بررسی آماری ویژگی پروزودیک تأکید در زبان فارسی و تأثیر آن بر فرکانس پایه، طول و انرژی واکه ها

داود غرویان
دانشجوی دکتر

سید محمد احدی
استادیار

دانشکده مهندسی برق، دانشگاه صنعتی امیرکبیر

چکیده

یکی از خواص مهم پروزودیک گفتار، تأکید می‌باشد. بازشناسی گفتار با تأکید، با توجه به تغییر خواص آکوستیکی و پروزودیکی گفتار همواره به عنوان یک مسئله مهم مطرح بوده است. از طرفی ایجاد یک دادگان بزرگ که تمام حالتهای گفتار با تأکید را شامل باشد کاری بسیار مشکل است. در زبان فارسی نیز چنین دادگانی وجود ندارد. هدف این تحقیق بررسی اثر تأکید بر پارامترهای پروزودیکی نظیر طول زمانی، فرکانس پایه، شیب آن و انرژی می‌باشد تا بتوان از آنها بعداً به عنوان منبع اطلاعات جدید برای سیستم بازشناسی استفاده کرد. از آنجا که بیشترین تأثیر تأکید بر واکه می‌باشد در این تحقیق تأثیر تأکیدگذاری بر روی پارامترهای طول زمانی، انرژی، فرکانس پایه و شیب آن برای واکه‌ها بررسی شده است.

کلمات کلیدی

پروزودی، تاکید، طول زمانی(دیرش)، فرکانس پایه، شیب فرکانس پایه، انرژی.

Statistical Evaluation of the Influence of Stress on Pitch Frequency and Phoneme Durations and Energy in Farsi Language

D. Gharavian
Ph.D. Candidate

S.M. Ahadi
Assistant Professor

Electrical Engineering Department,
Amirkabir University of Technology

Abstract

Stress is known to be an important issue among the prosdic features of speech. The recognition of stressed speech has always been an important issue for speech researchers. Meanwhile, providing a large corpus, to be used in speech recognition and with the coverage of all different stressed conditions in a certain language, is a difficult task. Farsi (Persian) has been no exception to this. In this research, our aim has been to evaluate the effect of stress on prosodic features of Farsi language, such as phoneme duration, energy, pitch frequency and the pitch contour slope. These might be valuable in further research in speech recognition. As the main influence of stress is on vowels, the effects of stress on such parameters as duration, energy, pitch frequency and its slope for the vowels have been evaluated.

Keywords

Prosody, Stress, Duration, Pitch Frequency, Pitch Frequency Slope, Energy

مقدمه

پیشرفت‌های اخیر در بازشناسی گفتار نقش مهمی را در ارتباط انسان و کامپیوتر، برای این فن آوری در آینده ایجاد خواهد نمود. از طرفی، مشکلات بازشناسی، هنگامی که گفتار از حالت رسمی خارج شود بیشتر می گردد. یکی از حالتهای پروزودیک گفتار، تأکید می‌باشد که وجود آن سبب تغییر پارامترهای طول زمانی، فرکانس پایه و شیب آن می‌شود. اگر چه وجود تأکید در گفتار، نقش مهمی در هدایت شنونده برای دریافت مفهوم گفتار دارد[۲،۱] ولی عدم توجه به آن کار شناسایی گفتار را مشکل می‌سازد[۳]. به عنوان مثالی از بکارگیری پارامترهای پروزودیک در بازشناسی گفتار می‌توان گفت که: این اطلاعات در یک سیستم بازشناسی گفتار HMM می‌توانند به بردار ویژگی اضافه شوند[۴]، بصورت پس پردازش مورد استفاده قرار گیرند [۵]، همزمان با سیستم بازشناسی استفاده شوند (برخی از مسیرهای جستجوی ویتربی را حذف کنند) [۶] و یا در مدلهای ترکیبی و بازشناسی چند مرحله‌ای بکار روند[۷]. در زبانهای نواختی (Tonal) مهمترین پارامتر پروزودیک فرکانس پایه است، در حالی که در زبانهای تکیه گویشی (Stress Accent) پارامترهای طول زمانی و انرژی نیز مهم بوده و اعمال تأکید سبب افزایش طول و انرژی می‌گردد[۸]. در زبان فارسی مهمترین پارامتر تأثیر پذیر از تأکید، فرکانس پایه می‌باشد که مهمترین اثر تأکید، تغییر شیب پوش فرکانس پایه می‌باشد. از طرفی تأکید سبب افزایش طول و انرژی نیز می‌گردد[۹]. بررسی‌ها نشان می‌دهد که واکه‌ها بیش از سایر واج‌ها از تأکید تأثیر می‌پذیرند[۱۰]. ما تحقیقی در مورد طول زمانی واج‌های مختلف و خصوصاً واکه‌ها در زبان فارسی برای گفتار معمولی انجام داده و به این نتیجه رسیده‌ایم که طول واکه بستگی زیادی به نوع هجا دارد[۱۱]. در این تحقیق، با توجه به اینکه در زمینهٔ تأثیرگذاری تأکید در زبان فارسی کار آماری دقیقی انجام نشده، میزان تأثیرپذیری طول زمانی، فرکانس پایه و شیب آن را برای هر یک از واکه‌ها در اثر تأکید در مورد هجاهای مختلف بررسی کرده و قواعدی را برای این تغییرات ارائه خواهیم داد. نشان خواهیم داد که میزان تأثیر پذیری از تأکید به نوع هجا بستگی دارد.

۱ـ دادگان

تنها دادگان گفتار پیوسته موجود در بازار برای زبان فارسی، فارس دات[۱۲] می‌باشد. این دادگان شامل ۶۰۰۰ جمله از ۳۰۰ گوینده زن و مرد می‌باشد که هر یک از ۳۹۲ عبارت متفاوت، تعدادی را بصورت اتفاقی ادا کرده اند. این جملات بصورت عادی و بدون تأکید ادا شده‌اند. ما با استفاده از حدود ۱۸۰۰ جمله از جملات بالا، یک سیستم بازشناسی ناوابسته به گوینده بر اساس HMM تشکیل داده‌ایم.

برای تشکیل مدلهای ناوابسته، ابتدا از الگوریتم ویتربی[۱] با استفاده از تعداد محدودی جملات آموزشی که تقطیع زمانی شده‌اند، استفاده می‌شود. با این کار مدلهای HMM مقداردهی اولیه می‌شوند. تقطیع زمانی جملات آموزشی بصورت دستی صورت گرفته است. آموزش نهایی با استفاده از جملات آموزشی بیشتر(بدون تقطیع زمانی) و با استفاده از الگوریتم بام ـ ولش[۲] صورت می‌گیرد.

در این تحقیق از یک گوینده مرد برای ایجاد یک دادگان با تأکید استفاده شده است. این گوینده ۳۹۲ عبارت مذکور را ۳ بار بصورت عادی بیان کرده است. این جملات برای تطبیق مدلهای HMM با صدای گوینده مذکور و ایجاد یک سیستم وابسته به گوینده بکار برده شده است. این مدل را M1 می نامیم. از طرف دیگر این گوینده ۱۵۴ عبارت را در ۴۶۸ حالت مختلف از نظر نقطه اعمال تأکید ادا کرده است. مدلهای تأکیدی M2 با استفاده از مدلهای M1 ایجاد شده‌اند. برای تطبیق مدل ها در هر دو حالت از تطبیق بیزین[۳] ، یا MAP [۴] استفاده شده است[۱۳و ۱۴] .

برای بدست آوردن مرزهای واکه‌ها در جملات، نیاز به زمان بندی جملات بدون تأکید و با تأکید می‌باشد. از این مرزهای زمانی برای استخراج اطلاعات طول زمانی، فرکانس پایه و انرژی استفاده کرده‌ایم. مدلهای M1 و M2 برای بدست آوردن این مرزبندی‌ها با استفاده از ابزار HTK[۵] [۱۴] توسط الگوریتم ویتربی بدست آمده است. علاوه بر این، از HTK برای آموزش و تطبیق مدلها نیز استفاده شده است.

۲ـ واکه‌ها و هجاها

در زبان فارسی در حدود ۳۰ واج مختلف وجود دارد که شش عدد از آنها واکه می‌باشند که در جدول ۱ با استفاده از نمادهای IPA [۶] نشان داده شده‌اند.

جدول (۱) واکه‌های زبان فارسی

مثال	واکه
sæbr	/æ/
xɑb	/ɑ/
tʃɛrɑɢ	/ɛ/
diruz	/i/
gozæʃt	/o/
ruz	/u/

واجهای /æ/ ، /ɛ/ و /o/ را اصطلاحاً واکه‌های ضعیف و واجهای /ɑ/ ، /i/ و /u/ را واکه‌های قوی می‌نامیم.

واج /ow/ (نظیر صدای استفاده شده در کلمات "روشن" یا "شورا")، یک Diphthong است و خواصی مشابه واکه‌ها دارد. این تحقیق همراه با شش واکه، در مورد /ow/ نیز به بررسی نحوه تغییر پارامترهای پروزودیک می پردازد.

مطابق نظر اکثر زبان شناسان، در فارسی تنها سه نوع هجا وجود دارد که عبارتند از: CVC, VC و CVCC. در این میان V همان واکه است که یکی از شش واکهٔ بالا به همراه /ow/ می‌تواند باشد و C نیز بعنوان صامت، یکی از ۲۳ واج دیگر است.

۳ـ بررسی تغییرات پارامترهای پروزودیک بر اثر تأکید

۳ـ۱ـ فرکانس پایه

همانگونه که قبلاً گفته شد فرکانس پایه، مهمترین پارامتر پروزودیک زبان فارسی می‌باشد. استخراج فرکانس پایه با استفاده از روش Medan *et al.* [۱۵] و با استفاده از ابزار Pda (از مجموعهEdinburgh Speech Tools Library [۱۶]) صورت گرفته است. با استفاده از نتایج تقطیع زمانی، محدودهٔ پوش فرکانس پایه در مورد هر واکه بدست آمده و سپس مقدار متوسط این پوش محاسبه می‌شود. در این قسمت ما تغییرات این مقدار متوسط را بر اثر تأکید بررسی می‌کنیم. جدول ۲ تغییرات مقدار متوسط فرکانس پایه را بدون در نظر گرفتن نوع هجا نشان می‌دهد. در واقع نمونه‌های متعدد از هر قسمت جدول تشکیل توزیعی آماری را می‌دهند که ما در اینجا پارامترهای آن توزیع را ذکر کرده‌ایم. Std بیانگر انحراف معیار این توزیع می‌باشد. معیار Std/Mean نیز نزدیک بودن مقادیر پوش فرکانس پایه را به مقدار میانگین نشان می‌دهد. S/U نیز میزان این تغییرات را در حالت با تأکید به حالت بدون تأکید نشان می‌دهد. در شکل ۱ نیز، تغییرات میانگین فرکانس پایه واکه ها در هجاهای مختلف را می‌توان مشاهده کرد.

از مجموعهٔ این نتایج می‌توان به نکات زیر پی برد:

ـ تأکید سبب افزایش متوسط فرکانس پایه برای تمام واکه‌ها می‌شود.

ـ تأکید سبب کاهش نسبت Std/Mean می شود. یعنی فرکانس پایه در حالت با تأکید تغییرات کمتری خواهد داشت.

ـ نسبت افزایش فرکانس پایه حدود ۱۰٪ است.

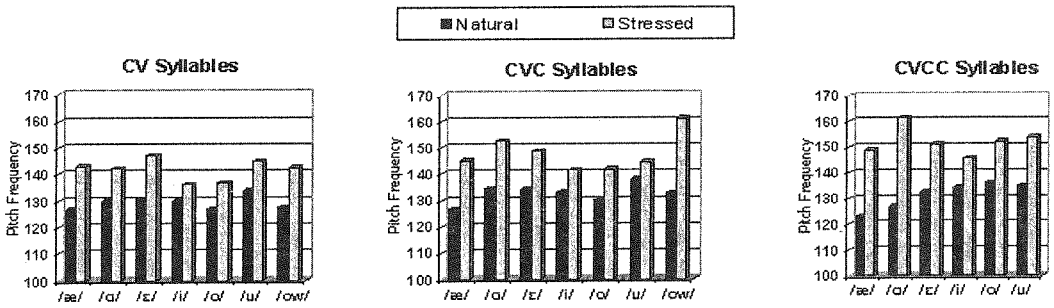
ـ در مورد واکه های ضعیف نسبت افزایش کمی بیشتر است.

ـ مقدار افزایش برای هجاهای CV و CVC نزدیک به هم و برای هجاهای CVCC بیش از این دو گروه می‌باشد. اگرچه با توجه به تعداد محدود هجاهای CVCC در دادگان، نتایج این گروه کاملاً قابل اعتماد نیست.

ـ نتایج مربوط به /ow/ تغییراتی بیش از واکه‌ها را در انواع متفاوت هجا نشان می‌دهد.

جدول (۲) نتایج تغییرات متوسط پوش فرکانس پایه برای واکه‌های مختلف بدون در نظر گرفتن نوع هجا.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	۱۲۵/۹۱	۲۰/۵۷	۰/۱۶۳۲	۱۴۵/۲۶	۱۹/۱۴	۰/۱۳۱۱	۰/۱۵۳۷
/ɑ/	۱۳۱/۳۶	۲۰/۶۲	۰/۱۵۷۱	۱۴۶/۵۷	۱۷/۸۲	۰/۱۲۱۷	۰/۱۱۵۷
/ɛ/	۱۳۱/۵۱	۲۰/۵۶	۰/۱۵۵۹	۱۴۷/۴۷	۲۰/۹۱	۰/۱۴۱۴	۱/۱۲۱۴
/i/	۱۳۱/۲۰	۲۰/۶۸	۰/۱۵۷۷	۱۳۸/۹۳	۱۸/۲۲	۰/۱۳۰۵	۱/۰۵۸۹
/o/	۱۲۹/۳۰	۲۰/۰۷	۰/۱۵۲۲	۱۴۰/۸۲	۱۸/۴۷	۰/۱۳۱۳	۰/۰۸۹۱
/u/	۱۳۶/۰۶	۲۰/۲۷	۰/۱۴۹۱	۱۴۵/۲۳	۲۰/۸۵	۰/۱۴۳۷	۱/۰۶۷۴
/ow/	۱۲۹/۱۳	۲۰/۰۳	۰/۱۵۵۲	۱۵۰/۶۳	۱۷/۹۸	۰/۱۱۹۵	۱/۱۶۶۵



شکل (۱) نتایج تغییرات متوسط پوش فرکانس پایه برای واکه‌ها در هجاهای مختلف.

۳ـ۲ـ شیب پوش فرکانس پایه

نقش شیب پوش فرکانس پایه در زبان فارسی مهمتر از مقدار میانگین پوش می‌باشد[۹]. برای بدست آوردن این شیب، پوش فرکانس پایه با یک خط تخمین زده و شیب این خط را به عنوان پارامتر پروزودیک مذکور در نظر گرفته‌ایم. با بررسی مقادیر مربوط به این شیب مشخص شد که برای دست آوردن نتایج معتبر برخی از داده‌های پراکنده باید حذف شوند. در واقع بررسی هیستوگرام شیب فرکانس پایه نشان می‌دهد که اکثر داده‌ها در نزدیک صفر (مقدار متوسط شیب، عددی منفی و کوچک است) تشکیل یک هیستوگرام گوسی داده و تنها تعداد محدودی مقادیر پراکندهٔ دور از توزیع وجود دارند. برای حذف مقادیر پراکنده از روش آنالیز فرکانس استفاده می‌کنیم. در این روش داده‌هایی که از میانگین توزیع به اندازه نصف انحراف معیار فاصله دارند حذف می‌شوند. ضریب ۰٫۵ بصورت تجربی انتخاب شده است. در جدول ۳ نتایج مربوط به شیب پوش فرکانس پایه بدون در نظر گرفتن نوع هجا و در جداول ۴ و ۵ و۶ نتایج مربوط به این شیب در انواع مختلف هجا بررسی شده است.

جدول (۳) نتایج تغییرات شیب پوش فرکانس پایه برای واکه های مختلف بدون در نظر گرفتن نوع هجا.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-۰،۳۰۲۵	۰،۴۸۹۷	-۱،۶۱۹۰	-۰،۲۱۷۹	۰،۳۷۹۸	-۱،۷۴۲۸	۰،۷۲۰۰
/ɑ/	-۰،۱۱۲۱	۰،۱۷۵۰	-۱،۵۶۱۵	-۰،۰۱۸۲	۰،۰۱۶۹	-۰،۹۲۸۶	۰،۱۶۲۴
/ɛ/	-۰،۷۰۴۷	۰،۷۴۴۵	-۱،۰۵۶۵	-۰،۴۶۰۹	۰،۵۷۷۴	-۱،۲۵۲۶	۰،۶۵۴
/i/	-۰،۲۲۴۹	۰،۳۳۲۲	-۱،۴۷۷۴	-۰،۱۰۹۶	۰،۲۰۹۷	-۱،۹۱۳۶	۰،۴۸۷۳
/o/	-۰،۳۸۸۱	۰،۵۲۴۹	-۱،۳۵۲۶	-۰،۱۴۸۲	۰،۰۵۲۶	-۰،۳۵۵۵	۰،۳۸۱۸
/u/	-۰،۰۸۷۲	۰،۱۳۱۰	-۱،۵۰۲۴	-۰،۰۱۸۶	۰،۰۰۹۷	-۰،۵۲۳۶	۰،۲۱۳۳
/ow/	-۰،۰۲۴۰	۰،۰۰۶۷	-۰،۲۸۰۶	-۰،۰۰۶۹	۰،۰۰۱۲	-۰،۱۷۴۶	۰،۲۸۷۵

جدول (٤) نتایج تغییرات شیب فرکانس پایه برای واکه‌های مختلف در هجای CV بعد از آنالیز واریانس.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-۰،۴۹۷۹	۰،۵۸۰۸	-۱،۱۶۶۳	-۰،۲۴۴۲	۰،۲۱۲۷	-۰،۸۷۱۲	۰،۴۹۰۰
/ɑ/	-۰،۱۲۰۴	۰،۱۶۷۵	-۱،۳۹۱۸	-۰،۰۲۱۷	۰،۰۱۱۸	-۰،۵۴۳۴	۰،۱۸۰۲
/ɛ/	-۰،۷۸۳۶	۰،۷۴۰۶	-۰،۹۴۵۱	-۰،۵۴۳۱	۰،۶۰۳۸	-۱،۱۱۱۷	۰،۶۹۳۲
/i/	-۰،۱۹۱۷	۰،۲۶۵۱	-۱،۳۸۲۸	-۰،۱۱۵۵	۰،۲۵۶۸	-۲،۲۲۳۵	۰،۶۰۲۰
/o/	-۰،۶۳۸۵	۰،۶۱۴۶	-۰،۹۶۲۵	-۰،۲۹۲۹	۰،۰۶۹۲	-۰،۲۳۶۱	۰،۴۵۸۷
/u/	-۰،۱۵۱۸	۰،۲۰۶۰	-۱،۳۵۶۹	-۰،۰۱۸۸	۰،۰۰۷۵	-۰،۳۹۶۲	۰،۱۲۳۸
/ow/	-۰،۰۳۱۵	۰،۰۰۵۶	-۰،۱۷۶۵	-۰،۰۰۸۳	۰،۰۰۸	-۰،۰۹۶۸	۰،۲۶۳۱

جدول (٥) نتایج تغییرات شیب فرکانس پایه برای واکه‌های مختلف در هجای CVC بعد از آنالیز واریانس.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-۰،۲۸۱۲	۰،۴۰۴۱	-۱،۴۳۷۲	-۰،۲۷۲۱	۰،۴۶۴۱	-۱،۷۰۵۴	۰،۹۶۷۶
/ɑ/	-۰،۰۸۸۳	۰،۱۵۱۴	-۱،۷۱۷۸	-۰،۰۱۸۵	۰،۰۱۹۴	-۱،۰۴۶۰	۰،۲۰۹۵
/ɛ/	-۰،۴۰۱۲	۰،۴۲۷۵	-۱،۰۶۵۷	-۰،۲۲۷۶	۰،۰۸۰۲	-۰،۳۵۲۴	۰،۵۶۷۳
/i/	-۰،۲۰۸۲	۰،۲۳۲۹	-۱،۱۱۸۶	-۰،۱۱۱۱	۰،۰۲۶۹	-۰،۲۴۲۵	۰،۵۳۳۶
/o/	-۰،۲۲۳۹	۰،۳۴۳۲	-۱،۵۳۲۷	-۰،۱۰۳۶	۰،۰۳۶۹	-۰،۳۵۶۲	۰،۴۶۲۷
/u/	-۰،۰۵۳۵	۰،۰۷۷۴	-۱،۴۴۵۹	-۰،۰۲۰۳	۰،۰۱۱۳	-۰،۵۵۶۸	۰،۳۷۹۴
/ow/	-۰،۰۱۰۰	۰،۰۰۲۵	-۰،۲۵۶۴	-۰،۰۰۵۵	۰،۰۰۱۳	-۰،۲۳۷۶	۰،۵۵۰۰

جدول (٦) نتایج تغییرات شیب فرکانس پایه برای واکه‌های مختلف در هجای CVCC بعد از آنالیز واریانس.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-۰،۰۴۶۵	۰،۰۸۲۵	-۱،۷۷۱۷	-۰،۰۱۴۵	۰،۰۱۰۸	-۰،۷۴۶۸	۰،۳۱۲
/ɑ/	-۰،۲۶۶۷	۰،۰۳۰۷	-۱،۱۵۲۳	---	---	---	---
/ɛ/	-۰،۰۴۱۹	۰،۰۴۵۳	-۱،۰۷۹۱	---	---	---	---
/i/	-۰،۰۳۷۹	۰،۰۵۳۸	-۱،۴۱۸۶	---	---	---	---
/o/	-۰،۰۲۵۹	۰،۰۱۱۷	-۰،۴۵۲۸	---	---	---	---
/u/	-۰،۰۲۲۹	۰،۰۰۴۷	-۰،۲۰۵۹	---	---	---	---

از بررسی جدول ۳ ، ۴، ۵، ۶ می‌توان گفت که اکثراً:

ـ قدر مطلق شیب کانتور Pitch با اعمال تأکید کم می‌شود.

ـ شیب در مورد واکه های ضعیف بیشتر از واکه‌های قوی است.

ـ نسبت Std/Mean پس از اعمال تأکید کم می‌شود و در واقع پراکندگی مقادیر شیب کمتر می‌شود.

ـ بزرگتر شدن هجا، معمولاً نسبت S/U را افزایش می دهد. افزایش نسبت در واقع مبین میزان تغییرات شیب فرکانس پایه کمتر است. این نکته در مورد هجاهای CV و CVC قابل مشاهده است. تعداد داده‌های مربوط به هجاهای CVCC بدلیل اعمال روش آنالیز واریانس، کم شده که از لحاظ آماری معتبر نیست و بنابراین ذکر نشده است.

ـ هر چه هجا بزرگتر می شود نسبتها در مورد واکه های کوتاه و بلند به هم نزدیکتر می شوند.

/ow/ کمترین شیب را دارد اما معمولاً به صورت نسبی بیش از واکه‌های قوی تغییر شیب دارد.

مشاهدهٔ نتایج نشان می‌دهد که پارامتر شیب، پراکندگی بیشتری نسبت به متوسط فرکانس پایه دارد. از طرفی چند مورد در جداول وجود دارد که نتیجه گیری های بالا را تأیید نمی‌کند. این تفاوتها در چند مورد، برای واکه‌های /æ/ ، /ɛ/ و /i/ قابل مشاهده است. بررسی هیستوگرام پارامتر شیب در دو حالت بدون تأکید و با تأکید برای این موارد تناقض، مطالبی را

روشن کرد. پس از اعمال تأکید، در موارد مذکور هیستوگرام توزیع، تبدیل به یک تابع توزیع با دو مد[۷] می‌شود. لذا کاهش محدودهٔ آنالیز واریانس در یک توزیع دو مد منجر به حذف داده‌های زیادی می‌شود. بررسی دقیقتر داده‌ها در توزیع، منجر به نتیجه‌گیری‌های زیر شد.

۱ـ در مورد /ɑ/ در هجاهای CVC آمدن واج‌های Plosive (/p/, /d/، /b/) در انتهای هجا سبب کاهش بیشتر میزان شیب در مقایسه با زمان وقوع صامت‌های از نوع Unvoiced fricative (/s/ و/ʃ/) می‌شود.

۲ـ در مورد /ɛ/ در هجاهای CV آمدن صامت های Plosive (/p/, /d/، /b/) سبب کاهش بیشتر شیب فرکانس پایه می‌شود.

۳ـ در مورد /i/ در هجاهای CV آمدن صامت های Plosive (/p/, /d/، /b/) سبب کاهش بیشتر میزان شیب در مقایسه با زمان وقوع صامت‌های از نوع Unvoiced fricative (/s/ و/ʃ/) و یا Nasal (/m/ و/n/) و یا Voiced fricative (/z/ و /v/) می شود.

۴ـ در مورد /i/ در هجاهای CVC آمدن صامت‌های Plosive (/p/, /d/، /b/) در انتهای هجا سبب کاهش بیشتر میزان شیب در مقایسه با زمان وقوع صامت‌های از نوع glide (/j/) و liquid (/l/) می‌شود.

با توجه به موارد بالا، بصورت کلی می‌توان گفت که وقوع یکی از صامتهای Plosive (/p/, /d/، /b/) می‌تواند سبب شود که شیب پوش فرکانس پایه بیش از حد معمول تغییر کند.

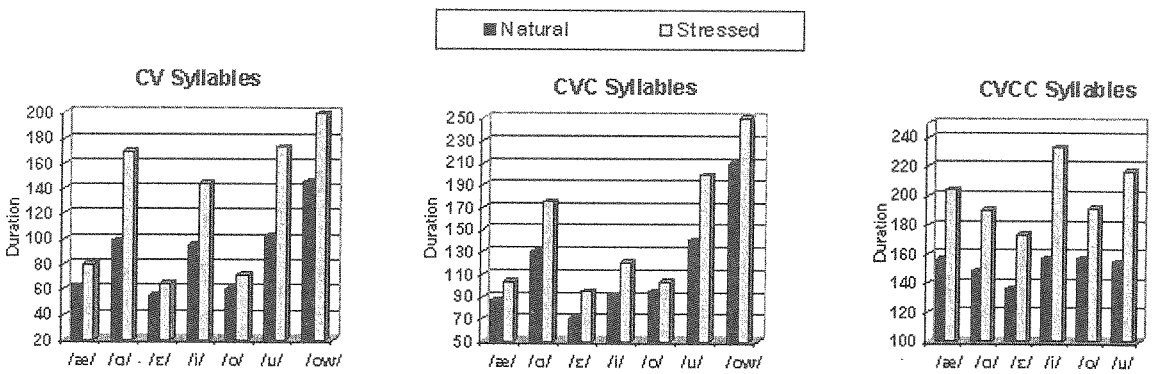
۳-۳ـ طول زمانی (دیرش)[۸]

۳-۳-۱ ـ بررسی کلی طول زمانی

در این قسمت به بررسی تأثیر تأکید بر طول واکه‌ها خواهیم پرداخت. مشاهده نتایج در جدول ۷ و شکل ۲ حاکی از تأثیر پذیری زیاد و قانونمند طول زمانی از تأکید بوده که قابل توجه می‌باشد. در جدول ۷، نتایج برای واکه‌های مختلف بدون در نظر گرفتن نوع هجا و در شکل ۲ نیز خلاصه نتایج این بررسی در هجاهای مختلف ارائه شده است.

جدول (۷) نتایج تغییرات طول زمانی واکه‌ها بر اثر تأکید.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	۹۵.۳۹	۴۵.۳۶	۰.۴۷۵	۱۱۷.۸۲	۵۹.۲۸	۰.۵۰۳	۱.۲۳۵
/ɑ/	۱۱۰.۴۸	۴۶.۰۰	۰.۴۱۶	۱۷۲.۲۷	۵۷.۷۷	۰.۳۳۵	۱.۵۵۹
/ɛ/	۵۹.۹۳	۲۹.۸	۰.۴۹۷	۷۳.۵۶	۳۳.۸۶	۰.۴۶۰	۱.۲۲۷
/i/	۹۷.۹۵	۴۶.۸	۰.۴۷۷	۱۴۰.۱۶	۶۲.۵۸	۰.۴۴۶	۱.۴۳۱
/o/	۸۶.۳۱	۴۲.۳۴	۰.۴۹۰	۹۸.۴۹	۴۲.۹۶	۰.۴۳۶	۱.۱۴۱
/u/	۱۲۴.۳۲	۵۲.۳۷	۰.۴۲۱	۱۸۹.۲۵	۴۸.۳۴	۰.۲۵۵	۱.۵۲۲
/ow/	۱۶۵.۰۶	۵۰.۳۷	۰.۳۰۵	۲۴۲.۰۰	۴۹.۴۵	۰.۲۰۴	۱.۴۶۶



شکل (۲) نتایج تغییرات طول زمانی برای واکه‌ها در هجاهای مختلف.

با توجه به شکل و جدول ارائه شده می‌توان نتیجه گرفت که:

ـ تأکید گذاری، طول واکه را افزایش می‌دهد.

ـ تأکید بر روی واکه‌های بلند بیشتر تأثیر دارد.

ـ نسبت تغییرات طول برای واکه‌های ضعیف حدود ۲۰٪ و برای واکه‌های بلند حدود ۵۰٪ است.

ـ هنگام تأکیدگذاری نسبت Std/Mean برای هر واکه کاهش می‌یابد. بنابراین می‌توان نتیجه گرفت که افزایش تأکید، پراکندگی طول واکه‌ها را کمتر می‌کند.

و با بررسی نتایج مربوط به هجاهای مختلف در شکل ۲ می‌توان گفت که:

ـ هر چه هجا بزرگتر می‌شود نسبت در واکه‌های بلند و کوتاه به هم نزدیکتر می‌شود.

ـ نسبت افزایش در مورد CV ها برای واکه‌های کوتاه حدود ۲۰٪ و برای واکه‌های بلند در حدود ۶۰٪ است.

ـ نسبت افزایش در مورد CVC ها برای واکه‌های کوتاه حدود ۱۰٪ و برای واکه‌های بلند در حدود ۳۰٪ است.

ـ نسبت افزایش در مورد CVCC ها برای واکه‌های کوتاه و بلند حدود ۳۰٪ است.

ـ میزان تغییرات طولی /ow/ در انواع مختلف هجا، کمتر از واکه‌های قوی می‌باشد و بیش از واکه‌های ضعیف است.

لازم به اشاره است که بجز اندازه هجا و سایر موارد ذکر شده، بطور کلی طول واکه از پارامترهای دیگری نیز تاثیر می‌پذیرد. از جمله در بررسی‌های انجام شده مشخص شد که مکث (Pause) نقش مهمی در طول واکه هجای بعد از آن دارد. در واقع در برخی موارد، اعمال تأکید بر قسمتی از جمله، مستلزم سکوت قبل از ادای آن می‌باشد. همچنین موقعیت واکه در کلمه بر طول آن تأثیر گذار است که در ادامه، این اثر بررسی می‌شود.

۳-۳-۲ ـ بررسی تأثیر موقعیت واکه در کلمه بر طول زمانی واکه

نتایج اولیه نشان می‌داد که موقعیت واکه در کلمه بر میزان تغییرات طولی با تأکید، مؤثر است. به همین دلیل بررسی‌های دقیق‌تری صورت گرفت که در ادامه به آن خواهیم پرداخت.

با بررسی‌هایی که صورت گرفت مشخص شد که در جملاتی که انتخاب شده حدود ۲۰۹۳ کلمه که ممکن است تکرار شده باشند وجود دارد، اگر چه تعداد کلمات متفاوت ۱۲۰۰ عدد می باشد. ۲۰۹۳ کلمه مذکور مطابق جدول زیر توزیع شده‌اند:

جدول (۸) تعداد کلمات موجود در دادگان بر اساس تعداد هجا.

تعداد هجا	۱	۲	۳	۴	۵	۶	۷
تعداد کلمه	۸۶۴	۷۳۳	۳۷۳	۱۱۳	۸	۱	۱

با توجه به مشکلاتی که از نظر دسترسی به دادهٔ کافی وجود داشت، حداکثر تا کلمات سه هجایی مورد بررسی قرار گرفتند. بنابراین هجا در سه مکان متفاوت در کلمه می‌تواند قرار گیرد. معیار محل هجا را نیز از انتهای کلمه در نظر گرفته و سپس در این موقعیتهای مکانی، تغییرات طولی واکه‌ها مورد بررسی قرار گرفته است. در جداول ۹ تا ۱۴ این نتایج در کلماتی که حداکثر سه هجا دارند آورده شده است. همچنین بررسی ها صرفاً شامل هجاهای CV و CVC می‌باشد.

جدول (۹) نتایج تغییرات طول زمانی برای واکه‌های مختلف- موقعیت واکه اول از انتها در هجاهای CV.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	۳۹.۱۷	۱۵.۰۵	۰.۳۸۴۳	---	---	---	---
/ɑ/	۱۰۴.۲۵	۳۵.۲۴	۰.۳۳۸۱	۱۴۷.۰۶	۴۲.۲۵	۰.۲۸۷۳	۱.۴۱۰۶
/ɛ/	۵۴.۳۸	۲۹.۲۴	۰.۵۳۷۶	۶۴.۰۷	۳۱.۳۶	۰.۴۸۹۴	۱.۱۷۸۲
/i/	۱۱۵.۱۹	۵۶.۷۱	۰.۴۹۲۳	۱۳۲.۶۳	۴۵.۵۶	۰.۳۴۳۵	۱.۱۵۱۴
/o/	۶۶.۲۴	۳۱.۹۷	۰.۴۸۲۶	۷۵.۵۶	۲۲.۹۷	۰.۳۰۴۱	۱.۱۴۰۶
/u/	۱۳۲.۴۲	۴۴.۵۱	۰.۳۳۶۱	۲۰۳.۳۳	۳۵.۰۲	۰.۱۷۲۳	۱.۵۳۵۵
/ow/	۱۸۵.۲۲	۳۸.۳۶	۰.۲۰۷۱	۲۴۵.۰۰	۳۸.۷۳	۰.۱۵۸۱	۱.۳۲۲۲

جدول (۱۰) نتایج تغییرات طول زمانی برای واکه‌های مختلف- موقعیت واکه اول از انتها در هجاهای CVC.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	۸۸.۲۱	۳۳.۵۲	۰.۳۸۰۰	۹۹.۴۱	۳۸.۰۲	۰.۳۸۲۵	۱.۱۲۶۹
/ɑ/	۱۳۵.۳۵	۶۰.۱۲	۰.۴۴۴۲	۱۷۴.۴۱	۶۶.۴۴	۰.۳۸۰۹	۱.۲۸۸۶
/ɛ/	۷۷.۸۸	۲۸.۹۵	۰.۳۷۱۷	۸۳.۳۳	۲۹.۷۲	۰.۳۵۶۷	۱.۰۶۹۹
/i/	۹۱.۴۸	۴۶.۷۸	۰.۵۱۱۴	۱۲۰.۰۰	۵۷.۷۴	۰.۴۸۱۱	۱.۳۱۱۷
/o/	۱۰۸.۶۲	۳۸.۱۵	۰.۳۵۱۲	۱۰۴.۲۹	۲۰.۶۵	۰.۱۹۸۰	۰.۹۶۰۰
/u/	۱۴۳.۲۳	۵۴.۸۲	۰.۳۸۲۷	۲۰۰.۸۶	۵۲.۶۰	۰.۲۶۱۹	۱.۴۰۲۳
/ow/	۲۰۹.۱۷	۴۸.۵۵	۰.۲۳۲۱	۲۷۸.۳۳	۴۱.۶۷	۰.۱۴۹۷	۱.۳۳۰۷

جدول (۱۱) نتایج تغییرات طول زمانی برای واکه‌های مختلف- موقعیت واکه دوم از انتها در هجاهای CV.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	۶۱.۴۴	۱۷.۷۵	۰.۲۸۸۹	۸۱.۵۷	۱۸.۴۸	۰.۲۲۶۶	۱.۳۲۷۷
/ɑ/	۹۵.۰۹	۳۲.۱۲	۰.۳۳۷۸	۱۷۲.۰۳	۵۴.۹۵	۰.۳۱۹۴	۱.۸۰۹۰
/ɛ/	۵۲.۹۰	۲۰.۰۹	۰.۳۷۹۷	۶۰.۳۴	۱۸.۰۲	۰.۲۹۸۷	۱.۱۴۰۷
/i/	۹۳.۹۱	۳۹.۸۹	۰.۳۹۲۹	۱۵۷.۲۴	۷۴.۰۶	۰.۴۷۱۰	۱.۶۷۴۳
/o/	۶۰.۳۲	۱۴.۵۵	۰.۲۴۱۳	۷۴.۳۸	۲۰.۹۷	۰.۲۸۱۹	۱.۲۳۳۰
/u/	۹۲.۹۰	۳۲.۳۰	۰.۳۴۷۷	۱۶۹.۳۸	۳۳.۹۵	۰.۲۰۰۵	۱.۸۲۳۳
/ow/	۱۴۷.۹۴	۳۲.۰۸	۰.۲۱۶۸	---	---	---	---

جدول (۱۲) نتایج تغییرات طول زمانی برای واکه‌های مختلف- موقعیت واکه دوم از انتها در هجاهای CVC.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	۸۵.۱۲	۲۸.۶۳	۰.۳۳۶۴	۱۰۲.۱۲	۳۸.۲۶	۰.۳۷۴۷	۱.۱۹۹۷
/ɑ/	۱۱۱.۵۴	۴۲.۳۶	۰.۳۷۹۸	۱۸۷.۱۴	۶۶.۹۶	۰.۳۵۷۸	۱.۶۷۷۷
/ɛ/	۶۸.۶۸	۲۱.۵۲	۰.۳۱۳۴	۱۰۶.۱۵	۲۶.۲۴	۰.۲۴۷۲	۱.۵۴۵۶
/i/	۱۰۰.۷۵	۳۶.۷۳	۰.۳۶۴۶	---	---	---	---
/o/	۸۵.۳۴	۲۳.۸۴	۰.۲۷۹۳	۱۰۳.۵۳	۲۹.۷۸	۰.۲۸۷۶	۱.۲۱۳۲
/u/	۸۳.۷۵	۱۵.۴۳	۰.۱۸۴۳	---	---	---	---

جدول (۱۳) نتایج تغییرات طول زمانی برای واکه‌های مختلف- موقعیت واکه سوم از انتها در هجاهای CV.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	۶۵.۶۶	۲۱.۰۸	۰.۳۲۱۱	۷۶.۲۵	۲۳.۹۲	۰.۳۱۳۷	۱.۱۶۱۳
/ɑ/	۱۰۰.۶۲	۳۰.۶۵	۰.۳۰۴۷	۱۹۵.۴۲	۴۶.۳۴	۰.۲۳۷۲	۱.۹۴۲۲
/ɛ/	۵۳.۳۳	۱۶.۶۵	۰.۳۱۲۱	۶۸.۸۹	۲۱.۳۹	۰.۳۱۰۵	۱.۲۹۱۷
/i/	۷۸.۹۱	۳۲.۸۷	۰.۴۱۶۶	۱۳۲.۱۴	۴۲.۴۶	۰.۳۲۱۳	۱.۶۷۴۵
/o/	۵۷.۱۰	۱۸.۸۷	۰.۳۳۰۵	۵۷.۵۷	۱۳.۴۵	۰.۲۲۹۷	۱.۰۲۵۸
/u/	۸۵.۰۰	۳۶.۹۲	۰.۴۳۴۳	۱۴۸.۰۰	۳۱.۱۴	۰.۲۱۰۴	۱.۷۴۱۱
/ow/	۱۱۲.۸۵	۲۰.۹۱	۰.۱۸۵۳	---	---	---	---

جدول (١٤) نتایج تغییرات طول زمانی برای واکه‌های مختلف- موقعیت واکه سوم از انتها در هجاهای CVC.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	٨٥،١٨	٢٩،٩٠	٠،٣٥١٠	١١٥،٨٨	٥٢،٥٧	٠،٤٥٣٦	١،٣٦٠٤
/ɑ/	١٠٣،٦٣	٤٨،٩٣	٠،٤٧٢١	---	---	---	---
/ɛ/	٦١،٨٦	٢٤،٢٤	٠،٣٩١٨	٨٤،٠٠	٤٥،٦١	٠،٥٤٢٩	١،٣٥٧٨
/i/	٨١،٠٥	٢٢،٣٣	٠،٢٧٥٦	---	---	---	---
/o/	٧٦،٤٦	٢١،٠٩	٠،٢٧٥٩	---	---	---	---
/u/	١٤٤،٥٤	٣٣،٢٨	٠،٢٣٠٢	---	---	---	---

از مجموعهٔ این جداول، اگر چه موارد نقض نیز وجود دارد، می‌توان نتیجه‌گیری کرد که اکثراً:

- هر چه هجا از انتهای کلمه دورتر می‌شود طول واکه کاهش می‌یابد (در مورد گفتار معمولی). در این مورد هجاهای CVC تبعیت بیشتری دارند.
- در مورد گفتار با تأکید، موقعیت واکه دوم از انتها، بیشترین طول را بصورت تقریبی دارد.
- از طرفی موقعیت دوم و سوم از انتها بیشترین تأثیر پذیری را از تأکید دارند.
- نسبت تأثیر پذیری واکه‌های قوی با دور شدن از انتهای کلمه بیشتر می‌شود.
- هر چه هجا بزرگتر می شود نسبت افزایش طول بر اثر تأکید کمتر می‌شود.

٣ـ٤ـ انرژی

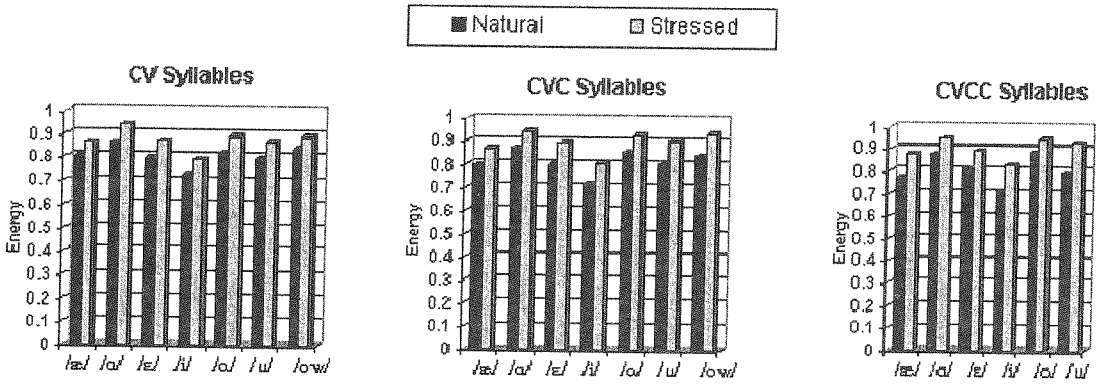
بدون شک تأکید بر انرژی تأثیر دارد که در ادامه به بررسی آن خواهیم پرداخت. در مورد طول زمانی و یا فرکانس پایه شدت ادای جملات بر نتیجه‌گیری مؤثر نیست اما در مورد انرژی، پارامترهایی نظیر فاصلهٔ گوینده از میکروفن و شدت ادای جملات در نتیجه‌گیری مؤثر است. اگر چه تا حد ممکن سعی شده یکنواختی رعایت شود، اما برای حداقل کردن این مشکلات، نرمالیزاسیون انرژی ضروری بنظر می‌رسد. این مطلب واضح است که تأکید، انرژی را افزایش می‌دهد بنابراین در جملات تأکیدی قسمتهای تأکید دار در سطح بالاتری نرمالیزه شده و می‌توانند با قسمتهای دیگر جمله که بدون تأکید است در کل دادگان تأکیدی مقایسه شوند. اما برای جملات عادی بنظر می‌رسد که راهی منطقی برای مقایسه با قسمتهای تأکید دار وجود ندارد. بنابراین در ادامه نتایج صرفاً از دادگان تأکیدی بدست آورده می‌شود.در جدول ١٥ نتایج تأثیر تأکید بر انرژی بدون در نظر گرفتن نوع هجا ارائه شده است و همچنین شکل ٣ نتایج تأثیر تأکید با توجه به نوع هجا است.

از این نتایج می‌توان به نکات زیر اشاره کرد:

- تأکید انرژی را افزایش می‌دهد.
- حدوداً ١٠٪ انرژی اضافه شده است.
- نسبت Std/Mean با اعمال تأکید کم شده است بنابراین پراکندگی مقادیر انرژی کاهش یافته است.
- هرچه هجا بزرگتر می‌شود، افزایش مقدار انرژی بیشتر است.
- اکثراً تغییرات انرژی واکه‌های بلند بیشتر است.

جدول (١٥) نتایج تغییرات انرژی برای واکه های مختلف بدون در نظر گرفتن نوع هجا.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	٠،٧٩١١	٠،٠٨٠١	٠،١٠١٣	٠،٨٦٩٧	٠،٠٧١٣	٠،٠٨١٩	١،٠٩٩٤
/ɑ/	٠،٨٦٨٦	٠،٠٧٦٢	٠،٠٨٧٧	٠،٩٤٢٦	٠،٠٣٦٩	٠،٠٣١٩	١،٠٨٥٠
/ɛ/	٠،٧٩٤٨	٠،٠٨٧٢	٠،١٠٩٧	٠،٨٧٦٨	٠،٠٨٠٢	٠،٠٩١٥	١،١٠٣٠
/i/	٠،٧٢٢٣	٠،٠٧٥٨	٠،١٠٤٩	٠،٧٩٩٢	٠،٠٧٩٣	٠،٠٩٩٢	١،١٠٦٥
/o/	٠،٨٤٢١	٠،٠٧٨٥	٠،٠٩٣٢	٠،٩١٤١	٠،٠٥٧٥	٠،٠٦٢٩	١،٠٨٥٦
/u/	٠،٧٩٩٨	٠،٠٦٨٤	٠،٠٨٥٥	٠،٨٩١٥	٠،٠٥٥٢	٠،٠٦١٩	١،١١٤٦
/ow/	٠،٨٤٠٤	٠،٠٦٦٧	٠،٠٧٩٤	٠،٩٠٩٩	٠،٠٣٥٦	٠،٠٩٣١	١،٠٨٢٧



شکل(۳) نتایج تغییرات انرژی برای واکه‌ها در هجاهای مختلف.

٤ـ نتیجه گیری

از مجموعه بررسی‌هایی که برای پارامترهای پروزودیکی نظیر طول زمانی، فرکانس پایه و شیب آن و انرژی صورت گرفت مشخص شد که تأکید بر هر چهار پارامتر پروزودیک مورد بحث تأثیر دارد. از طرفی نسبتها نشان می‌دهد که بیشترین تأثیر پذیری را شیب پوش فرکانس پایه دارد و بعد از آن طول زمانی ، متوسط پوش فرکانس پایه و انرژی تحت تأثیر قرار می‌گیرند. اما با نتایج گرفته شده نمی‌توان پیش‌بینی کرد که اولاً میزان تأثیرگذاری هر یک از این پارامترها در بازشناسی گفتار با تأکید چقدر است و دوم اینکه کدام پارامتر در بازشناسی مؤثرتر است. در واقع جواب این سئوالات پس از بکارگیری این نتایج در یک سیستم بازشناسی مشخص خواهد شد. از طرفی برای اینکه نتایج میزان تأثیرپذیری پارامترهای پروزودیک بهتر مشخص شود، می بایست که برچسب تأکید دارای چند مقدار باشد، زیرا معمولاً در یک قسمت با تأکید از جمله، همهٔ هجاهای آن به یک نسبت تأکید نمی‌پذیرند.

زیرنویس‌ها

1-Viterbi
2- Baum-Welch
3- Bayesian
4-Maximum *a Posteriori*
5- Hidden Markov Model Toolkit
6- International Phonetic Alphabet
7- Bi-modal
8- Duration

مراجع

[1] X. Huang, et.al, Spoken language processing, Prentice-Hall, 2001.
[2] E. Shriberg, et.al, Prosody-based automatic segmentation of speech into sentences and topics, Speech Communication, 32(1-2), September 2000.
[3] B. Shneilerman, The limits of speech recognition, Communication of the ACM, vol. 43, no. 9, September 2000.
[4] Y. Cheng and H.C. Leung, Speaker verification using fundamental frequency, In Proc ICSLP'98, Sydney.
[5] M. Kemal Sonmez, et.al , A lognormal tied mixture model of pitch for prosody-based speaker recognition, In Proc. Eurospeech'97, Rhodos, Greece.
[6] C. Wang and S. Seneff, Lexical stress modeling for improved speech recognition of spontaneous telephone speech in the JUPITER domain, In Proc Eurospeech 2001, Aalborg, Denmark.
[7] K. Hirose and K. Iwano, Detection of prosodic word boundaries by statistical modeling of MORA transitions of fundamental frequency contours and its use for continuous speech recognition, In Proc. ICASSP 2000, Istanbul, III-1763.
[8] R. Silipo and S. Greenberg, Automatic transcription of prosodic stress for spontaneous English discourse, The 14th International Congress of Phonetic Sciences, San Francisco, August, 1999.
[9] F. Almasganj, Structural analysis of the Farsi utterance using prosodic features, Phd Thesis, Tarbiyat Modarres University, Tehran, 1998.
[10] L. Hitchcock and S. Greenberg, Vowel height in intimately associated with stress accent in spontaneous American English discourse, In Proc Eurospeech 2001, Aalborg, Denmark.
[11] D. Gharavian, H. Sheikhzadeh and S.M. Ahadi, An experimental multi-speaker study on Farsi phoneme duration rules using automatic alignment, In Proc SST2000, Canberra, Australia.

[12] M. Bijankhan *et al.*, The speech database of Farsi spoken language, In Proc. SST'94, Perth, Australia.
[13] J-L.Gauvain and C-H.Lee, Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains, IEEE Trans. Speech and Audio Proc., vol. SAP-2, No. 2, April 1994, pp.291-298.
[14] S.J. Young *et al.*, The HTK Book, Cambridge University Eng. Dept., 2001.
[15] Y. Medan, E. Yair and D. Chazan, Super resolution pitch determination of speech signals, IEEE Trans. Sig. Proc., Vol. 39, No.1, January 1991.
[16] Edinburgh Speech Tools Library, available at http://festvox.org/docs/speech_tools-1.2.0/x2152.htm