پنجمین کنفرانس مهندسی برق ایران

۱۹-۱۷ اردیبهشت ۱۳۷۶

مجموعه مقالات ـ ۵

مخابرات (الف)

دانشگاه صنعتی شریف

دانشکده مهندسی برق

ICEE - 97

پنجمین کنفرانس مهندسی برق ایران ۱۹-۱۷ اردیبهشت ۱۳۷۶
Fifth Iranian Conference on Electrical Engineering May 7-9, 1997

جداسازی دو سیگنال صحبت درهم

مسعود بابایی‌زاده و محمود تبیانی
دانشکدهٔ برق دانشگاه صنعتی شریف

چکیده: در این مقاله روش جدیدی برای جداکردن دو سیگنال صحبتِ مخلوط ارائه شده است. بر خلاف روشهای قبلیِ موجود، که دو صحبت را در حوزهٔ فرکانس از هم جدا می‌کردند، کلیهٔ محاسباتِ این روش در حوزهٔ زمان انجام می‌گردد. این روش بر مبنای قضیهٔ جالب و ساده‌ای قرار دارد، که آن هم برای نخستین بار در این مقاله بیان می‌شود. برای آنالیز و سنتز صحبت نیز، از مدل سینوسیِ صحبت استفاده شده است.

کلمات کلیدی: سیگنال صحبت، جداسازی، مدل سینوسیِ صحبت، حداقل مربعات.

۱ - مقدمه

مسألهٔ جداکردن دو سیگنالِ صحبتِ مخلوط، با وجودِ بیش از بیست سال تحقیق، هنوز به جوابِ مطلوبی منتهی نشده است. شکل موج پیچیدهٔ سیگنالِ صحبت، و طبیعت غیرایستانِ آن، باعث عدم موفقیتِ کلیهٔ تلاشهای انجام شده برای حل این مسأله گردیده است. نخستین بار شیلدز و اپنهایم[1] به این مسأله پرداختند و استفاده از فیلتر شانه‌ای را برای حل این مسأله بررسی نمودند[۱]. ایدهٔ انتخاب دامنهٔ هارمونیکها برای حل این مسأله، نخستین بار توسط پارسُنز[2] مطرح گردید[۲] و بعد

---

[1] V.C. Shields and A.V. Oppenheim

[2] T.W. Parsons

توسط دیگران کاملتر شد[۳و۴و۵]. کواتیری و دانیسویچ[۳] استفاده از مدل سینوسیِ صحبت را در حل این مسأله آزمودند[۶] و برخی کار آنها را دنبال کردند[۷و۸]. ویژگی همهٔ این روشها آنست که عمل جداسازی، در حوزهٔ فرکانس انجام می‌شود و محدود بودن طول فریمها (ناشی از طبیعت غیرایستانِ صحبت)، باعثِ کاهش قدرت تفکیک در حوزهٔ فرکانس و همپوشانی پیکها و بالاخره دشواریِ عمل جداسازی می‌گردد.

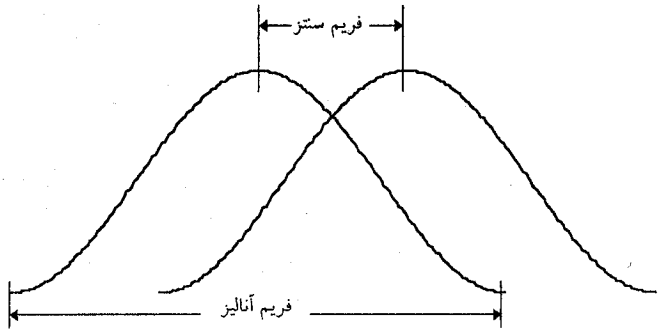
روش معرفی شده در این مقاله، عمل جداسازی را در حوزهٔ زمان انجام می‌دهد و در نتیجه مشکل قدرت تفکیک کم در حوزهٔ فرکانس، تا حد زیادی کاهش می‌یابد. همچنین، قضیه‌ای مطرح خواهیم کرد که با استفاده از آن و پذیرفتن درصدی خطا، می‌توان حجم محاسبات را به مقدار زیادی کاهش داد. نتایج بدست‌آمده نیز حاکی از موفقیت این روش در عمل هستند.

۲ - مدل سینوسی صحبت

در این مدل، سیگنال صحبت در هر لحظه، بصورت جمع تعدادی سینوسی با دامنه، فاز و فرکانسهای مختلف مدل می‌شود[۹]، یعنی:

$$x(t) = \sum_{k=1}^{M} a_k \cos(\omega_k t + \varphi_k) \qquad (۱)$$

یکی از مزایای این مدل، مستقل بودن طول فریمهای سنتز و آنالیز از یکدیگر است (شکل ۱). هنگام تجزیهٔ صحبت، پارامترهای مدل در طول یک فریم آنالیز محاسبه شده و به نقطهٔ مرکزیِ فریم نسبت داده می‌شود. سپس در هنگام سنتز، با توجه به معلوم بودن پارامترها در دو انتهای فریمِ سنتز، مقادیر آنها در نقاط میانی درونیابی می‌گردد. نشان داده شده است[۹] که برای درونیابیِ مقادیر دامنه، می‌توان از درونیابی خطی استفاده کرد، در حالیکه برای محاسبهٔ فاز و فرکانس در نقاطِ میانی، باید از یک چندجمله‌ایِ درجهٔ سه بهره برد.



شکل (۱): فریمهای سنتز و آنالیز در مدل سینوسی صحبت

توجه شود که برای نواحیِ واکدارِ صحبت، فرکانسهای این مدل، هارمونیکهای فرکانسِ اصلیِ (گامِ) صحبت هستند.

---

[3] T.F. Quatieri and R.G. Danisewics

مدل فوق را می‌توان به سادگی برای دو صحبتِ نیز تعمیم داد[۶]. در اینحالت داریم:

$$x[n]=\sum_{k=1}^{M_a} a_k \cos(\omega_{a,k} n+\varphi_{a,k})+\sum_{k=1}^{M_b} b_k \cos(\omega_{b,k} n+\varphi_{b,k}) \qquad (۲)$$

استخراج فرکانسها از روی صحبتِ مخلوط، عملی دشوار است و در این مقاله (همچون بیشترِ مقاله‌های قبلی)، بمنظور شبیه‌سازیِ روش ارائه شده، آنها را *از قبل* معلوم فرض می‌نماییم.

۳- جداسازی دو صحبتِ مخلوط

برای جداکردن دو صحبت از یکدیگر، قضیهٔ زیر را بکار می‌گیریم[۱۰] :

قضیه: *فرض کنید سیگنال $x[n]$ جمع تعدادی سیگنال سینوسی با فرکانسهای متفاوت باشد، یعنی:*

$$x[n]=\sum_{k=1}^{M_a} a_k \cos(\omega_{a,k} n+\varphi_{a,k})+\sum_{k=1}^{M_b} b_k \cos(\omega_{b,k} n+\varphi_{b,k})$$

علاوه بر آن فرض کنید فرکانسِ تعدادی از سینوسی‌های تشکیل دهندهٔ سیگنال فوق، مثلاً $\omega_{a,k}$ها را می‌دانیم و می‌خواهیم دامنه و فازهای آنها را بدست آوریم. در این صورت اگر سیگنال زیر را با دامنه و فازهای مجهول تشکیل دهیم:

$$y[n]=\sum_{k=1}^{M_a} a'_k \cos(\omega_{a,k} n+\varphi'_{a,k})$$

و مقادیرِ این دامنه و فازهای مجهول را بگونه‌ای تعیین کنیم که میانگین $|x[n]-y[n]|^2$ حداقل شود، آنگاه همان مقادیر دامنه و فاز واقعی را بدست خواهیم آورد، یعنی:

$$a'_k=a_k \ , \ \varphi'_{a,k}=\varphi_{a,k} \qquad ,k=1,\ldots,M_a$$

با توجه به قضیهٔ بالا، برای یافتنِ یکی از صحبتها، کافی است یک ترکیب خطی از فرکانسهای تشکیل دهندهٔ آن صحبت را، با دامنه و فازهای مجهول، تشکیل داده و مقادیر این دامنه و فازها را بگونه‌ای تعیین کنیم که میانگین مربعات خطای سیگنالِ حاصل با سیگنال مخلوط حداقل شود. در اینصورت طبق قضیهٔ بالا، سیگنال بدست آمده، همان صحبتِ موردِ نظر خواهد بود.

برای محاسبهٔ این دامنه و فازها، از فرم تربیعیِ[۴] سیگنال سینوسی استفاده کرده و ابتدا سعی می‌کنیم خطا را صفر نماییم. در اینصورت به دستگاه معادلات زیر خواهیم رسید:

---

[4] Quadrature Form

$$\mathbf{H}\cdot\mathbf{c}=\mathbf{s}, \tag{۳}$$

که در آن:

$$\mathbf{H}=\begin{bmatrix} 1 & 1 & 1 & \cdots & 0 & 0 & \cdots \\ 1 & \cos(\omega_{a,1}\cdot 1) & \cos(\omega_{a,2}\cdot 1) & \cdots & \sin(\omega_{a,1}\cdot 1) & \sin(\omega_{a,2}\cdot 1) & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ 1 & \cos(\omega_{a,1}\cdot(N-1)) & \cos(\omega_{a,2}\cdot(N-1)) & \cdots & \sin(\omega_{a,1}\cdot(N-1)) & \sin(\omega_{a,2}\cdot(N-1)) & \cdots \end{bmatrix} \tag{۴}$$

و:

$$\mathbf{c}=\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \beta_1 \\ \beta_2 \\ \vdots \end{bmatrix}, \quad \mathbf{s}=\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} \tag{۵}$$

دستگاه بالا را باید به مفهوم میانگین مربعات حل کنیم، یعنی $\mathbf{c}$ را طوری بدست آوریم که $\|\mathbf{H}\cdot\mathbf{c}-\mathbf{s}\|^2$ حداقل شود. می‌توان ثابت کرد[۱۱] که در اینصورت $\mathbf{c}$ از رابطهٔ زیر بدست می‌آید:

$$\mathbf{c}=\left(\mathbf{H}^*\mathbf{H}\right)^{-1}\mathbf{H}^*\cdot\mathbf{s} \tag{۶}$$

پس از محاسبهٔ $\mathbf{c}$، داریم:

$$\begin{aligned} a_k &= \sqrt{\alpha_k^2+\beta_k^2} \\ \varphi_{a,k} &= \tan^{-1}(\beta_k/\alpha_k) \end{aligned} \tag{۷}$$

و در نتیجه صحبتِ اول بازسازی می‌شود.

۴ - درونیابی چندفریمی

گاهی اوقات برخی از فرکانسهای صحبت دوم، بیش از اندازه به فرکانسهای صحبتِ اول نزدیک می‌شوند و این پدیده باعث ایجاد خطا در محاسبهٔ دامنه و فاز متناظر می‌گردد. برای رفع این مشکل، فاصلهٔ فرکانسهای دو نفر را همواره تحت مراقبت قرار می‌دهیم و در صورت وقوع پدیدهٔ فوق، فرکانس مربوطه را به نشانهٔ غیرقابل اعتماد بودن دامنه علامت‌گذاری می‌کنیم. در نهایت هنگام بازسازی صحبت، وقتی به چنین بلوکهایی رسیدیم، بجای آنکه تنها یک فریم سنتز را در نظر

بگیریم، فریمهای قبلی و بعدی را نیز تا جاییکه به دامنهٔ قابل اعتمادی برسیم با هم به عنوان یک فریمِ سنتزِ طولانی‌تر در نظر گرفته و درونیابیِ پارامترها را در این فریم طولانی انجام می‌دهیم. به این عمل درونیابی چندفریمی می‌گوییم. استفاده از این روش باعث افزایش کیفیت صداها می‌گردد.

۵- نتایج

با استفاده از روش مطرح شده، سیستمی برای جداکردن دو صحبت پیاده‌سازی گردید. برای این منظور از صحبتها بطور جداگانه و با فرکانس ۱۰KHz نمونه‌برداری کردیم. سپس فرکانسهای مدل سینوسیِ آنها محاسبه گردیده و در فایلهای جداگانه‌ای ذخیره شدند. همچنین گام صحبتها را نیز بطور جداگانه محاسبه نمودیم.

در مرحلهٔ اول با استفاده از فرکانسهای مدل سینوسیِ هرکدام از صحبتها، یکبار بدون استفاده از روش درونیابی چندفریمی و بار دیگز با استفاده از آن، مخلوط آنها را جداکردیم. سپس چندین نفر، به عنوان داور، به این صحبتها گوش کرده و دو نمره، یکی برای میزان جداسازیِ صحبتها، و دیگری برای کیفیت صحبتهای سنتزشده ثبت کردند. همگی افراد به میزان جداسازی نمرهٔ کامل (۴) دادند. میانگین نمرهٔ داده شده به کیفیت صحبتها، در حالت استفاده از درونیابی چندفریمی، ۲/۹ و در حالت استفاده نکردن از آن، ۲/۱ بود. این نکته، کارآیی درونیابی چند فریمی را در بهبود کیفیت، نشان می‌دهد.

در مرحلهٔ بعد بجای معلوم بودنِ همهٔ فرکانسهای مدل سینوسی، تنها فرکانسِ اصلی (گام) آنها معلوم فرض شد و بقیهٔ فرکانسهای مدل، هارمونیکهای آن فرض گردید. نتیجهٔ حاصل از بررسی این صحبتها نشان داد که میزان جداسازی از ۴ به ۲/۳ افت پیدا کرده است.

در مرحلهٔ سوم، بجای استفاده از قضیهٔ مطرح شده، ترکیب خطی از کلیهٔ فرکانسها بکار برده شد (حجم محاسبات در اینحالت حدود ۸ برابر بیشتر است). هیچکدام از داوران تفاوتی بین کیفیت صحبتهای بازسازی شده با صحبتهای مرحلهٔ اول تشخیص ندادند و این، کارآیی قضیهٔ مطرح شده را در کاهش حجم محاسبات نشان می‌دهد.

۶- خاتمه

در این مقاله پس از معرفی مدل سینوسی، روشی برای جداسازی دو صحبتِ درهم معرفی گردید. ویژگی این روش، انجام همهٔ محاسبات در حوزهٔ زمان است. همچنین قضیه‌ای برای افزایش سرعت محاسبه بیان شد. در نهایت یک مطالعهٔ subjective برای بررسی کارآیی روش انجام گردید که این مطالعه، کارآیی روش معرفی شده را بخوبی نشان داد.

مراجع

[1] V.C. Shields, "Separation of Added Speech Signals by Digital Comb Filtering," M.S. Thesis, MIT, 1970.

[2] P.W. Parsons, "Separation of Simultaneous Vocalic Utterances of Two Talkers," Ph.D. Thesis, Department of Elec. Eng., Polytech. Inst. of New York, 1975.

[3] B.A. Hanson and D.Y. Wong, "The Harmonic Magnitude Supression (HMS) Technique for Intelligibility Enhan 1984, 18A.5.1-18A.5.4.

[4] D.G. Childers and C.K. Lee, "Co-Channel Speech Separation," ICASSP 1987, 6.4.1-6.4.4.

[5] J.A. Naylor ans
Cahnnel Speech," ICASSP 1987, 6.12.1-6.12.4.

[6] T.F. Quatieri and R.G. Danisewics, "An Approach to Co-Channel Talker Interference Supression Using a Sinusoidal Model for Speech," IEEE Trans. on ASSP, Vol. 38, pp.56-69, January 1990.

[7] F.M. Silva and L.B. Almeida, "Speech Separation by Means of Stationary Least Squares Harmonic Estimation," ICASSP 1990, pp. 809-812.

[8] J.A. Naylor and J. Porter, "An Effective Speech Separation System witch Requires No Priori Information," ICASSP 1991, pp. 937-940.

[9] R.J. McAulay and T.F. Quatieri, "Speech Analysis/Synthesis Based on A Sinusoidal Representation," IEEE trans. on ASSP, vol. 34, pp. 744-754, Aug. 1986.

[۱۰] بابایی‌زاده، مسعود، "جداسازی دو سیگنال صحبت مخلوط"، رسالهٔ کارشناسی ارشد، دیماه ۱۳۷۵، دانشکدهٔ برق دانشگاه صنعتی شریف.

[11] Kay, S.M., *Modern Spectral Estimation*, Prentice Hall, 1988.