دوازدهمین کنفرانس بین‌المللی انجمن کامپیوتر ایران
دانشگاه شهید بهشتی، دانشکده مهندسی برق و کامپیوتر ،تهران، ایران، ۱ تا ۳ اسفند ۱۳۸۵

ایجاد واحدهای ناموجود در دادگان برای سنتز با انتخاب واحد

مجید نم نبات[†], محمد مهدی همایونپور[*]

چکیده

سنتز با روش انتخاب واحد از جمله روشهای سنتز پیوندی می باشد که کیفیت خروجی حاصل از آن مناسب و طبیعی و از جمله روشهای بالقوه برای تحقیق و توسعه برای بهبود کیفیت خروجی سیستمهای سنتز می باشد. از جمله عوامل تاثیر گذار بر کیفیت خروجی این سیستمها واحد سنتز می باشد. استفاده از واحد سنتز واج در سنتز با انتخاب واحد ، موجب افزایش میزان طبیعی بودن خروجی می گردد، هر چند که امکان ایجاد گسستگی در مرز میان واجها در سیگنال خروجی نیز بطور بالقوه افزایش مي یابد. در این مقاله، ما از ایده ایجاد واحدهای ناموجود در دادگان، برای حذف گسستگی های موجود در مرزهای اتصال هنگام استفاده از واحد سنتز واج استفاده کرده ایم. در الگوریتم پیشنهادی، گسستگی های موجود در سیگنال خروجی پیدا و سپس یکی از واجهای دو سمت گسستگی با یک نمونه واج مناسب و ناموجود در دادگان جایگزین می گردد. نمونه واج ناموجود در دادگان با ترکیب نیم واج چپ و نیم واج سمت راست دو نمونه واج موجود در دادگان ایجاد می شود. ارزیابی های شنوایي، حاکی از بهبود کیفیت خروجی سیستم سنتز و کاهش گسستگی های موجود در آن می باشد.

کلمات کلیدي

سنتز پیوندی، انتخاب واحد، هزینه اتصال، هزینه هدف، واحد سنتز، ایجاد واحدهای ناموجود، زبان فارسی.

Unseen Unit Generation for Unit Selection-based Concatenative Speech Synthesis

Majid Namnabat, M. Mehdi Homayounpour
Laboratory for Intelligent Sound and Speech Processing Systems
Computer Engineering and Information Technology Department,
Amirkabir University of Technology (Tehran Polytechnics), Tehran, IRAN

Abstract

The unit selection speech synthesis method is accounted of concatenative speech synthesis methods that it produces more natural quality speech than other approaches. So developing and improvement of this approach attracts great attraction to produce more natural output speech. Synthesis unit is one of important factors that have great effect on output quality of synthesizer systems. Using phones as synthesis unit can cause more natural output, however the discontinuity at the boundaries of concatenated phonemes can be potentially increased in output speech. In this paper, we proposed using unseen unit generation idea to eliminate discontinuities at phoneme boundaries when phone instances are used as synthesis units. In proposed method, existence discontinuities at phoneme boundaries in output speech are firstly found and then if it possible, one of adjacent phonemes of these boundaries are replaced by a suitable unseen unit. To synthesizing every unseen unit, left half phone of one existence instance of corpus are merged by right half phone of other existence instance. Perceptual Evaluations show quality output improvement due to decreasing discontinuities at boundaries.

Keywords

Concatenative Speech Synthesis, Unit Selection, Concatenation Cost, Target Cost, Synthesis Unit, Unseen Unit Generation, Persian Language.

[†] کارشناس ارشد مهندسی کامپیوتر و فن آوری اطلاعات، دانشگاه صنعتی امیرکبیر، آزمایشگاه سیستمم هاي هوشمند صوتي-گفتاري، تهران. namnabat@ce.aut.ac.ir

[*] عضو هیات علمي دانشکده مهندسی کامپیوتر و فن آوری اطلاعات، دانشگاه صنعتی امیرکبیر، آزمایشگاه سیستمم هاي هوشمند صوتي-گفتاري، تهران. homayoun@aut.ac.ir

۱- مقدمه

سنتز با انتخاب واحد یکی از روشهای ارائه شده برای سنتز صحبت می باشد که از یک دادگان بزرگ شامل نمونه های مختلف استفاده می نماید. این روش سنتز از جمله روشهای سنتز پیوندی می باشد که در آن برای سنتز هر رشته آوایی، بهترین نمونه ها از میان نمونه های موجود انتخاب و گفتار خروجی با متصل نمودن این نمونه ها به یکدیگر ساخته می شود. اجزای اصلی این روش یک دادگان بزرگ شامل نمونه های صوتی مختلف، دو معیار برای ارزیابی نمونه ها بنامهای هزینه هدف و اتصال و در انتها یک الگوریتم جستجو برای انتخاب بهترین نمونه ها می باشد. هزینه هدف، نشان دهنده میزان شباهت یک نمونه از دادگان با واحد هدف مورد نظر می باشد. مقدار این هزینه با استفاده از مجموع خطی وزندار یکسری زیرهزینه مشخص می شود. هر زیر هزینه مربوط به یک ویژگی می باشد که مقدار آن با استفاده از تابع متریک (D_j^t) انتخاب شده برای آن ویژگی (j) محاسبه می شود. در صورتیکه t_i واحد هدف و u_i یک نمونه از دادگان، W_t^j وزن تابع فاصله ویژگی j و P تعداد زیرهزینه ها باشد. هزینه هدف بصورت زیر محاسبه میشود:

$$C^t = \sum_{j=1}^{P} w_j^t D_j^t(t_i, u_i) \qquad (۱)$$

در حالیکه هزینه اتصال، مشخص کننده میزان گسستگی است که از اتصال دو نمونه از دادگان بوجود می آید. برای محاسبه مقدار این هزینه، مشابه هزینه هدف، یک سری ویژگی نشان دهنده میزان گسستگی انتخاب و سپس برای هر ویژگی یک تابع فاصله (D^c) انتخاب میشود. تابع هزینه اتصال بصورت جمع خطی وزندار مقادیر فاصله محاسبه میشود. در صورتیکه u_{i-1} نمونه انتخاب شده قبلی و u_i یک نمونه از دادگان، q تعداد زیرهزینه های انتخاب شده و W_j^c را وزن تابع فاصله مربوط به ویژگی j در نظر بگیریم، فرمول محاسبه این هزینه بصورت زیر می باشد:

$$C^c = \sum_{j=1}^{q} w_j^c D_j^c(u_{i-1}, u_i) \qquad (۲)$$

برای یافتن مسیر بهینه شامل بهترین نمونه ها برای یک رشته آوایي هدف, گرافی شامل نمونه های موجود برای رشته آوایي هدف ساخته می شود. در این گراف هزینه عبور از هر یال برابر با هزینه اتصال نمونه های مربوط به دو طرف یال و هزینه هر گره معادل هزینه هدف آن نمونه در نظر گرفته میشود. با کمک الگوریتم جستجوی دینامیک ویتربی در این گراف, مسیر شامل نمونه های بهینه با کمترین هزینه پیدا مي شود[1,2,3,4] .

روش سنتز با انتخاب واحد ابتدا توسط آزمایشگاه ATR با پیاده سازی در سیستم سنتز v-Talk معرفی گردید[5]. در این سیستم نمونه های بهینه بر اساس کمینه نمودن فاصله طیفی انتخاب می شدند. در این سیستم برای محاسبه فاصله طیفی از ویژگیهای آواهای مجاور بطور دقیق استفاده گردید. سیستم CHATR برای انتخاب نمونه های مناسبتر علاوه بر ویژگیهای طیفی، از ویژگیهای نوایي نیز برای انتخاب نمونه ها استفاده نمود[6]. سپس هانت و همکاران، با اضافه کردن هزینه اتصال، از الگوریتم جستجوی ویتربی برای پیدا کردن مسیر شامل نمونه های بهینه استفاده نمودند. امروزه این روش در سیستم تبدیل متن به گفتار AT&T [8] مورد استفاده قرار گرفته است و ارزیابیها نشان می دهند که این سیستم در زمره بهترین سیستمهای تبدیل متن به گفتار برای زبان انگلیسی می باشد. همچنین این روش در سیستم سنتز IBM نیز مورد استفاده قرار گرفته است. منتها در این سیستم بجای استفاده از واج یا دایفون بعنوان واحد سنتز، از قطعات صحبت به اندازه یک حالت مدل HMM بعنوان واحد سنتز استفاده شده است. این سیستم از درختهای تصمیم گیری و جستجوی ویتربی برای تعیین ویژگیهای مناسب برای قطعه صحبت بهینه استفاده می نماید[7]. علاوه بر این اخیرا برای زبان ژاپنی با بهره گیری از دادگانهای بسیار بزرگ (در حدود ۱۰-۲۰ ساعت) سیستمهای تبدیل متن به گفتار با کیفیت خروجی بسیار طبیعی ارائه شده است[9]. از معایب این روش، می توان به ثابت نبودن کیفیت خروجی برای تمامی متون ورودی اشاره کرد. عموما در صورتیکه متن ورودی با متن دادگان گفتاری مورد استفاده توسط سیستم شباهت موضوعی داشته باشد کیفیت خروجی طبیعی و در غیر اینصورت از کیفیت خروجی کاسته می شود. در بخش ۲ این مقاله تاریخچه ایجاد واحدهای ناموجود و در بخش ۳ الگوریتم پیشنهادی شرح داده می شود. بخش ۴ به مسائل مربوط به پیاده سازی و ارزیابی سیستم می پردازد و در انتها در بخش ۵ نتیجه گیری و ارائه پیشنهادات آورده شده است.

۲- تاریخچه

ایده ایجاد واحدهای ناموجود در دادگان با استفاده از نمونه های موجود در دادگان، ابتدا توسط واثقی و همکاران [10] برای سیستمهای سنتز پیوندی نرمال ارائه گردید. در سیستمهای سنتز پیوندی افزایش اندازه واحد سنتز، موجب بهبود کیفیت خروجی می شود اما به دلیل افزایش تعداد نمونه های آن، ساخت دادگان با پوشش کامل برای نمونه های واحد سنتز بسیار دشوار می باشد. در اینحالت عموما دادگان مورد استفاده تمامی نمونه های واحد سنتز را پوشش نمی دهد و در صورت نیاز به نمونه های ناموجود در دادگان، از یک روش جایگزین استفاده می شود. بطور مثال به هنگام استفاده از واحد واج وابسته به محتوا، می توان با استفاده از یک درخت تصمیم گیری، بجای نمونه اصلی از یک نمونه با محتوای آوایي مشابه موجود در دادگان استفاده نمود. در مقاله [10] با انتخاب واج وابسته به محتوا به عنوان واحد سنتز، یک روش ادغام و همپوشانی همزمان با پیچ، برای ایجاد واحدهای ناموجود با استفاده از واحدهای موجود در دادگان ارائه شده است. در این روش ساخت واحدهای ناموجود دو مرحله دارد: اول انتخاب واحدهای کاندید مناسب و دوم عملیاتهای پردازش سیگنالی.

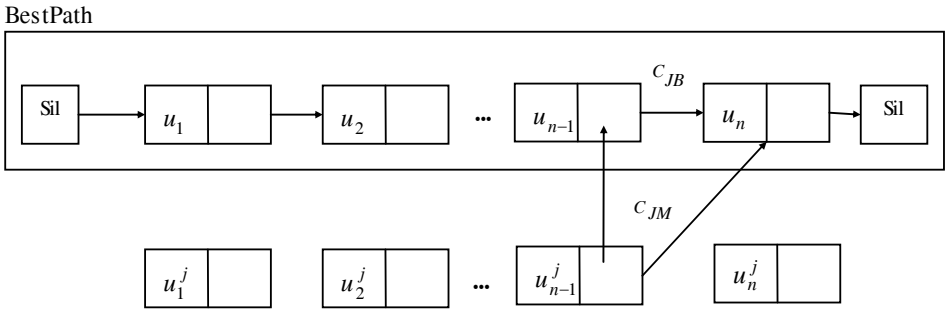
برای واج وابسته به محتوا، دو واحدی که واج سمت چپ و مرکزی یکی و واج مرکزی و سمت راست دیگری با واحد ناموجود یکی باشد، بعنوان کاندید انتخاب می شوند. در صورت وجود چندین واحد کاندید با شرایط فوق، با توجه به مشابهت های آوایي، دو واحدی که از نظر محتواي آوایي به واحد ناموجود مشابه تر باشند، انتخاب می شوند. ایجاد واحد ناموجود با استفاده از دو واحد کاندید بصورت زیر انجام می شود:

$$L-C+R = L-C+NN(R) \otimes NN(L)-C+R \qquad (۳)$$

در رابطه بالا، L محتوای واجی سمت چپ، R محتوای واجی سمت راست، C واج مرکزی برای واحد ناموجود و و NN تابع بیشترین شباهت و عملگر $\otimes$ ترکیب را نشان می دهد. در اینحالت، با استفاده از روش TD-PSOLA فرکانس پیچ دو واحد کاندید ابتدا با یکدیگر یکسان و سپس دو واحد از نظر زمانی با یکدیگر تراز می شوند. در انتها این دو واحد بصورت مجموع وزنی با یکدیگر ادغام می شوند. وزنها به گونه ای انتخاب می شوند که نیم واج چپ واحد جدید بیشتر مشابه $L-C+NN(R)$ و نیم واج راست بیشتر مشابه $NN(L)-C+R$ باشد. واثقی و همکاران در تحقیق خود به خاطر انتخاب یک واحد سنتز بزرگ و عدم پوشش کامل دادگان، از ایده ایجاد واحدهای ناموجود برای کامل ساختن پوشش دادگان استفاده کرده اند. اما ما در سیستم سنتز خود [12] از واحد سنتز واج استفاده کرده ایم و نمونه های زیادی از هر واج در دادگان سیستم موجود می باشد. مرجع [13] به بررسی کیفیت خروجی سنتز با انتخاب واحد در حالت استفاده از واحدهای سنتز مختلف همچون واج، دایفون و نیم واج می پردازد. بررسی های این مقاله نشان می دهد که استفاده از واحد واج بدلیل وجود نمونه های زیاد برای آن امکان تولید خروجی طبیعی را افزایش می دهد، اما بدلیل ناپایداری طیف سیگنال در مرز واجها، امکان ناپایداری کیفیت خروجی و ایجاد گسستگی در مرزهای اتصال افزایش می یابد. استفاده از دایفون، میزان پایداری و قابل فهم بودن خروجی را افزایش می دهد ولی به دلیل کاهش تعداد نمونه های موجود بازای هر دایفون، میزان طبیعی بودن خروجی می تواند کاهش یابد، مگر اینکه از یک دادگان بزرگ استفاده شود. نتایج این مقاله و دیگر مقالات [8,14,15] نشان می دهند که استفاده از نیم واج به عنوان واحد سنتز با وجود دو برابر شدن تعداد اتصالات منجر به بهترین کیفیت خروجی می گردند. البته استفاده از نیم واج در سیستمهای سنتز با انتخاب واحد، موجب دو برابر شدن حجم محاسبات و زمان لازم برای سنتز نیز می گردد. این در حالی است که یکی از مشکلات سنتز با انتخاب واحد، پیچیدگی محاسباتی بالای آن می باشد که تاکنون برای کاهش این پیچیدگی و زمان سنتز تحقیقات مختلفی همچون [11] صورت گرفته است. در این پژوهش، ما برای بررسی ایده مورد نظر خود از سیستم سنتز با انتخاب واحد پیاده سازی شده برای زبان فارسی[12] استفاده کرده ایم. همانطور که اشاره شد این سیستم سنتز از واحد واج استفاده می کند و خروجی حاصل از آن دارای کیفیت مناسب می باشد، اما تعداد گسستگی های موجود در آن تقریبا زیاد می باشد. این گسستگی ها ناشی از اتصال دو واج با یکدیگر می باشد، به طوریکه این دو نمونه در مرز اتصال از نظر طیفی یا نوایي همچون پیچ و انرژی با یکدیگر مشابه نمی باشند. در اینحالت جایگزین نمودن یکی از این نمونه ها با یک نمونه مناسب تر و مشابه تر به نمونه دیگر در بخش مرزی خود، می تواند به رفع این گسستگی و ایجاد یک اتصال هموار کمک نماید. این نمونه مناسب و جدید با یافتن یک نمونه مناسب که موجب ایجاد اتصال هموار می شود صورت می گیرد. نمونه جدید با اتصال نیم واجهای نمونه بهینه کنونی و نمونه مناسب پیدا شده ساخته می شود.

۳- ایجاد واحدهای ناموجود

همانطور که اشاره شد در انتخاب واحد برای یافتن بهترین نمونه ها برای یک دنباله واجی هدف یا مسیر بهینه، گرافی شامل نمونه های موجود برای واحدهای سازنده دنباله واجی هدف ساخته می شود. در این گراف هزینه عبور از هر یال برابر با هزینه اتصال نمونه های دو سمت یال و هزینه اشغال هر گره برابر با هزینه هدف نمونه معادل در نظر گرفته می شود. بعد از ساخت درخت از الگوریتم جستجوی ویتربی برای یافتن مسیر بهینه با کمترین هزینه استفاده می شود. هزینه های هدف و اتصال بواقع مجموع وزندار زیرهزینه های مختلفی می باشند. در سیستم کنونی (فارس بیان) برای هزینه هدف، زیرهزینه های آوایي، مربوط به شباهتهای آوایي و زیرهزینه های نوایي برای تعیین میزان شباهت نوایي نمونه با واحد هدف در نظر گرفته شده است. برای هزینه اتصال نیز سه زیرهزینه مختلف برای سنجش میزان گسستگی پیچ، انرژی و طیف ناشی از اتصال میان دو نمونه در نظر گرفته شده است. برای تولید واحدهای ناموجود، در گام اول لازم است که در مسیر بهینه نمونه های نامناسبِ انتخاب شده مشخص گردند. لذا در فرایند تولید واحدهای ناموجود در مرحله اول سعی می شود اتصالات نامناسب پیدا شوند. در این حالت بعد از جستجوی ویتربی و مشخص شدن مسیر بهینه، اتصالات غیر هموواری که هزینه اتصال مربوط به آنها از مقدار خاصی بزرگتر باشد، بعنوان اتصالات نامناسب در نظر گرفته می شوند. برای ساخت نمونه های جدید و ناموجود در دادگان از اتصال نیم واج چپ و راست دو نمونه موجود در دادگان استفاده می شود. بعد از تعیین یک اتصال غیر هموار در مسیر بهینه، امکان جایگزین کردن نمونه سمت چپ اتصال با یک نمونه جدید مورد بررسی قرار می گیرد. این نمونه جدید از اتصال نیم واج چپ نمونه بهینه با نیم واج راست نمونه ای که هزینه اتصال کمینه با نمونه بعدی در مسیر بهینه دارد، بوجود می آید. در شکل ۳، مسیر بهینه انتخاب شده شامل نمونه های $u_1u_2...u_{n-1}u_n$ می باشد. همچنین نمونه های $u_1^j, u_2^j, ..., u_3^j, u_n^j$ با نمونه بعدی در مسیر بهینه، هزینه اتصال کمینه دارند. بطور مثال هزینه اتصال u_2 با u_1^j کمینه می باشد.



شکل (۳): الگوریتم تولید واحدهای ناموجود در دادگان

همانطور که در این شکل مشاهده می شود، در فرایند تولید واحدهای ناموجود، نیم واج راست u_{n-1}^j جایگزین نیم واج راست u_{n-1} به دلیل وجود اتصال نامناسب میان u_{n-1} و u_n در مسیر بهینه شده است. برای بهبود کیفیت خروجی سیستم به کمک این فرایند، لازم است که اولا هزینه هدفِ نمونه با هزینه اتصال کمینه تا حد امکان کوچک و نزدیک به هزینه هدفِ نمونه بهینه باشد و دوما مجموع هزینه اتصال دو نیم واج به یکدیگر (C_{JH}) و مقدار کمینه هزینه اتصال (C_{JM}) کوچکتر از هزینه اتصال دو نمونه بهینه به یکدیگر (C_{JB}) باشد.

$$C_{JB} > C_{JH} + C_{JM} \qquad (٤)$$

هزینه اتصال میان دو نیم واج، مشابه با هزینه اتصال دو نمونه از دادگان محاسبه می شود، در این حالت، برای سنجش میزان گسستگی طیف از ضرایب MFCC فریمهای دو طرف مرز اتصال بعنوان معیار استفاده شده است. یافتن نمونه هایي که با نمونه بهینه بعدی هزینه اتصال کمینه می سازند، بدون افزایش پیچیدگی محاسباتی هنگام جستجوی ویتربی صورت می گیرد. در حالیکه یافتن نمونه هایي که با نمونه بهینه قبلی هزینه اتصال کمینه می سازند، پیچیده و ممکن است حتی نیازمند محاسبه دوباره هزینه اتصال تمامی نمونه های یک واحد هدف با نمونه بهینه واحد هدف بعدی باشد. لذا در این سیستم برای جلوگیری از افزایش بیش از حد پیچیدگی محاسباتی تنها امکان جایگزینی نمونه جدید بجای نمونه چپ اتصالات نامناسب مورد بررسی قرار گرفته است.

۴- پیاده سازی و ارزیابی سیستم

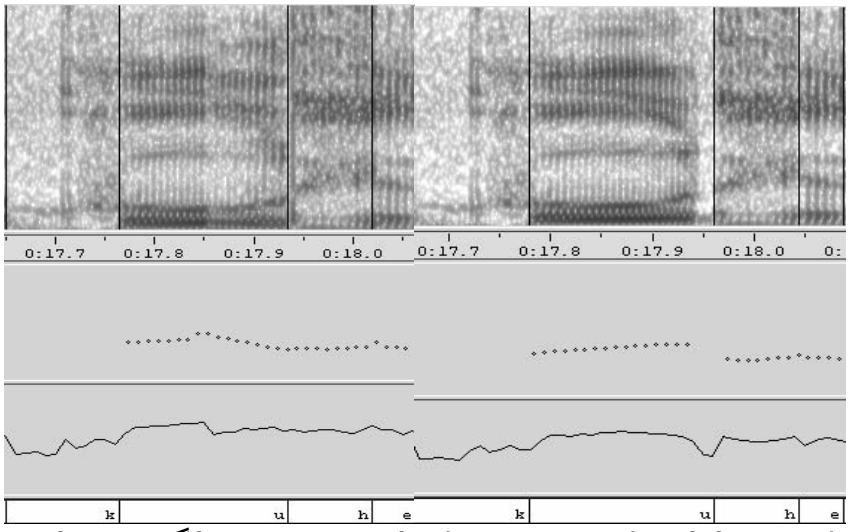
سیستم سنتز کنونی، از حدود ۲ ساعت گفتار ضبط شده شامل ۶۳۰۰۰ واجگونه بعنوان دادگان صحبت استفاده می نماید. متن این گفتار، بخشهایي از متون دادگان فارس دات بزرگ می باشد. همچنین برای ارزیابی این سیستم از تست MOS استفاده شده است که به ترتیب برای معیارهای کیفیت کلی، میزان قابل فهم بودن، طبیعی بودن و خوشایند بودن مقادیر ۳/۸، ۳/۹، ۳/۶ و ۳/۵ بدست آمده است. برای ارزیابی الگوریتم ایجاد واحدهای ناموجود، از دنباله واجی حدود ۴۵ دقیقه گفتار شامل ۱۷۰۰۰ واج و ۸۲۰ پاره گفتار خارج از مجموعه گفتار موجود در دادگان بعنوان مجموعه تست استفاده شده است. پارامترهای نوایي مورد نیاز برای سنتز این دنباله واجی، از خود سیگنال گفتار آن استخراج گردیده است که به اینکار اصطلاحا سنتز مجدد۱ گفته می شود. مجموعه تست توسط سیستم سنتز و اطلاعات آماری مربوط به اتصالات موجود در سیستم سنتز در جدول ۱ نشان داده شده است. در اتصالات طبیعی، دو نمونه مجاور در دادگان به عنوان بهینه انتخاب شده اند. از میان اتصالات غیر طبیعی حدود ۱۵% از اتصالات به عنوان اتصالات نامناسب با گسستگی زیاد در فرایند ایجاد واحدهای ناموجود شناخته شده است. از میان این اتصالات نیز حدود ۳۹% اتصالات حایز شرایط شناخته شده و نمونه سمت چپ اتصال آنها با نمونه جدید ناموجود در دادگان جایگزین شده است. علاوه بر این بررسی زمان لازم برای سنتز داده تست نشان می دهد که سرعت سیستم سنتز تقریبا به اندازه ۲۰% کاهش یافته است.

در شکل ۴ در سمت راست یک اتصال نامناسب بین نمونه واجهای /u/ و /h/ نشان داده شده است. در سمت چپ برای حذف این اتصال، یک نمونه ناموجود ایجاد و جایگزین واج /u/ شده است. همانطور که مشاهده می شود، این جایگزینی موجب کاهش میزان گسستگی منحنی پیچ و انرژی شده است.

جدول ۱: اطلاعات آماری مربوط به اتصالات

درصد	اتصالات
%۴۰	طبیعی
%۳۵	میان همخوانهای واکدار یا واکه ها و واکدارها
%۲۵	دیگر موارد

ارزیابی های شنوایي غیر رسمی و مقایسه دقیق کیفیت اتصالات جدید نسبت به حالت قبل، حاکی از کاهش گسستگی های واضح در سیگنال خروجی و بهبود جزیي کیفیت خروجی سیستم می باشد. البته به دلیل نسبت کم تعداد اتصالات تغییر یافته به کل اتصالات، میزان بهبود کیفیت توسط شنوندگان معمولی قابل لمس نبوده و لذا ارزیابی های شنوایي رسمی صورت نگرفته است.



شکل (۴): سمت راست: نمونه نامناسب واج /u/ ، سمت چپ: واحد ناموجود سنتز شده و جایگزین شده بجای نمونه نامناسب

۵- نتیجه گیری و پیشنهادات

در این مقاله ما از ایده ایجاد واحدهای سنتز ناموجود در دادگان برای بهبود کیفیت خروجی سیستم سنتز با انتخاب واحد استفاده کرده ایم. در این حالت، بعد از یافتن نمونه های واج بهینه، اتصالات نامناسب مشخص و سپس با استفاده از واحد نیم واج سعی می شود که میزان گسستگی موجود در این اتصالات کاهش یابد. استفاده از این ایده، امکان استفاده از واحد نیم واج برای بهبود کیفیت خروجی بدون افزایش قابل توجه پیچیدگی محاسباتی لازم برای سنتز گفتار را فراهم می نماید. نتایج حاصل از ارزیابی سیستم، حاکی از بهبود کیفیت خروجی ناشی از بکارگیری از این ایده می باشد. برای بهبود این الگوریتم، می توان امکان جایگزینی نمونه جدید بجای نمونه راست اتصالات نامناسب نیز مورد بررسی قرار داد، هر چند که این بهبود، موجب افزایش قابل توجه پیچیدگی محاسباتی سیستم نیز می گردد.

قدرداني

این مقاله در قالب طرح پژوهشي ماده ۴۵ از طرف مرکز تحقیقات مخابرات ایران و سازمان مدیریت و برنامه ریزي مورد حمایت قرار گرفته است.

مراجع

[1] Black, A., and Campbell, N., *Optimizing Selection of Units from Speech Databases for Concatenative Synthesis*, ICSLP, pp. 581-584, 1995.

[2] Hunt, A., and Black, A., *Unit Selection in A Concatenative Speech Synthesis System Using A Large Speech Database*, ICASSP, pp. 373-376, 1996.

[3] Campbell, N., and Black, A., *CHATR: A multi-lingual speech re-sequencing synthesis system*, Institute of Electronic, Information and Communication Engineers, 1996.

[4] Black, A. W., and Taylor, P., *Automatically Clustering Similar Units for Unit Selection in Speech Synthesis*, ICASSP, Vol. 2, pp. 601-604, 1997.

[5] Sagisaka, Y., Kaiki, N., Iwahashi, N., and Mimura, K., *ATR v-TALK speech synthesis system*, ICSLP, Vol. 1, pp. 483–486, 1992.

[6] Black, A.W., and Taylor, P., *CHATR: A Generic Speech Synthesis System*, International Conference on Computational Linguistics, Vol. 2, pp. 983-986, 1994.

[7] Donovan, R.E., *Current Status of the IBM Trainable Speech Synthesis System*, 4th ISCA Tutorial and Research on Speech Synthesis, Edinburgh, 2001.

[8] Beutnagel, M., Conkie, A., J., Schroeter, Stylianou, Y. and Syrdal, A., *The AT&T Next-Gen TTS System*, 137th Acoustical Society of America Metting, 1999.

[9] Nukaga, N., Kamashida, R. and Nagamatsu, K., *Unit Selection Using Pitch Synchronous Correlation for Japanese Concatenative Speech Synthesis*, 5th ISCA Speech Synthesis Workshop, pp. 43-48, 2005.

[10] Low, P. H., and Vaseghi, S., *Synthesis of Unseen Context and Spectral and Pitch Contour Smoothing in Concatenated Text to Speech Synthesis*, ICCASP, Vol. 1, pp. 469-472, 2002.

[11] Beutnagel, M., Mohri, M., and Riley, M., *Rapid Unit Selection from a Large Speech Corpus for Concatenative Speech Synthesis*, Speech Communication and Technology, Vol. 2, pp. 607-610, 1999.

[۱۲] همایونپور، م. م. ، نم نبات، م. ، **فارس بیان: گامی اساسی در سنتز گفتار فارسی**، مقاله برگزیده دومین کارگاه پژوهشی زبان فارسی و رایانه، ص ۳۶۷-۳۷۵، ۱۳۸۵.

[13] Kishore, S. P., and Black, A. W., *Unit Size in Unit Selection Speech Synthesis*, ICASSP, pp. 1317-1320, 2003.

[14] Conkie, A., *A Robust Unit Selection System for Speech Synthesis*, 137th meet. ASA/Forum Acusticum, 1999.

[15] Clark, R. A. J., Richmond, K., and King, S., *Festival 2 - Build Your Own General Purpose Unit Selection Speech Synthesiser*, 5th ISCA Speech Synthesis Workshop, pp. 173-178, 2004.

زیرنویس‌ها

[1] Resynthesis