

مدلسازی منحنی گام گفتار فارسی با استفاده از مدل فوجی ساکی برای سیستم‌های تبدیل متن به گفتار

معصومه بحرینی، محمد مهدی همایونپور، سینا ایران نژاد

دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، آزمایشگاه سیستم‌های هوشمند صوتی-گفتاری

masoomah_bahreini@yahoo.com, homayoun@aut.ac.ir

چکیده - یکی از روش‌های تولید کنتور گام، برای سیستم‌های تبدیل متن به گفتار، روش فوجی ساکی می‌باشد. روش فوجی ساکی برای تولید کنتور گام، از جمع آثار دو بخش فرمان تکیه و فرمان عبارت‌ترا استفاده می‌نماید. این روش برای هر فرمان تعدادی پارامتر در نظر می‌گیرد، در این مقاله، به منظور استخراج پارامترهای فوجی ساکی، یک روش مبتنی بر گرادینان را به کنتور پیچ، اعمال می‌کنیم. نحوه استخراج پارامترهای مدل، به منظور ساخت بردارهای آموزشی و تخمین این پارامترها با استفاده از شبکه عصبی و ماشین پشتیبان بردار، برای یک متن داده شده، در این مقاله آورده شده‌اند. میزان کورولیشن بین منحنی اصلی گام و منحنی سنتز شده با استفاده از پارامترهای تخمینی مدل فوجی ساکی بیش از ۷۰٪ بدست آمده است.

کلید واژه- تبدیل متن به گفتار، شبکه عصبی، فوجی ساکی، ماشین پشتیبان بردار، منحنی گام.

۱- مقدمه

مانند ژاپنی، ثابت شده است که این فرامین، تناظر خوبی با اطلاعات زبانشناسی متن دارند. [2] در مورد زبان فارسی هم، می‌توان صحت این مطلب را تأیید کرد. هر کنتور گام در گفتار فارسی از یکسری پیک‌های محلی تشکیل شده است که این پیک‌ها را می‌توان معادل فرامین تکیه در مدل فوجی ساکی در نظر گرفت. همچنین گوینده در هنگام ادای گفتار در محدوده یک عبارت نوایی، آهنگ گفتار خود را تقریباً ثابت نگه می‌دارد و این متناظر با فرمان عبارت در مدل فوجی ساکی می‌باشد.

از بخش‌های مهم و تقریباً پیچیده مدل فوجی ساکی، استخراج اتوماتیک پارامترهای آن، از روی کنتور پیچ می‌باشد که برای آن روش‌های متعددی ارائه شده است. در [3] برای استخراج فرامین تکیه و عبارت، کنتور پیچ درونبایی و هموار شده را از فیلتر بالاگذر عبور می‌دهد، تا جزء فرکانس بالای آن، که متناظر با فرامین تکیه است بدست آید، سپس کنتور بدست آمده را از کنتور اصلی تفریق می‌کند. باقیمانده آن مجموع فرکانس پایه گوینده و فرامین عبارت است.

یکی از بخش‌های اساسی در سیستم‌های تبدیل متن به گفتار، تولید نوای گفتار و کنترل ویژگی‌های نوایی می‌باشد. نوای گفتار شامل انرژی، کشش و فرکانس پایه آن می‌باشد که از بین این موارد، در اکثر زبان‌ها، فرکانس پایه مهمتر از بقیه موارد می‌باشد. فرکانس پایه، به صورت یک منحنی در نظر گرفته می‌شود که برای تخمین این منحنی، روش‌های مختلفی مانند فوجی ساکی، تیلت، ToBI، و ... ارائه شده است. از روش‌های پرکاربرد آن می‌توان به مدل فوجی ساکی اشاره نمود که ابتدا در سال ۱۹۸۳ توسط فوجی ساکی و همکارانش معرفی گردید [1] و سعی می‌کند، برای یک جمله، با استفاده از یک رابطه ریاضی، منحنی گام را تولید نماید. این روش ابتدا برای تولید کنتور گام زبان ژاپنی [1] مورد استفاده واقع شد و سپس برای زبان‌های دیگر مانند چینی، انگلیسی، آلمانی، یونانی و ... اعمال شد [8].

مدل فوجی ساکی فرض می‌کند که کنتور گام، از دو نوع فرمان تکیه و عبارت، تشکیل شده است. برای بعضی زبان‌ها،