https://jrl.ui.ac.ir/?lang=en

Journal of Researches in Linguistics
E-ISSN: 2322-3413
15(1), 119-136
Received: 07.02.2023 Accepted: 05.08.2023

Research Paper

Phonetics and phonology of the F0 valley in Persian intonation

Maedeh Ejei
Linguistics department, Faculty of Literature and Human sciences, Tehran University, Tehran, Iran
maedeh.ejei@gmail.com
Vahid Sadeghi*
English and Linguistics Department, Imam Khomeini International University, Qazvin, Iran
vsadeghi@hum.ikiu.ac.ir
Mahmood Bejankhan
University of Tehran, Tehran, Iran
mbjkhan@ut.ac.ir

Abstract
This research aims to improve speech synthesis in Persian and investigate the pitch contour fall between two H* peaks carried out in the framework of the Autosegmental Metrical (AM) theory of intonational phonology. In this paper, we tested two main hypotheses. Firstly, the F0 falls through phonetic interpolation and, secondly, it involves tone spreading. The present paper argues that the first hypothesis is proven to be correct. The data for the present research includes 45 enclitic phrases that were produced using voiced consonants between the two peaks. It means that in the course of producing each phrase, it was taken into consideration that all of the consonants placed between the two peaks would be deliberately and certainly voiced. This is mainly because if the consonants were not voiced, the data analysis results would be affected in Praat. It has to be noted that the distance between the two H* was increased by adding zero to two unstressed syllables. Totally, we recorded 1350 utterances from 30 native Persian speakers. We employed Praat software so as to analyse the utterances. Besides, we utilized SPSS for further analysis. The results of acoustic and statistical analyses showed that the distance between the first peak and the following F0 valley is increased with the addition of the unstressed syllables, demonstrating the alignment of this valley with the beginning of the stressed syllable. Results of statistical analyses revealed that the distance between the H and the following L target is increased significantly with the addition of the unstressed syllables. Furthermore, the normalized pitch contour was computed for all data. This indicates that the fall of the pitch contour between the two H peaks is realized through phonetic interpolation.
Keywords: Speech Synthesis, Phonetic Interpolation, Tone Spreading, Autosegmental Metrical theory (AM), Pitch Accent

Introduction

Text-to-speech technology can be used in many different ways such as allowing blind or visually impaired people to read texts. It also assists people with speech impairments to establish verbal communication and/or receive information from a text through listening. Speech intonation and the prosodic structure play a pivotal role in the process of synthetic speech production. The aim of this research is to investigate the pattern of the F0 declination between the H* and L+H* tonal targets in the Persian language. Being systematic, this declination is possible to predict F0 in the same tonal environments and use these predictions in the process of synthetic speech production of Persian phrases.

The current research is carried out employing Autosegmental Metrical (AM) theory of intonational phonology. According to Autosegmental Metrical theory, H and L tones are regarded as abstract phonological elements. The

*Corresponding author

2322-3413 © University of Isfahan
This is an open access article under the CC BY-NC-ND/4.0/ License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.22108/jrl.2023.136677.1714

occurrence of a L* pitch accent in a word stands for the fact that its stressed syllable is produced with a low tone and the occurrence of a H* pitch accent means that a stressed syllable with a high tone is produced. In bitonal pitch accents, the starred tone is aligned with the stressed syllable and the un-starred tone appears immediately before or after the starred one (Sadeghi, 2018). The difference between the monotonal H* accent and the bitonal L*+H accent can be shown in terms of the placement of the F0 peak and valley. In the H* accent, the F0 peak is placed on the stressed syllable. In the L*+H accent, the F0 valley is placed on the stressed syllable and the F0 peak occurs slightly after it. The pitch accent in Persian is defined as a bitonal L+H* pitch accent (Mahjani, 2003; Sadat-Tehrani, 2009). This is the consecutive combination of low and high tones. It is to note that both are aligned with the stressed syllable.

This research seeks to determine the type of the F0 declination, which can happen gradually or sharply. The gradual fall of F0 is due to phonetic interpolation. This is while the sudden and sharp decline in F0 is because of tone spreading. Given these considerations, two main hypotheses of this research run as follows: Firstly, the first H peak is interpolated to the L target. Secondly, the first H falls until the beginning of the second word in the phrase since the L in the second pitch accent spreads itself to the begnning of the second word.

Materials and Methods

The data for this research comes from phrases that are featured with two pitch accents (H* L+H*). Thus, 45 enclitic phrases were produced based on this pattern, using voiced consonants between the two peaks, and the distance between the two H* was increased by adding zero to two unstressed syllables. In order to collect these utterances, thirty native Persian speakers, including 15 male and 15 female speakers, took part in our project. Below are examples of each group of data:

(a) [ʔɑ.be **rud**] ɟelɑlud bud
1st word 2nd word

(b) tʃand mɑh piʃ dar [mɑziɕe.je mɑ.**li**] budand
1st word 2nd word

(c) ʔaz [mahal.le.je ɢa.di.**mi**] rafte budand
1st word 2nd word

We employed Praat and ProsodyPro software for analysing so as to calculate the distances between the tonal targets. Using SPSS, we carried out other tests with the obtained data in the previous step. These tests included ANOVA and post-hoc tests and the correlation between each group of data.

According to the data analysis, the L tone is commonly placed before the beginning of the stressed syllable. That is to say, the more the distance between the H tone in the first word and the stressed syllable, the more the L tone's distance becomes. This finding supports the first hypothesis pointing to the phonetic interpolation between the two tonal targets. In order to inspect the overall changes in the pitch contour, the normalized pitch contour was also computed for all the research data.

Discussion of Results and Conclusions

The results of analysing data confirm the validity of the phonetic interpolation hypothesis, according to which the L tone, being placed on the stressed syllable of the word, is where the F0 downtrend from the first H tone ends. To put it differently, the H tone in the first word and the L tone in the second word are interpolated to each other through a steady fall. Therefore, the F0 slope is distinct from the two tones. This distinction depends on the distance between the first H and the L tone. All the conducted tests in this research indicate that the L tone is aligned with the beginning of the stressed syllable, confirming that the F0 fall is realized through phonetic interpolation. As a result of the F0 fall between the first peak and the valley, the pitch contour's slope depends on the placement of the stressed syllable or the L tone.

The findings of this research can be used for synthetic speech production in Persian and in the prosodic pattern determination step of the text-to-speech systems. Predicting the precise changes in the F0 in the produced prosodic structure will result in a more natural production of the synthetic speech which will, in turn, enhance the functionality of the text-to-speech systems.

https://doi.org/10.22108/jrl.2023.136677.1714

نشریه پژوهش‌های زبان‌شناسی
سال پانزدهم، شماره اول، شماره ترتیبی۲۸، بهار و تابستان ۱۴۰۲، ص ۱۱۹-۱۳۶
تاریخ وصول: ۱۴۰۱/۱۱/۱۸ تاریخ اصلاحات: ۱۴۰۲/۰۴/۱۸ تاریخ پذیرش: ۱۴۰۲/۰۵/۱۴

مقاله پژوهشی

آواشناسی و واج‌شناسی دره زیروبمی در آهنگ گفتار فارسی

مائده اژه‌ای *
وحید صادقی **
محمود بی‌جن خان ***

چکیده

پژوهش حاضر با هدف بهبود بازسازیِ گفتار در زبان فارسی و بررسی افت منحنی فرکانس پایه بین دو قلهٔ زیروبمی در چارچوب نظریهٔ خودواحد عروضی آهنگ انجام شده است. دو فرضیهٔ اصلی این پژوهش افت فرکانس پایه از طریق وقوع قاعدهٔ درون‌یابی آوایی و یا از طریق وقوع قاعدهٔ گسترش نواخت است و پیش‌بینی می‌شود که فرضیهٔ اول پژوهش تأیید شود. داده‌های پژوهش، شامل ۴۵ ساخت پی‌بستی، با توجه به واکدار بودن همخوان‌ها در تمامی هجاها بین دو قلهٔ زیروبمی طراحی شده و فاصلهٔ دو قله با افزودن به تعداد هجاهای بدون تکیه از مقدار صفر تا دو هجا افزایش داده شد. تعداد ۱۳۵۰ داده از ۳۰ گویشور فارسیِ معیار ضبط و جمع‌آوری شد. این داده‌ها در نرم‌افزار پرات تحلیل آوایی شد و نتایج این تحلیل در برنامهٔ اس‌پی‌اس‌اس بررسی شد. نتایج تحلیل‌ها نشان داد فاصلهٔ قلهٔ اول تا درهٔ منحنی با افزایش تعداد هجاها افزایش یافته که نشان‌گر برهم‌نهادگیِ این دره با ابتدای هجای تکیه‌بر است. اختلاف معنادارِ این فاصله میان سه گروهِ داده‌ها به‌صورت دوبه‌دو نیز نشان‌دهندهٔ افزایش آن به نسبت افزایش تعداد هجاهای بدون تکیه است. همچنین منحنی فرکانس پایه به‌طور میانگین برای داده‌ها رسم شد که نشان داد این افتِ میان دو قله تدریجی است و از طریق وقوع درون‌یابی رخ می‌دهد.

کلیدواژه‌ها : بازسازی گفتار، درون‌یابی آوایی، گسترش نواخت، نظریهٔ خودواحد عروضی آهنگ، تکیهٔ زیروبمی

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

* دانش‌آموختهٔ کارشناسی ارشد گروه زبان‌شناسی، دانشکدهٔ ادبیات و علوم انسانی، دانشگاه تهران، تهران، ایران maedeh.ejei@gmail.com
**دانشیار گروه زبان انگلیسی و زبان‌شناسی دانشگاه بین المللی امام خمینی، قزوین، ایران (نویسنده مسئول) vsadeghi@hum.ikiu.ac.ir
*** استاد گروه زبان‌شناسی دانشگاه تهران، تهران، ایران mbjkhan@ut.ac.ir

2322-3413 © University of Isfahan
This is an open access article under the CC BY-NC-ND/4.0/ License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.22108/jrl.2023.136677.1714

۱. مقدمه

فناوری تبدیل متن به گفتار می‌تواند در زمینه‌های متعددی مانند خواندن متون مختلف توسط افراد نابینا و کم‌بینا، برقراری ارتباط کلامی توسط افراد با ناتوانی گفتاری و یا دریافت اطلاعات از یک متن از طریق شنیداری (کتاب یا متنِ گویا) راه‌گشا باشد. آهنگ گفتار و ساخت نوایی در فرایند بازسازیِ گفتار مصنوعی نقش بسزایی را ایفا می‌کنند. در بازسازی‌کننده‌های گفتاری کافی است که بازسازی‌کننده بتواند واحدهای نوایی[1] اصلی شامل زیروبمی، دیرش[2] و شدت[3] را تغییر بدهد تا با ترکیب آن با قواعد مربوط به آن زبان، بتوان مابقی واحدهای نوایی گفته‌شده را به‌دست آورد (همایون‌پور، ۱۳۹۰). فرکانس پایهٔ گفتار از اهمیت بسیاری در پردازش خودکار سیگنال گفتار برخوردار است. اطلاعات نوایی زیروبمی گفتار عمدتاً توسط این پارامتر مشخص می‌شوند.

سیستم تبدیل متن به گفتار شامل سه بخش است که یکی از آن، مرحلهٔ تحلیل متنی و آوایی است. در این مرحله، بخش مرتبط با نوای گفتار در خروجی‌ها مدنظر است. جایگاه‌های تکیه و سایر اطلاعات لازم برای انجام تحلیل‌های نوایی و تعیین الگوهای شدت و زیروبمی، تکیه و آهنگِ جملات از نتایج حاصل از این بخش است. پیش‌بینی نوع نوای گفتار، به علت عدم وضوح اطلاعات معنایی و ساختاری گفتار در ورودی‌های سیستم‌های متن به گفتار، مشکل است. درک نوای مناسب حتی با وجودِ این نوع از اطلاعات با مشکل مواجه است و علت آن عدمِ وجود تحقیقات کافی در زمینهٔ تعاملِ بین مشخصه‌های نوایی و همچنین تحت تاثیرِ عوامل ادراکی و متنی بودنِ تبدیل مقوله‌های زبانی به پارامترهای آوایی و نوایی است. تعریف پارامترهای مناسب در یک فضای نواختیِ پیش‌بینی شده لازمهٔ تولید صحیح گفتارِ بازسازی‌شده است. یکی از این پارامترها، جابه‌جایی‌های محلی نقاطی خاص بر روی منحنی فرکانس پایه در ساخت نوایی نمادین است که در پژوهش حاضر به آن پرداخته شده است.

تمرکز این پژوهش بر روی بررسی قاعده‌مند بودن افت فرکانس پایه در حدفاصل اهداف نواختی H* و L+H* در پاره‌گفتارهای زبان فارسی است. در صورت قاعده‌مند بودن این افت، می‌توان فرکانس پایه را در فضاهای مشابه نواختی به‌طور دقیق‌تری پیش‌بینی کرد و از آن در فرایند تولید پاره‌گفتارهای مصنوعی در زبان فارسی استفاده کرد.

پرسش اصلیِ این پژوهش تشخیص نوعِ افتِ منحنیِ فرکانس پایه است که می‌تواند به دو صورت تدریجی و یا با شیب تندتر انجام گیرد. در صورت رخ دادن نوع افت اول، درون‌یابی آوایی[4] رخ داده و در صورت صحیح بودن نوع افت دوم، قاعدهٔ گسترش رخ داده است. دو فرضیهٔ اصلی این پژوهش براساس این دو نوع افت است. در فرضیهٔ اول، قاعدهٔ درون‌یابی آوایی که براساس آن قلهٔ H اول به نواخت L تکیهٔ زیروبمی دوم درون‌یابی شده، مدنظر است. در فرضیهٔ دوم، قاعدهٔ گسترشِ نواخت[5] اعمال شده است که طبق آن، قلهٔ H اول تا ابتدای کلمهٔ دوم نزول می‌کند؛ زیرا نواخت L تکیهٔ زیروبمی دوم خود را تا ابتدای کلمه گسترده می‌کند.

الگوی نواختی مدنظر در طراحی داده‌ها، H* L+H* و محل هجای تکیه‌بر در کلمهٔ دوم مدنظر بوده است. پارامترهای اندازه‌گیری‌شده برای تحلیل و بررسی داده‌ها متمرکز بر فاصلهٔ H اول تا اهداف نواختی مدنظر در L+H* است. داده‌ها به‌طور کلی به دو صورت نرم‌افزاری و آماری تحلیل شدند. در بخش نرم‌افزاری فواصل اهداف نواختی اندازه‌گیری و در بخش آماری، آزمون‌های تحلیل واریانس و تعقیبی و همچنین، محاسبهٔ ضریب همبستگی و میانگین منحنی فرکانس پایه بین تمامی داده‌ها به تفکیک گروه انجام شد. نتایج تحلیل‌ها برمبنای داده‌های گردآوری‌شده بیانگر صحت فرضیهٔ مبنی بر درون‌یابی آوایی بین اهداف نواختی است.

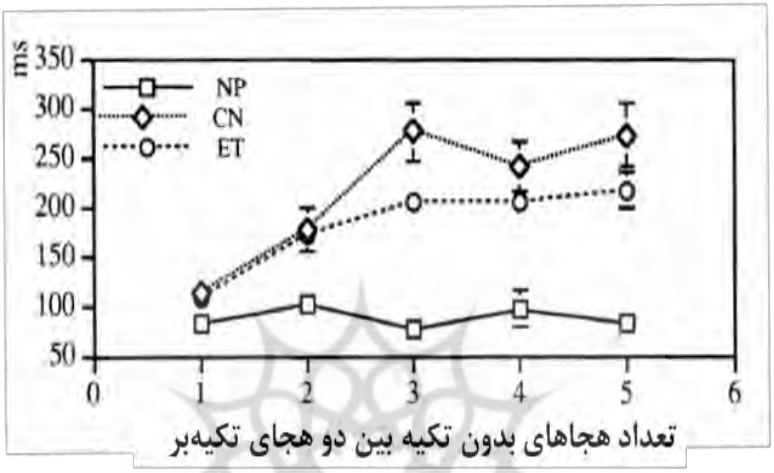
۲. پیشینهٔ پژوهش

افت فرکانس پایه بین دو قلهٔ H در زبان‌های دیگر بررسی شده است. آروانیتی[6] و لد[1] (1995) در مقاله‌ای، ترکیب نواختی و برهم‌نهادگی

[1] tone unit
[2] duration
[3] intensity
[4] phonetic interpolation
[5] tone spreading
[6] A. Arvaniti

https://doi.org/10.22108/jrl.2023.136677.1714

تکیه‌های زیروبمی پیش‌هسته را در زبان یونانی بررسی کردند. داده‌های آزمایش براساس فاصلهٔ دو تکیهٔ زیروبمیِ هدف طراحی شدند طوری که این فاصله با افزودن هجاهای بدونِ تکیه بین این دو هدف نواختی بیشتر می‌شد. فرضیهٔ مورد پژوهش، چگونگی افت فرکانس پایه در ابتدای تکیهٔ زیروبمیِ دوم بود. چنین فرض شد که اگر این افت به علت نزول منحنی بین دو H* رخ دهد، عمق درهٔ آن با افزایش تعداد هجاهای بدون تکیه میان دو تکیهٔ زیروبمی، بیشتر می‌شود؛ و اگر این افت به علت حضور یک نواختِ L مشخص در یک تکیهٔ زیروبمیِ دونواختی باشد، برهم‌نهادگی و مقدار این نواخت حتی با افزایش تعداد هجاهای بدون تکیه پایدار می‌ماند. نتایج این پژوهش نشان داد بازنویسیِ تکیهٔ زیروبمی پیش‌هسته در این زبان به‌صورت L*+H است؛ زیرا نواخت L به‌طور مشخص نتیجهٔ نزول منحنی نیست و از لحاظ مقدار و برهم‌نهادگی از نواخت H پایدارتر است.



شکل ۱– فاصلهٔ ابتدای هجای تکیه‌بر تا نواخت *H* در تکیهٔ اول به‌صورت تابعی از تعداد هجاهای بدون تکیه بین دو هجای تکیه‌بر در سه گویشور زبان یونانی.(Cited in Arvaniti & Ladd, 1995)

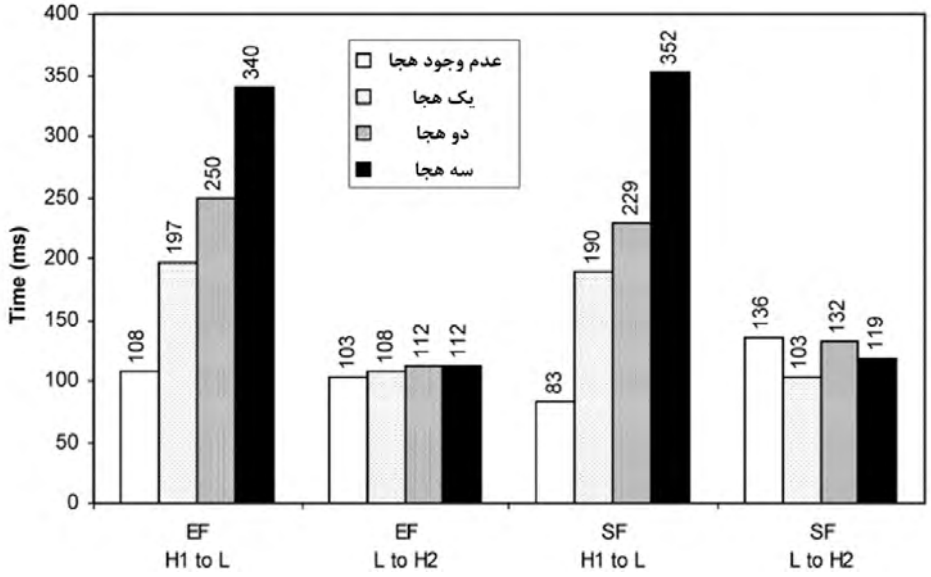
Figure 1 – The distance of the H tone from the beginning of the first accented syllable as a function of the number of unaccented syllables between accents, for 3 Greek speakers (Cited in Arvaniti & Ladd, 1995).

در شکل (۱) مشاهده می‌شود که فاصلهٔ نواخت H از ابتدای هجای تکیه‌بر با افزایش تعداد هجاهای بدون تکیه بیشتر می‌شود. شیب این منحنی پس از رسیدن به هجای سوم و یا پیش از هجای ماقبل‌پایانی کمتر می‌شود؛ اما به‌طور کلی با افزایش تعداد هجاهای بدون تکیه بین این دو هدف نواختی تا مقدار سه هجا، نواختِ H از ابتدای هجای تکیهٔ زیروبمی دورتر واقع می‌شود.

لد و شپمن[۲] (2003) نیز در آزمایش‌هایی مشابه افت فرکانس پایه را میان دو تکیهٔ زیروبمی بالا (H*) در زبان انگلیسی بررسی کردند. هدف آنان بررسی میزان فرورفتگی بین این دو تکیه و ارزیابی مدل ناحیهٔ گذار بود که توسط پی‌یرهامبرت[۳] (1980) ارائه شده بود. سه آزمایش در این پژوهش انجام شد. اولین آزمایش مربوط به برهم‌نهادگیِ L بر شروع هجای تکیه‌بر بود. نتیجهٔ این آزمایش نشان داد نقطهٔ حداقلی فرکانس پایه با یک نواختِ پایینِ مشخص مطابقت دارد. در آزمایش دوم نشان داده شد که برهم‌نهادگیِ نقطهٔ حداقلی، یک سرنخ ادراکی برای تشخیص عبارت‌های مشابه توسط شنونده است. آزمایش سوم ادعای پیرهامبرت دربارهٔ فرورفتگی بین دو تکیهٔ زیروبمی بالا را به محک آزمون قرار داد. طبق گفتهٔ پی‌یرهامبرت، فرورفتگی بین این دو تکیه فاقد هدف نواختی پایین (L) است و این درّه با افزایش فاصله بین دو قلهٔ فرورفتگی، عمیق‌تر می‌شود. در این آزمایش تعداد هجاها در فاصلهٔ بین دو دو قلهٔ زیروبمی از تعداد صفر تا سه هجا افزوده شده و نقطهٔ حداقلی فرکانس پایه در این ناحیهٔ گذار اندازه‌گیری گردید.

[1] D. R. Ladd
[2] A. Schepman
[3] J. B. Pierrehumbert

https://doi.org/10.22108/jrl.2023.136677.1714



شکل ۲ – داده‌های زمانی دو گویشور زبان انگلیسی در آزمایش سوم که نشان‌گر طول فواصل بین دره و دو قلهٔ زیروبمی است (Cited in Ladd & Schepman, 2003)

Figure 2 – The durational data for two English speakers in Experiment 3 demonstrating the distance between the valley and the two peaks (Cited in Ladd & Schepman, 2003)

همان‌طور که در شکل (۲) مشاهده می‌شود، فاصلهٔ زمانی بین L و H2 تغییر چندانی با افزایش تعداد هجا نمی‌کند که این نشان‌گر ثابت بودن محل وقوع L قبل از هجای تکیه‌بر است؛ اما فاصلهٔ L از H1 با افزایش تعداد هجاها به‌طور نظام‌مند افزایش می‌یابد. نتایج کلی این پژوهش نشان داد فرورفتگی ناحیهٔ گذار بین دو قلهٔ زیروبمی ناظر بر یک نواخت L مشخص است.

۳. چارچوب نظری پژوهش

پژوهش حاضر در چارچوب نظریهٔ خودواحد عروضی[1] انجام شده است. هدف اصلی این نظریه توصیف و تبیین تغییرات آوایی پیوسته در منحنی زیروبمی گفتار است. لد (2008) این هدف کلی را به دو رسالت واجی و رسالت آوایی تقسیم کرده است. رسالت واجیِ این نظریه، ارائه تحلیلی کارآمد از تغییرات پیوستهٔ زیروبمی براساس زنجیره‌ای از عناصر مقوله‌ای و تمایزدهنده و رسالت آواییِ آن تبیین نحوهٔ نگاشت عناصر ناپیوستهٔ واجی به پارامترهای آکوستیکیِ پیوسته است.

H و L در این نظریه، دو عنصر واجی با ماهیت انتزاعی هستند و ماهیت آوایی ندارند. این دو عنصر ممکن است در شرایط مختلف به گونه‌های متفاوت ظاهرشوند. در جدول (۱) به‌طور خلاصه به انواع نواخت و محل آن‌ها پرداخته شده است.

[1] Autosegmental Metrical Theory

https://doi.org/10.22108/jrl.2023.136677.1714

جدول ۱ - انواع نواخت در نظریهٔ خودواحد عروضی آهنگ

Table 1 – Different types of tones in Autosegmental Metrical Theory

نوع	نام	نماد	محل وقوع
-	نواخت بالا	H	-
	نواخت پایین	L	-
تکیه زیروبمی	تکیه زیروبمی زیر	H*	تولید هجای تکیه‌بر با نواخت بالا
	تکیه زیروبمی بم	L*	تولید هجای تکیه‌بر با نواخت پایین
نواخت‌های مرزی	نواخت کناری زیر	H-	ناظر بر بخشی از زنجیرهٔ گفتار بین تکیه زیروبمی هسته پاره‌گفتار و پایان پاره‌گفتار
	نواخت کناری بم	L-	
	نواخت مرزنما زیر	H%	ناظر بر خیز پایانیِ منطبق با مرز پایانی پاره‌گفتار
	نواخت مرزنما بم	L%	ناظر بر نبودِ خیز پایانی در انتهای پاره‌گفتار

منحنی زیروبمی در نظریهٔ خودواحد عروضی شامل دو نوع ویژگی زیروبمی است. برخی ویژگی‌ها محلی و در نقاط خاصی از منحنی هستند و برخی دیگر، به صورت گسترهٔ زیروبمی بین رویدادهای نواختی دیگر واقع شده و محدودهٔ گذار هستند. از بین این دو، فقط رویدادهای نواختیِ محلی اهمیت زبان‌شناختی دارند و نوع دوم در این نظریه فاقد اهمیت است. مهم‌ترین رویدادهای محلی ساخت نواختی گفتار، تکیهٔ زیروبمی و نواخت‌های مرزی است. تکیهٔ زیروبمی با هجاهای برجسته در ساخت زنجیره‌ای و نواخت‌های مرزی با مرز واحدهای نوایی منطبق هستند. با توجه به اینکه تکیهٔ زیروبمی و نواخت‌های مرزی در نواحی نسبتاً مشخصی در زنجیرهٔ گفتار رخ می‌دهند، به این رویدادها «اهداف نواختی»[1] گفته می‌شود. تمایز بین این دو رویداد تمایزی اساسی در انگارهٔ خودواحد عروضی است (صادقی، ۱۳۹۷).

تکیه‌های زیروبمی برحسب جایگاهشان درگروه آهنگ به دو دستهٔ «تکیهٔ زیروبمی هسته»[2] و «تکیهٔ زیروبمی پیش‌هسته»[3] تقسیم می‌شوند. در اغلب زبان‌ها، آخرین تکیهٔ زیروبمی گروه آهنگ، تکیهٔ زیروبمی هسته و به تکیه‌های زیروبمی قبل از آن تکیهٔ زیروبمی پیش‌هسته گفته می‌شود.

نشانهٔ ستاره در نظریهٔ خودواحد عروضی به مفهوم انطباق نواخت با هجای تکیه‌برِ کلمه یا گروه تکیه‌ای است. گروه تکیه‌ای[4] در ساخت سلسله‌مراتبی آهنگ گفتار از کلمه بزرگتر و از گروهِ آهنگ کوچک‌تر است و در آن یک تکیهٔ زیروبمی وجود دارد که روی قوی‌ترین هجای گروه قرار می‌گیرد (Beckman & Pierrehumbert, 1986).

وقوع تکیهٔ زیروبمی L* روی یک کلمه به معنای آن است که هجای تکیه‌بر آن با نواخت پایین تولید شده و وقوع تکیه زیروبمی H* به معنی تولید هجای تکیه‌برِ کلمه با نواخت بالاست. در تکیه‌های زیروبمی دونواختی، نواخت ستاره‌دار، با هجای تکیه‌بر کلمه انطباق دارد و نواخت بدون ستاره با فاصله کمی قبل و یا بعد آن ظاهر می‌شود (صادقی، ۱۳۹۷).

تفاوت تکیهٔ تک نواختی H* و تکیهٔ دونواختیِ L*+H در محل قله و درهٔ فرکانس پایه است. در H* قلهٔ فرکانس پایه روی هجای تکیه‌بر است. در L*+H درهٔ فرکانس پایه روی هجای تکیه‌بر قرار دارد و قله کمی بعد از این هجا قرار دارد. لد (1983) و گوسن‌هافن[5] (1984) تمایز بین این دو تکیه را صرفا آواشناختی و غیرمقوله‌ای بیان کرده‌اند. یعنی قلهٔ H تکیهٔ تک‌نواختیِ H* منطبق بر هجای تکیه‌بر و وقوعِ به-

[1] tonal targets
[2] nuclear pitch accent
[3] pre-nuclear pitch accent
[4] accentual phrase
[5] C. Gussenhoven

https://doi.org/10.22108/jrl.2023.136677.1714

هنگام دارد ولی قلهٔ H تکیه دونواختی L*+H بعد از هجای تکیه‌بر قرار دارد و وقوعِ دیرهنگام دارد. بنابراین L*+H گونه ای ازهمان الگوی نواختی H* است. در تکیه دونواختیِ L+H* نقطهٔ شروع خیز، هجای قبل از هجای تکیه‌بر و محل پایان آن، ناحیه‌ای از هجای تکیه‌بر است. در حالی که در L*+H، شروع خیز از هجای تکیه‌بر و در ناحیه‌ای از هجای بدون تکیه بعد پایان می‌یابد.

۱-۳. ساخت آهنگ گفتار فارسی

سادات‌تهرانی (2009 & 2007) ساخت آهنگ فارسی را متشکل از دو سطح نوایی گروه تکیه‌ای[1](AP) و گروه آهنگی[2](IP) تعریف می‌کند. گروه تکیه‌ای از یک کلمهٔ محتوایی و واژه‌بست‌های وابسته به آن تشکیل می‌شود و گروه آهنگی از یک یا چند گروه تکیه‌ای در ترکیب با هم شکل می‌گیرد. تکیهٔ زیروبمی در زبان فارسی به‌صورت تکیهٔ دونواختی L+H* تعریف شده است (Sadat-Mahjani, 2003 ; Tehrani, 2009). این تکیه ترکیبِ متوالی یک نواخت پایین(L) و یک نواخت بالا (H) است که هردو با هم با هجای تکیه‌بر کلمه انطباق دارد. تکیه دونواختی L+H* در بازنویسی ساخت نواختی کلمات چندهجایی با تکیهٔ غیرآغازی استفاده می‌شود. در کلمات یک‌هجایی یا چندهجایی با تکیهٔ آغازی، L+H* به‌صورت تکیهٔ تک‌نواختی H* ظاهر می‌شود؛ زیرا فضای کافی برای تظاهر آوایی نواخت L وجود ندارد. اسلامی (۱۳۸۴) معتقد است تکیهٔ زیروبمی در زبان فارسی هم به‌صورت بسیط L*، H* و هم به‌صورت مرکب L+H*، L*+H بازنویسی و توصیف می‌شود.

گروهِ آهنگ دارای یکی از نواخت‌های مرزی L% یا H% است. L% در پایان جملات خبری و امری و H% در پایان جملات پرسشی ظاهر می‌شود. در هر گروه تکیه‌ای یک نواخت کناری به‌صورت -L یا -H ظاهر می‌شود که بین محل وقوع تکیهٔ زیروبمی و مرز پایانی گروه قرار دارد (اسلامی، ۱۳۸۴؛ Sadat-Tehrani, 2009). اگر پایان تکیهٔ زیروبمی منطبق بر پایان گروه باشد، نواخت کناری روی هجای تکیه‌بر قرار می‌گیرد. اگر بین محل وقوع تکیهٔ زیروبمی و پایان گروه تکیه‌ای به دلیل وجود واژه‌بست، یک یا چند هجا وجود داشته باشد، نواخت کناری روی تمام هجاهای بدون تکیه بعد از هجای تکیه‌بر تا پایان گروه تکیه‌ای گسترده می‌شود (Sadat-Tehrani, 2009).

سادات‌تهرانی (2009) علت وجود نواخت کناری در ساخت آهنگ گفتار فارسی را تفاوت بین نواخت کناریِ گروه تکیه‌ای هسته و پیش‌هسته می‌داند. نواخت کناری گروه تکیه‌ای هسته در جملات سادهٔ بی‌نشان -L و نواخت کناری گروه تکیه‌ای پیش‌هسته به‌صورت -H است. بنابراین، تفاوت بین گروه تکیه‌ای هسته با گروه تکیه‌ای پیش هسته، نوع نواخت کناری گروه تکیه‌ای است. وی ترادف اهداف نواختی H و L در تکیه‌های زیروبمی دونواختی L+H* در زبان فارسی را در جایگاه‌های نوایی پیش‌هسته، هسته و کانونی بررسی کرده است و نشان می‌دهد که نقاط گذار L در تمامی گروه‌های تکیه‌ای به‌طور نظام‌مند مترادف با آغاز هجای تکیه‌بر است؛ ولی محل ترادف نقاط گذار H بسته به نوع تکیهٔ زیروبمی تغییر می‌کند، بدین معنی که وقوع قله‌ها در تکیه‌های زیروبمی پیش هسته نسبت به تکیهٔ زیروبمیِ هسته یا کانونی دیرهنگام‌تر است. البته در ادامه بیان کرده است که نقاط پایانیِ خیز در تمام گروه‌ها صرف‌نظر از نوع تکیهٔ زیروبمی، در هجای بعد از هجای تکیه‌بر واقع می‌شود و این به معنای وقوع دیرهنگام قله و یا دیرکرد قلهٔ هجا در زبان فارسی است. در تکیه‌های زیروبمی هسته یا کانونی، H با همخوان آغازیِ هجای بعد از هجای تکیه‌بر مترادف است؛ ولی در تکیهٔ زیروبمی پیش‌هسته با آغاز واکهٔ این هجا مترادف است. سادات‌تهرانی (2009:14) این تفاوت را به نواخت‌های مرزی متفاوتِ این تکیه‌ها نسبت داده است.

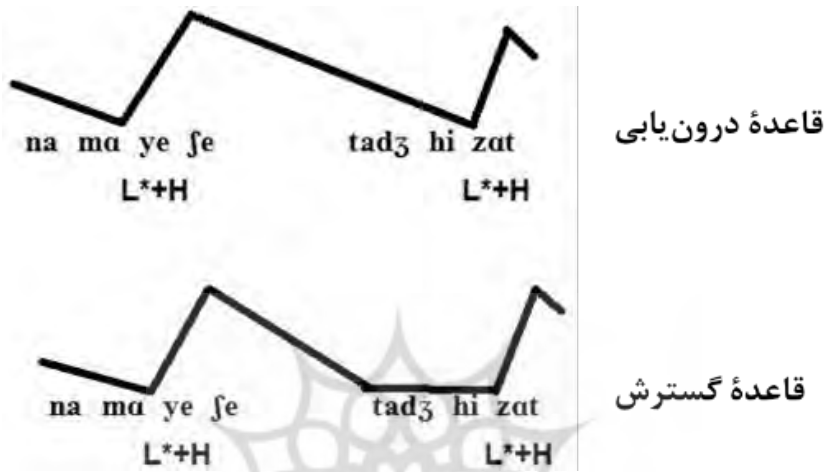
صادقی (۱۳۹۷) طی پژوهش‌هایی نشان داده است که هر دو نواخت تکیهٔ زیروبمیِ پیش‌هسته در زبان فارسی با نواحیِ مشخص در ساخت زنجیره‌ایِ گفتار انطباق پایدار دارد. L به‌طور منظم همیشه روی آغازهٔ هجای تکیه‌بر و H همیشه در مرز آغازی واکه بعد از هجای تکیه‌بر قرار دارد. نتایج مشاهدات صادقی، فرضیهٔ اتصال زنجیره‌ای را اثبات کرده است و نشان داد که دیرش و شیب تغییرات زیروبمی در زبان فارسی ثابت نیست و به نسبت فاصلهٔ زمانی بین نقاط انطباق زنجیره‌ای تغییر می‌کند. صادقی (۱۳۹۷) مطابق این فرض که اطلاق نشانهٔ ستاره به یک نواخت صرفاً به معنی انطباق آن نواخت با زنجیرهٔ آواییِ هجای تکیه‌بر است، تکیهٔ زیروبمی پیش‌هستهٔ فارسی را به‌صورت L*+H بازنویسی کرده است. طبق این بازنویسی نواختی، نواخت L همواره روی هجای تکیه‌بر واقع شده است و H همواره با تأخیر روی هجای بعد از هجای تکیه‌بر واقع می‌شود.

[1] Accentual Phrase
[2] Intonational Phrase

https://doi.org/10.22108/jrl.2023.136677.1714

۴. روش اجرای پژوهش

داده‌های نواییِ مدنظر این پژوهش، عبارت‌هایی هستند که توالی دو تکیه زیروبمی به‌صورت H* L+H* (یک نواخت H* و یک گروه زیروبمیِ دونواختیِ L+H*) را شامل می‌شوند. دو احتمال برای ناحیهٔ گذار از H* اول تا H* دوم فرض شده است. در فرضیهٔ اول، قاعدهٔ درون‌یابی آوایی پیش‌بینی شده است که براساس آن قلهٔ H اول به نواخت L تکیهٔ زیروبمی دوم درون‌یابی شده است. در فرضیهٔ دوم، قاعدهٔ گسترش اعمال شده است که طبق آن، قلهٔ H اول تا ابتدای کلمهٔ دوم نزول می‌کند؛ زیرا نواخت L تکیهٔ زیروبمی دوم تا ابتدای کلمه خود را گسترده می‌کند. الگوی تغییرات فرکانس پایه براساس این دو فرضیه در شکل (۳) ارائه شده است.



شکل ۳ – دو قاعدهٔ مدنظر فرضیات پژوهش

Figure 3 – Phonological rules underlying the research hypotheses

۴-۱. داده‌های پژوهش و شرکت‌کنندگان

داده‌های این پژوهش شامل ۴۵ عبارت دوکلمه‌ای است. هر داده از یک پاره‌گفتار متشکل از دو کلمهٔ پشت‌سرهم تشکیل شده است که با پی‌بستِ کسرهٔ اضافه به هم متصل می‌شوند. تمامی داده‌ها براساس واکدار بودن همخوان‌ها در محل حضور محل هجای تکیه‌بر و اهداف نواختی در پاره‌گفتار طراحی شده‌اند. این پاره‌گفتارهای دوکلمه‌ای براساس محل هجای تکیه‌بر در کلمهٔ دوم به سه گروه تقسیم شده‌اند. مدنظر بود که همخوان‌ها در دو هجای آخر کلمهٔ اول و در تمام همخوان‌های کلمهٔ دوم واکدار باشند. در گروه اول تکیه روی هجای اول کلمهٔ دوم قرار دارد، در این حالت کلمهٔ دوم یا به‌صورت یک کلمهٔ یک‌هجایی و یا یک ساختار پی‌بستیِ دوهجایی است. در گروه دوم، تکیه روی هجای دوم کلمهٔ دوم قرار دارد و در گروه سوم تکیه روی هجای سوم کلمهٔ دوم قرار دارد. تمامی سه گروه داده در پایان پژوهش حاضر پیوست شده‌اند. یک نمونه از هر گروه داده در ادامه آورده شده است.

(۱) آبِ رود گل‌آلود بود.

(1) [ʔɑ.be **rud**] ɟelɑlud bud
کلمهٔ دوم کلمهٔ اول

(۲) زیرِ نورِ چراغ بهتر معلوم بود.

(2) [zi.re **nu**.re] tʃeraɢ behtar maʔlum bud
کلمهٔ دوم کلمهٔ اول

(۳) چند ماه پیش در مضیقهٔ مالی بودند.

(3) tʃand mɑh piʃ dar [mazi**ɢ**e.je mɑ.**li**] budand
کلمهٔ دوم کلمهٔ اول

(۴) از محلهٔ قدیمی رفته بودند.

(4) ʔaz [mahal.le.je ɢa.di.**mi**] rafte budand
کلمهٔ دوم کلمهٔ اول

https://doi.org/10.22108/jrl.2023.136677.1714

تعداد ۳۰ نفر شرکت‌کننده (۱۵ زن و ۱۵ مرد) در این پژوهش از بین رده سنی ۳۰-۷۰ سال انتخاب شدند. تمام این افراد گویشور بومیِ زبان فارسی معیار و دارای تحصیلات دانشگاهی بودند. عباراتی که در مرحله اول طراحی شده بود داخل جملات خبری مناسب قرار داده شدند تا گویشور به‌طور طبیعی و بدون تأکید بر روی کلمات خاصی آن‌ها را بخواند. جملات سه گروه داده به‌طور تصادفی و نامنظم و بدون اطلاع شرکت‌کنندگان از نوع و یا تفاوت آن‌ها بر روی کاغذ به آن‌ها ارائه شد. به شرکت‌کنندگان توضیح داده شد که این جملات صرفاً خبری هستند و به‌صورت پرسشی یا تعجبی خوانده نمی‌شوند. همچنین، به آنان چند دقیقه فرصت داده شد تا جملات را مطالعه کنند که در صورت نیاز دربارهٔ کلماتی که شاید برایشان نامفهوم باشد، سؤال بپرسند. به گویشوران توضیحات کامل دربارهٔ نحوهٔ ضبط داده‌ها ارائه شد. علامت مکث کردن، علامت خواندنِ دادهٔ بعدی و علامت تکرارِ مجدد داده در صورت خطای کلامی و یا نوفه ناگهانی از فضای بیرون برای آن‌ها شرح داده شد. داده‌ها برای هر گویشور در سه قسمت ضبط شدند. هر ۱۵ داده در یک پروندهٔ صوتی با فاصلهٔ زمانیِ مکث ۵ ثانیه‌ای بین هر داده ضبط شد که با حرکت دست به گویشور برای مکث کردن و یا خواندنِ داده بعدی علامت داده می‌شد. هر جا گویشور دچار اشتباه کلامی و یا مکث طولانی در میان عبارتِ مدنظر می‌شد، مجدداً پس از ۵ ثانیه مکث، آن داده را تکرار می‌کرد. در پایان، پژوهشگر یک‌بار پرونده‌های صوتی را برای هر گونه خطای احتمالی بررسی کرد تا در صورت نیاز به‌صورت جداگانه داده‌ها مجدداً ضبط شوند.

۲-۴. روش اندازه‌گیری و تحلیل داده‌ها

برای تحلیل داده‌ها از نرم‌افزار پرات نسخهٔ ۶/۱/۱۶ (Boersma & Weenink, 2020) استفاده شد. اندازه‌گیری فواصل مدنظر به‌صورت دستی بر روی منحنی فرکانس پایه به کمک شکلِ موج[1] و طیف‌نگاشت[2] برای یافتن دقیق‌تر اهداف نواختی انجام شد. مرزهای فواصل در چهار لایهٔ فاصله‌ای[3] زیر تعیین شدند:

۱) A: فاصلهٔ نواخت H در کلمهٔ اول تا نواخت L در کلمهٔ دوم

۲) B: فاصلهٔ نواخت H در کلمهٔ اول تا مرکز هجای تکیه‌بر کلمهٔ دوم

۳) C: فاصلهٔ نواخت H در کلمهٔ اول تا اول کلمهٔ دوم

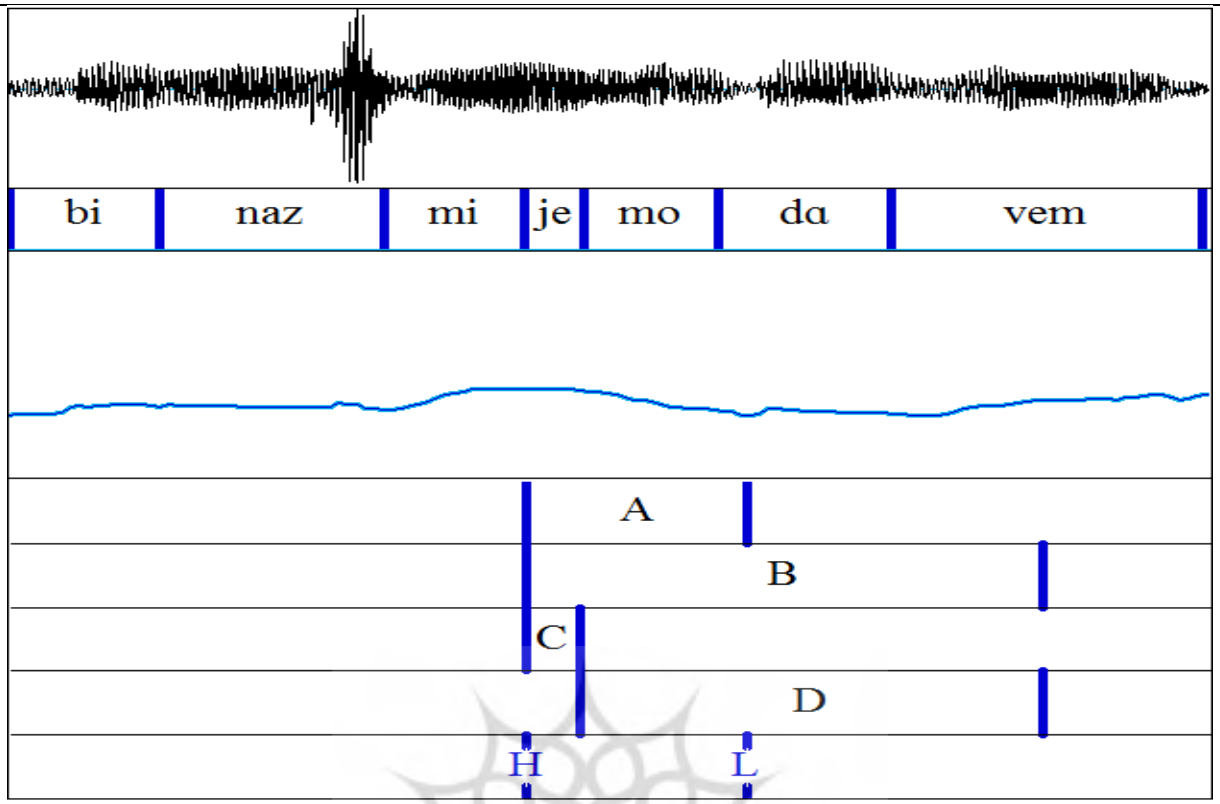
۴) D: فاصلهٔ اول کلمهٔ دوم تا مرکز هجای تکیه‌بر کلمهٔ دوم

یک لایهٔ نقطه‌ای[4] نیز در لایهٔ پنجمِ تحلیل تعریف شد که نشان‌گر مرزهای لایهٔ اول یعنی H و L است.

پس از برچسب‌دهی و تعیین همهٔ فواصل مدنظر در برنامهٔ پرات، به کمک برنامهٔ ProsodyPro نسخه ۵/۷/۸/۶، (Xu, 2013) فواصلِ تعیین‌شده در هر لایه اندازه‌گیری شدند و مقادیر آن‌ها در سه پروندهٔ اکسل طبق گروه داده‌های مدنظر طبقه‌بندی شدند. در شکل (۴) یک نمونه از برچسب‌دهی لایه‌ای داده‌ها مشاهده می‌شود.

[1] waveform
[2] spectrogram
[3] interval tier
[4] point tier

https://doi.org/10.22108/jrl.2023.136677.1714



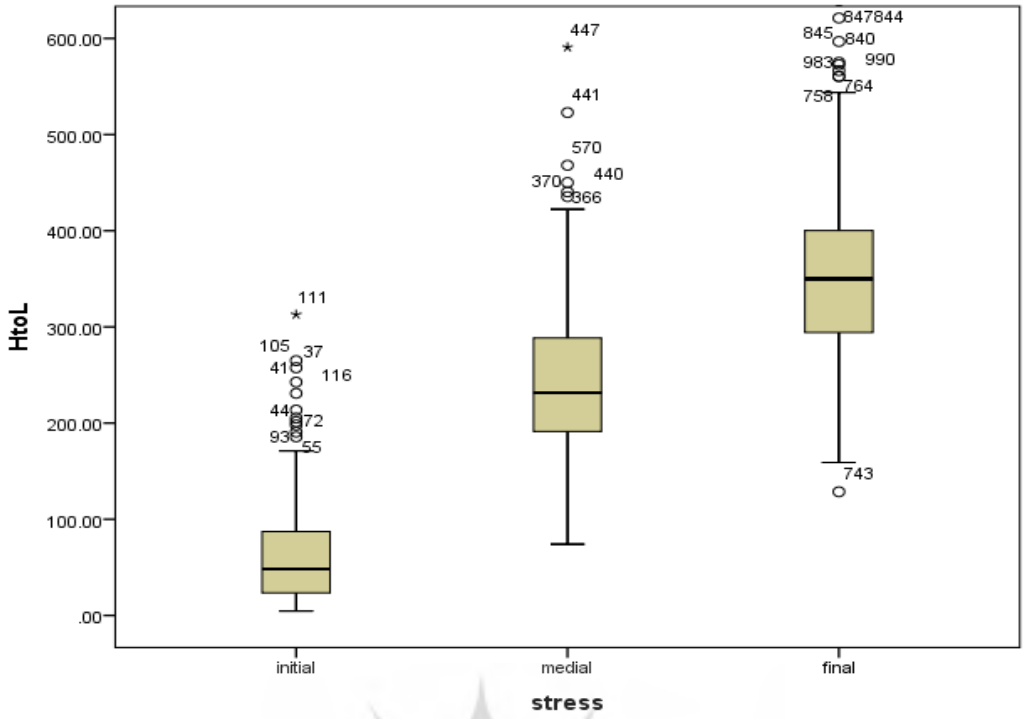
شکل ۴ – برچسب‌دهی لایه‌ای پاره‌گفتار «بی‌نظمیِ مداوم» از گروه سوم داده‌ها

Figure 4 – Annotaion of the utterance “binazmi.je mo.dɑ.vem” from the third group of data

۵. نتایج

شکل (۵) میانگین فاصلهٔ زمانی بین H در کلمهٔ اول و L در کلمهٔ دوم را به تفکیک هر گروه نشان می‌دهد. همان طور که مشاهده می‌شود، میانگین میزان این فاصله در داده‌های گروهِ اول در بازهٔ حدودی صفر تا یک‌دهم ثانیه قرار دارد و در گروه دوم و سوم به ترتیب بین بازهٔ حدودی دودهم ثانیه تا سه‌دهم ثانیه و سه‌دهم ثانیه تا چهاردهم ثانیه قرار دارد. این تفاوت بینِ میانگین‌ها، نشان‌دهندهٔ تغییراتِ این فاصله به نسبت محل وقوعِ تکیهٔ زیروبمی است. این نسبت در تغییرات، به افت تدریجی فرکانس پایه تا ابتدای هجای تکیه‌بر اشاره دارد؛ زیرا در غیر این صورت، L در هر سه گروه تا ابتدای کلمهٔ دوم گسترده شده است و این فاصلهٔ زمانی با تفاوت معنادار میان گروه‌های مختلف همراه نمی‌شد.

https://doi.org/10.22108/jrl.2023.136677.1714



شکل ۵- میانگین فاصلهٔ زمانی بین H در کلمهٔ اول و L در کلمهٔ دوم در فواصل هجایی مختلف در سطح تمامیِ داده‌ها

Figure 5 – The mean distance in time between H (the first peak) and L (the valley) across all data and speakers

همان طور که در شکل بالا مشاهده می‌شود، تعدادی از داده‌ها در محدودهٔ میانگین گروه خود نیستند. علت این امر آن است که علی‌رغم اینکه تلاش شده بود داده‌هایی انتخاب شود که از زنجیرهٔ آواهای واکدار تشکیل شده باشند، با این حال در برخی داده‌ها منحنی فرکانس پایه در محل منطبق با همخوان‌های گرفتهٔ واکدار دچار پارگی شده بود، طوری که کار اندازه گیری فواصل زمانی را با مشکل مواجه می‌ساخت. بنابراین، برای این داده‌ها گاهاً اندازه‌گیری‌ها با در نظر گرفتن تقریبیِ محل وقوع نواخت‌ها انجام شده است.

برای بررسی سطح معناداری تفاوت‌ها، آزمون تحلیل واریانس (آنوا[1]) بر روی میزان فاصلهٔ H در کلمهٔ اول تا L در کلمهٔ دوم در سه گروه داده‌های پژوهش انجام شد. در این آزمون، محل وقوع تکیهٔ زیروبمی (عامل تفکیک گروه‌ها از یکدیگر) به‌عنوان عامل مستقل و HtoL به‌عنوان متغیر وابسته انتخاب شدند. نتایج آزمون تحلیل واریانس (جدول ۳) نشان داد فاصلهٔ زمانی H در کلمهٔ اول تا L در کلمهٔ دوم در گروه‌های مختلف با یکدیگر تفاوت معنادار دارد.

جدول۲ - نتایج آزمون تحلیل واریانس در محاسبهٔ سطحِ معناداری فاصلهٔ زمانی H در کلمهٔ اول تا L در کلمهٔ دوم در تمامی داده‌های آزمایش

Table 2 – Results of ANOVA analysis for calculating the significance level of the distance in time between H (the first peak) and L (the valley) across all data and speakers

	df	F	مقدار معناداری
محل وقوع تکیهٔ زیروبمی	2	475.163	.000

پس از محاسبهٔ کلیِ مقدارِ معناداری، آزمون‌های تعقیبی نیز بر روی داده‌ها انجام شد تا با مقایسه‌های دوبه‌دو، مقدارِ معناداری میان گروه‌ها نیز به‌دست آید.

[1] ANOVA

https://doi.org/10.22108/jrl.2023.136677.1714

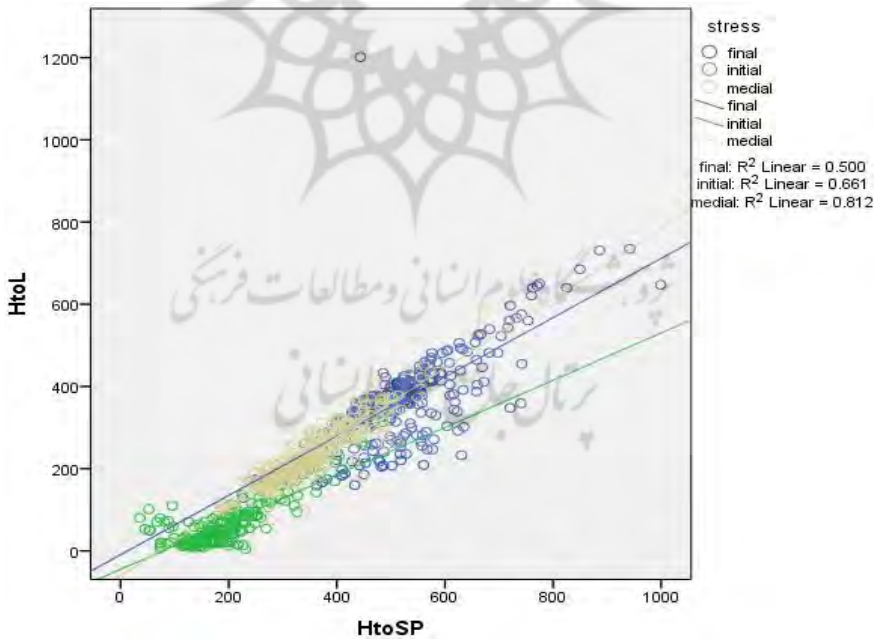
جدول ۳ – نتایج آزمون تعقیبی بر روی داده‌های پژوهش به صورت دوبه‌دو در میان هر سه گروه

Table 3 – Results of post-hoc pairwise tests between each group of data

	گروه اول		گروه دوم		گروه سوم	
	گروه دوم	گروه سوم	گروه اول	گروه سوم	گروه اول	گروه دوم
مقدار معناداری	.000	.000	.000	.000	.000	.000

طبق نتایج آزمون‌های تعقیبی که در جدول (۳) ارائه شده است، مقدارِ معناداری بین همهٔ گروه‌ها به‌صورت دوبه‌دو معادل «۰/۰۰» است که به معنی معنادار بودن اختلاف این فاصلهٔ زمانی بین تمام گروه‌های داده‌های آزمایش است.

سپس در گامی دیگر، ضریبِ همبستگی[1] بین فاصلهٔ زمانیِ HtoL و فاصلهٔ زمانیِ HtoSP در هر سه گروه اندازه‌گیری شد. به‌طور کلی نتایج آزمون همبستگی پیرسون نشان داد ارتباطی مستقیم و قوی بین فاصلهٔ زمانیِ H در کلمهٔ اول تا L در کلمهٔ دوم و فاصلهٔ زمانیِ H در کلمهٔ اول تا مرکز هجای تکیه‌بر در کلمهٔ دوم در هر سه گروه از داده‌ها وجود دارد (شکل ۶). به بیان دیگر، تغییراتِ میزان فاصلهٔ تکیهٔ زیروبمی در کلمهٔ دوم از H در کلمهٔ اول، مشابهِ تغییراتِ میزان فاصلهٔ L در کلمهٔ دوم از H در کلمهٔ اول است. طبق مشاهدات و اندازه‌گیری‌های انجام‌شده، محل وقوع L در اکثر مواقع قبل از شروع هجای تکیه‌بر قرار دارد. بر این اساس، با دور شدن هجای تکیه‌بر از H در کلمهٔ اول، L نیز از آن دور می‌شود. این یافته، فرضیهٔ اول پژوهش را که مبتنی بر درون‌یابی آوایی بین این دو هدفِ نواختی است، تأیید می‌کند؛ زیرا طبق این فرضیه در صورت گسترش L تا ابتدای کلمهٔ دوم، محل وقوع آن به صورت یک فلات[2] پایینِ گسترده صورت می‌گیرد و فاصلهٔ H تا L به‌طور یکسان برابر با فاصلهٔ H تا ابتدای کلمهٔ دوم است.



شکل ۶ – نمودار رگرسیون فاصلهٔ H در کلمهٔ اول تا L در کلمهٔ دوم (HtoL) به‌صورت تابعی از فاصلهٔ H در کلمهٔ اول تا مرکز هجای تکیه‌بر در کلمهٔ دوم (HtoSP) در تمامی داده‌های آزمایش به تفکیک گروه

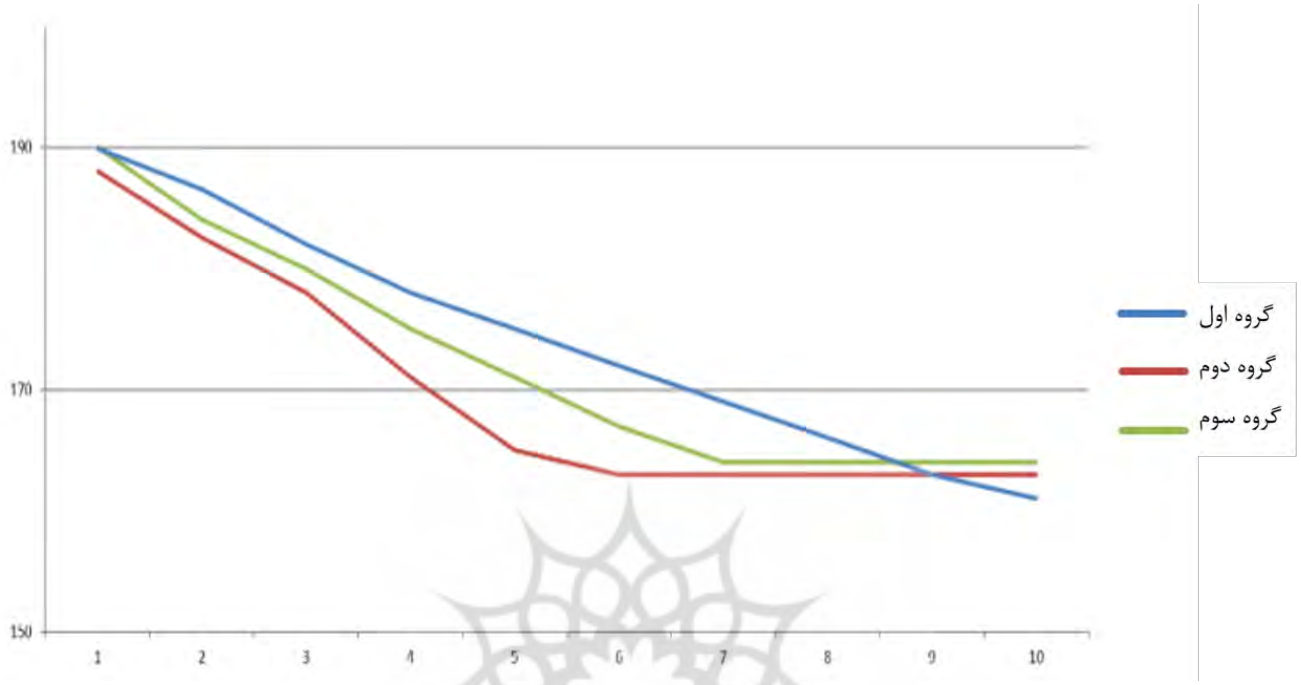
Figure 6 – The regression of the distance in time between H (the first peak) and L (the valley) as a function of the distance in time between H (the first peak) and the middle of the accented syllable in each group of data

[1] مقادیر ضریب همبستگی پیرسون بین مقادیر ۰/۵ تا ۱ (رابطهٔ مستقیم) و یا ۰/۵- تا ۱- (رابطهٔ معکوس) رابطه‌ای متوسط تا قوی را بین این پارامترها نشان می‌دهد.

[2] plateau

https://doi.org/10.22108/jrl.2023.136677.1714

سپس، در گام پایانی برای بررسی الگوی کلی تغییرات زیروبمی در جملات آزمایش، منحنی میانگینِ فرکانس پایه در گروه‌های جداگانه به کمک نرم‌افزار ProsodyPro محاسبه شد.[۱] این نرم‌افزار منحنی میانگینِ فرکانس پایه را با استخراج مقادیر F0 از ده نقطه با فواصل منظم در هر پاره‌گفتار محاسبه می‌کند.



شکل ۷ – میانگین منحنیِ فرکانس پایه به تفکیک گروه در سطح تمامیِ داده‌ها

Figure 7 – The normalized pitch contour in each group of data

همان طور که در شکل (۷) مشاهده می‌شود، منحنیِ فرکانس پایه در حدِ فاصل H در کلمهٔ اول تا L در کلمهٔ دوم افتی تدریجی دارد تا به محل نواخت L برسد. در داده‌های گروه اول که در آن تکیه بر روی هجای اول کلمات قرار دارد، F0 افتی نسبتاً تند دارد؛ یعنی با شیبی نسبتاً تند از قلهٔ H تکیهٔ زیروبمی اول به نواخت L کلمه دوم درون‌یابی شده است. در گروه دوم، که در آن تکیه بر روی هجای قبل پایانی قرار دارد، افت منحنی فرکانس پایه شیب کندتری دارد و بالأخره اینکه در گروه سوم که در آن تکیه بر روی هجای پایانی کلمات قرار دارد، فرکانس پایه با شیبی بسیار ملایم از قله H کلمه اول به نواخت L کلمه دوم درون‌یابی شده است.

۶- بحث و نتیجه‌گیری

نتایج به‌دست‌آمده به‌طور کلی به تأیید فرضیهٔ اول پژوهش اشاره دارد. در این فرضیه، افت فرکانس پایه، تدریجی پیش‌بینی شده است که براساس قاعدهٔ درون‌یابی آوایی رخ می‌دهد. نتایج به‌دست‌آمده نشان داد فاصلهٔ زمانی بین قلهٔ H تکیه زیروبمی اول تا نواخت L کلمه دوم (HtoL) در گروه اول کمتر از گروه دوم و در گروه دوم کمتر از گروه سوم است. به بیان دیگر، هرقدر تکیه نسبت به آغاز کلمه (کلمه دوم) دورتر باشد، به همان اندازه فاصله زمانی HtoL بیشتر می‌شود. نتایج آزمون تحلیل واریانس نشان داد فاصلهٔ زمانی HtoL برای هر سه گروه با یکدیگر اختلاف معنادار دارد. در محاسبهٔ ضریب همبستگی بین فاصلهٔ قلهٔ اول تا L بین دو قلهٔ زیروبمی (HtoL) و فاصلهٔ قلهٔ اول تا مرکز هجای تکیه‌بر کلمهٔ دوم (HtoSP) مشاهده شد که بین مقادیر این دو فاصله ارتباط مستقیم وجود دارد. یعنی با افزایش فاصلهٔ HtoL به فاصلهٔ HtoSP نیز بیشتر می‌شود.

[۱] تعدادی از فواصل زمانی در داده‌های هر گروه برای ProsodyPro قابل اندازه‌گیری نبود. این داده‌ها از این مرحلهٔ تحلیل حذف شده‌اند. علت آن عدم امکان محاسبهٔ دستیِ نقاطِ موردنظر روی منحنی فرکانس پایه در حدِ فاصل H در کلمهٔ اول تا L در کلمهٔ دوم است.

https://doi.org/10.22108/jrl.2023.136677.1714

همچنین، در بررسی و مقایسه منحنی میانگین فرکانس پایه در سه گروه هدف آزمایش مشاهده کردیم که شیب افت فرکانس پایه از قله H تکیه زیروبمی اول به نواخت L کلمه دوم در گروه‌های مختلف با یکدیگر تفاوت درخورتوجهی دارد به این صورت که هر قدر محل تکیه (یا محل وقوع تکیه زیروبمی) از آغاز کلمه دورتر می‌شود، شیب افت فرکانس پایه کندتر می‌شود.

در مجموع شواهد به‌دست‌آمده از تحلیل داده‌ها، فرضیهٔ اول پژوهش، یعنی فرضیه درون‌یابی آوایی، را تأیید می‌کند. گفتیم که بر مبنای قاعدهٔ درون‌یابی آوایی، قلهٔ H اول به نواخت L تکیهٔ زیروبمی دوم درون‌یابی می‌شود. بر مبنای این فرضیه، نواخت L تکیهٔ زیروبمی دوم که بر روی هجای تکیه‌بر این کلمه واقع می‌شود، محل فرود منحنی نزولی فرکانس پایه از قلهٔ H تکیه زیروبمی اول است. یعنی دو نواخت H از کلمه اول و L از کلمه دوم با استفاده از یک خط گذار نزولی به یکدیگر درون‌یابی می‌شوند. به این ترتیب، شیب تغییرات F0 بین دو نواخت با توجه به فاصله بین قله H تکیه زیروبمی اول و نواخت L کلمه دوم متفاوت است. تمامی آزمون‌های این پژوهش، با تأیید فرضیه درون‌یابی آوایی، نشان می‌دهد نواختِ L با ابتدای هجای تکیه‌بر ترادف دارد و در نتیجه گذار نزولی F0 بین قله زیروبمی کلمه اول و دره زیروبمی کلمه دوم شیبی دارد که مقدار تندی آن بسته به محل وقوع تکیه (و یا محل وقوع نواخت L) در سطح کلمه متفاوت است.

در مقابل، یافته‌های این پژوهش فرضیهٔ دوم یعنی قاعدهٔ گسترش نواخت را تأیید نمی‌کند. گفتیم که بر مبنای این فرضیه انتظار داریم فاصلهٔ HtoL با افزایش تعداد هجاها در گروه‌های تکیه‌ای مختلف بدون تغییر باقی بماند و بین دو متغیر زمانی فاصلهٔ HtoL و فاصلهٔ HtoW (فاصلهٔ قلهٔ اول تا ابتدای کلمهٔ دوم) همبستگی بالایی وجود داشته باشد؛ زیرا این فرضیه ادعا می‌کند که نواخت L همواره در ابتدای کلمهٔ دوم واقع می‌شود. این در حالی است که نتایج آزمون‌های آماری نشان داد فاصله زمانی HtoL در گروه‌های تکیه‌ای مختلف بسته به محل وقوع تکیه در کلمه دوم به‌طور معناداری با یکدیگر متفاوت است. همچنین، تحلیل‌های رگرسیون نشان داد رابطهٔ دو پارامتر زمانی HtoL با HtoW در گروه‌های تکیه‌ای مختلف همبستگی ضعیفی دارد.

به‌طور کلی، نتایج این پژوهش دربارهٔ رفتار فرکانسی فاصلهٔ زمانی بین قله H تکیه زیروبمی اول تا نواخت L کلمه دوم (HtoL) با نظریهٔ خودواحد عروضی آهنگ مطابقت دارد. این نظریه منحنی آهنگ را به‌صورت توالی از اهداف نواختی L و H (یعنی نقاط گسستهٔ بالا و پایین) توصیف می‌کند که سطح ارتفاع و انطباق آن‌ها با ساخت زنجیری کاملاً پیش‌بینی‌پذیر است. ایدهٔ اصلی در مدل خودواحد عروضی این است که تمامی هجاها در سطح آهنگ پاره‌گفتار از نظر ویژگی زیروبمی مشخصه‌دهی نمی‌شوند. اهداف نواختی در منحنی آهنگ یک پاره‌گفتار با فاصله بر روی نقاط زبان‌شناختی مشخصی در ساخت زنجیری تظاهر می‌یابند و فاصله بین هر دو هدف با استفاده از یک خط پیوندی (که لزوماً صاف یا مستقیم نیست) به یکدیگر متصل می‌شود. بر این اساس، تعداد اهداف نواختی با تعداد هجاها برابر نیست، بلکه به طرز معناداری از آن کمتر است. یعنی اغلب هجاها در سطح واجی مشخص‌دهی نمی‌شوند، بلکه ویژگی زیروبمی خود را از مختصات خطوط پیوندی بین اهداف به دست می‌آورند.

صادقی (۱۳۹۷) طی پژوهشی تطابق این نظریه با زبان فارسی را بررسی کرده است که نتایج آن نشان داد حضور درهٔ فرکانس پایه بین دو تکیهٔ زیروبمی ناشی از یک نواخت مشخص L است. بدین معنی که ترادف و سطح ارتفاع دره صرف‌نظر از فاصلهٔ زمانی بین قله‌ها ثابت و پایدار است. همچنین، درهٔ فرکانس پایه همواره به شکل منظم و پایدار با آغاز هجای تکیه‌بر ترادف دارد. در پژوهش حاضر نیز نتایج مبنی بر همین امر است و این نشان‌گر قابلیت مدل‌سازی منحنی فرکانس پایه در آهنگ گفتار فارسی به‌صورت تعدادی هدف نواختی طبق این نظریه است.

از یافته‌های این پژوهش می‌توان در زمینهٔ بازسازی مصنوعی گفتار در بخش تحلیل متنی و آوایی سیستم‌های تبدیل متن به گفتار نیز استفاده کرد. پیش‌بینیِ دقیق‌تر شکلِ محلی جابه‌جایی‌های منحنی فرکانس پایه در ساخت نوایی نمادین با استفاده از نتایج به‌دست‌آمده منجر به بازسازی گفتار به طور طبیعی‌تر می‌شود که در نهایت، بهبود عملکرد سیستم‌های تبدیل متن به گفتار را شامل می‌شود.

https://doi.org/10.22108/jrl.2023.136677.1714

منابع فارسی

اسلامی، محرم. (۱۳۸۴). *واج شناسی: تحلیل نظام آهنگ زبان فارسی*. تهران: سمت.

صادقی، وحید. (۱۳۹۷). *ساخت نوایی زبان فارسی: تکیه واژگانی و آهنگ*. تهران: سمت.

همایون‌پور، محمدمهدی. (۱۳۹۰). *پژوهشنامهٔ تبدیل متن به گفتار*. تهران: دبیرخانه شورای عالی اطلاع‌رسانی.

References

Arvaniti, A., & Ladd, D. R. (1995). Tonal alignment and the representation of accentual targets. *Proceedings of the 13th International Congress of Phonetic Science 4*, 220-223.

Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook* 3, 255-310.

Boersma, P. & Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program]. Version 6.1.16, retrieved December 2020 from http://www.praat.org/

Eslami, M. (2005). *Phonology: Analyzing the intonation system of Persian*. Tehran: SAMT. [In Persian]

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*, Dordrecht: Foris.

Homayounpour, M. (2011). *Text to speech system*. Tehran: Supreme Council of Information and Communications Technology. [In Persian]

IBM SPSS Statistics for Windows, version 25, IBM Corp., Armonk, N.Y., USA.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics 31*, 81-112.

Ladd, D. R. (1983). Phonological features of intonational peaks. *Language 59*, 721-759.

Mahjani, B. (2003). *An instrumental study of prosodic features and intonations in Modern Farsi (Persian)*. MS thesis, University of Edinburgh.

Pierrehumbert, J. B. (1980). *The Phonology and phonetics of English intonation*. Ph.D. Dissertation, Massachusetts Institute of Technology.

Sadat-Tehrani, N. (2009). The alignment of L+H* pitch accents in Persian intonation. *Journal of the International Phonetic Association 39*(2), 205-230.

Sadat-Tehrani, N. (2007). *The Intonational grammar of Persian*. Ph.D. Dissertation, the university of Manitoba.

Sadeghi, V. (2018). *The Prosodic structure of Persian language: Lexical stress and intonation*. Tehran: SAMT. [In Persian]

Xu, Y. (2013). ProsodyPro — A Tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France. 7-10.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

https://doi.org/10.22108/jrl.2023.136677.1714

پیوست‌ها:

داده‌های گروه اول پژوهش

seda.je **zan**.ɟe dar ʔumad (۱) صدای زنگِ در اومد.

ru.je **dʒi**.be lebɑs dokme duxt (۲) روی جیبِ لباس دکمه دوخت.

manba.ʔe **ʔɑb** tʃek.ke mikone (۳) منبعِ آب چکه می کنه.

ru.je **ɢɑ**.be nɑɢ.ɢɑʃi tʃɑsb rixt (۴) روی قابِ نقاشی چسب ریخت.

ʔaz bu.je **ɟol** badam ʔumad (۵) از بوی گل بدم اومد.

ʃed.date dʒar.jɑ.ne **barɢ** ɢavi bud (۶) شدتِ جریانِ برق قوی بود.

dandɑ.ne **lɑɢ**.ɢe ʔaɢabi rɑ kandam (۷) دندانِ لقِ عقبی را کندم.

natidʒe.je **dʒanɟ** maʔlum ʃod (۸) نتیجهٔ جنگ معلوم شد.

nomraʃ ʔaz dafʔe.je **ɢabl** behtar ʃode bud (۹) نمره ش از دفعهٔ قبل بهتر شده بود.

zi.re **nu**.re tʃeraɢ behtar maʔlum bud (۱۰) زیرِ نورِ چراغ بهتر معلوم بود.

bar.ɟe **zar**.de deraxtan ruje zamin rixte (۱۱) برگِ زردِ درختان روی زمین ریخته.

ʔɑ.be **rud** ɟelalud bud (۱۲) آبِ رود گل آلود بود.

divɑ.re **bɑ**.ɢe kenari taxrib ʃod (۱۳) دیوارِ باغِ کناری تخریب شد.

tar.rahije ru.je **dʒel**.de ketab zibɑ bud (۱۴) طراحیِ روی جلدِ کتاب زیبا بود.

zuze.je **ɟorɟ** az xɑb bidareʃan kard (۱۵) زوزهٔ گرگ از خواب بیدارشان کرد.

داده‌های گروه دوم پژوهش

ʔaz ru.je ʔol.**ɟu**.je xaj.jati borid (۱) از روی الگوی خیاطی بُرید.

je ɟu.ni.je be.rendʒ ʔɑvord (۲) یه گونیِ برنج آورد.

bɑ tamɑ.me vo.**dʒud** xoʃhal ʃod (۳) با تمامِ وجود خوشحال شد.

havɑ.je ʔɑ.**li**.je sobh sarehalam ʔɑvord (۴) هوای عالیِ صبح سرحالم آورد.

mavad.de lɑ.**ze**.me poxte ɢazɑ ro xarid (۵) موادِ لازمِ پختِ غذا رو خرید.

je ɢuti ran.ɟe ʔɑ.**bi**.je ʔasemani ɟereftam (۶) یه قوطی رنگِ آبیِ آسمانی گرفتم.

ʔemzɑ.je mo.**di**.re porodʒe lazeme (۷) امضای مدیرِ پروژه لازمه.

natidʒe.je ʔa.**mal** movaf.faɢijatamiz bud (۸) نتیجه ی عمل موفقیت آمیز بود.

ʔɑr.de ɟan.**dom** tuje jaxtʃal hast (۹) آردِ گندم توی یخچال هست.

dar dʒarijɑ.ne bɑ.**zi**.je final masdum ʃod (۱۰) در جریانِ بازیِ فینال مصدوم شد.

daʃt por az ɟo.le lɑ.**le** bud (۱۱) دشت پر از گلِ لاله بود.

tʃand mah piʃ dar mazi**ɢ**e.je mɑ.**li** budand (۱۲) چند ماه پیش در مضیقه ی مالی بودند.

taʔmi.re dʒɑ.**ru**.je ɢadimi az ʔohdaʃ xaredʒ bud (۱۳) تعمیرِ جاروی قدیمی ازعهده ش خارج بود.

bɑdbɑ.ne ɢɑ.**jeɢ** ʃekast (۱۴) بادبانِ قایق شکست.

danehɑje ɢerme.ze ʔa.**nɑr** ruje farʃ rixte (۱۵) دانه های قرمزِ انار روی فرش ریخته.

داده‌های گروه سوم پژوهش

ʔaz tari**ɢ**e marɑ.**dʒe**.ʔe ɢanuni ʔeɢdam kardim (۱) از طریقِ مراجعِ قانونی اقدام کردیم.

ban.de ʔɑ.ɢɑ.**zi**.ne ɢarardad moʃkel daʃt (۲) بندِ آغازینِ قرارداد مشکل داشت.

da.re ʔa.ɢa.**bi**.je mɑʃin bɑz bud (۳) درِ عقبیِ ماشین باز بود.

nazarije.je ri.jɑ.**zi**.je mohem ʔesbat ʃod (۴) نظریه ی ریاضیِ مهم اثبات شد.

tavanɑji.je za.bɑ.**ni**.je zijadi dare (۵) توانایی زبانیِ زیادی داره.

sedɑ.je ba.rɑ.**da**.re ʔali ʔɑmad (۶) صدای برادرِ علی آمد.

niru.je dʒɑj.ɟo.**zi**.ne xub dɑrand (۷) نیروی جایگزینِ خوب دارند.

https://doi.org/10.22108/jrl.2023.136677.1714

nijaz be ʔeʔla.me mo.dʒad.**dad** darim	(۸) نیاز به اعلامِ مجدد داریم.
toli.de ʔe.ner.**ʒi**.je xorʃidi ʔaz tavaneʃan xaredʒ bud	(۹) تولیدِ انرژیِ خورشیدی از توانشان خارج بود.
ʔaz binazmi.je mo.dɑ.**vem** ʃaki bud	(۱۰) از بی نظمیِ مداوم شاکی بود.
ruje naza.re mo.ʔɑ.**ve**.ne modir taʔsir dare	(۱۱) روی نظرِ معاونِ مدیر تاثیر داره.
dʒelo.je ma.ɢɑ.**ze** ʔistadim	(۱۲) جلوی مغازه ایستادیم.
ʔaz mahal.le.je ɢa.di.**mi** rafte budand	(۱۳) از محله ی قدیمی رفته بودند.
barande.je vɑ.ɢe.**ʔi** kase diɟari bud	(۱۴) برنده ی واقعی کسِ دیگری بود.
dar dore.je ma.dʒɑ.**zi** ʃerkat kardam	(۱۵) در دوره ی مجازی شرکت کردم.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

https://doi.org/10.22108/jrl.2023.136677.1714