

## ارزیابی کیفیت سیستم های تبدیل متن به گفتار فارسی بصورت مفعولی و مستقل از متن با استفاده از مدل پنهان مارکوف

علی عابدی<sup>۱</sup>، محمد مهدی همایون پور<sup>۲</sup>، و محمد صنیعی آبا<sup>۳</sup>

<sup>۱</sup> ایران، تهران، دانشگاه تربیت مدرس، دانشکده مهندسی برق و کامپیوتر، ali.abedi@modares.ac.ir

<sup>۲</sup> ایران، تهران، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر و فن آوری اطلاعات، homayoun@aut.ac.ir

<sup>۳</sup> ایران، تهران، دانشگاه تربیت مدرس، دانشکده مهندسی برق و کامپیوتر، saniee@modares.ac.ir

### چکیده

در این مقاله، یک روش مفعولی و مستقل از متن برای ارزیابی کیفیت گفتار سنتز شده توسط سیستم های تبدیل متن به گفتار فارسی ارائه شده است. ابتدا، مدل های پنهان مارکوف را با گفتار طبیعی و بصورت مستقل از متن آموزش می دهیم. سپس گفتار سنتز شده را به مدل های پنهان مارکوف، بعنوان مدل های مرجع، ارائه می کنیم. یک شباهت لگاریتمی، بین گفتار سنتز شده و مدل های مرجع بدست می آوریم. از طرفی، گفتار سنتز شده را با معیارهای ارزیابی مبتنی بر فرد چندبُعدی هم امتیازدهی می کنیم. در نهایت، برای بررسی میزان شباهت روش مفعولی با روش های مبتنی بر فرد، میزان همبستگی بین شباهت های لگاریتمی بدست آمده از روش مفعولی و امتیازهای بدست آمده از روش های مبتنی بر فرد را بدست می آوریم. آزمایش های انجام شده نشان می دهد که روش مفعولی ارائه شده، تخمین قابل قبولی از بیشتر معیارهای مبتنی بر فرد می باشد.

### کلمات کلیدی

مدل پنهان مارکوف، گفتار سنتز شده، تبدیل متن به گفتار، ارزیابی کیفیت مفعولی، ارزیابی کیفیت چندبُعدی.

مشارکت دارند. برای کاربردهایی مثل تنظیم کردن پارامترهای سیستم تبدیل متن به گفتار، آزمایش های زیادی ممکن است. در چنین حالت هایی، ارزیابی مبتنی بر فرد امکان پذیر نخواهد بود، و ارزیابی کیفیت مفعولی مورد نیاز است.

تاکنون، روش ارزیابی کیفیت مفعولی بطور کامل پذیرفته شده ای برای گفتار سنتز شده ارائه نشده است [1]. بیشتر روش های ارائه شده، برای سیستم های تبدیل متن به گفتار الحاقی<sup>۰</sup> مبتنی بر پایگاه داده، که پایگاه داده گفتار طبیعی موجود است می باشند. در [2]، یک تابع هزینه میانگین الحاقی، برای ارزیابی میزان طبیعی بودن سنتزکننده های الحاقی مورد استفاده قرار گرفته است. بیشتر بودن تعداد الحاق ها، متناظر با پایین تر بودن کیفیت گفتار سنتز شده است. در [3]، برای ارزیابی کیفیت، فاصله طیفی بین گفتار سنتز شده، و معادل گفتار طبیعی آن، بعنوان مرجع محاسبه می گردد. چنین روش هایی زمانی مفید خواهند بود که مجموعه گفتارهای طبیعی مرجع موجود باشند، که این لزوما در عمل امکان پذیر نخواهد بود. برای رفع چنین محدودیت هایی، یک معیار مستقل از متن<sup>۱</sup>، مانند روش ارائه شده در [1] لازم است، که در آن دو مدل مرجع آموزش دیده با گفتار طبیعی، برای گفتار مذکر و مونث تولید می گردد. سپس

### ۱ - مقدمه

هدف سیستم های تبدیل متن به گفتار<sup>۱</sup> این است که متن دلخواه ورودی به گفتار طبیعی و قابل فهم تبدیل گردد. بعضی از کاربردهای سیستم های تبدیل متن به گفتار عبارت اند از سیستم های دستیار افراد نابینا، سیستم های ایمیل خوان، سیستم های پیام خوان، سیستم های آموزش زبان های خارجی، و سیستم های ترجمه همزمان. با پیشرفت و افزایش کاربرد های سیستم های تبدیل متن به گفتار، نیاز به روش هایی که کیفیت سیستم های تبدیل متن به گفتار را مورد ارزیابی قرار دهند نیز افزایش یافته است. همچنین، آزمایش و ارزیابی گفتار سنتز شده، مرحله مهمی در فرآیند طراحی یک سیستم تبدیل متن به گفتار است. روش های ارزیابی کیفیت گفتار را می توان در دو دسته قرار داد. دسته اول، روش های ارزیابی مبتنی بر فرد<sup>۲</sup> هستند که در آنها میانگین امتیاز نظرات<sup>۳</sup> افراد مختلف (MOS) برای گفتار سنتز شده بدست می آید. دسته دوم، روش های ارزیابی مفعولی<sup>۴</sup> هستند که بیشتر بر اساس مدل کردن ویژگی های ادراکی گفتار عمل می کنند.

یکی از مشکلات عمده روش های ارزیابی مبتنی بر فرد، هزینه های بالا و نیروی انسانی زیادی است که در آزمایش ها