامانه پیشنهادی برای تعیین هویت از ماتریس ویژگی‌ها استفاده شده که این ماتریس برای هر گوینده یکتا است تا سامانه به‌درستی به شناسایی گوینده بپردازد. از ترکیب سه ویژگی ضریب کپسترال بسامد مل، ضریب پیشگویی خطی و ضریب کپستروم برای شناسایی گوینده استفاده شده و برای آزمون از مجموعه داده‌ای ای‌بی‌ال‌اس‌دی آر بهره‌برداری شده است. برای پذیرش یا عدم‌پذیرش سامانه، مقدار حد آستانه‌ای ۵/۵ تعیین شده است. شکل ۱۷ نمایانگر بخش تشخیص هویت سامانه است. با استفاده از این روش سامانه قادر است داده‌های با نوفه را نیز بازیابی کند که به دقت تشخیص هویت گوینده بیش از ۸۰ درصد رسیده است. در صورتی که اگر از سازوکار حذف نوفه استفاده نشود و داده‌های نوفه‌ای وارد سامانه شوند دقت تشخیص

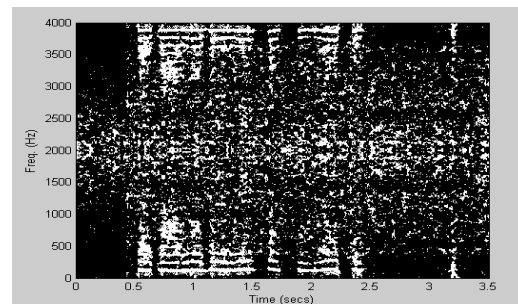
از ویژگی مقاوم تشخیص داده شده در طراحی ماسک استفاده می‌شود و خروجی به دست‌آمده اسپکتروگرام علامت گفتار در شکل ۱۶ نشان داده شده است.



(الف)



(ب)



(ج)

شکل ۱۶ اسپکتروگرام علامت گفتار؛ (الف) علامت گفتار بدون حضور نوفه، (ب) علامت گفتار در حضور نوفه، (ج) علامت گفتار پس از اعمال ماسک.

علامت گفتار بدون نوفه در شکل ۱۶- الف، در حضور نوفه در شکل ۱۶- ب و با اعمال ماسک در شکل ۱۶- ج ارائه شده‌اند. مجموعه داده‌ای آرورا از علامت‌های بدون نوفه و نوفه موجود در محیط‌های مختلف تشکیل شده است. به منظور آزمایش به هر داده، نوفه‌ای افزوده شده و علامت حاصل نوفه‌ای مورد بررسی و آزمایش قرار گرفته است. به منظور حذف نوفه از ماسک دودویی استفاده شده که در صورتی که طیف علامت، انرژی بیش‌تر از حد آستانه تعریف شده داشته باشد مقدار آن یک و در غیر این

منتشر نشده است. نویسندگان هیچ منافع مالی یا روابط شخصی شناخته شده‌ای که به‌تواند بر انتشار این مقاله تأثیرگذار باشد ندارند.

#### فهرست منابع

- [1] R. Ram, A. Das, S.K. Mohapatra, "Speech enhancement using wiener filter based on voiced speech probability," *Advances in Intelligent Computing and Communication*, vol. 1, pp. 161-170, 2020.
- [2] G. Krishna, C. Tran, J. Yu, A.H. Tewfik, "Speech recognition with no speech or with noisy speech," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.
- [3] H.P.P. Win, P.T.T. Khine, "Speech enhancement techniques for noisy speech in real world environments," *7th International Conference on Computer Applications (ICCA)*, 2019.
- [4] M. Forsberg, "Why is speech recognition difficult," *Chalmers University of Technology*, 2003.
- [5] M.H. Soni, N. Shah, H.A. Ptil, "Time-frequency masking-based speech enhancement using generative adversarial network," *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 5039-5043, 2018.
- [6] G. Theodoros, P. Aggelos, "Introduction to audio analysis: A MATLAB approach," *Academic Press*, 2014.
- [7] A.S. Ajibola, N.K. Alang Rashid, "Some commonly used speech feature extraction algorithms," *IntechOpen*, pp. 2-19, 2018.
- [8] R. Rajeev, T. Abhishek, "Analysis of feature extraction techniques for speech recognition system," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 7C2, pp. 197-200, 2019.
- [9] S. Nisar, M. Tariq, A. Adeel, M. Gogate, A. Hussain, "Cognitively inspired feature extraction and speech recognition for automated hearing loss testing," *Cognitive Computation*, vol. 11, no. 4, pp. 489-502, 2019.
- [10] K. Gajanan, K. Birajdar, D. Mukesh, "Speech/music classification using visual and spectral chromagram features," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 329-347, 2020.
- [11] T. Drugman, Y. Utsuke, Y. Kida, "Voice activity detection: Merging source and filter-based information," *IEEE SIGNAL*

هویت گوینده به زیر ۴۰ درصد می‌رسد. لذا استفاده از این روش با وجود داده‌های نوفه‌ای و بدون نیاز به پیش پردازش‌های سنتی توانسته قدرت تشخیص تعیین هویت گوینده را به صورت چشمگیری افزایش دهد.

#### ۵. نتیجه‌گیری

ویژگی‌های مختلف حوزه زمان و بسامد مورد بررسی و آزمایش در محیط‌های نوفه‌ای مختلف قرار گرفتند. از ویژگی انرژی که در مقایسه با دیگر ویژگی‌ها مقاوم‌تر است در طراحی ماسک استفاده شده تا از آن بتوان در سامانه‌هایی از شناسایی گوینده مانند سمک استفاده شود. با استفاده از ماسک دودویی طراحی شده میزان علامت به نوفه در حدود ۲ الی ۳ دسی‌بل افزایش می‌یابد. سامانه شناسایی گوینده نیز از ماتریس ویژگی‌ها ساخته شده تا بتواند هر گوینده را تشخیص دهد. البته در حضور نوفه سامانه تشخیص گوینده دچار خطا می‌شود لذا روش حذف نوفه مورد بررسی قرار گرفت. می‌توان این سامانه را در میکروکنترلر به صورت بی‌درنگ قرار داد و در صنعت استفاده نمود. ایده اصلی ارایه طبقه‌بندی ویژگی‌های مهم برای هر علامت به عنوان ابزار لازم برای مقاوم‌سازی سامانه به‌ازای شرایط علامت در تهدیدات مختلف است. روش پیشنهادی با وجود داده‌های نوفه‌ای و بدون پیش پردازش‌های سنتی توانسته قدرت تشخیص تعیین هویت گوینده را به صورت چشمگیری افزایش دهد. در این راستا برای مقاوم‌سازی بی‌درنگ پویا در شرایط زمانی مختلف به یک الگوی مقاوم‌سازی زمانی وابسته به علامت تحت شرایط هر لحظه نیاز است که در کارهای آینده، این الگو و افزایش سرعت برای پردازش بی‌درنگ و قابل‌کاربرد در تجهیزات اینترنت اشیا دنبال خواهد شد.

#### ۶. تقدیر و تشکر

از مرکز تحقیقات آکوستیک دانشگاه شاهد جهت کمک به نمونه‌برداری داده‌های مورد استفاده این مقاله تشکر و قدردانی می‌شود.

#### ۷. تضاد منافع

این مقاله اصلی و توسط نویسندگان نوشته شده و قبلاً

- PROCESSING LETTERS, vol. 23, no. 2, pp. 252-256, 2015.
- [12] Y. Shi, J. BAI, P. Xue, D. Shi, "Fusion feature extraction based on auditory and energy for noise-robust speech recognition," IEEE Access, vol. 7, pp. 81911-81922, 2019.
- [13] D. Deshwal, P. Sangwan, D. Kumar, "Feature Extraction Methods in Language Identification: A Survey," Wireless Personal Communications, vol. 107, no. 4, pp. 2071-103, 2019.
- [14] M.T. Hosain, A.A. Arif, A.I.Pritom, M.R. Rahman, M. Zahidul Islam, "Development of a tangent based robust speech feature extraction model," International Conference on Cyber Security and Computer Science, pp. 414-425, 2020.
- [15] N. Saleem, M.I. Khattak, "Deep neural networks for speech enhancement in complex noisy environments," IJIMAI journal, vol. 6, no.1, pp. 84-90, 2019.
- [16] S. Lakshmikanth, "Noise cancellation in speech signal processing-A review," International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 1, pp. 5175-5186, 2014.
- [17] C. Shraddha, M.L. Chayadevi, M.A. Anusuya, "Noise cancellation and noise reduction techniques: A review," 1st International Conference on Advances in Information Technology (ICAIT), IEEE, pp. 159-166, 2019.
- [18] NOIZEUS, <https://ecs.utdallas.edu/loizou/speech/noizeus/>.
- [19] ILCDR, <http://www.imm.dtu.dk>.
- [20] Z. Heydari, A. Mahabadi, A. Ranjbar, "The effect of energy and interpolation combination on the accuracy of indoor localization based on the speech signal time difference of the arrival," Journal of Acoustical Engineering Society of Iran, vol. 8, no. 2, pp. 60-78, 2021.

(Research Article)

## Classification of noise-robust speech features in the speaker authentication system

M. Mirbeygi<sup>1,3</sup>, A. Mahabadi<sup>1,3\*</sup>, A. Ranjbar<sup>2,3</sup>

1. Computer Engineering Department, Shahed University

2. Electrical Engineering Department, Shahed University

3. Acoustic Research Center, Shahed University

Received: 2020/09/10, Accepted: 2021/07/22

DOR 20.1001.1.23455748.1400.9.1.3.5

### Abstract

Automatic speaker recognition has a wide range of applications in industrial and security systems and requires the extraction of speech signal features. The use of the feature matrix is very important in real-time recognition of the speaker, and the presence of environmental and processing noise leads to a violation in the characteristics of the features and the production of recognition errors. Increasing the accuracy of recognition detection requires the noise removal process to correctly determine the energy characteristics, energy entropy, zero-crossing rate, spectral centroid, spectral spread, spectral entropy, spectral flux, and spectral roll off the signal. In designing real-time and reliable algorithms, there are critical processes of correct speech extraction, sensitivity detection, and measuring the robustness of signal parameters to eliminate noise and improve speech quality, which play a key role in improving the signal-to-noise ratio. In this paper, the classification of speech signal features for designing real-time and noise-robust speaker recognition algorithms in measuring its robustness are investigated. The proposed method of noise removal uses a binary mask with a robust feature and the experimental results of the experiments on the standard data show the rate of signal improvement to the noise of approximately 2 to 3 db. The feature matrix evaluation for the authentication system consists of mel frequency coefficient, linear prediction coefficient and, cepstrum coefficient, which has been evaluated by the Euclidean distance method in another experimental standard data set. Our proposed method achieves on overall 80% real-time recognition accuracy in noisy data set.

**Keywords:** Classification of speech features, Speaker authentication, Noise-robust, Speech signal, Noise.

pp. 28-39 (In Persian)

---

\* Corresponding author E-mail: mahabadi@shahed.ac.ir