

# یک ویژگی جدید برای استفاده در تعیین شباهت در سیگنال گفتاری: نرمالیزه کردن سیگنال زمان کوتاه گفتار نسبت به فاز

سید احمد معتمدی

استادیار دانشکده مهندسی برق، دانشگاه صنعتی امیرکبیر

سیدعلی سید صالحی

مربی دانشکده برق، دانشگاه صنعتی امیرکبیر و دانشجوی دکترای مهندسی پزشکی دانشگاه تربیت مدرس

**چکیده:**

تعیین میزان شباهت مابین تکه‌های زمان کوتاه از سیگنال گفتاری در طبقه‌بندی آنها کاربرد دارد. طبقه‌بندی سیگنال زمان کوتاه گفتار در: تشخیص گفتار، تشخیص گوینده، و اخیراً در بهبود کیفیت گفتار آلوده به نویز مورد استفاده قرار می‌گیرد. برای تعیین کمی این میزان شباهت، ویژگی (feature) های مختلفی تاکنون ارائه گردیده‌اند. در اینجا ما چند نوع ویژگی موجود برای این تعیین کمی میزان شباهت بین سیگنالهای زمان کوتاه گفتار را مورد بحث قرار داده و سپس یک ویژگی جدید که براساس تخمین سیگنال زمان کوتاه گفتار نرمالیزه شده نسبت به فاز می‌باشد را معرفی و آنرا با سایر ویژگی‌های رایج مقایسه نموده‌ایم. نتایج عملی، حاکی از برتری کلی این روش جدید از لحاظ کیفیت نتایج و سرعت الگوریتم، بر سایر روشهای اشاره شده می‌باشند.

کلید واژه: سیگنال زمان کوتاه نرمالیزه نسبت به فاز، استخراج ویژگی، کمی‌کردن شباهت، تشخیص گفتار.

## "Phase Normalized Short Time Periods of Speech Signal , as a new Feature for Similarity Measurement"

S. A. Motamedi, Ph.D.

Assist. Prof. Elec. Eng . Dept . Amirkabir University

S. A. Seyyed Salehi, M.Sc.

Lecturer Elec. Eng . Dept . Amirkabir University  
Graduate Student Tarbiat Modarres University

### ABSTRACT

*Similarity measurement among short time periods of speech signal (STPSS) is used in the related classification. The application of the classification of STPSS is used in areas such as: speech recognition, speaker identification and recently in speech enhancement.*

*In order to quantify the similarities, different features are already proposed. Here, we discuss about some of these features using for similarity quantifications. Then, we propose a new features which is based on normalizatin of STPSS with respect to phase , and compare it with other conventional features. Experimental results show*

some preference of this new feature to other methods, with respect to the quality of results and the computational efficiency.

Keywords: Phase normalization, Feature extraction, Similarity measurement Speech recognition.

مقدمه:

تمامی تکه سیگنال دوم (مرکز تابع کورلیشن) دورتر باشد، از وزن کمتری برخوردار خواهد بود. در صورتی که بخواهیم با قراردادن ضرایب وزنی متفاوت به تناسب محل، برای  $R(m)$  این اشکال را جبران نماییم، در هر حال، مقادیر انتهایی تابع کورلیشن مربوط به موقعی است که تنها بخش کوچکی از دو تکه سیگنال مورد مقایسه، برای ارزیابی شباهت روی هم منطبق گردیده‌اند و کمیت شباهت حاصله، در این حالت از اعتبار کمی برخوردار خواهد بود. روش تعیین میزان شباهت بر پایه محاسبه مستقیم کورلیشن زمان کوتاه را می‌توان متناظر با روش محاسبه شباهت بین دو بردار در فضای چند بعدی دانست که در آن ابتدا هر بردار بوسیله تقسیم بر قدر مطلق آن نرمالیزه شده و سپس کسینوس زاویه بین دو بردار به عنوان کمیتی که میزان شباهت (تطابق) دو بردار نرمالیزه شده را تعیین می‌نماید، محاسبه گردد.

$$r = \frac{\langle \bar{a} \cdot \bar{b} \rangle}{|\bar{a}| \cdot |\bar{b}|} = \langle \bar{a}_n \cdot \bar{b}_n \rangle = \cos \alpha$$

که  $r$  میزان شباهت،  $\bar{a}$  و  $\bar{b}$  دو بردار در فضای چند بعدی،  $\bar{a}_n$  و  $\bar{b}_n$  بردارهای یک (نرمالیزه شده) در جهت‌های  $\bar{a}_m$  و  $\bar{b}_n$  و  $\alpha$  زاویه بین دو بردار است.

### روش دوم: یکسان سازی وزنها

برای رفع اشکال روش قبل، اینبار تکه سیگنال اولی یعنی  $x_1$  را بطول  $N$  نمونه انتخاب نموده و کورلیشن آن را با  $2N$  نمونه از سیگنال گفتاری دوم محاسبه می‌کنیم. این بدین معنی است که تکه سیگنال گفتاری دوم  $x_2$  را نگاه داشته و  $x_1$  را روی آن بلغزانیم و در هر گام از لغزش مجموع حاصل ضرب آن دو را محاسبه نماییم. در اینجا  $N$  باید حداقل یک پیوند از پیچ گفتار را در بر داشته باشد. با تعیین مقدار ماکزیمم این تابع کورلیشن (نقطه حداکثر شباهت)، مقدار آنرا با استفاده از انرژی سیگنال  $x_1$  و همچنین انرژی آن مقدار از سیگنال  $x_2$  که در نقطه ماکزیمم تابع کورلیشن روی  $x_1$  منطبق گردیده است، نرمالیزه می‌نماییم.

$$r = \frac{\{ \max_{n=1}^N [ \sum_{n=1}^N x_1(n) \cdot x_2(n+m) ] \}^2}{( \text{انرژی } x_1 ) \cdot ( \text{انرژی } x_2 )}$$

تعیین میزان شباهت مابین تکه‌های کوتاه زمانی از سیگنال‌های گفتاری در طبقه‌بندی آنها مورد استفاده قرار می‌گیرد. از طبقه‌بندی تکه‌های کوتاه زمانی سیگنال گفتار در تشخیص گفتار، تشخیص گوینده، و اخیراً در بهبود کیفیت گفتار آلوده به نویز استفاده می‌شود. روشهای متنوعی برای کمی کردن این میزان شباهت وجود دارد که با توجه به ماهیت سیگنال گفتار، غالباً بر ویژگیهای ترکیبی آن، استوار می‌باشد. در اینجا ما ابتدا به چند روش عملی موجود برای تعیین کمی میزان شباهت بین سیگنالهای زمان کوتاه گفتار اشاره نموده و سپس یک روش جدید استخراج ویژگی که مجموعاً بر روشهای اشاره شده، برتری دارد را ارائه می‌نماییم. در ارزیابی روشها، ما بر دقت تعیین شباهت بهتر و مقرون بصرفه تر بودن از نظر حجم محاسبات، تکیه داشته‌ایم.

### روش اول: تعیین میزان شباهت بر پایه محاسبه مستقیم کورلیشن زمان کوتاه

اولین روش، محاسبه کورلیشن زمان کوتاه مابین دو تکه سیگنال گفتاری است که برای مستقل نمودن نتیجه از انرژی سیگنالها، تابع کورلیشن زمان کوتاه حاصله را بر انرژی دو تکه سیگنال تقسیم می‌نماییم (نرمالیزه):

$$R_{x_1 x_2}(m) = \frac{\sum_{n=1}^N x_1(n) \cdot x_2(n+m)}{\sqrt{ \left[ \sum_{n=1}^N x_1^2(n) \right] \cdot \left[ \sum_{n=1}^N x_2^2(n) \right] }}$$

$$r = \max [ R_{x_1 x_2}(m) ]$$

$R_{x_1 x_2}(m)$  مقادیر نرمالیزه تابع کورلیشن زمان کوتاه دو تکه سیگنال گفتاری هم طول بوده و دارای  $(2N-1)$  نمونه می‌باشد.  $x_1$  و  $x_2$  مقادیر دو تکه سیگنال گفتاری بصورت توابع گسسته  $N$  طول هر تکه سیگنال می‌باشند. در اینجا  $R(m)$  مقادیری بین ۱ و ۱- را دارا بوده و ماکزیمم مقدار آن برای تعیین شبیه‌ترین نقطه انطباق و همچنین میزان کمی شباهت بین دو تکه سیگنال، مورد استفاده قرار می‌گیرد. این روش در عین سادگی دارای این اشکال هست که کمیت میزان شباهت حاصله به روش فوق، به موقعیت شبیه‌ترین نقطه انطباق در سیگنال کورلیشن، وابسته بوده و هر چه که این نقطه از محل متناظر با انطباق تمامی تکه سیگنال اول روی

بدین ترتیب هنگامیکه تکه سیگنال  $x_2$  در روی  $x_1$  می‌لغزد، در ناحیه مشخص شده  $0 \leq m \leq N$  تابع کورلیشن زمان کوتاه حاصله، ناشی از مجموع حاصلضرب طولهای مساوی از  $x_1$  روی  $x_2$  (بطول  $N$  نمونه) خواهد بود و بدین طریق وزنه‌های یکسان و اعتبار مساوی برای کمیت‌های شباهت حاصله، در روی ناحیه‌ای از تابع کورلیشن زمان کوتاه، ایجاد نموده‌ایم. این روش بخصوص برای موافقی که شباهت قسمتهای مختلف سیگنال گفتار مربوط به یک گوینده واحد را بررسی می‌کنیم مفید خواهد بود که در آن امکان انطباق کامل پر یود پیچ گفتار فراهم می‌شود.

### روش سوم: محاسبه فاصله (دستانس) بین پارامترهای کورلیشن نرمالیزه شده

در دو روش فوق، برای تعیین میزان شباهت دو تکه سیگنال گفتاری از رابطه کورلیشن استفاده می‌کنیم، اگر بخواهیم در یک مجموعه‌ای که از  $M$  تکه سیگنال تشکیل شده، میزان شباهت بین تک تک تکه سیگنال‌های درون مجموعه با یکدیگر را تعیین نماییم، به تعداد  $M(M-1)$  دفعه می‌بایستی عمل کورلیشن زمان کوتاه را انجام دهیم، که حتی اگر از FFT با روشهای مناسب برای محاسبه تابع کورلیشن استفاده گردد باز هم به زمان محاسبات بالایی نیاز خواهد بود. اگر بتوانیم مقایسه سیگنالها را به نحوی انجام دهیم که بخش مهمی از محاسبات خارج از حلقه اصلی مقایسه و صرفاً برای هر تکه سیگنال یکبار انجام گردد آنگاه برای تمامی تکه سیگنالها فقط  $M$  بار محاسباتی با حجم در حدود کورلیشن خواهیم داشت و صرفه جویی مطلوبی در زمان مورد نیاز، پدید می‌آید. برای رسیدن به این هدف، بایستی از روشهای تعیین فاصله (دستانس) استفاده نماییم. سؤالی که در اینجا مطرح می‌گردد این است که انتخاب چه ویژگیهایی از دو سیگنال برای محاسبه فاصله، در تعیین کمی میزان شباهت مطلوبتر خواهد بود؟ به عنوان اولین کاندیدایی که می‌تواند معرف محتوای طیفی دو تکه سیگنال مورد مقایسه باشد، روش استفاده از مقادیر لگاریتم چگالی طیفی قدرت، مطرح می‌گردد. در طیف فرکانسی گفتار، پیچ بصورت یک شکل دندانداره‌ای بروز نموده و روی مقایسه سیگنالها برای تعیین شباهت، اثر سوء می‌گذارد. برای رفع این اثر منفی روشهایی وجود دارد. از این روشها، روشهای صاف کردن چگالی طیفی قدرت لگاریتمی، توسط فیلتر مستقیم با همومورفیک، و یا استفاده از پیشگویی خطی می‌باشد.

از انواع دیگر، ویژگی که برای محاسبه فاصله، از تکه سیگنالهای گفتار می‌تواند استخراج گردند، مقادیر کپستروم است که به وسیله تبدیل فوریه گرفتن از لگاریتم چگالی طیفی قدرت بدست می‌آید. این مشخصه‌ها دارای قدرت تفکیک و ویژگیهای مربوط به ناحیه صوتی از سیگنال پیچ صدا هستند. با انتخاب مناسب مقادیر از تابع کپستروم، می‌توان اثر حضور سیگنال تارهای صوتی را در مقادیر ویژگی‌های حاصله کاهش داد. پارامترهای

LPC نیز می‌تواند به عنوان یک نوع ویژگی برای سیگنالها مطرح باشد. روشهای استخراج ویژگی که تا اینجا ذکر شدند، از حجم محاسباتی نسبتاً بالایی برخوردارند. ما علاقمندیم که حتی المقدور به روشهایی دست یابیم که به کمترین میزان محاسبات نیاز داشته باشد. در اینجا ما پارامترهای کورلیشن را که محاسبه آنها به حجم محاسباتی کمتری نیاز دارد، جهت تعیین شباهت مورد بررسی قرار می‌دهیم. روش محاسبه بدینصورت است که ابتدا اتوکورلیشن زمان کوتاه هر تکه سیگنال گفتاری را محاسبه می‌کنیم. طول هر تکه سیگنال مورد استفاده در این روش نیز باید حداقل برابر یک پر یود از پیچ گفتار باشد.

$$R_{x_1}(m) = \frac{1}{N} \sum_{n=1}^N x_1(n) x_1(n+m)$$

به عنوان بردار ویژگی،  $K$  نمونه از مقادیر تابع اتوکورلیشن را، بعد از نرمالیزه کردن بر می‌گیریم. برای نرمالیزه کردن از مقدار اتوکورلیشن در لحظه  $m=0$  که با انرژی سیگنال متناظر است استفاده می‌کنیم.

$$F_{x_1} = \frac{R_{x_1}(m)}{R_{x_1}(0)}; m = 1 \rightarrow k$$

که  $F_{x_1}$  بردار مقادیر ویژه مربوط به تکه سیگنال  $x_1$  می‌باشد. تابع اتوکورلیشن عملاً اطلاعات مربوط به فاز سیگنال را حذف می‌نماید و به عبارت دیگر سیگنالها را از نقطه نظر تفاوت‌های فازی، نرمالیزه می‌نماید. میزان کمی شباهت بین دو تکه سیگنال را می‌توان از طریق ضرب داخلی دوبردار ویژگی در یکدیگر بدست آورد.

$$r = \sum_{n=1}^k F_{x_1}(n) \cdot F_{x_2}(n)$$

که  $r$  عددی برابر ۱ برای  $x_1=x_2$  (حالت کاملاً شبیه) و دارای حداقل ۱- خواهد بود. اگر بخواهیم میزان کمی شباهت را به روش تعیین فاصله بین دوبردار ویژگی محاسبه کنیم:

$$d = \text{norm}(\bar{F}_{x_1} - \bar{F}_{x_2})$$

(norm در اینجا منظور جذر مجموع مؤلفه‌ها است).

این فاصله  $d$  دارای مقدار صفر برای شبیه‌ترین حالت خواهد بود. با دور شدن مقدار  $d$  از صفر، شباهت بین دوبردار ویژگی، یعنی دو سیگنال، کمتر می‌گردد. این مقدار  $d$  را می‌توان طول بردار حاصل از کم کردن بردارهای ویژگی نرمالیزه از یکدیگر، در فضای  $K$  بعدی، دانست.

## روش چهارم: محاسبه فاصله (دستانس) بین مقادیر

### زمانی سیگنالهای نرمالیزه شده نسبت به فاز

روشی که در این مقاله معرفی می‌گردد بر پایه تلاش برای مقایسه خود سیگنالهای زمانی گفتار با یکدیگر، استوار است. به عبارت دیگر سعی می‌شود که زمینه‌ای فراهم گردد که بتوان از روشهای شناسایی الگوی آماری برای تعیین کمی شباهت دو تکه سیگنال گفتاری بهره جست. از محاسن این کار این خواهد بود که عملاً ما در این روش در واقع به جای آنکه فاصله بین مقادیر ویژگی متناسب با مجذور مقادیر سیگنال (کورلیشن یا مجذور طیف) را محاسبه کنیم، فاصله مابین مقادیر ویژگی متناسب با خود مقادیر زمانی سیگنال را محاسبه می‌کنیم و این از جهت تعیین شباهت، به وضعیت طبیعی تکه سیگنالهای گفتاری نزدیکتر است. مسئله‌ای که اینجا در روبروی ما وجود دارد عدم تطابق مقادیر فاز در دو سیگنال مورد مقایسه است و برای تعیین شباهت نیازمندیم که به نحوی ابتدائاً مقادیر زمانی دو سیگنال را نسبت به فاز نرمالیزه نماییم. این در مورد یک سیگنال پیروی یک به معنای آن است که تمامی هارمونیکهای آنرا در محور زمان به قدری شیفت دهیم تا در لحظه  $t=0$  فاز همه آنها برابر صفر گردد (دارای مقدار ضرایب فوریه حقیقی مثبت باشند) و بدین طریق تمامی مؤلفه‌های فوریه آن از نظر مقادیر فاز نرمالیزه می‌شوند.

در مورد سیگنال شبه پیروی یک متغیر با زمان گفتار، یک روش آن است که تبدیل فوریه زمان کوتاه سیگنال را محاسبه نموده و فقط مقادیر دامنه طیف سیگنال را به عنوان بردار ویژگی مورد استفاده قرار دهیم. شرط وجود دقت کافی در این روش این است که در تکه زمان کوتاه سیگنال، به تعداد زیادی از پیروی پیچ گفتار حضور داشته باشد. از آنجا که عملاً مشخصات سیگنال گفتار با زمان متغیر است، غالباً امکان انتخاب فوق وجود ندارد. یک راه حل، همان حذف اثر پیروی پیچ در روی طیف به روشهایی است که قبلاً ذکر شدند و غالباً به حجم محاسبات نسبتاً بالا نیاز دارند، لیکن روش مقرون به صرفه‌تر این است که در اینجا شرح داده می‌شود.

برای تعیین میزان شباهت بین تکه سیگنالهای گفتاری؛

(۱) از هر یک بطول زمانی حدود ۱۵ الی ۲۵ میلی‌ثانیه

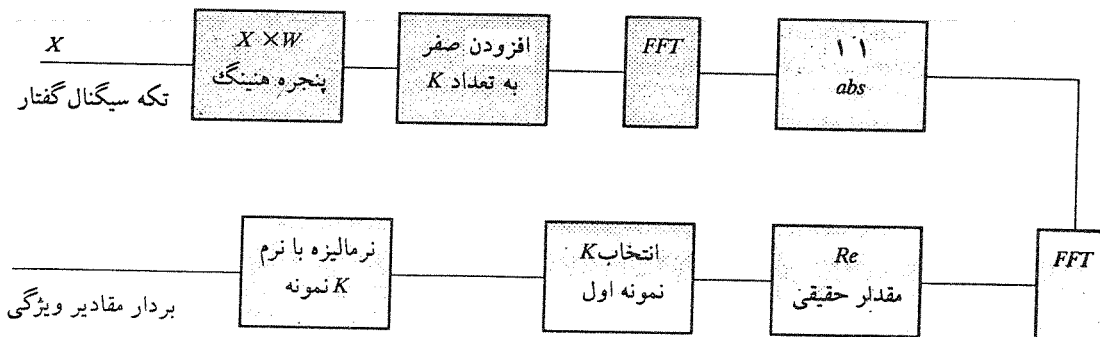
انتخاب و آنرا در یک پنجره مناسب زمانی (مثلاً هنینگ) ضرب می‌نماییم.

(۲) سپس با افزودن تعداد  $K$  (تعداد مطلوب عناصر بردار ویژگی) عدد صفر به دنبال سیگنال، از نتیجه حاصله تبدیل فوریه می‌گیریم.

(۳) فاز سیگنال را در دامنه فرکانس، صفر نموده و تبدیل فوریه معکوس آن را محاسبه می‌نماییم. سیگنال حاصله، معادل تکه سیگنال زمانی است با این تفاوت که تمامی مؤلفه‌های آن باندازه‌ای در زمان شیفت یافته‌اند که همگی از نظر فاز نرمالیزه شده‌اند. با این عمل، امکان انطباق سیگنالها در دامنه زمان را جهت یافتن سیگنالهایی با محتوای مشابه فراهم نموده‌ایم. مقادیر این سیگنال زمانی نرمالیزه نسبت به فاز را با ایمپالس ریسپانس اعضای صوتی، بخصوص در نمونه‌های ابتدای سیگنال، می‌توان قابل مقایسه دانست.

به عنوان مقادیر ویژگی،  $k$  نمونه اول این سیگنال زمانی نرمالیزه نسبت به فاز را انتخاب و مقادیرشان را با نرم این  $k$  نمونه نرمالیزه می‌نماییم. در اینجا به چند نکته باید توجه داشت، اول برای داشتن دقت هر چه بیشتر در نتیجه، در دامنه فرکانس بایستی تخمین هر چه بهتری از طیف فرکانسی سیگنال را داشته باشیم تا صفر کردن فاز هارمونیکها در طیف زمان کوتاه گفتار به واقعیت بیشتر نزدیک باشد، لذا استفاده از پنجره‌ای که این مقصود را هر چه بهتر حاصل نماید (هنینگ یا همینگ) ضروری است. نکته دوم اینکه در محاسبه بردار ویژگی در اینجا برای اینکه مقادیر سیگنال زمانی نرمالیزه شده نسبت به فاز را نسبت به انرژی سیگنالها نرمالیزه کنیم، آنها را بر نرم (جذر مجموع مجذور) مقادیر  $k$  نمونه انتخابی تقسیم می‌کنیم، زیرا در اینجا نرم نمونه‌ها با انرژی سیگنال متناسبتر است.

نکته سوم اینکه، در این روش اضافه کردن تعداد کافی صفر قبل از گرفتن تبدیل فوریه از سیگنال زمانی اولیه ضروری است، زیرا در غیراینصورت شیفت زمانی مؤلفه‌های مختلف فرکانسی سیگنال، موجب تداخل آنها با پیرودهای تکراری بعدی تبدیل فوریه ناپیوسته شده (تقریباً مشابه با کانولوشن دایره‌ای) و در مقادیر واقعی، خطا ایجاد می‌نماید. لذا اضافه کردن حداقل به



تعداد  $k$  (تعداد عناصر بردار مقادیر ویژه) عدد صفر به دنبال سیگنال زمانی اولیه قبل از گرفتن تبدیل فوریه از آن، ضروری است، تا حداقل تعداد  $k$  نمونه اول سیگنال زمانی نرمالیزه شده نسبت به فاز دارای اعتبار باشد.

از ویژگیهای این روش کم‌رنج کردن اثر پیروی با انتخاب مناسب مقادیر از سیگنال زمانی نرمالیزه شده نسبت به فاز می‌باشد و به همین دلیل دیگر لزومی ندارد که تعداد زیادی پیروی زمانی پیچ‌گفتار در تکه سیگنالها حضور داشته باشد و فقط باندازه یک الی سه پیروی پیچ کافی است.

در عمل ما برای سیگنالهای گفتاری نمونه‌برداری شده با فرکانس  $10\text{KHZ}$  طول هر تکه سیگنال گفتاری انتخابی را  $192$  نمونه انتخاب و در پنجره نوع هنینگ ضرب کرده‌ایم. سپس به تعداد  $K=64$  نمونه صفر به آن افزوده‌ایم. از سیگنال  $256$  نمونه‌ای حاصله  $FFT$  گرفته و فاز آنرا صفر کرده‌ایم. بجای عکس تبدیل فوریه، معادل ساده شده آن یعنی مقدار حقیقی نتیجه اعمال  $FFT$  روی دامنه طیف را استفاده می‌نماییم. (دامنه طیف سیگنال مقارن و مقداری حقیقی است).

$k$  نمونه اول این سیگنال را انتخاب و با نرم این  $k$  نمونه نرمالیزه می‌نماییم. بدین طریق بردار مقادیر ویژگی حاصل می‌شود که برای تعیین کمی شباهت بین سیگنالها می‌توان فاصله بین این بردارهای ویژگی را محاسبه نمود.

در صورتیکه سیگنالهای مورد مقایسه برای تعیین شباهت، آلوده به نویز سفید نیز باشند، در مقادیر ویژگی بهتر است از نمونه اول استفاده نشود. با این عمل اثر منفی وجود نویز را می‌توان تا حدی کاهش داد.

مبانی تئوری نرمالیزه کردن سیگنال زمان کوتاه گفتار نسبت به فاز:

فرض کنید  $S_1(t)$  یک سیگنال دوره‌ای بطول بی‌نهایت باشد، در صورتیکه ضرایب سری فوریه این سیگنال را روی یک پیروی حساب کنیم و سپس مقادیر فاز ضرایب فوریه مختلط سیگنال را برابر صفر قرار داده و مجدداً آنرا به دامنه زمان برگردانیم، سیگنال حاصله در واقع همان سیگنال اولیه است که مؤلفه‌های زمانی فوریه آن در زمان طوری شیفت پیدا کرده‌اند که مقدار فاز تمامی این مؤلفه‌ها در لحظه صفر، برابر صفر شده است ( $S_1(t)$ ).

$$S_1(t) = \sum_{n=1}^{\infty} c_n e^{j(n\omega t + \varphi_n)}$$

(مقدار متوسط  $S_1(t)$  صفر فرض می‌شود).

$$\bar{S}_1(t) = \sum_{n=1}^{\infty} c_n e^{jn\omega t}$$

ضرایب  $C_n$  (با فرض  $n$  محدود) می‌توانند یک مقادیر ویژه مناسبی برای سیگنال  $S_1(t)$  (برای مواقعی که اطلاعات فازی

سیگنال مهم نباشد) باشند. حال اگر مقادیر  $a_{n1}, \dots, a_{nm}$  مقدار نمونه برداری شده از مؤلفه فوریه  $m$  ( $\varphi_n = 0$ ) از سیگنال دوره‌ای باشد که بر روی یک دوره انتخاب شده است (تعداد  $m$  با توجه به شرایط نایکوئیست می‌تواند تعیین گردد)، این مقادیر  $a_{n1}$  و  $a_{nm}$  یک بیان از مؤلفه  $n$  ام هستند. در اینحال، مجموع مقادیر:

$$[(a_{1m} + a_{nm} + \dots) \quad (a_{11} + \dots + a_{n1} + \dots)] \\ = [f_1, \dots, f_m] = F$$

بردار  $F$  نیز یک بیان از سیگنال  $\bar{S}_1(t)$  است که مقادیر فوق عبارت از  $m$  مقدار نمونه‌برداری شده از سیگنال  $S_1(t)$  روی یک دوره می‌باشند. این بیان حاصله از سیگنال  $\bar{S}_1(t)$  در واقع یک بردار مقادیر ویژگی بطول  $m$  از سیگنال  $\bar{S}_1(t)$  است که مستقل از اطلاعات فازی مؤلفه‌ها می‌باشد. روشن است که برای یک سیگنال با پهنای باند محدود، در صورتیکه شرط نایکوئیست در انتخاب تعداد  $m$  مراعات شده باشد، بردار ویژگی  $F$  یک تبدیل خطی از بردار  $C$  می‌باشد. حال فرض کنیم  $S_2(t)$  یک سیگنال شبه دوره‌ای متغیر با زمان (مثلاً گفتار) باشد. برای چنین سیگنالی اگر بخواهیم که بردار مقادیر ویژگی برای محل معینی از سیگنال تعریف نماییم، ابتدا طولی از سیگنال را که بتوان مشخصات سیگنال را در آن فاصله با تقریب قابل قبولی ثابت فرض نمود انتخاب و تبدیل فوریه آنرا محاسبه می‌کنیم. فرض کنیم  $S_2^*(t)$  سیگنال دوره‌ای است که از تکرار  $w(t)$ .  $S_2(t)$ ،  $W(t)$  یک پنجره زمانی با طول مناسب در زمان ایجاد شده باشد.

$$S_2^*(t) = \sum_{k=-\infty}^{\infty} S_2(t - kT) w(t - kT)$$

( $T$  طول هر دوره است) محاسبه تبدیل فوریه  $w(t)$ .  $S_2(t)$  بر پایه محاسبه ضرایب فوریه سیگنال دوره‌ای  $S_2^*(t)$  می‌باشد. در این حالت نیز می‌توان بردار مقادیر ویژگی سیگنال فاز صفر شده را برای این سیگنال دوره‌ای  $S_2^*(t)$  مشابه قبل حساب نمود.

$$F^* = [f_1^*, \dots, f_m^*]$$

حال اگر بخواهیم که این بردار  $F^*$  فوق را به عنوان یک بردار مقادیر ویژگی قابل قبولی برای سیگنال  $w(t)$ .  $S_2^*(t)$  نیز تلقی نماییم و به عبارت دیگر بردار  $F$  را به عنوان مقادیر ویژگی موضعی برای سیگنال متغیر با زمان  $S_2(t)$ ، مورد استفاده قرار دهیم، چه شرایطی بایستی موجود باشند؟ پاسخ این است که:

(۱) بمنظور جلوگیری از تولید خطای ناشی از مرز دوره‌ها در مؤلفه‌های فرکانسی، هنگام پنجره نهادن روی سیگنال و تکرار دوره‌ای آن برای تولید، بایستی مقادیر سیگنال در نزدیکی مرز

شباهتها در سیگنال گفتاری کارآمدتر است، در یافتن تکه بعدی برای هر تکه از سیگنال گفتاری، موفق تر باشد.

بمنظور مقایسه بهتر، از ۱۵ شایه سیگنال گفتاری نمونه برداری شده که مقداری آلوده به نویز و نیز دارای کمی اعوجاج است استفاده نموده ایم. ملاک ارزیابی روشها، درصد تعداد تکه‌های بعدی صحیح تشخیص داده شده در کل سیگنال می باشد. جدول زیر نتایج این ارزیابی را نشان می دهد:

جدول ۱-

روش دوم	روش سوم	روش چهارم (نرمالیزه فاز)	
۰/۷۵	۰/۷۶	۰/۹۸	درصد تشخیص صحیح
۴	۱	۱	زمان کل (نسی)

جدول ۱ - درصد تعداد تکه‌های بعدی که با هر روش بطور صحیح تعیین شده‌اند. روش چهارم تقریباً در تمامی نقاطی از سیگنال که حاوی گفتار است پاسخ صحیح را انتخاب نموده است. ارزیابی عملی نشان می دهد که روش نرمالیزه کردن سیگنال زمان کوتاه گفتار نسبت به فاز (روش چهارم) برای تعیین کمی شباهت، از مزایای دقت تعیین شباهت بالا و حجم محاسباتی مورد نیاز نسبتاً کمتری، برخوردار می باشد.

### جمع بندی:

در این مقاله به منظور کمی کردن میزان شباهت بین دو تکه سیگنال گفتاری، دو روش بر پایه استفاده از کورلیشن زمان کوتاه و دو روش بر اساس محاسبه فاصله بین بردارهای ویژگی سیگنال گفتاری مورد بحث و مقایسه قرار گرفتند. روش استخراج بردار ویژگی با استفاده از نرمالیزه کردن سیگنال گفتار زمان کوتاه نسبت به فاز که در این مقاله پیشنهاد گردید (روش چهارم)، جمعاً از جنبه‌های کیفیت تعیین شباهت و زمان لازم برای اجرای الگوریتم، برتری چشمگیری را از خود نشان داده است.

دوره‌ها، خیلی کوچک گردد (مثلاً پنجره نوع هنینگ).  
۲) با توجه به اینکه هنگام صفر شدن فاز مؤلفه‌ها، آنها روی محور زمان شیفت پیدا می کنند، مقادیری از دوره‌های مجاور که با این شیفت زمانی روی  $m$  نقطه دوره اصلی می افتند، می بایستی صفر باشند. در چنین حالتی است که می توان گفت که مقادیر سیگنال فاز صفر شده در دوره اصلی، تخمین خوبی از مقادیر  $S_0^*(t)$  از سیگنال متغیر با زمان اصلی  $S_0(t)$ ، می بایستی از پنجره‌ای استفاده نمود که دارای مشخصات زیر باشد:

$$w(t) = \begin{cases} 0 - \frac{T}{2} < t \leq \frac{T}{4} \\ w'(t) - \frac{T}{4} < t < \frac{T}{4} \\ 0 + \frac{T}{2} < t \leq \frac{T}{2} \end{cases}$$

که  $w'(t)$  یک پنجره از نوع هنینگ، بارتلت و یا مشابه اینها می باشد. در صورتیکه معتبر بودن فقط تعدادی از  $m$  مؤلفه بردار ویژگی  $F^*$  مورد نظر باشد، متناسب آن، می توان طول زمانی محلتهای صفر در پنجره  $w(t)$  را کوتاهتر انتخاب نمود. با این ملاحظات، بردار مقادیر ویژگی  $F^*$  حاصله از اینطریق را می توان یک بیان قابل قبول از سیگنال زمان کوتاه پنجره نهاده شده، به عنوان مشخصات موضعی سیگنال متغیر با زمان اصلی تلقی نمود.

### ارزیابی عملکرد روشهای تعیین کمی شباهت

عملکرد روشهای تعیین میزان شباهت که قبلاً ذکر شدند را به وسیله آزمایش زیر عملاً مورد مقایسه قرار داده ایم. در این آزمایش، سیگنال صوتی را به تکه‌های کوتاه زمانی تقسیم نموده و توانایی روشهای تعیین شباهت فوق را در یافتن صحیح تکه بعدی، برای هر تکه سیگنال گفتاری می آزماییم. از آنجا که تکه‌های مجاور در سیگنال گفتاری، بدلیل خواص دینامیکی ناحیه صوتی، دارای شباهتهای زیادی در نزدیکی مرزهایشان بایکدیگر می باشند انتظار می رود الگوریتمی که در تعیین

### مراجع:

- 1- "Digital signal processing", Oppenheim & Shafer.
- 2- "Digital methods for signal analysis", Beau Champ, 1979.
- 3- "Digital filters: analysis and design", Antoniou, 1979.
- 4- "Digital processing of speech signals" Rabiner & Shafer.
- 5- "Total least square approach for frequency estimation using Linear prediction", IEEE
- 6- "A weighted Cepstral distance measure for speech recognition", IEEE Trans. of Assp, Oct. 87.
- 7- "Methodologies in pattern recognition and image analysis, a brief survey", Mantas J., Pattern recognition, Vol 20, No.1, 1987.
- 8- "Neural nets for adaptive Filtering and adaptive pattern reconition", IEEE Copmputer Mag., March 88.