

بهسازی گفتار با تخمین گر کمترین میانگین مربعات

خطا بر پایه توزیع مخلوط لاپلاس برای گفتار

زینب محمدپوری، حسین مروی و امیدرضا معروضی
دانشگاه شاهرود، دانشکده مهندسی برق، شاهرود، ایران.

چکیده

در این مقاله یک روش بهسازی گفتار آماری با فرض توزیع مخلوط لاپلاس برای گفتار، جهت تخمین سیگنال گفتار تمیز (بدون نوفه) از سیگنال گفتار نوفه‌ای ارائه شده است. در روش پیشنهادی، ضرایب تبدیل فوریه زمان کوتاه گسسته سیگنال گفتار با استفاده از تخمین گر کمترین میانگین مربعات خطا، به دست می‌آید. در این تخمین، فرض می‌شود تابع چگالی احتمال ضرایب تبدیل فوریه سیگنال تمیز و نوفه به ترتیب، مخلوط لاپلاس و گوسی با میانگین صفر است. نتایج حاصل از معیارهای SNR قطعه‌ای، LLR و PESQ نشان می‌دهد که روش پیشنهادی عملکرد بهتری نسبت به دو روش مبتنی بر توزیع گوسی و روش مبتنی بر توزیع لاپلاس دارد.

واژگان کلیدی: تخمین گر MMSE، الگوریتم EM، توزیع مخلوط لاپلاس، توزیع گوسی.

۱- مقدمه

در سال‌های اخیر روش‌های زیادی برای بهسازی سیگنال گفتار در حوزه زمان و فرکانس ارائه شده است؛ زیرا بهسازی گفتار کاربردهای وسیع و متفاوتی در حوزه‌های مختلف مانند تشخیص خودکار گفتار^۱، تشخیص هویت گوینده^۲، کدینگ، وسایل کمک شنوایی مانند سمعک، سامانه‌های مخابرات هوا به زمین، تلفن‌های عمومی و ... دارند. از مهم‌ترین و معروف‌ترین این روش‌ها می‌توان به تفریق طیفی^۳ (بول، ۱۹۷۹)، زیرفضای سیگنال^۴ (افرایم و همکاران، ۱۹۹۵) و روش‌های آماری^۵ (افرایم و همکاران ۱۹۸۹، افرایم و همکاران، ۱۹۸۴؛ مک آلی و همکاران، ۱۹۸۰؛ لیم و همکاران، ۱۹۷۹) اشاره کرد که روش‌های آماری به لحاظ عملکرد بهتر و ایجاد اعوجاج کم‌تر در سیگنال نهایی نسبت به سایر روش‌ها بیشتر مورد توجه هستند (لوئیزو، ۲۰۰۷، ص ۸۷).

در این روش‌ها یک توزیع آماری برای گفتار تمیز و یک توزیع آماری هم برای نوفه در نظر گرفته می‌شود؛ و سپس با استفاده از یک تخمین گر، مانند بیش‌ترین شباهت (ML) (مک آلی و همکاران، ۱۹۸۰)، کم‌ترین میانگین مربعات خطا^۶ (مارتین، ۲۰۰۵؛ افرایم و همکاران ۱۹۸۴؛ افرایم و همکاران ۱۹۸۵، ...) یا بیش‌ترین احتمال پسین (MAP)^۷ (لاتر و همکاران، ۲۰۰۵؛ ولف و همکاران، ۲۰۰۱، ولف و همکاران، ۲۰۰۳a)، سیگنال تمیز از روی سیگنال نوفه‌ای تخمین زده می‌شود.

یکی از فرض‌های ابتدایی و اساسی برای تخمین سیگنال تمیز در حوزه فرکانس، گوسی بودن تابع توزیع چگالی احتمال ضرایب طیف سیگنال تمیز بوده است (افرایم و همکاران ۱۹۸۴؛ افرایم و همکاران ۱۹۸۵). فرض گوسی بودن ضرایب براساس قضیه حد مرکزی^۸ است؛ چون هر ضریب فوریه زمان کوتاه در یک قاب^۹ با طول N ، برابر

⁶ - Maximum Likelihood

⁷ - Minimum Mean Square Error (MMSE)

⁸ - Maximum A Posteriori

⁹ - Central Limit Theorem

¹⁰ - Frame

¹ - Automatic Speech Recognition

² - Speaker Identification

³ - Spectral Subtraction

⁴ - Signal Subspace

⁵ - Statistical Method

برای به دست آوردن پارامترهای مدل مخلوط لاپلاس از الگوریتم EM⁴ استفاده شده و برای محاسبه واریانس توزیع نوفه هم روش آمارگان کمینه⁵ به کار رفته است. بخش بندی این مقاله بدین صورت است: بخش بعدی به بررسی توزیع سیگنال گفتار و فرض‌های به کاررفته در روش پیشنهادی می‌پردازد. بخش سوم نیز درباره محاسبه تخمین گر MMSE با فرض تابع چگالی احتمال مخلوط لاپلاس برای گفتار است. در بخش چهارم هم به بررسی الگوریتم EM برای محاسبه پارامترهای توزیع مخلوط لاپلاس پرداخته می‌شود و بخش پنجم هم به ارزیابی نتایج و مقایسه روش پیشنهادی با روش‌های دیگر می‌پردازد.

۲- فرضیات پایه در روش پیشنهادی

فرض اصلی در روش‌های آماری این است که سیگنال و نوفه با هم جمع شده‌اند و این فرض در بسیاری از محیط‌های واقعی فرض درستی است. یعنی:

$$y(i) = s(i) + n(i) \quad (1)$$

که در آن y سیگنال نوفه‌ای، s سیگنال بدون نوفه، n نوفه و i شماره نمونه هستند. همچنین فرض می‌شود که سیگنال و نوفه از هم مستقل‌اند؛ سپس سیگنال‌ها به وسیله تبدیل فوریه گسسته زمان کوتاه به حوزه فرکانس برده می‌شوند؛ یعنی سیگنال‌ها به قاب‌هایی که با هم هم‌پوشانی⁶ دارند تقسیم شده و سپس قاب‌ها در یک پنجره ضرب شده و از آنها تبدیل فوریه گرفته می‌شود؛ چون خاصیت جمع‌شوندگی در حوزه فرکانس هم‌چنان صحیح است، می‌توان نوشت:

$$Y(\mu, k) = S(\mu, k) + N(\mu, k) \quad (2)$$

که در آن μ شماره قاب و k اندیس متناظر با فرکانس $2\pi k/L$ هستند. (L تعداد نمونه‌های یک قاب است).

فرض دیگر که باعث ساده شدن تخمین می‌شود، مستقل بودن مؤلفه‌های طیف سیگنال است؛ زیرا به خاطر مستقل بودن مؤلفه‌ها، تخمین MMSE ضرایب طیف سیگنال تمیز یعنی $S(k, \mu)$ فقط از روی $Y(k, \mu)$ انجام می‌شود (مارتین، ۲۰۰۳).

مجموع وزن دار N متغیر تصادفی در آن قاب است. از طرفی این فرض هنگامی صحیح است که طول قاب زیاد (بیش‌تر از صد میلی ثانیه برای سیگنال گفتار) باشد؛ ولی در کاربردهای پردازش گفتار، طول هر قاب کوتاه (بین بیست تا چهل میلی ثانیه) است. هم‌چنین با توجه به هیستوگرام بخش حقیقی و موهومی ضرایب تبدیل فوریه گفتار، استفاده از توزیع‌های غیرگوسی مانند گاما و لاپلاس منجر به نتایج بهتری خواهد شد. البته فرض گوسی بودن برای ضرایب تبدیل فوریه نوفه، فرض مناسبی است (پاپیولیس و همکاران، ۲۰۰۲، ص ۲۵۳؛ بریلینگر، ۱۹۸۱، ص ۱۲۰).

این مسأله نخستین بار در (پورتر و همکاران، ۱۹۸۴) مطرح شد و سپس در مقالات زیادی به کار گرفته شد. در نخستین پژوهش‌ها در این زمینه، توزیع لاپلاس یا گاما برای ضرایب مختلط تبدیل فوریه گفتار تمیز در نظر گرفته شده و سپس با استفاده از تخمین گر MMSE، این ضرایب تخمین زده شدند (مارتین، ۲۰۰۵؛ مارتین و همکاران، ۲۰۰۳). در برخی دیگر از مقالات، با فرض توزیع گاما یا لاپلاس برای ضرایب مختلط طیف گفتار، دامنه ضرایب تبدیل فوریه گفتار تمیز با تخمین گرهای MMSE یا MAP تخمین زده شده است. (لاتر و همکاران، ۲۰۰۵؛ چن و همکاران، ۲۰۰۷)

بعضی از پژوهش‌گران نیز توزیع‌های مخلوط مانند توزیع مخلوط گوسی^۱، توزیع مخلوط رایلی^۲ و ... را برای مدل کردن توزیع ضرایب یا دامنه طیف در نظر گرفته‌اند. (هوانگ و همکاران، ۲۰۰۷؛ اکلنس و همکاران، ۲۰۰۷؛ چاترجی و همکاران، ۲۰۰۸)

در برخی از مقالات، برای نوفه هم از توزیع‌های غیر گوسی استفاده شده است (مارتین و همکاران، ۲۰۰۳).

در روش پیشنهادی ما از توزیع مخلوط لاپلاس (LMM)^۳ (مجموع چند لاپلاس با واریانس‌ها و میانگین‌های متفاوت) برای مدل کردن توزیع ضرایب مختلط طیف گفتار استفاده شده است؛ زیرا این توزیع نسبت به توزیع‌های ذکر شده، به هیستوگرام بخش حقیقی و موهومی ضرایب تبدیل فوریه گفتار نزدیک‌تر بوده و بهتر و دقیق‌تر بر آن برآزنده می‌شود. طیف نوفه نیز با توزیع گوسی مدل شده و از تخمین گر MMSE برای تخمین ضرایب طیف استفاده شده است.

⁴ - Expectation Maximization

⁵ - Minimum Statistic

⁶ - overlap

¹ - Gaussian Mixture Model

² - Rayleigh Mixture Model

³ - Laplacian Mixture Model

همچنین با افزایش تعداد لاپلاس‌های توزیع مخلوط لاپلاس (N)، مقدار این معیار کاهش می‌یابد که منطقی نظر می‌رسد؛ اما پس از $N=30$ ، تغییرات مقدار این معیار کاهش می‌یابد و به نظر می‌رسد که این تعداد، مقدار مناسبی می‌باشد؛ زیرا هرچند با افزایش N تقریب دقیق‌تری داریم ولی حجم و زمان محاسبات نیز افزایش می‌یابد. پس باید مقداری از N را به‌طور تجربی یافت که در آن علاوه‌بر نتایج مناسب، پیچیدگی محاسباتی و زمان اجرای الگوریتم نیز مطلوب باشند.

شکل (۱-۲) هیستوگرام بخش حقیقی ضرایب تبدیل فوریه سیگنال گفتار تمیز (۵۰ جمله از پایگاه داده TIMIT از ۲۵ مرد و ۲۵ زن) و توزیع‌های لاپلاس، مخلوط شش لاپلاس و مخلوط شش گوسی تقریب زده شده از روی این داده‌ها را نشان می‌دهد. با توجه به شکل مخلوط شش لاپلاس با دقت بهتری توزیع واقعی را تقریب می‌زند. توزیع لاپلاس به‌صورت زیر نمایش داده می‌شود:

$$L(x; c, m) = ce^{-2c|x-m|} \quad (4)$$

که در آن $\frac{1}{c}$ بیان‌گر واریانس و m بیانگر میانگین توزیع است. توزیع مخلوطی از لاپلاس‌ها نیز بدین صورت تعریف می‌شود:

$$p(\theta) = \sum_{i=1}^N \alpha_i c_i e^{-2c_i|\theta_n - \theta_i|} \quad (5)$$

که در آن α_i ، θ_i و c_i به ترتیب ضریب، میانگین، و واریانس هر کدام از لاپلاس‌ها و N تعداد لاپلاس‌هاست و داریم:

$$\sum_{i=1}^N \alpha_i = 1 \quad (6)$$

متداول‌ترین روش برای تخمین این پارامترها، الگوریتم EM است (دمپستر و همکاران، ۱۹۷۷؛ بیلمز، ۱۹۹۸) که در بخش بعدی توضیح داده می‌شود. با توجه به توضیحات، می‌توان نوشت:

$$p(S_R) = \sum_{i=1}^N \alpha_{1i} c_i e^{-2c_{1i}|S_R - m_{1i}|} \quad (7)$$

$$(S_I) = \sum_{j=1}^N \alpha_{2j} c_j e^{-2c_{2j}|S_I - m_{2j}|} \quad (8)$$

برای ساده‌شدن نمایش روابط، تبدیل فوریه زمان کوتاه سیگنال نوفه‌ای، سیگنال تمیز و نوفه در قاب μ ام و فرکانس k ام، به ترتیب با S, Y و N نمایش داده می‌شوند. Y_I و Y_R ، S_I و S_R ، N_I و N_R به ترتیب نمایش‌دهنده بخش حقیقی و موهومی سیگنال نوفه‌ای، سیگنال تمیز و نوفه هستند.

۱-۲- مدل مخلوط لاپلاس برای ضرایب DFT سیگنال گفتار تمیز

همان‌طور که در قبل گفته شد، تابع چگالی احتمال ضرایب طیف سیگنال گفتار برای قاب‌های کوتاه (۲۰-۴۰ میلی ثانیه) بهتر است با توزیع‌های غیرگوسی مدل شود. در روش پیشنهادی فرض می‌شود که این توزیع، مخلوطی از لاپلاس‌هاست. زیرا هیستوگرام‌های بخش‌های حقیقی و موهومی ضرایب تبدیل فوریه زمان کوتاه گفتار تمیز شباهت زیادی به توزیع لاپلاس دارند؛ ولی مانند توزیع لاپلاس حول محور عمودی متقارن نیستند؛ پس نمی‌توان آنها را با یک لاپلاس به‌طور دقیق تقریب زد. به همین دلیل استفاده از مجموع چند لاپلاس برای این تقریب پیشنهاد می‌شود.

با معیار I_{KL} هم می‌توان نشان داد که توزیع مخلوط لاپلاس از توزیع گوسی، مخلوط گوسی و لاپلاس تنها برای تقریب ضرایب طیف گفتار مناسب‌تر است. این معیار بدین صورت تعریف می‌شود:

$$I_{KL} = \sum_x p_H(x) \log\left(\frac{p_H(x)}{p(x)}\right) \quad (3)$$

که در آن $p_H(x)$ هیستوگرام داده‌ها و $p(x)$ توزیع در نظر گرفته برای تقریب‌زدن آنهاست. در مقایسه دو توزیع، هرچه مقدار این معیار کمتر باشد، نشان می‌دهد که توزیع مفروض تقریب بهتری برای هیستوگرام داده‌هاست و برای مدل کردن توزیع واقعی مناسب‌تر است (کول بک، ۱۹۹۷).

پس از محاسبه این معیار برای چهار توزیع گوسی، مخلوط شش گوسی، یک لاپلاس و مخلوط شش لاپلاس، و هیستوگرام بخش حقیقی و موهومی ضرایب تبدیل فوریه زمان کوتاه تعدادی از جملات پایگاه داده TIMIT، کم‌ترین مقدار به‌دست آمده مربوط به توزیع مخلوط شش لاپلاس است که نشان می‌دهد به‌طور دقیق‌تری طیف سیگنال تمیز را مدل می‌کند. این مقادیر برای توزیع گوسی ۴۲۴۹۱، برای مخلوط شش گوسی ۱۹۸۷۵، برای شش لاپلاس ۱۴۵۵۰ و برای یک لاپلاس ۲۸۲۵۰ است.

¹ Kullback-Leibler discrimination information

سیگنال بهسازی شده در حوزه زمان به دست آید. سیگنال به دست آمده به صورت قاب‌هایی با هم‌پوشانی است که با اعمال روش هم‌پوشانی-جمع^۱ (حذف هم‌پوشانی قاب‌ها و ادغام آنها) بر روی آنها، سیگنال بهسازی شده نهایی به دست خواهد آمد (مارتین، ۲۰۰۳؛ مارتین، ۲۰۰۵). تخمین MMSE بخش حقیقی به صورت زیر است:

$$E(S_R|Y_R) = \frac{\int_{-\infty}^{+\infty} S_R P(Y_R|S_R) P(S_R) dS_R}{P(Y_R)} \quad (12)$$

که با جای‌گذاری تابع چگالی احتمال نوفه و گفتار تمیز (۹ و ۷) در آن، رابطه زیر به دست می‌آید:

$$E(S_R|Y_R) = \frac{1}{\sqrt{\pi}\sigma_n P(Y_R)} \int_{-\infty}^{+\infty} S_R \exp\left(-\frac{(Y_R - S_R)^2}{\sigma_n^2}\right) \sum_{i=1}^N \alpha_{i1} c_{i1} \exp(-2c_{i1}|s_R - m_{i1}|) ds_R \quad (13)$$

با استفاده از قضیه 3.462,1 در (گراشدتاین و همکاران، ۱۹۹۴)، و تغییر متغیرهای زیر:

$$S_R - m_{i1} = t_i \Rightarrow S_R = t_i + m_{i1} \quad \text{و} \quad ds_R = dt_i \quad (14)$$

پس از انجام محاسبات، رابطه زیر به دست می‌آید:

$$E(S_R|Y_R) = \frac{1}{2P(Y_R)} \sum_{i=1}^N \alpha_{i1} c_{i1} \exp(c_{i1}^2 \sigma_n^2) \{ \sigma_n [L_{R1}(Y_R - m_{i1}) \exp(2c_{i1}(Y_R - m_{i1})) \operatorname{erfc}(L_{R1}(Y_R - m_{i1})) - L_{R2}(Y_R - m_{i1}) \exp(-2c_{i1}(Y_R - m_{i1})) \operatorname{erfc}(L_{R2}(Y_R - m_{i1}))] + m_{i1} [\exp(2c_{i1}(Y_R - m_{i1})) \operatorname{erfc}(L_{R1}(Y_R - m_{i1})) + \exp(-2c_{i1}(Y_R - m_{i1})) \operatorname{erfc}(L_{R2}(Y_R - m_{i1}))] \} \quad (15)$$

که در آن $P(Y_R)$ برابر است با:

$$P(Y_R) = \int_{-\infty}^{+\infty} P(S_R) P(Y_R|S_R) dS_R \quad (16)$$

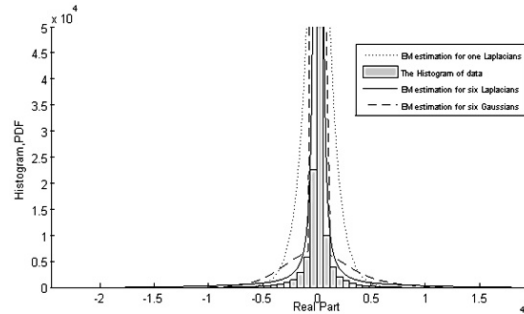
رابطه قبل با استفاده از قضیه 3.322,2 در (گراشدتاین و همکاران، ۱۹۹۴)، تغییر متغیر قبل و انجام محاسبات به صورت زیر خواهد بود:

^۱ -Overlap-Add

برای نوفه نیز توزیع گوسی با میانگین صفر در نظر گرفته می‌شود:

$$p(N_R) = \frac{1}{\sqrt{\pi}\sigma_n} \exp\left(-\frac{(N_R)^2}{\sigma_n^2}\right) \quad (9)$$

$$p(N_I) = \frac{1}{\sqrt{\pi}\sigma_n} \exp\left(-\frac{(N_I)^2}{\sigma_n^2}\right) \quad (10)$$



شکل (۲-۱): هیستوگرام بخش حقیقی ضرایب تبدیل فوریه سیگنال گفتار تمیز و توزیع‌های لاپلاس، مخلوط شش لاپلاس و مخلوط شش گوسی تقریب‌زده شده از روی این داده‌ها: خط مشکی توزیع مخلوط شش لاپلاس، نقطه‌چین توزیع لاپلاس و خط چین توزیع مخلوط شش گوسی

که در آن $\frac{\sigma_n^2}{2}$ واریانس بخش حقیقی و موهومی ضرایب تبدیل فوریه زمان کوتاه گسسته سیگنال نوفه را مشخص می‌کند.

۳- محاسبه سیگنال تمیز با تخمین گر MMSE

تخمین MMSE سیگنال تمیز (S) از روی سیگنال نوفه‌ای (Y) از رابطه $E(S|Y)$ به دست می‌آید (عمل گر E بیان‌گر امید ریاضی است) که با فرض مستقل بودن بخش حقیقی و موهومی ضرایب تبدیل فوریه می‌توان این تخمین را به صورت دو تخمین جداگانه بخش حقیقی و موهومی به صورت زیر در نظر گرفت:

$$E(S|Y) = E(S_R|Y_R) + jE(S_I|Y_I) \quad (11)$$

در نتیجه با تخمین جداگانه هر کدام از بخش‌ها و ترکیب آنها، تخمین نهایی به دست می‌آید؛ سپس از این مقادیر تبدیل معکوس فوریه زمان کوتاه گرفته می‌شود تا

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt = 1 - \operatorname{erfc}(x) \quad (18)$$

و $L_{R1}(Y_R)$ و $L_{R2}(Y_R)$ به صورت زیر تعریف می شوند:

$$L_{R1}(Y_R) = c\sigma_n + \frac{Y_R}{\sigma_n} \quad (19)$$

$$L_{R2}(Y_R) = c\sigma_n - \frac{Y_R}{\sigma_n} \quad (20)$$

$$E(S_R|Y_R) =$$

$$\frac{\sum_{i=1}^N \alpha_i c_i \exp(c_i^2 \sigma_n^2) \left[\sigma_n \left(L_{R1}(Y_R - m_{1i}) \exp(2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R1}(Y_R - m_{1i})) - L_{R2}(Y_R - m_{1i}) \exp(-2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R2}(Y_R - m_{1i})) \right) + m_{1i} \left(\exp(2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R1}(Y_R - m_{1i})) + \exp(-2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R2}(Y_R - m_{1i})) \right) \right]}{\sum_{i=1}^N \alpha_i c_i \exp(c_i^2 \sigma_n^2) \left[\exp(2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R1}(Y_R - m_{1i})) + \exp(-2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R2}(Y_R - m_{1i})) \right]} \quad (21)$$

$$E(S_I|Y_I) =$$

$$\frac{\sum_{i=1}^N \alpha_i c_i \exp(c_i^2 \sigma_n^2) \left[\sigma_n \left(L_{R1}(Y_I - m_{2i}) \exp(2c_{2i}(Y_I - m_{2i})) \operatorname{erfc}(L_{R1}(Y_I - m_{2i})) - L_{R2}(Y_I - m_{2i}) \exp(-2c_{2i}(Y_I - m_{2i})) \operatorname{erfc}(L_{R2}(Y_I - m_{2i})) \right) + m_{2i} \left(\exp(2c_{2i}(Y_I - m_{2i})) \operatorname{erfc}(L_{R1}(Y_I - m_{2i})) + \exp(-2c_{2i}(Y_I - m_{2i})) \operatorname{erfc}(L_{R2}(Y_I - m_{2i})) \right) \right]}{\sum_{i=1}^N \alpha_i c_i \exp(c_i^2 \sigma_n^2) \left[\exp(2c_{2i}(Y_I - m_{2i})) \operatorname{erfc}(L_{R1}(Y_I - m_{2i})) + \exp(-2c_{2i}(Y_I - m_{2i})) \operatorname{erfc}(L_{R2}(Y_I - m_{2i})) \right]} \quad (22)$$

$$\alpha_i = \frac{1}{T} \sum_{n=1}^T p(i|\theta_n) \quad (23)$$

$$\theta_i = \frac{\sum_{n=1}^T \frac{\theta_n}{|\theta_n - \theta_i|} p(i|\theta_n)}{\sum_{n=1}^T \frac{1}{|\theta_n - \theta_i|} p(i|\theta_n)} \quad (24)$$

$$c_i = \frac{\sum_{n=1}^T p(i|\theta_n)}{2 \sum_{n=1}^T |\theta_n - \theta_i| p(i|\theta_n)} \quad (25)$$

که در آنها $p(i|\theta_n)$ احتمال تعلق داشتن نمونه θ_n به لاپلاس نام است و از رابطه زیر به دست می آید:

$$P(i|\theta_n) = \frac{\alpha_i c_i e^{-2c_i|\theta_n - \theta_i|}}{\sum_{i=1}^N \alpha_i c_i e^{-2c_i|\theta_n - \theta_i|}} \quad (26)$$

برای مقداردهی اولیه به الگوریتم EM، ابتدا این الگوریتم با مقادیر اولیه تصادفی اجرا و جواب های آن به عنوان مقدار اولیه به الگوریتم EM اصلی داده شد. سپس پاسخ نهایی این الگوریتم به عنوان پارامترهای توزیع مخلوط لاپلاس سیگنال صحبت به تخمین گر MMSE داده شد.

نمودار جریانی روش پیشنهادی در شکل (۴-۱) آورده شده است. در روش پیشنهادی، ابتدا پارامترهای توزیع مخلوط لاپلاس برای سیگنال تمیز با استفاده از الگوریتم EM محاسبه می شود؛ سپس از سیگنال نوفه ای تبدیل فوریه زمان کوتاه گرفته شده و از روی این ضرایب واریانس نوفه محاسبه می شود. در انتها با پارامترهای محاسبه شده در مراحل قبل، از روی سیگنال نوفه ای تخمین MMSE ضرایب

$$P(Y_R) = \frac{1}{2} \sum_{i=1}^N \alpha_i c_i \exp(c_i^2 \sigma_n^2)$$

$$\left[\exp(2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R1}(Y_R - m_{1i})) + \exp(-2c_{1i}(Y_R - m_{1i})) \operatorname{erfc}(L_{R2}(Y_R - m_{1i})) \right]$$

در این روابط erfc متمم تابع خطا^۱ است و از رابطه زیر محاسبه می شود:

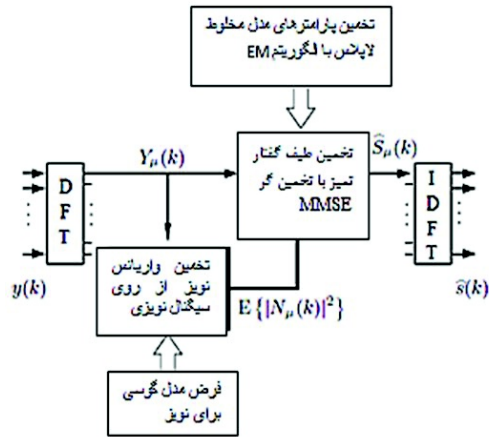
بخش موهومی طیف سیگنال تمیز هم به همین طریق تخمین زده می شود و سپس تخمین MMSE نهایی گفتار تمیز با استفاده از رابطه $E(S|Y) = E(S_R|Y_R) + E(S_I|Y_I)$ به دست می آید که $E(S_I|Y_I)$ و $E(S_R|Y_R)$ از روابط (۲۱) و (۲۲) محاسبه می شوند. با توجه به این دو رابطه، محاسبه تخمین مستلزم محاسبه میانگین ها، واریانس ها، ضرایب لاپلاس ها و واریانس نوفه است که در بخش های بعدی توضیح داده خواهد شد.

۴- تخمین پارامترهای توزیع مخلوط لاپلاس با الگوریتم EM

یکی از مهم ترین و گسترده ترین کاربردهای الگوریتم EM، به دست آوردن پارامترهای مخلوطی از چگالی هاست (دمپستر و همکاران، ۱۹۷۷؛ بیلمز، ۱۹۹۸). با توجه به رابطه (۵) برای توزیع مخلوط لاپلاس و روش ذکر شده در (میتیانودیس و همکاران، ۲۰۰۴)، θ_i و c_i برای هر کدام از لاپلاس ها به ترتیب از روابط زیر محاسبه می شوند: (θ_n نمونه های آموزشی و T تعداد آنها است).

¹ complementary error function

طیف سیگنال صحبت محاسبه شده و به حوزه زمان برده می‌شوند.



شکل (۴-۱): بلوک دیاگرام روش پیشنهادی

و ۱۰ دسی بل به ۱۰ جمله (پنج مرد و پنج زن) اضافه شدند. این جملات در ابتدا به قاب‌هایی با طول 32ms و با هم‌پوشانی ۵۰٪ تقسیم شده و پس از ضرب در پنجره همینگ، از آنها تبدیل فوریه گرفته شد.

برای ارزیابی روش پیشنهادی از سه معیار عینی^۳ سیگنال به نوفه^۴ قطع‌های^۵ (هیو و همکاران، ۲۰۰۶) (SNR_{seg}) ، LLR^۶ (هیو و همکاران، ۲۰۰۶) و PESQ^۷ (لوئیزو، ۲۰۰۷، ص ۴۳۷) استفاده شده است.

سیگنال به نوفه^۸ قطع‌های به صورت زیر تعریف می‌شود:

$$SNR_{seg} = \frac{10}{M} \sum_{f=1}^M \log_{10} \left[\frac{\sum_{n=L.F}^{L.F+L-1} s^2(n)}{\sum_{n=L.F}^{L.F+L-1} [s(n) - \hat{s}(n)]^2} \right] \quad (28)$$

که در آن M تعداد قاب‌ها، L اندازه هر قاب، s(n) سیگنال گفتار بدون نوفه و $\hat{s}(n)$ سیگنال بهسازی شده است.

برای محاسبه سیگنال به نوفه قطع‌های، قاب‌های سکوت در نظر گرفته نمی‌شوند؛ در واقع قاب‌هایی که سیگنال به نوفه آنها کمتر از 10dB است، در محاسبه لحاظ نمی‌شوند. همچنین به سبب اینکه گوش انسان بین قاب‌هایی با مقادیر SNR بالاتر از 35 dB تمایز چندانی نمی‌گذارد، یک آستانه بالا برای متعادل کردن SNR هایی که بسیار بالا هستند، به کار می‌رود. بنابراین قاب‌هایی که سیگنال به نوفه آنها از 35 dB بیشتر است نیز در محاسبه سیگنال به نوفه قطع‌های حذف می‌شوند (هیو و همکاران، ۲۰۰۶).

از آنجایی که این معیار با معیارهای ذهنی همبستگی پایینی دارد، از دو معیار LLR و PESQ هم که همبستگی قوی‌تری با معیارهای ذهنی دارد، برای ارزیابی روش پیشنهادی استفاده شده است (هانس و همکاران، ۱۹۹۸). معیار LLR که یکی از مرسوم‌ترین معیارهای ارزیابی تمام قطب است، به صورت زیر تعریف می‌شود:

$$LLR = \log \left(\frac{a_e^T R_x a_e}{a_x^T R_x a_x} \right) \quad (29)$$

که در آن a_e و a_x به ترتیب ضرایب LPC سیگنال تمیز و سیگنال بهسازی شده و R_x ماتریس همبستگی^۸ سیگنال

۵- آزمایش‌ها و نتایج

۵-۱- ارزیابی روش پیشنهادی

برای پیاده‌سازی روش پیشنهادی از پایگاه داده TIMIT استفاده شده که در آن جملات با نرخ 16kHz نمونه‌برداری شده‌اند. بدین صورت که برای به دست آوردن پارامترهای توزیع مخلوط لاپلاس با استفاده از الگوریتم EM، از دویست جمله که توسط صد مرد و صد زن بیان شده بودند، به عنوان داده‌های آموزشی استفاده شد. برای تخمین واریانس نوفه هم از روش آمارگان کمینه^۱ (مارتین، ۱۹۹۴) استفاده شده است. در این روش، کمینه توان طیف در یک پنجره محدود (به طول قسمت آنالیز) دنبال می‌شود. در روش آمارگان کمینه بعد از محاسبه پرپودوگرام^۲ سیگنال نوفه‌ای، توان این سیگنال با توجه به میزان نوسان بالای پرپودوگرام، به صورت رابطه (۲۷) محاسبه می‌شود. در این رابطه $0 \leq \alpha \leq 1$ یک ضریب هموارسازی است. با ردیابی کمینه^۳ $p(\mu, k)$ بر روی یک پنجره با طول محدود، تخمینی از توان نوفه به دست می‌آید. طول پنجره مورد نظر در (مارتین، ۲۰۰۶) بین ۰/۸ تا ۱/۴ ثانیه پیشنهاد شده است.

$$p(\mu, k) = \alpha p(\mu - 1, k) + (1 - \alpha) |Y(\mu, k)|^2 \quad (27)$$

سپس برای پیاده‌سازی روش پیشنهادی، سه نوفه سفید، همهمه و هواپیمای F16 در سه سیگنال به نوفه ۵، ۰، ۰

³ - Objective Measure

⁴ - Segmental Signal to Noise

⁵ - Log-Likelihood Ratio

⁶ - Perceptual Evaluation of Speech Quality

⁷ - Linear Predictive Coding

⁸ - Autocorrelation

¹ - Minimum Statistic

² - Periodogram

در حالی که پیچیدگی محاسباتی، زمان اجرای الگوریتم و حجم حافظه مورد نیاز افزایش می‌یابد.

۵-۲- مقایسه روش پیشنهادی با روش‌های دیگر

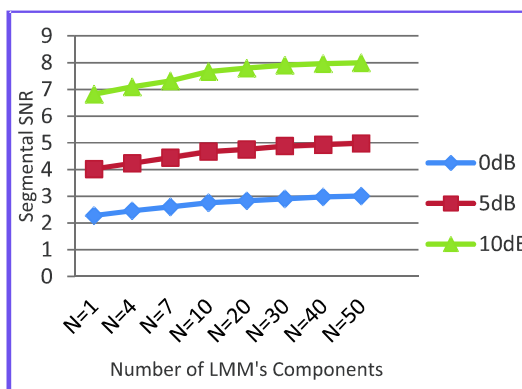
در این بخش روش پیشنهادی با دو روش مبتنی بر توزیع گوسی و روش مبتنی بر توزیع لاپلاس مقایسه می‌شود. در یکی از دو روش مبتنی بر توزیع گوسی فرض می‌شود که توزیع ضرایب طیف سیگنال گفتار گوسی است و با این فرض دامنه و مرحله طیف تبدیل فوریه زمان کوتاه گفتار با تخمین گر MMSE تخمین زده می‌شود (افرایم و همکاران، ۱۹۸۴). این روش با MMSE نشان داده شده است. در روش دیگر با فرض گوسی بودن ضرایب طیف سیگنال گفتار، لگاریتم دامنه طیف تخمین زده می‌شود (افرایم و همکاران، ۱۹۸۵). این روش با Log-MMSE مشخص شده است. در روش مبتنی بر توزیع لاپلاس نیز با فرض توزیع لاپلاس برای ضرایب طیف گفتار تمیز، این ضرایب با تخمین گر MMSE تخمین زده می‌شوند (مارتین، ۲۰۰۳). این روش با Lap-MMSE و روش پیشنهادی (N=30) با LMM-MMSE مشخص می‌شوند.

جدول (۱-۵) مقادیر سیگنال به نوفه قطعه‌ای را برای چهار روش ذکر شده، در سه سیگنال به نوفه ۵/۰ و ۱۰ و دسی بل و دو نوفه همهمه و F16 نشان می‌دهد. با توجه به این جدول بالاترین مقادیر سیگنال به نوفه قطعه‌ای در همه سیگنال به نوفه‌ها و نوفه‌ها مربوط به روش پیشنهادی است و این نشان می‌دهد که روش پیشنهادی نسبت به دو روش دیگر، موفقیت بیشتری در حذف نوفه داشته است. به عنوان مثال برای نوفه همهمه و در همه سیگنال به نوفه‌ها، روش پیشنهادی در مقایسه با دو روش مبتنی بر توزیع گوسی، 1 dB بهبود در سیگنال به نوفه قطعه‌ای ایجاد کرده که مقدار قابل ملاحظه‌ای است.

جدول (۲-۵) مقادیر LLR را برای سیگنال بهسازی شده با چهار روش قبل و نوفه‌ها و سیگنال به نوفه‌های مذکور نشان می‌دهد. با توجه به جدول، پایین‌ترین مقادیر LLR در همه حالات، مربوط به روش پیشنهادی است که نشان می‌دهد سیگنال تخمین زده شده با این روش نسبت به سه روش دیگر شباهت بیشتری به سیگنال اصلی دارد.

تمیز است. معیار LLR برای همه پنجره‌های یک جمله حساب می‌شود. در واقع این مقدار برای هر قاب از سیگنال محاسبه شده و در نهایت از ۹۵٪ پایین‌ترین مقادیر به دست آمده برای قاب‌ها میانگین گرفته می‌شود. در مقایسه دو سیگنال بهسازی شده، هر کدام که LLR پایین‌تری داشت، شباهت بیشتری به سیگنال گفتار بدون نوفه دارد (هیو و همکاران، ۲۰۰۶).

معیار PESQ که از لحاظ محاسباتی پیچیده‌ترین معیار عینی است، نتایج آزمایش‌ها ارزیابی شنیداری را پیش‌بینی می‌کند. در این روش سیگنال بهسازی شده با سیگنال اصلی مقایسه می‌شود و نمره‌ای بین ۰/۵- تا ۴/۵ داده می‌شود. هر چقدر که این معیار بیشتر باشد به معنای کیفیت بالاتر سیگنال بهسازی شده است.



شکل (۱-۵): نمودار سیگنال به نوفه قطعه‌ای سیگنال بهسازی شده: به ازای Nهای متفاوت، پس از اعمال روش پیشنهادی روی سیگنال آغشته به نوفه سفید

شکل (۱-۵) نتایج سیگنال به نوفه قطعه‌ای را برای سیگنال بهسازی شده با روش پیشنهادی، به ازای Nهای متفاوت، در سه سیگنال به نوفه مختلف و برای نوفه سفید نشان می‌دهد. با توجه به نمودار با افزایش N، سیگنال به نوفه قطعه‌ای افزایش می‌یابد که به معنای حذف نوفه بیشتر و بهتر شدن نتایج است؛ زیرا همان‌طور که در قبل نیز گفته شد، با افزایش N توزیع مخلوط لاپلاس مفروض به‌طور دقیق‌تری بر توزیع واقعی سیگنال گفتار تمیز براننده می‌شود؛ پس نتایج تخمین گر MMSE هم که به‌طور مستقیم با این توزیع رابطه دارد، بهتر می‌شوند. با توجه به شکل، تغییرات N در بازه ۱-۵۰ صورت گرفته است؛ زیرا نتایج بعد از N=30 تغییرات قابل توجهی ندارند،

جدول (۵-۱): مقایسه چهار روش MMSE، Log-MMSE، Lap-MMSE و LMM-MMSE بر حسب سیگنال به نوفه قطعه‌ای و برای نوفه همهمه و F16.

Noises Estimators	Babble noise			F-16 Cockpit noise		
	-9.48/0 dB	-5.48/5 dB	-1.4/10 dB	-9.51/0 dB	-6.1/5 dB	-2.1/10 dB
MMSE	1.341	3.892	6.42	1.57	4.132	6.731
Log-MMSE	1.773	4.251	7.123	1.873	4.052	6.69
Lap-MMSE	2.874	4.808	7.582	2.81	4.613	7.288
LMM-MMSE (Proposed method)	2.993	4.985	7.996	2.97	4.82	7.912

جدول (۵-۲): مقایسه چهار روش MMSE، Log-MMSE، Lap-MMSE و LMM-MMSE بر حسب LLR و برای نوفه همهمه و F16.

Noises Estimators	Babble noise			F-16 Cockpit noise		
	1.44/0 dB	1.103/5 dB	0.915/10 dB	1.391/0 dB	1.112/5 dB	0.957/10 dB
MMSE	0.981	0.751	0.572	1.021	0.812	0.619
Log-MMSE	1.213	0.984	0.852	1.196	0.935	0.734
Lap-MMSE	0.81	0.644	0.507	0.879	0.696	0.57
LMM-MMSE (Proposed method)	0.746	0.587	0.487	0.82	0.621	0.532

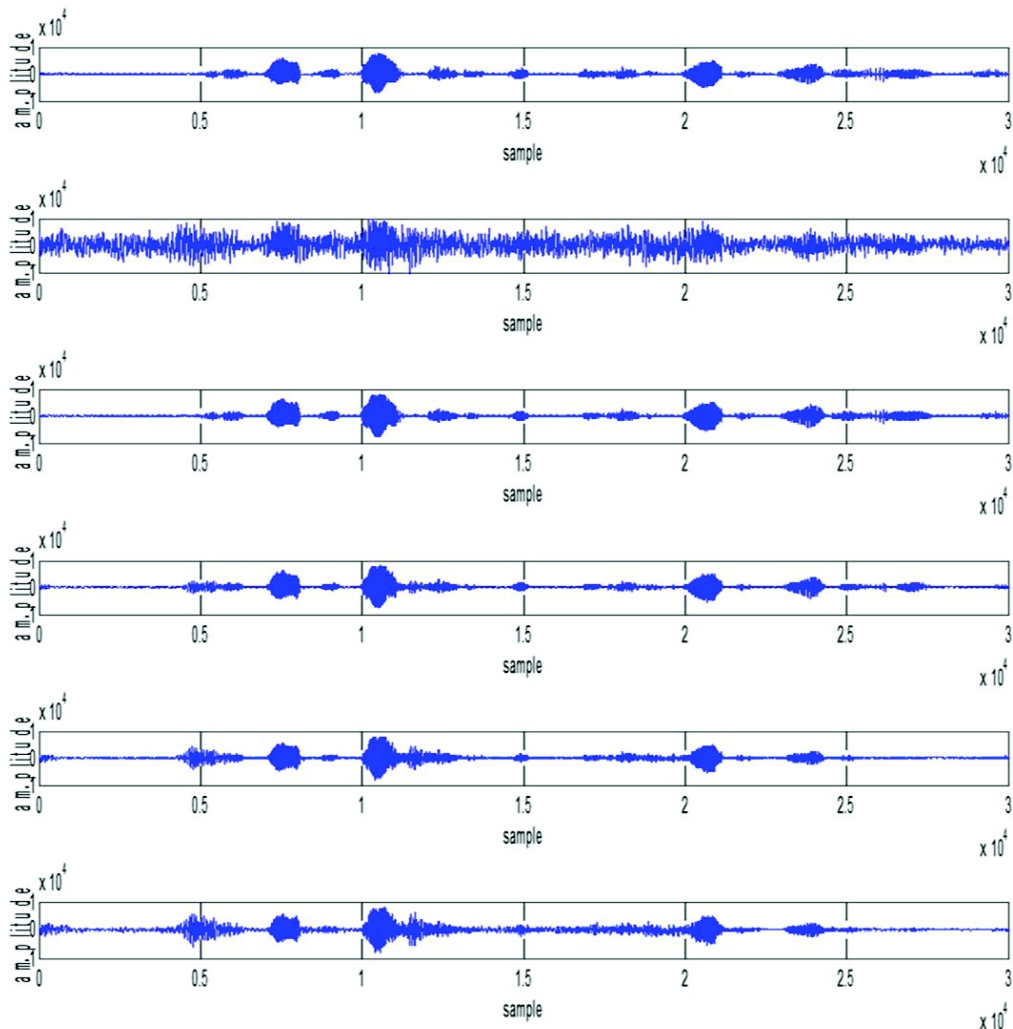
جدول (۵-۳): مقایسه چهار روش MMSE، Log-MMSE، Lap-MMSE و LMM-MMSE بر حسب PESQ و برای نویز همهمه و F16.

Noises Estimators	Babble noise			F-16 Cockpit noise		
	1.67/0 dB	2.211/5 dB	2.697/10 dB	1.423/0 dB	2.037/5 dB	2.412/10 dB
MMSE	2.114	2.456	2.781	2.081	2.396	2.712
Log-MMSE	2.172	2.487	2.833	2.035	2.19	2.585
Lap-MMSE	2.183	2.52	2.857	2.136	2.314	2.508
LMM-MMSE (Proposed method)	2.241	2.66	3.11	2.17	2.623	2.9

شکل (۵-۲) یک نمونه سیگنال گفتار خالص از پایگاه داده TIMIT، سیگنال آغشته به نوفه همهمه با $SNR=-5$ dB و سیگنال‌های بهسازی شده با چهار روش Log-MMSE، Lap-MMSE و LMM-MMSE را نشان می‌دهد. با توجه به شکل‌ها سیگنال بازسازی شده با روش پیشنهادی، نوفه باقیمانده و اعوجاج کم‌تری دارد. با توجه به نتایج می‌توان گفت، روش پیشنهادی نوفه را به خوبی کاهش می‌دهد و کیفیت سیگنال بهسازی شده نیز بالاست و اعوجاج و نوفه باقیمانده کمی دارد. پس در مقایسه با سه روش دیگر، روش پیشنهادی، روشی مناسب‌تر و با کارایی بالاتر در بهسازی گفتار است.

با توجه به جدول (۵-۳) که مقادیر PESQ را نشان می‌دهد، برای دو نوفه همهمه و F16 در سیگنال به نوفه‌های مفروض، روش پیشنهادی بالاترین مقادیر PESQ را داراست و این بیان‌گر کیفیت بالاتر سیگنال بهسازی شده توسط این روش است.

همچنین در هر سه جدول و در همه حالات، دو روش مبتنی بر توزیع لاپلاس نتایج بهتری از دو روش مبتنی بر توزیع گوسی داشته‌اند که این نشان می‌دهد، توزیع‌های غیرگوسی برای تقریب توزیع گفتار تمیز مناسب‌ترند. مقایسه دو روش مبتنی بر یک لاپلاس و روش پیشنهادی نیز نشان می‌دهد که روش پیشنهادی نتایج بهتری ایجاد می‌کند؛ یعنی مجموع چند لاپلاس برای تقریب گفتار تمیز، مناسب‌تر از یک لاپلاس است.



شکل (۵-۲): یک جمله از TIMIT که با نوفه همهمه در سیگنال به نوفه -5 dB آغشته شده و با روش‌های MMSE، Lap-MMSE، LMM-MMSE و MMSE بهسازی شده است. به ترتیب از بالا به پایین: سیگنال اصلی، سیگنال نوفه‌ای، سیگنال بهسازی شده با روش MMSE و LMM-MMSE، سیگنال بهسازی شده با روش Lap-MMSE، سیگنال بهسازی شده با روش Log-MMSE و سیگنال بهسازی شده با روش MMSE

بیان‌گر عملکرد بهتر این روش در حذف نوفه و کیفیت بالاتر سیگنال بهسازی شده است؛ پس می‌توان توزیع مخلوط لاپلاس را به‌عنوان مدلی مناسب در روش‌های بهسازی گفتار آماری در نظر گرفت.

۷- مراجع

A. Gradshteyn and I. Ryzhik, "Table of Integrals, Series, and Products". Academic Press, 5th ed., 1994.

A. Kundu, S. Chatterjee, A. S. Murthy and T. V. Sreenivas, "GMM Based Bayesian Approach To Speech Enhancement In Signal / Transform Domain", IEEE

۶- نتیجه‌گیری

در حال حاضر روش‌های آماری با فرض توزیع‌های غیرگوسی برای سیگنال گفتار تمیز، یکی از مهم‌ترین و بهترین روش‌ها برای حذف نوفه سیگنال گفتار هستند. در این مقاله نیز از توزیع مخلوط لاپلاس برای مدل کردن سیگنال گفتار تمیز استفاده شد و پارامترهای این مدل با الگوریتم EM تخمین زده شد؛ سپس با استفاده از تخمین گر MMSE ضرایب طیف سیگنال گفتار تمیز تخمین زده شد. نتایج ارزیابی این روش و مقایسه آن با دو روش مبتنی بر توزیع گوسی و روش مبتنی بر توزیع لاپلاس، بر حسب سه معیار ارزیابی عینی،

- R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation", *Signal Processing*, vol.86, pp.1215-1229, 2006
- R. Martin, "Speech enhancement based on Minimum Mean-Square Error Estimation and super-gaussian priors", *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 5, pp. 845–856, Sept. 2005.
- R.Martin and C.Breithaupt, "Speech Enhancement in the DFT domain using Laplacian Speech Prior", In *International Workshop on Acoustic Echo and noise control*, pp. 87-90, Kyoto, Japan.2003
- R.Martin, (1994), *Spectral Subtraction Based on Minimum Statistic*, European signal processing Conference, pp 1182-85, Edinburgh, Scotland
- R.Martin, (2003), *Statistical Method for Enhancement of Noisy Speech*, In *International Workshop on Acoustic Echo and noise control*, pp. 43-65, Kyoto, Japan.
- R.McAulay and M.Malpass, (1980), *Speech Enhancement Using a Soft-Decision Noise Suppression Filter*, *IEEE Transaction on Acoustic, Speech and Signal Processing*, Vol. 9, No. 5, pp. 504-512
- S. Kullback, *Information Theory and Statistics*. New York: Dover, 1997.
- S.Boll ,(1979), *Suppression of Acoustic Noise in Speech Using Spectral Subtraction*, *IEEE Transaction on Acoustic, Speech and Signal Processing*, Vol. 27, No. 2, pp. 113-120
- T. Lotter, and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model," *EURASIP Journ. Appl. Signal Proc.*, vol.7, pp. 1110–1126, 2005
- Wolf. P. J and Godsill. S. J, (2003a), *Efficient Alternative to the Ephraim and Malah Suppression Rule for Audio Signal Enhancement*, *EURASIP Journal on Applied Signal Processing*, Vol. 2003, No. 10, pp. 1043-1055
- Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
- Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.
- Y. Ephraim and D. Malah, Juang.B.H, (1989), *On the Application of Hidden Markov Models for Enhancing Noisy Speech*, *IEEE Transaction on Acoustic, International Conference on Acoustics, Speech and Signal Processing*, pp. 4893-4896, 2008
- A.P.Dempster, N.M.Laird and D.B.Rubin. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*, *Journal of the Royal Statistical Society*, Vol. 39, No.1, pp. 45-67
- A.Papoulis and S.Pillai, (2002), *Probability, Random Variables and Stochastic Processes*, New York, McGraw-Hill
- B.Chen, P.C.Loizou, "A Laplacian-Based MMSE Estimator for Speech Enhancement", *Speech Communication*, Vol. 49, No. 2, pp. 134-143, 2007
- D. Brillinger, "Time Series: Data Analysis and Theory", Holden-Day, 1981.
- H.L.Hansen and B.Pellom, "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithm", In: *Proc. ICSLP*, vol. 7, pp. 2819–2822. 1998.
- J.A.Bilmes, (1998), *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*, Technical Report, ICSI-Technical Report-97-021
- J.Lim and A.Oppenheim, (1979), *Enhancement and Bandwidth Compression of Noisy Speech*, *Proceeding of the IEEE*, Vol. 67, No. 12, pp. 1586-1604
- J.S.Eekelens, J.Jensen and R.Heusdens, "Speech Enhancement Based on Rayleigh Mixture Modeling of Speech Spectral Amplitude Distributions", In *European Signal Processing Conference*, pp. 65-9. 2007b
- N.Mitianoudis and T. Stathaki, "Overcomplete source separation using Laplacian Mixture Model", *IEEE Signal Processing Letters*, Vol. 1, No. 2, pp. 1-9, 2004
- P.C.Loizou, "Speech Enhancement: Theory and Practice", CRC Press, (2007), pp. 87, 437
- P.J.Wolf and S.J.Godsill, (2001), *Simple Alternative to the Ephraim and Malah Suppression Rule for Speech Enhancement*, *IEEE Workshop on Statistical Signal Processing*, Vol. 2, pp.496-499
- Porter, J. and Boll, s. (1984). *Optimal Estimation for Spectral Restoration of Noisy Speech*, In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'84)*. San Diego, California, USA, pp. 18A.2.1-18A.2.4.
- Q.Huang, J.Yang, Y.Zhou, "Variational Bayesian method for speech enhancement", *Neur-computing*, vol. 70, pp. 3063–3067, 2007

سال ۱۳۸۵ از دانشگاه تربیت مدرس در مهندسی مخابرات دریافت کردند و در حال حاضر استادیار دانشگاه شاهرود است. زمینه‌های تحقیقاتی مورد علاقه ایشان پردازش سیگنال، شبکه‌های نوری، ارتباطات ماهواره‌ای و پروتکل‌های ارتباطات پرسرعت داده‌هاست. نشانی رایانامه ایشان عبارت است از:

ormzi@yahoo.com

Speech and Signal Processing, Vol. 37, No. 12, pp. 1846-1856

Y.Ephraim and T.Van, (1995), A Signal Subspace Approach for Speech Enhancement, IEEE Transaction on Acoustic, Speech and Signal Processing, Vol. 3, No. 12, pp. 252-266

Y.Hu and P.Loizou, "Evaluation of objective measures for speech enhancement", Proc. Of INTER-SPEECH, Philadelphia, PA. 2006.



زینب محمدپوری مدرک کارشناسی خود را در رشته مهندسی پزشکی از دانشگاه شاهد تهران در سال ۱۳۸۶ و مدرک کارشناسی ارشد خود را در رشته مهندسی الکترونیک از دانشگاه صنعتی

شاهرود در سال ۱۳۸۹ دریافت کرد. وی در حال حاضر دانشجوی دکترای رشته مهندسی پزشکی دانشگاه حکیم سبزواری است. زمینه‌های پژوهشی مورد علاقه ایشان، پردازش سیگنال، پردازش تصویر، شناسایی الگو و شبکه‌های عصبی است.

نشانی رایانامه ایشان عبارت است از:

zmohzmmzdpoor@gmail.com



حسین مروی مدرک کارشناسی را از دانشگاه فردوسی مشهد در مهندسی الکترونیک، و کارشناسی ارشد را از دانشگاه شیراز در مهندسی مخابرات اخذ کردند. ایشان مدرک دکترای

تخصصی را در زمینه تشخیص گفتار از مرکز CVSSP دانشگاه Surrey در سال ۲۰۰۵ اخذ کردند. وی در حال حاضر استادیار دانشگاه شاهرود است. زمینه‌های تحقیقاتی مورد علاقه ایشان پردازش سیگنال‌ها، پردازش صوت، تشخیص گفتار، تشخیص گوینده و بهسازی گفتار است.

نشانی رایانامه ایشان عبارت است از:

h.marvi@shahroodut.ac.ir



امیدرضا معروضی مدرک کارشناسی خود را در سال ۱۳۶۹ از دانشگاه صنعتی شریف در رشته مهندسی الکترونیک و مدرک کارشناسی ارشد را در سال ۱۳۷۲ از همان دانشگاه در رشته مهندسی

مخابرات اخذ کرد. ایشان مدرک دکترای تخصصی را در