



بازیابی میسینگ فیوچر برای سیگنال های باند محدود گفتار

سید علی حسینی شریف^۱، مجید پوراحمدی^۲

۱- گروه برق، دانشکده فنی شهید صدوقی، دانشگاه فنی و حرفه ای یزد، ایران

۲- گروه برق، دانشکده فنی و مهندسی، واحد یزد، دانشگاه آزاد اسلامی، یزد، ایران

، Sadashti222@yahoo.com

چکیده

استفاده از اطفیف نگاره ناقص قابل انجام نیست و میتوان آن را از دیگر مشکلات استفاده از بردارهای طیف به حساب آورد. با توجه به مشکلات استفاده از روش هایی که بصورت مستقیم از اطفیف نگاره ناقص تشخیص گفتار انجام می شود روش های میسینگ فیوچر برای تشخیص گفتار ارائه شده اند.

سیستم تشخیص گفتار مغشوش در نویز دارای عملکرد ضعیفی است. در خصوص جبران سازی با حذف مولفه های شامل نویز نشان داده شده در اطفیف نگاره از روش های میسینگ فیوچر انجام می پذیرد و تشخیص گفتار تنها با مولفه های قابل اعتماد صورت می پذیرد. اکثر روش های ارائه شده برای تشخیص گفتار مغشوش در نویز تلاش بر این دارد که نویز را کرکترایز کرده و اثر نویز را بر شفافیت گفتار مدل کند. عملکرد این روش ها شدیداً وابسته به توانایی و دقت اندازه گیری نویز است.

۲- سیگنال های باند محدود گفتار

با توجه به افزایش جمع آوری داده ها دیجیتالی بصورت اتوماتیک نیاز به بازیابی داده ها بصورت اتوماتیک رو به گسترش می باشد. برای بازیابی اطلاعات بصورت اتوماتیک بخصوص برای داده ها مالتی مدیا که دارای اطلاعات حیاتی بوده، نیاز به پردازش سیگنال ها بشدت حس می شود. یکی از بزرگترین جمع آوری کنندگان داده ها مالتی مدیا (National Gallery Spoken Word) NGSW که دارای دو فاز ثبت نام و بازیابی برخط می باشد. در فاز ثبت نام، برچسب گذاری بصورت اتوماتیک بر روی سیگنال های گفتار که شامل قطعه قطعه کردن سیگنال و برچسب گذاری با استفاده از ماشین تشخیص گفتار پیوسته با تعداد لغات زیاد صورت می گیرد. در فاز دوم بازیابی اطلاعات برچسب گذاری شده با استفاده از سیستم $MG[1],[2]$ اطلاعات موجود در این پایگاه داده دارای پوشش گسترده ای از لحاظ طیفی و محتوایی فرمت ها می باشد. از محدودیت های پهنای باند، ضعیف بودن صدا، تفاوت در نوع میکروفون ها، اعوجاج و اکو ها در اماکن عمومی، ضبط صدا از تلفن، نویز های پس زمینه طیف گسترده ای از سبک و لهجه ها بعنوان مشکلات پردازش گفتاری این داده ها بحساب می آیند.

اطفیف نگاره که در شکل پایین نشان داده شده اند مربوط به سیگنال گفتار ضبط شده با فرکانس نمونه 16KHz از Thomas Edison و Former President Bill Clinton می باشد که از داده ها NGSW گرفته شده است. بخاطر شرایط ضبط صدا سیگنال گفتار ادیسون دارای پهنای باند محدود بین 1.5-2.5 KHz بوده است. در این مقاله تلاش بر این است تا ویژگی ها طوری بازیابی شود تا عملکرد تشخیص چنین سیگنال های تحریف شده با پهنای باند محدود بالا رود.

کلید واژه: سیگنال، نویز، تشخیص گفتار، باند، SNR

۱- مقدمه

سیگنال های گفتار دارای درجه آزادی redundancy بالایی هستند. بطور مشابه گفتار که تحت برش باند طیفی و یا مناطق زمان کوتاه هستند براحتی قابل فهم می باشند. بنابراین تشخیص گفتار تنها با استفاده از بخشی از طیف سیگنال قابل انجام باشد. به این منظور برای بکار گیری میسینگ فیوچر از اسپکتروگراف سیگنال استفاده می شود که شامل رشته هایی از انرژی یا log طیف استفاده می شود. هنگامی که سیگنال نویزی می شود، برخی از مولفه های فرکانس-زمان آن بیشتر مغشوش در نویز در قیاس با بقیه مولفه های مولفه ها با SNR پائین غیر قابل اعتماد پنداشته میشوند و برای تشخیص تنها از مولفه های قابل اعتماد استفاده می شود. بنابراین مولفه های غیر قابل اعتماد به عنوان missing در نظر گرفته می شوند و مولفه های غیر قابل قبول از روی مولفه ها با SNR بالاتر تخمین زده می شوند. در اکثر الگوریتم ها توزیع بردارهای طیف سیگنال تمیز بصورت GMM مدل شده اند.

در روش های تشخیص گفتار بر اساس بردارهای cepstral می توان از state-based imputation و marginalization استفاده کرد زیرا توزیع بردارهای طیف را نمی توان از بردارهای cepstral بخاطر پایین بودن تعداد آنها قابل تعیین نمی باشند. برای تشخیص بهتر نیازه نرمالایز کردن میانگین و واریانس است که با