

تشخیص دیداری-شنیداری حروف صدا دار فارسی

علی حریمی^۱ و زینب اسماعیلیان^۲

^۱ دانشگاه آزاد اسلامی، واحد شاهرود، گروه برق، شاهرود، ایران، a.harimi@gmail.com

^۲ دانشگاه آزاد اسلامی، واحد علوم و تحقیقات شاهرود، گروه برق، شاهرود، ایران، z.esmaileyan@gmail.com

چکیده

در این مقاله به منظور بهبود سیستم های تشخیص گفتار بویژه در محیط های نویزی از اطلاعات سیگنال گفتار و تصویر لب بطور همزمان استفاده شده است. بدین منظور تصویر لب بوسیله ی الگوی دودویی محلی ناحیه بندی شده و ویژگی های مناسب از تصویر لب استخراج شده است. این ویژگی ها با ویژگی های استخراج شده از سیگنال گفتار ترکیب شده و بردار ویژگی را تشکیل می دهند. سپس توسط یک الگوریتم انتخاب ویژگی دو مرحله ای ویژگی های نویزی حذف شده و بردار ویژگی نهایی شکل می گیرد. در انتها شش حرف صدا دار فارسی بوسیله ی الگوریتم کلاسه بند از یکدیگر تفکیک می شوند. نتایج حاکی از کارآمد بودن روش پیشنهادی بویژه در محیط های نویزی می باشد. متوسط نرخ تشخیص برای سیگنال بدون نویز ۹۴/۲٪ و برای سیگنال آغشته به نویز با نسبت سیگنال به نویز چهار برابر ۸۷/۵٪ محاسبه شده است.

کلمات کلیدی

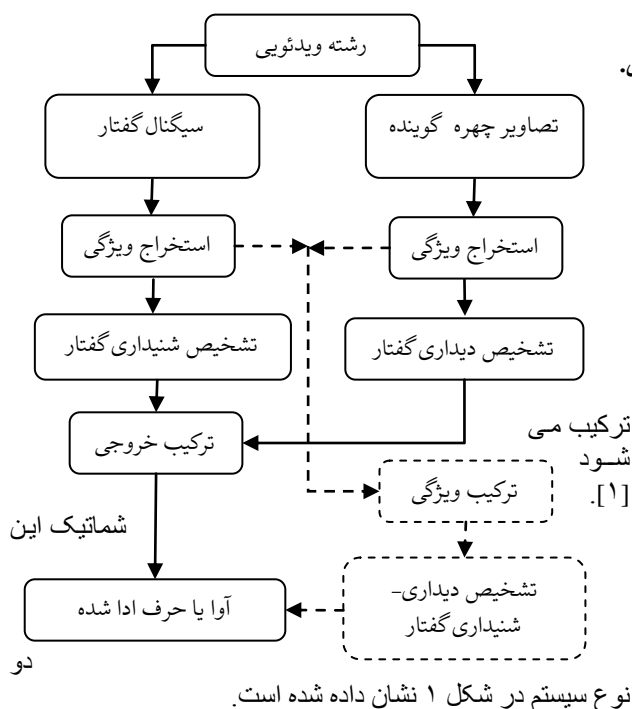
سیستم دیداری-شنیداری تشخیص گفتار، ناحیه بندی لب، الگوی دودویی محلی.

۱- مقدمه

با پیشرفت سریع تکنولوژی سیستم های تعاملی بین انسان و کامپیوتر جایگاه ویژه ای پیدا کرده اند. از آن جا که گفتار سریعترین و پرکاربردترین راه انتقال اطلاعات بین انسان ها می باشد، آنالیز سیگنال گفتار در این سیستم ها نیز از اهمیت خاصی برخوردار می باشد. انسان ها در تعاملات خود برای انتقال بهتر پیام از حرکات دست، بدن و صورت بعنوان مکمل اطلاعات زبانی استفاده می کنند. بسیاری از محققین نیز به منظور افزایش دقت سیستم های تشخیص گفتار استفاده از این اطلاعات تکمیلی را پیشنهاد کرده اند [۱ و ۲].

سیستم های شنیداری-دیداری گفتار را می توان از دیدگاه تشخیص الگو در سه بخش استخراج ویژگی، انتخاب ویژگی و کلاسه بندی مورد بررسی قرار داد. در اولین مرحله ویژگی های مناسب از سیگنال گفتار و تصویر گوینده استخراج می شوند. در مرحله دوم ویژگی های موثر تر در سیستم توسط الگوریتم انتخاب ویژگی شناسایی شده و ویژگی های نویزی نیز حذف می شوند. در آخرین مرحله نیز الگوریتم کلاسه بند با توجه به ویژگی های انتخاب شده آوای ادا شده را مشخص می کند.

سیستم های مختلف شناسایی دیداری-شنیداری را می توان در دو گروه مختلف بررسی نمود. در سیستم های نوع اول که سیستم های ترکیب ویژگی نام دارند بردار ویژگی از ترکیب ویژگی های صوتی و تصویری حاصل می شود و سیستم با استفاده از این بردار آوای ادا شده را تشخیص می دهد اما در سیستم های نوع دوم که سیستم های ترکیب خروجی نام دارند دو بردار ویژگی مختلف داریم که یکی از آن ها حاوی اطلاعات تصویری و دیگری حاوی اطلاعات صوتی می باشد. در چنین سیستم هایی دو بردار ویژگی به دو زیر سیستم تشخیص شنیداری و تشخیص دیداری گفتار اعمال می شوند و در نهایت خروجی این دو سیستم با یکدیگر



نوع سیستم در شکل ۱ نشان داده شده است.

شکل (۱): شماتیک سیستم ها تشخیص دیداری-شنیداری گفتار.

در این شکل سیستم های با ترکیب خروجی با خطوط ممتد و سیستم های با ترکیب ویژگی با خط چین نشان داده شده اند.