http://ui.ac.ir/en

Journal of Researches in Linguistics
E-ISSN: 2322-3413
Document Type: Research Paper
12(2), 241-262
Received: 16.05.2021 Accepted: 22.09.2021

Speaker-specific features of simple vowels in Persian based on the source-filter theory

Homa Asadi *
Assistant Professor, Department of Linguistics, Faculty of Foreign Languages, University of Isfahan, Isfahan, Iran
h.asadi@fgn.ui.ac.ir
Batool Alinezhad
Associate Professor, Department of Linguistics, Faculty of Foreign Languages, University of Isfahan, Isfahan, Iran
b.alinezhad@fgn.ui.ac.ir

Abstract
Based on source-filter theory, the present research attempts to investigate between- and within-speaker variability in simple vowels of Persian using experimental phonetics tools. This research aims to discover which of the simple vowels of Persian represent more speaker-specific information and which acoustic parameters can better distinguish Persian speakers. To test between- and within-speaker variability, two types of acoustic parameters, one related to the larynx, i.e. fundamental frequency, and the other related to the vocal tract, i.e. formant frequencies, were selected. Fundamental frequency as well as formant values were extracted from the steady state point of the vowels uttered by twelve Persian-speaking male speakers. Speech data were recorded non-contemporaneously in laboratory environment on two different occasions separated by one to two weeks, thereby allowing for analyzing occasion-to occasion within-speaker variability. Speech tokens were acoustically measured with PRAAT version 5.2.34 and statistical analyses were carried out with SPSS version 21 and R version 3.3.3. Results of the study indicated that the low front vowel /a/ and the third formant frequency convey more speaker-specific information compared to the other vowels and formant frequencies. In addition, discriminatory power of fundamental frequency was reported to be stronger than formant frequencies. The results also revealed that fundamental frequency is correlated with the first formant frequency which is subsequently indicative of interdependence between the source and filter sections.
Keywords: Acoustic phonetics, Forensic phonetics, Formant frequency, Fundamental frequency, Speaker identification, Source-filter theory

Introduction

Verbal communication is an integral part of human social interactions. Everyday experience tells us that humans are able to recognize easily familiar speakers through their voice. This indicates that speech sounds contain specific information which could be reflected in the acoustic characteristics of speech signals. Vowels are among those speech sounds which have always been the center of attention in the field of forensic voice comparison. In a study by Gold and French (2011), vowels have been reported as one of the most analyzed segments among forensic practitioners. They also reported that F0 and formant structures are two acoustic parameters which are commonly used in forensic voice comparison. Earlier studies on particularities of vowels were primarily focused on calculation of the average values of formant and fundamental frequency over a long stretches of a speech recordings. However, long term extraction of F0 and formant values represent solely the discriminatory power of formant structures without allowing us to measure the strengths of vowels in separation. In this study we aim to examine the discriminatory role of

Copyright©2021, University of Isfahan. This is an Open Access article distributed under the terms of the Creative Commons Attribution License https://creativecommons.org/licenses/by-nc-nd/4.0/, which permits others to download this work and share it with others as long as they credit it, but they can't change it in any way or use it commercially.
http://dx.doi.org/10.22108/jrl.2021.128697.1577

simple vowels in Persian with focus on the extent of source-filter independence or interdependence within the context of speaker identification. This study aims to determine which of the vowels in Persian can better distinguish speakers and which acoustic parameters of the source and filter sections represent more speaker-specificity. We also aim to examine whether source and filter features have potential in capturing complementary information about speakers that can be used to improve speaker discrimination.

Materials & Methods

To test between-and within-speaker variability, twelve male Persian speakers were recorded on two different sessions, separated by a time-lapse of one to two weeks. Speakers were asked to read the 54 sentences one by one, with a pause, and in a natural way, without any marked intonation. Speech tokens were analyzed using Praat (version 5.2.34, Boersma and Weenink, 2013). For this study, mean values of the fundamental frequency (F0) and the first four formants i.e. F1, F2, F3 and F4 were measured at the central points of six simple vowels in Persian. Statistical analysis of data was carried out using R (R core Team 2014) version 3.3.3 and SPSS (IBM Corp. 2012) version 21.

Discussion of Results and Conclusions

In this section, we provide the results of different acoustical models i.e. univariate analysis of variance, multinomial logistic regression and principal component analysis that were employed on the collected speech data of Persian. In the present study, we explored potential speaker-specific acoustic parameters of simple vowels in Persian based on the source-filter theory. Statistical analysis of speech data revealed that selected acoustic parameters i.e. F0, F1, F2, F3 and F4 of the all vowels, except for the F2 of the vowel /u/, were able to discriminate between Persian speakers. The current findings showed that the low front vowel /a/ appear to convey the highest between-speaker discrimination power. In terms of formant structures, for most vowels, effects of speaker were stronger on F3 and F4 compared to F1 and F2. Additionally, fundamental frequency was reported to be more discriminatory than formant frequencies. The results also revealed a significant correlation between F0 and F1 which show a considerable interdependence between the source and filter sections.

References

Alinezahd, A & Hosseinibalam, F. (2013). *Fundamentals of Acoustic Phonetics*. University of Isfahan. [In Persian].

Alderman, T. (2005). *Forensic speaker identification: A likelihood ratio-based approach using vowel formants*. LINCOM Studies in Phonetics.

Asadi, H., Hosseini-Kivanani, N & Nourbakhsh, M. (2018 a). *Speaker-specificity in spectral moments of fricative /s/ in Persian*. TABU Dag, The Netherlands.

Asadi, H., Nourbakhsh, M, Sasani, F & Dellwo, V. (2018 b). Examining long-term formant frequency as a forensic cue for speaker identification: An experiment on Persian. In M. Nourbakhsh, H. Asadi, & M. Asiaee (Eds.), *Proceedings of the First International Conference on Laboratory Phonetics and Phonology* (pp. 21-28). Neveesh Parsi Publications.

Bijankhan, M. (2013). *Phonetic system of the Persian language*. Samt. [In Persian]

Boersma, P. & Weenink, D. (2013) *Praat: Doing phonetics by computer* (version 5.2.34). http://www.praat.org.

Fant, G. (1960). *Acoustic theory of speech production*. Mouton.

Gold, E., French, J.P. (2011). International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law*, *18*(2), 293-307.

Gold, E., French, J.P & Harrison, P. (2013). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. In *Proceedings of Meetings on Acoustics*, Montreal, Canada, (pp. 1-8).

Goldstein, U. (1976). Speaker-identifying features based on formant tracks. *The Journal of the Acoustical*

Society of America 59 (3), 176-182.

Gordon, M. & P. Ladefoged. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29 (4), 383–406.

Hughes, V., Cardoso, A., Foulkes, P., French, P., Gully, A. & Harrison, P. (2019). The contribution of source and filter to speaker characterization. *The 28th Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA).* Istanbul, Turkey.

Hughes, V., Harrison, P., Foulkes, P., French, P., Kavanagh, C., Segundo, E.S. (2017). Mapping across feature spaces in forensic voice comparison: The contribution of auditory-based voice quality to (semi-) automatic system testing. *INTERSPEECH,* (3892-3896).

Jessen, M. & Becker, T. (2010). Long-term formant distribution as a forensic phonetic feature. *Conference of the Acoustical Society of America*, Cancun, Mexico.

Kahn, J., Audibert, J.F.B., & Rossato, S. (2011). Inter and intra-speaker variability in French: An analysis of oral vowels and its implication for automatic speaker verification. *International Congress of Phonetic Sciences (ICPhS)*, 17 (pp. 1002-1005).

Kavanagh, C. M. (2012). *New consonantal acoustic parameters for forensic speaker comparison* [Ph.D. dissertation]. University of York.

Kinoshita, K. (2001). *Testing realistic forensic speaker identification in Japanese: A likelihood ratio based approach using formants* [Ph.D. dissertation]. Australian National University.

Kinoshita, Y. (2002). Use of likelihood ratio and Bayesian approach in forensic speaker identification. In *Proceedings of the 9th Australian International Conference on Speech Science and Technology*. Melbourne, Australia, (pp. 297-302).

Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustic Society of America* 87 (2), 820-857.

Ladefoged, P. (2006). *A course in phonetics*. Wadsworth Cengage Learning.

McDougall, K. (2004). Speaker-specific formant dynamics: an experiment on Australian English /aɪ/. *International Journal of Speech, Language and the Law* 11(1), 103-130.

McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: Toward a new approach using formant frequencies. *International Journal of Speech, Language and the Law* 13 (1), 89-126.

Modarresi Ghavami, G. (2011). *Phonetics: The scientific study of speech*. Samt. [In Persian].

Nolan, F & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech Language and the Law 12* (2), 143-173.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge University Press.

Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). The DyViS database: Style controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech Language and the Law 16* (1), 31-57.

Nourbakhsh, M. (2013). *Acoustic phonetics using computer*. Nashre Elm. [In Persian].

Rose, P. (2002). *Forensic speaker identification*. Taylor & Francis.

Rose, P. (2007). Forensic speaker discrimination with Australian English vowel acoustics. In *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, (pp. 1817-1820).

Rose, P. & E. Winter (2010). Traditional forensic voice comparison with female formants: Gaussian mixture model and multivariate likelihood ratio analyses. In *Proceedings of the 13th Australian International Conference on Speech, Science and Technology*, pp. 42–45.

Rose, P., Osanai, T., & Kinoshita, Y. (2003). Strength of forensic speaker identification evidence: multispeaker formant- and cepstrum-based segmental discrimination with a Bayesian likelihood ratio as threshold. *Forensic Linguistics* 10, 179-202.

Sambur, S. (1975). Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustic, Speech and Signal Processing* 23 (2), 176-182.

San Segundo, E., Tsanas, A., Gómez-Vilda, P. (2017). Euclidean distances as measures of speaker similarity

including identical twin pairs: A forensic investigation using source and filter voice characteristics. *Forensic Science International* 270, 25-38.

Strange, W. (1989). Dynamic specification of co-articulated vowels spoken in sentence context. *Journal of the Acoustical Society of America* 85 (5), 2135-2153.

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *Journal of the Acoustical Society of America 123* (5), 2733–2749.

West, P. (1999). The extent of coarticulation of English liquids: an acoustic and articulatory study. *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, US, (1901-1904).

Wolf, J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America* 51 (6B), 2044-2056.

نشریه پژوهش‌های زبان‌شناسی
سال دوازدهم ، شماره دوم، شماره ترتیبی۲۳، پاییز و زمستان ۱۳۹۹
نوع مقاله: پژوهشی
تاریخ وصول: ۱۴۰۰/۰۲/۲۶
تاریخ اصلاحات: ۱۴۰۰/۰۶/۱۳
تاریخ پذیرش: ۱۴۰۰/۰۶/۳۱
صص ۲۴۱-۲۶۲

بررسی ویژگی‌های فردویژهٔ واکه‌های سادهٔ زبان فارسی بر اساس نظریهٔ منبع-صافی

* هما اسدی
** بتول علی‌نژاد

چکیده

پژوهش حاضر با استفاده از ابزارهای آواشناسی آزمایشگاهی و براساس نظریهٔ منبع-صافی به بررسی تغییرات بین-گوینده و درون-گوینده در واکه‌های سادهٔ زبان فارسی می‌پردازد. این پژوهش در پیِ آن است تا مشخص کند کدام‌یک از واکه‌های سادهٔ زبان فارسی اطلاعات فردویژهٔ بیشتری را نشان می‌دهد و کدام پارامتر آکوستیکی بهتر می‌تواند گویندگان فارسی‌زبان را از هم متمایز سازد. به‌منظور بررسی تغییرات بین-گوینده و درون-گوینده، دو دسته پارامتر آکوستیکی یکی مربوط به حنجره یعنی فرکانس پایه و دیگری پارامترهای مربوط به دستگاه گفتار یعنی فرکانس‌های سازه انتخاب شد. فرکانس پایه و فرکانس سازه‌های اول، دوم، سوم و چهارم از منطقهٔ ثبات واکه‌های تولیدشده توسط دوازده گویشور مرد فارسی‌زبان استخراج شد. نمونه‌های آوایی با استفاده از برنامهٔ Praat ویرایش ۳۴، ۲، ۵ مورد تجزیه و تحلیل آکوستیکی قرار گرفت و تحلیل آماری داده‌ها و مقادیر به‌دست‌آمده از بررسی آکوستیکی نمونه‌های آوایی با استفاده از نرم‌افزار SPSS ویرایش ۲۱/۰ و نرم افزار R ویرایش ۳. ۳. ۳ انجام شد. نتایج نشان داد واکهٔ پیشین افتادهٔ /a/ در مقایسه با سایر واکه‌ها و فرکانس سازهٔ سوم در مقایسه با سایر سازه‌ها اطلاعات فردویژهٔ بیشتری را نشان می‌دهند. براساس نتایج، فرکانس پایه نسبت به فرکانس سازه‌ها پارامتر قدرتمندتری در نشان‌دادن تغییرات بین-گوینده است. همچنین، نتایج حاکی از این است که فرکانس پایه با فرکانس سازهٔ اول همبستگی دارد که این موضوع به نوعی نشانگر رابطهٔ متقابل میان بخش‌های منبع و صافی است.

کلیدواژه‌ها

آواشناسی قضایی، نظریهٔ صافی-منبع، تشخیص هویت گوینده، فرکانس سازه، فرکانس پایه

* استادیار گروه زبان‌شناسی، دانشکدهٔ زبان‌های خارجی دانشگاه اصفهان، اصفهان، ایران (نویسنده مسئول) h.asadi@fgn.ui.ac.ir
** دانشیار گروه زبان‌شناسی، دانشکدهٔ زبان‌های خارجی دانشگاه اصفهان، اصفهان، ایران b.alinezhad@fgn.ui.ac.ir

Copyright©2021, University of Isfahan. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/BY-NC-ND/4.0), which permits others to download this work and share it with others as long as they credit it, but they can't change it in any way or use it commercially.

http://dx.doi.org/10.22108/jrl.2021.128697.1577

۱. مقدمه

تجربهٔ روزمرهٔ مکالمات افراد نشان می‌دهد که انسان‌ها قادرند تا افراد آشنا را از طریق صدایشان به‌راحتی تشخیص دهند. این امر نمود بارز خود را هنگام برقراری ارتباط از طریق خطوط تلفن نشان می‌دهد، زمانی‌که پس از شنیدن اولین هجاها و یا جمله‌ها می‌توان حدس زد که طرف مقابل چه کسی است. پس می‌توان ادعا کرد صداهای افراد دارای ویژگی‌های منحصربه‌فردی‌اند که شنونده را قادر می‌سازد تا فرد مخاطب را تنها از راه شنیدن صدایش شناسایی کند. فرایند تشخیص هویت افراد بر مبنای ویژگی‌های فیزیکی مستتر در گفتارشان در آواشناسی قضایی[1] و به‌طور خاص در تشخیص قضایی گوینده[2] نمود می‌یابد. آواهای گفتاری حاوی اطلاعاتی مربوط به آناتومی دستگاه آوایی فرد، فیزیولوژی، شرایط روحی و روانی و نیز پیش‌زمینهٔ اجتماعی وی هستند و این اطلاعات در ویژگی‌های آکوستیکی سیگنال‌های آوایی منعکس می‌شوند. با وجود این، آواهای گفتار به‌لحاظ مشخصه‌های فردویژه یکسان نیستند. به سخن دیگر، برخی از آواها نسبت به سایر آواها اطلاعات فردویژهٔ بیشتری در خود حمل می‌کنند (Kavanagh, 2012). بهترین پارامترهای آوایی تمایزدهنده آن دسته از پارامترها هستند که شامل بیشترین تغییرات بین-گوینده[3] و کمترین تغییرات درون-گوینده[4] باشند (Rose, 2002). به عبارت دیگر، پارامترهای آوایی که افراد مختلف یک زبان خاص را به بهترین نحو از هم متمایز می‌سازد و در عین حال کمترین میزان تغییر را درون یک گوینده واحد نشان می‌دهد.

نقش تمایزدهندهٔ واکه‌ها در آواشناسی قضایی همواره مورد توجه بوده است و این آواها جایگاه قابل‌ملاحظه‌ای در پژوهش‌های شناسایی گوینده دارند (Sambur, 1975 ; Goldstein,1976; Rose, 2002; McDougall, 2006). گلد[5] و فرنچ[6] (2011) در پژوهش خود با عنوان «روال‌های بین‌المللی در مقایسهٔ قضایی گوینده»[7] به این نتیجه رسیدند که ۹۴ درصد آواشناسان و محققان قضایی در پژوهش‌های خود واکه‌ها را بررسی می‌کنند. همچنین، در رتبه‌بندی پارامترهای آکوستیکی مهم، فرکانس پایه و فرکانس سازه‌ها دو متغیری بودند که بیشترین میزان تحلیل‌ها را به خود اختصاص داده بودند.

واکه‌ها آواهایی هستند که در تولید آن‌ها مجرای گفتار نسبتاً باز است و معمولاً براساس سه معیار تولیدی ارتفاع بدنهٔ زبان، موقعیت پیشین-پسین زبان و درجهٔ گردی لب‌ها توصیف می‌شوند. منبع واکه‌ها تکانه‌های حاصل از ارتعاش تارآواهاست که با خروج از صافی دستگاه گفتار به‌صورت موج صوتی، هوای اطراف گوینده را مرتعش می‌کنند (بی‌جن‌خان، ۱۳۹۲:۱۳۱). واکه‌ها در پژوهش‌های آواشناسی قضایی سهم قابل‌ملاحظه‌ای را به خود اختصاص داده‌اند. یکی از دلایل مورد توجه بودن واکه‌ها مربوط به دو ویژگی آکوستیکی مهم در این آواهاست: یکی فرکانس پایه و دیگری فرکانس سازه‌ها. فرکانس پایه همبستهٔ آکوستیکی میزان ارتعاش تارآواهاست. به عبارتی، فرکانس پایه تعداد دفعات بازوبسته شدن تارآواها را در هر ثانیه براساس مقیاس چرخه در ثانیه یا هرتز محاسبه می‌کند. گویندگان به‌لحاظ توزیع انرژی در فرکانس پایهٔ گفتارشان با هم متفاوتند. این امر به عوامل متعدد فیزیکی مانند طول و حجم تارآواها بستگی دارد. از سوی دیگر، فرکانس سازه‌ها به‌مثابهٔ یکی از واضح‌ترین

[1]. Forensic Phonetics
[2]. forensic speaker identification
[3]. between-speaker variability
[4]. within-speaker variability
[5]. E. Gold
[6]. P. French
[7]. International practices in forensic speaker comparison

http://dx.doi.org/10.22108/jrl.2021.128697.1577

همبسته‌های آکوستیکی رزونانس دستگاه گفتار اطلاعات فردویژهٔ بسیاری در خود رمزگذاری می‌کنند (Jessen and Becker, 2010 ; Gold et al., 2013).

نولان و گریگوراس (2005) فرکانس سازه و دینامیک آن را محصول تعامل میان دستگاه آوایی فرد با حرکات فردویژهٔ آنان برای تولید آوا می‌دانند. آن‌ها بر این باورند فرکانس سازه‌ها افزون بر بازتاب ویژگی‌های آناتومیک ساختار دستگاه گفتار فرد، می‌تواند عادت‌های زبانی گویندگان را نیز منعکس کند مانند کامی‌شدگی[1] که با بالارفتن فرکانس سازهٔ دوم همراه است (Nolan and Grigoras, 2005).

اگر از دیدگاه نظریهٔ آوایی منبع-صافی[2] به تولید واکه‌ها نگاه کنیم، فرکانس پایه و فرکانس سازه‌ها به‌عنوان مهم‌ترین انعکاس آکوستیکی تولید واکه از تعامل میان دو بخش مستقل، یعنی منبع و صافی حاصل می‌شوند. منبع تولید واکه ارتعاش تارآواها و صافی نیز مربوط به شکل دستگاه گفتار فوق حنجره است. تفاوت‌های فیزیکی میان طول و حجم تارآواهای افراد و همچنین، دستگاه گفتار افراد منجر به تغییر در مقادیر این پارامترهای آکوستیکی میان افراد می‌شود. بنابراین، این دو ویژگی آکوستیکی برجسته واکه‌ها را به آواهایی سودمند در مطالعات آوایی-قضایی تبدیل کرده است.

بدین ترتیب، در پژوهش پیشِ رو قصد داریم براساس چارچوب آواشناسی قضایی و با اتکا به نظریهٔ منبع-صافی به بررسی آکوستیکی ویژگی‌های معینی از واکه‌های تولیدی در گویندگان فارسی زبان بپردازیم و مشخصه‌های آکوستیکی فردویژه را در گفتار آن‌ها کشف کنیم. در پژوهش حاضر واکه‌های /a, e, o, ɑ, i, u/ در زبان فارسی به‌عنوان بخش‌های آواییِ منتخب مورد بررسی قرار خواهند گرفت و فرکانس سازهٔ اول، دوم، سوم، چهارم و فرکانس پایهٔ هریک از آن‌ها عنوان پارامترهای آکوستیکی منتخب اندازه‌گیری و بررسی خواهند شد.

رایج‌ترین شیوهٔ تحلیل واکه‌ها اندازه‌گیری فرکانس سازهٔ واکه‌ها (سازهٔ اول تا چهارم) در محدودهٔ مرکزی واکه است (Kahn et al., 2011 ; Rose: 2007; Kinoshita, 2002). به عبارتی، در این شیوه فرکانس سازه در محدودهٔ مرکزی واکه اندازه‌گیری می‌شود، جایی که گفته می‌شود واکه ثبات کافی دارد، تغییرات واکه در آن منطقه کم‌تر است و گویندگان هنگام تولید آوا قصد دارند به آن نقطهٔ هدف[3] برسند (Strange, 1989). به سخن دیگر، در این شیوه ویژگی‌های ایستای واکه مدنظر قرار می‌گیرد. یکی دیگر از شیوه‌های تحلیل واکه‌ها اندازه‌گیری فرکانس سازه‌ها و فرکانس پایه به روش بلندمدت است. در این روش که توسط نولان و گریگوراس (2005) ابداع شده است پارامترهای آکوستیکی از سطح کل زنجیرهٔ آوایی استخراج می‌شوند، بدین معنا که ابتدا زنجیرهٔ واکه‌ها از کل پاره‌گفتار استخراج می‌شود و در کنار هم قرار می‌گیرند. سپس، هر پنج یا ده‌هزارم ثانیه میانگین فرکانس پایه و سازه اندازه‌گیری می‌شود و در نهایت نیز یک میانگین و انحراف معیار برای کل واکه‌های پاره‌گفتار به دست می‌آید. اگرچه این روش در پژوهش‌های مربوط به تشخیص هویت گوینده کاربرد زیادی دارد؛ اما محدودیت‌های خاص خود را نیز دارد. در این روش تنها می‌توان قدرت فرکانس سازه‌ها را محک زد و نمی‌توان نشان داد کدام آوا نسبت به آوای دیگر فردویژه‌تر است. این درحالی است که آواها به‌لحاظ مشخصه‌های فردویژه تفاوت زیادی با هم دارند و همهٔ آواها به یک اندازه تمایزدهنده نیستند؛ برای مثال، پژوهش‌های مختلف آوایی-قضایی دربارهٔ همخوان‌های سایشی در

[1]. palatalization
[2]. source-filter theory
[3]. target

http://dx.doi.org/10.22108/jrl.2021.128697.1577

زبان‌های مختلف از جمله زبان فارسی نشان داده است که آوای /s/ به‌لحاظ نشان‌دادن مشخصه‌های فردویژه قدرت بسیار بالایی دارد (Asadi et.al., 2018 a ; Kavenagh, 2012). بنابراین، این احتمال نیز وجود دارد که واکه‌ها نیز با توجه به جایگاه و شیوهٔ تولید متفاوتی که دارند مشخصه‌های فردویژه‌شان نیز با هم متفاوت باشد. از این‌روی در پژوهش حاضر قصد داریم تا واکه‌های زبان فارسی را بر اساس شیوهٔ ایستا تجزیه و تحلیل کنیم تا از این رهگذر دریابیم که کدام واکه‌ها اطلاعات فردویژهٔ بیشتری در خود حمل می‌کنند و کدام پارامتر آکوستیکی مربوط به واکه‌ها (اطلاعات مربوط به حنجره یا اطلاعات مربوط به صافی دستگاه گفتار) اطلاعات فردویژهٔ بیشتری دارند؟ دستاوردهای این پژوهش افزون بر آواشناسی آزمایشگاهی قابلیت به‌کارگیری در زمینه‌های مختلفی از جمله آواشناسی قضایی و تشخیص اتوماتیک گوینده را نیز دارد.

۲. پیشینهٔ پژوهش

با توجه به نقش چشمگیر واکه‌ها در شناسایی گوینده، پژوهش‌های بسیاری تاکنون تلاش کرده‌اند تا جنبه‌های مختلف این آواها را با دیدگاه آوایی-قضایی بررسی کنند. کینوشیتا (2002) با اندازه‌گیری فرکانس سازه‌ها در محدودهٔ مرکزی واکه‌ها به بررسی پارامترهای آکوستیکی فرکانس سازهٔ دوم/i/، فرکانس سازهٔ دوم و سوم /e/ و فرکانس سازهٔ سوم /o/ در داده‌های صوتی مربوط به زبان ژاپنی پرداخت. فرکانس سازهٔ همهٔ نمونه‌های آوایی از محدودهٔ مرکزی کل دیرش واکه‌ها استخراج شد. استخراج داده‌ها در سه مرحله و به سه شیوهٔ متفاوت صورت گرفت. فرایند ضبط داده‌ها به‌صورت غیرهم‌زمان و با فاصلهٔ دو هفته انجام شد. نتیجهٔ پژوهش نشان داد مادامی‌که تعداد پارامترهای کافی در نظر گرفته شود، گویندگان ژاپنی‌زبان می‌توانند بر مبنای فرکانس سازه‌ها از هم متمایز شوند. این پژوهش در قالب رویکرد نسبت درست‌نمایی[1] صورت گرفت و از میان ۱۸۰ تشخیص درون-گوینده، ۵ مورد اشتباه و از میان ۹۰ تشخیص بین-گوینده، ۹ مورد اشتباه تخمین زده شد. این نتایج نشان داد حتی با وجود تعداد پارامترهای کم (شش پارامتر در این پژوهش) امکان تفکیک گویندگان وجود دارد.

رز[2] و همکاران (2003) در پژوهش خود به بررسی یک آزمایش آوایی-قضایی شناسایی گوینده انجام دادند که براساس آن جفت‌داده‌های مشابه آوایی تولیدشده توسط ۶۰ گویندهٔ مرد ژاپنی با جفت‌داده‌های غیرمشابه آوایی از همان گویندگان در قالب رویکرد درست‌نمایی با هم مقایسه شد. داده‌های مورد بررسی در دو جلسه به فاصلهٔ سه یا چهار ماه و از طریق خط تلفن ضبط شد. شرکت‌کنندگان این پیکرهٔ آوایی همگی از اعضای نیروی پلیس ژاپن بودند که محدودهٔ سنی بین ۲۰ تا ۵۰ سال داشتند. در این پژوهش سه بخش آوایی شامل یک خیشومی مورا، سایشی بیواک لثوی‌کامی و یک واکهٔ گرد میانی پسین کشیده به شیوهٔ ایستا تجزیه و تحلیل شدند. در کنار استخراج فرکانس سازه‌های اول تا پنجم از بخش‌های آوایی یادشده، آن‌ها یک پارامتر اتوماتیک یعنی کپستروم[3] را نیز در دستور کار خود قرار دادند تا در انتها پارامترهای مورد بررسی را با هم مقایسه کنند و قدرت شاهد[4] آن‌ها را بسنجند. نتایج نشان داد فرکانس سازه‌ها عملکرد مطلوبی در شناسایی گوینده داشته‌اند با این حال

[1]. likelihood ratio
[2]. P. Rose
[3]. cepstrum
[4]. strength of evidence

http://dx.doi.org/10.22108/jrl.2021.128697.1577

قدرت شاهد آن‌ها از کپستروم پایین‌تر گزارش شد که نویسندگان معتقدند علت این امر می‌تواند ناشی از عوامل تأثیرگذار بر فرکانس سازه‌ها باشد که بر آن‌ها پوشیده است.

آلدرمن[1] (2005) نیز در پژوهش خود به روش ایستا و از طریق اندازه‌گیری فرکانس سازه در محدودهٔ مرکزی واکه‌ها در دو پیکرهٔ آوایی قدیم و جدید نشان داد که می‌توان با استفاده از فرکانس سازهٔ پنج واکهٔ بلند انگلیسی استرالیایی نمونه‌های آوایی متعلق به افراد مشابه را از نمونه‌های آوایی افراد متفاوت متمایز ساخت. همچنین، استفاده از دو پیکرهٔ آوایی که در دو زمان مختلف ضبط شده است تأثیر چندانی در نتایج نداشته است که این موضوع نشان می‌دهد تغییر زبانی نمی‌تواند قدرت فرکانس سازه را به‌عنوان یک پارامتر شناسایی گوینده تحت تأثیر قرار دهد.

رز (2007) در دیگر پژوهش خود به روش ایستا و در قالب رویکرد نسبت درست‌نمایی به بررسی تغییرات بین-گوینده و درون-گوینده در پنج واکهٔ سخت[2] /i: ɨ: ə: a: o:/ و شش واکهٔ نرم[3] /ɪ e æ a o ʊ/ در زبان انگلیسی استرالیایی پرداخته است. نتایج نشان داد واکهٔ /i/ بهترین واکه در نشان‌دادن تمایز بین گویندگان بوده و فرکانس سازهٔ دوم /i/ نیز نسبت به سایر پارامترها شامل اطلاعات فردویژهٔ بیشتری بوده است. همچنین، عملکرد واکهٔ /æ/ نسبت به واکه‌های سخت /a:/ و /ə:/ بهتر بوده است.

رز و وینتر[4] (2010) در ادامه به بررسی صدای زن‌ها در چارچوب مقایسهٔ قضایی صدا پرداختند. به اعتقاد پژوهشگران امروزه تنها مردان مجرم نیستند و تعداد جرائمی که زنان مرتکب می‌شوند نیز رو به فزونی است. از این رو، آن‌ها در پژوهش خود سه فرکانس سازهٔ اول را از داده‌های آوایی تولیدشده توسط ۲۰ زن استرالیایی استخراج کردند. داده‌ها در دو جلسهٔ غیرهمزمان به فاصلهٔ یک تا پنج هفته ضبط شد و رویکرد تحلیلی این پژوهش نسبت درست‌نمایی بود. نتایج نشان داد برخلاف صدای مردان، فرکانس سازهٔ اول واکه‌های افراشته در صدای زن‌ها می‌تواند پارامتر سودمندی برای مقایسهٔ قضایی صدا باشد. همچنین، نتایج حاکی از آن بود که فرکانس سازهٔ سوم واکهٔ /i/ به نظر چندان در تمایز صدای زن‌ها مفید نباشد.

کان[5] و همکاران (2010) نیز با اتخاذ رویکرد ایستا در پژوهش خود با اعتقاد به اینکه مطالعهٔ تفاوت‌های بین-گوینده و درون-گوینده می‌تواند درک بهتری از صدا به عنوان داده‌ای بیومتریک به دست دهد به بررسی ۳۲۸,۰۱۶ نمونهٔ آوایی از ۱۰ واکهٔ دهانی زبانی فرانسه پرداختند. در این پژوهش مقدار فرکانس سازه از محدودهٔ میانی ده واکهٔ زبان فرانسه استخراج شد. نتایج این پژوهش نشان داد واکه‌های /æ/، /ɛ/ و /a/ حاوی اطلاعات فردویژهٔ بیشتری نسبت به سایر واکه‌های دهانی در زبان فرانسه بودند. به‌طور کل واکهٔ میانی /æ/ و /ɛ/ و واکهٔ افتادهٔ /a/ بیشترین قدرت تمایز میان گویندگان را داشتند.

تاکنون به پژوهش‌هایی اشاره کردیم که به شیوهٔ ایستا به بررسی تغییرات بین-گوینده و درون-گوینده در واکه‌ها پرداخته‌اند؛ اما ارتباط فرکانس سازه با فرکانس پایه در این پژوهش‌ها مدنظر نبوده است و تنها تمرکزشان بر فرکانس سازه‌ها بوده است. دسته‌ای دیگر از پژوهش‌ها استقلال میان پارامترهای مربوط به حنجره و پارامترهای مربوط به صافی گفتار را از زاویهٔ دید نظریهٔ

1. T. Alderman
2. tense vowels
3. lax vowels
4. E. Winter
5. J. Kahn

http://dx.doi.org/10.22108/jrl.2021.128697.1577

منبع-صافی بررسی کرده‌اند و به تبیین نقش این نظریه در آواشناسی قضایی پرداخته‌اند. اصل نظریهٔ منبع-صافی بر فرض استقلال مراحل تولید آوا در حنجره و در دستگاه گفتار استوار است. با این حال، در پاره‌ای از پژوهش‌ها به تعامل غیر-خطی میان منبع و صافی و رابطهٔ متقابل این دو بخش نیز اشاره شده است. تیتز[1] (2008) تعامل میان دو بخش منبع و صافی را بررسی می‌کند و معتقد است فشار هوای درون دستگاه گفتار بر جریان هوای فوق‌چاکنایی و نیز لرزش تارآوها تأثیر می‌گذارد. گوردون[2] و لدفوگد[3] (2001) نیز به تعامل میان منبع و صافی در فرایند واک‌سازی اشاره کرده‌اند و بیان می‌کنند که بالارفتن حنجره در تولید واک جیرجیری[4] و پایین‌آمدن آن درتولید واک نفسی[5] هر دو بر مقادیر فرکانس سازهٔ اول تأثیر می‌گذارند. با وجود این، پژوهش‌هایی که در بافت آوایی-قضایی انجام شده است به استقلال پارامترهای منبع و صافی تأکید می‌کنند و معتقدند ترکیب این دو پارامتر می‌تواند در شناسایی گویندگان برحسب صدایشان سودمند باشد. به‌عنوان نمونه، هیوز[6] و همکاران (2017) پارامترهای مربوط به منبع (کیفیت صدا)[7] و پارامترهای مربوط به صافی (فرکانس سازهٔ بلندمدت و MFCC)[8] را با سه رویکرد اتوماتیک، نیمه‌اتوماتیک و آوایی سنجیدند و به این نتیجه رسیدند که پارامترهای مربوط به صافی اطلاعات فردویژهٔ مشابهی نشان می‌دهند و در صورتی که در رویکرد اتوماتیک اطلاعات مربوط به حنجره نیز اضافه شود بازدهی سیستم‌های شناسایی گوینده نیز به موازات بالاتر می‌رود.

سن‌سگوندو[9] و همکاران (2017) در پژوهش خود شباهت گویندگان را براساس فاصلهٔ اقلیدسی[10] و با استفاده از مشخصه‌های منبع و صافی در دوقلوهای همسان بررسی کردند. نتایج این پژوهش استقلال میان مشخصه‌های مربوط به منبع و صافی را تأیید کرد و دیگر بار نشان داده شد که می‌توان به‌طور همزمان در تحلیل‌های آوایی-قضایی از ترکیب این مشخصه‌ها استفاده کرد بدون اینکه پارامتری اضافی در پژوهش دخیل باشد.

هیوز و همکاران (2019) در دیگر پژوهش خود داده‌های آوایی را از پیکرهٔ DyViS (Nolan et al., 2009) استخراج کردند و پارامترهای مربوط به منبع، یعنی فرکانس پایه و کیفیت صدا، و پارامترهای مربوط به صافی، یعنی فرکانس سازه‌ها و MFCC، را در این پیکره تجزیه و تحلیل کردند. نتایج حاکی از استقلال پارامترهای یادشده بود و نشان داد هر کدام از این پارامترها اطلاعات متفاوتی را دربارهٔ گوینده در خود حمل می‌کنند که در نتیجه، ترکیب آن‌ها می‌تواند به بهبود شناسایی گوینده در بافت‌های قضایی کمک کند.

طبق نتایج به‌دست‌آمده از مطالعات پیشین، انتظار می‌رود فرکانس پایه و فرکانس سازه به‌عنوان دو پارامتر مربوط به حنجره و دستگاه گفتار بتوانند اطلاعات فردویژهٔ متفاوتی دربارهٔ صدای گویندگان نشان بدهند و ترکیب آن‌ها در مطالعات آوایی-قضایی منجر به نتیجهٔ بهتری در تشخیص صدای گویندگان شود.

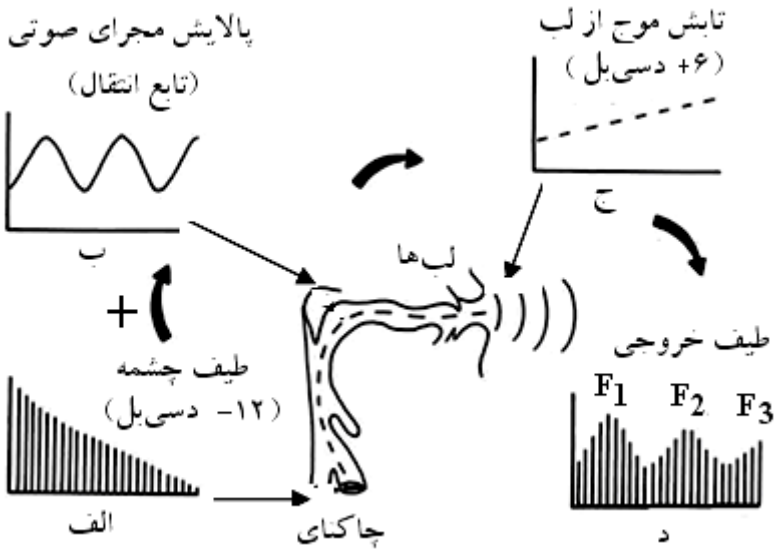
---

1. I.R. Titze
2. M. Gordon
3. P. Ladefoged
4. creaky voice
5. breathy voice
6. V. Hughes
7. voice quality
8. Mel-frequency cepstrum coefficient
9. E. San Segundo
10. Euclidean Distance

http://dx.doi.org/10.22108/jrl.2021.128697.1577

۳. نظریهٔ منبع-صافی

نظریهٔ منبع-صافی برای اولین بار توسط فانت[1] (1960) در کتاب *نظریهٔ صوتی تولید گفتار*[2] ارائه شد. در انگارهٔ پیشنهادی فانت، فرایند تولید، ارسال و درک گفتار با استفاده از مبانی صوتی انتقال امواج در لوله‌های صوتی به هم پیوند خورده‌اند (بی‌جن‌خان، ۱۳۹۲: ۹۳). انگارهٔ منبع-صافی از دو بخش به نام منبع و صافی تشکیل شده است که از هم مستقل‌اند و کارکردی متفاوت دارند. در واقع، گفتار حاصل تعامل یک منبع صوت و یک صافی است که به صوت ایجادشده شکل یا طنین می‌دهد (مدرسی قوامی، ۱۳۹۰: ۱۳۱). منظور از منبع، محلی از دستگاه گفتار است که بر اثر فعالیت‌های اندام‌های گویایی منجر به تولید صوت می‌شوند. پس از اینکه صدا در منبع تولید شد، وارد مرحلهٔ صافی می‌شود. صافیِ دستگاه گفتار همان حفره‌های بازخوانی، یعنی حلق و دهان، هستند که شکل و ابعاد این حفره‌ها در تعیین کیفیت واکه‌ها و مشخصه‌های آکوستیکی آن‌ها تأثیرگذار است. طبق انگارهٔ منبع-صافی، دو بخش منبع و صافی از هم مستقل‌اند و همین استقلال سبب می‌شود که بتوان ویژگی‌های مربوط به این دو بخش را جداگانه اندازه‌گیری و کمی‌سازی کرد. منبع صدا در تولید واکه‌ها ارتعاش منظم پرده‌های صوتی در حنجره است. پس از آن موج صوتی حاصل از فعالیت حنجره با عبور از دستگاه گفتار بر حسب شکل و اندازهٔ این دستگاه به صورت‌های مختلف فیلتر می‌شود؛ به این معنی که انرژی برخی از فرکانس‌ها تقویت و انرژی برخی فرکانس‌های دیگر تضعیف می‌شود و به این شیوه آواهایی با کیفیت‌های گوناگون تولید می‌شود (نوربخش، ۱۳۹۲: ۳۱). در شکل ۱ نحوهٔ تعامل میان منبع و صافی در تولید گفتار ترسیم شده است.



شکل ۱- نحوهٔ تعامل میان منبع و صافی در تولید گفتار (برگرفته از علی‌نژاد و حسینی‌بالام، ۱۳۹۲: ۷۰)

Figure 1- Source-filter interaction in speech production

[1]. G. Fant

[2]. Acoustic Theory of Speech Production

http://dx.doi.org/10.22108/jrl.2021.128697.1577

٤. روش‌شناسی پژوهش

در بخش‌های زیر اطلاعات مربوط به شرکت‌کنندگان، نحوهٔ ضبط داده‌های آوایی، تقطیع داده‌ها و نیز پارامترهای انتخاب‌شده برای پژوهش حاضر توضیح داده می‌شود.

٤-١. شرکت‌کنندگان و داده‌های آوایی

به‌منظور استخراج پارامترهای آکوستیکیِ تمایزدهنده میان گویشوران فارسی‌زبان، پیکره‌ای آوایی در محیط آزمایشگاهی ضبط شد. در این پیکرهٔ آوایی صدای دوازده گویشور مرد فارسی‌زبان با محدودهٔ سنی ۲۲ تا ۳۵ سال در دو جلسهٔ مجزا ضبط شد. علت ضبط داده‌های آوایی در دو جلسهٔ مجزا اندازه‌گیری میزان تغییرات درون-گوینده پارامترهای آکوستیکی با توجه به عامل گذر زمان بود. به‌منظور کنترل پیکره و محدودسازی عواملِ تأثیرگذاری مانند لهجهٔ منطقه‌ای و اجتماعی، تلاش شد گویشورانی انتخاب شوند که زبان مادری آن‌ها فارسی معیار است و دارای تحصیلات کارشناسی یا کارشناسی ارشد باشند. همچنین، گویشوران هیچ‌گونه سابقهٔ اختلال گفتاری و شنوایی نداشتند. آن‌ها تعداد ۵۴ جملهٔ فارسی را در دو جلسهٔ جداگانه به فاصلهٔ یک تا دو هفته تولید کردند. در مجموع تعداد ۱۲۹۶ (۱۲۹۶= ۲ تکرار×۵۴ جمله×۱۲ گویشور) نمونهٔ آوایی به دست آمد.

٤-٢. شیوهٔ ضبط و تقطیع داده‌ها

صدای آزمودنی‌ها با استفاده از میکروفون رولاند ۴۴۱۰۰ هرتز ضبط شد. میکروفون به‌صورت مورب و به فاصلهٔ ۲۰ سانتی‌متری از دهان شرکت‌کنندگان قرار گرفت. از آن‌ها درخواست شد که جمله‌ها را بدون آهنگ نشان‌دار به‌صورت طبیعی و با مکثی حدود ۳ ثانیه میان هر جمله تولید کنند. به‌منظور آشناسازی شرکت‌کنندگان با متن خوانداری، از آن‌ها درخواست شد تا پیش از شروع فرایند ضبط صدا متن را چند بار بخوانند. داده‌های آوایی با استفاده از نرم‌افزار Praat ویرایش ۳۴، ۲، ۵ (Boersma and Weenink, 2013) تقطیع و لایه‌بندی شد. متناسب با هر فایل صوتی، یک شبکهٔ متنی ایجاد شد. نمونه‌های آوایی با استفاده از علائم آواشناختی IPA[1] برچسب‌گذاری شد. واکه‌ها از طریق سازه‌هایشان که به‌صورت نوارهای پررنگ با پهنای نوار محدود در طیف‌نگاشت حضور دارند، مشخص شدند. پارامترهای آکوستیکی فرکانس پایه، فرکانس سازهٔ اول، فرکانس سازهٔ دوم، فرکانس سازهٔ سوم و فرکانس سازهٔ چهارم برای اندازه‌گیری تغییرات بین-گوینده و درون-گوینده در پیکرهٔ آوایی موردبررسی انتخاب شد. به‌منظور اندازه‌گیری این پارامترها ابتدا طیف‌نگاشت و موج‌صوتی واکهٔ مدنظر را به‌دست آوردیم و سپس، با انتخاب محدودهٔ مرکزی ۵۰ هزارم ثانیه مقادیر را از این منطقه استخراج کردیم. در این پژوهش مقادیر به‌دست‌آمده از واکه‌ها با استفاده از برنامهٔ اندازه‌گیری خودکار انجام شده است. تحلیل آماری داده‌ها و مقادیر به‌دست‌آمده از بررسی آکوستیکی نمونه‌های آوایی با استفاده از نرم‌افزار SPSS ویرایش ۲۱/۰ و نرم‌افزار R ویرایش ۳. ۳. ۳ صورت گرفت.

٥. گزارش نتایج

در بخش‌های زیر با توجه به پرسش‌های مطرح‌شده در بخش مقدمه به اجرای آزمون‌های آماری مرتبط پرداخته و نتایج مربوط به تجزیه و تحلیل داده‌ها گزارش خواهد شد. ابتدا نتایج آماری مربوط به تغییرات بین-گوینده در پارامترهای فرکانس پایه و

[1]. International Phonetic Association

http://dx.doi.org/10.22108/jrl.2021.128697.1577

فرکانس سازه‌های اول تا چهارم گزارش خواهد شد و سپس نتایج مربوط به تغییرات درون-گوینده در پارامترهای یاددشده ارائه می‌شود. در بخش نهایی نیز همبستگی میان پارامترهای منتخب و رابطهٔ میان این پارامترها بررسی و تحلیل خواهد شد.

۵-۱. تغییرات بین-گوینده در فرکانس پایه و فرکانس سازه‌های واکه‌ها

داده‌های آوایی جمع‌آوری‌شده از دوازده گویندهٔ مرد فارسی‌زبان وارد نرم‌افزار تحلیل گفتار Praat شد و سپس، با استفاده از برنامهٔ خودکار تحلیل سازه و فرکانس پایه، مقدار فرکانس سازه‌های اول تا چهارم و نیز فرکانس پایه از منطقهٔ ثبات واکه‌ها به دست آمد. در جدول (۱) میزان رخداد واکه‌ها به‌همراه میانگین، انحراف معیار بین-گوینده، انحراف معیار درون-گوینده و عدد مربع اتا[1] گزارش شده است.

به‌منظور بررسی تأثیر عامل گوینده بر متغیرهای آکوستیکی فرکانس پایه و فرکانس سازه‌های واکه‌های سادهٔ زبان فارسی آزمون تحلیل واریانس دوطرفه برای هر واکه در نرم‌افزار R اجرا شد. گوینده به‌عنوان متغیر مستقل و پارامترهای آکوستیکی فرکانس پایه و فرکانس سازه‌های اول، دوم، سوم و چهارم به عنوان متغیر وابستهٔ پژوهش در نظر گرفته شدند. توزیع تابع F در آزمون تحلیل واریانس امکان مقایسهٔ تغییرات بین-گوینده را نسبت به تغییرات درون-گوینده فراهم می‌سازد. هرچه نسبت F بالاتر باشد، میزان تمایز میان گویندگان نیز بیشتر است. براساس نتایج حاصل از اجرای آزمون تحلیل واریانس دوطرفه، اثر گوینده بر پارامترهای فرکانس پایه و فرکانس سازه‌های اول تا چهارم در همهٔ واکه‌های تحت آنالیز به جز یک مورد معنی‌دار بوده است ($p\leq0.05$). طبق نتایج به‌دست‌آمده، اثر گوینده تنها بر فرکانس سازهٔ دوم واکهٔ /u/ معنی‌دار نبوده است (F (11, 659) = 1.97, sig= 0.028). بالاترین نسبت F در فرکانس پایهٔ واکه‌ها به ترتیب در واکهٔ /ɑ/، واکهٔ /e/ و واکهٔ /a/ گزارش شده است.

F0 /ɑ/: F (11, 2745) =423.60, sig= 0.000
F0 /e/: F (11, 3563) =323.12, sig= 0.000
F0 /a/: F (11, 2963) =282.12, sig= 0.000

در رابطه با فرکانس سازه‌ها، بالاترین نسبت F به ترتیب در فرکانس سازهٔ اول و سوم واکهٔ /a/، فرکانس سازهٔ سوم /e/ و نیز فرکانس سازهٔ سوم /ɑ/ مشاهده شده است.

F1 /a/: F (11, 2963) =157.97, sig= 0.000
F3 /a/: F (11, 2963) =100.47, sig= 0.000
F3 /e/: F (11, 3563) =137.76, sig= 0.000
F3 /ɑ/: F (11, 2745) =88.03, sig= 0.000

آزمون تحلیل واریانس چندمتغیره با در نظرگرفتن گوینده به‌عنوان متغیر مستقل بر روی داده‌های آوایی نشان می‌دهد که در مجموع فرکانس سازه‌های بالاتر، یعنی سازهٔ سوم و چهارم، حاوی اطلاعات فردویژهٔ بیشتری نسبت به فرکانس سازهٔ اول و دوم هستند. در رابطه با واکه‌ها، اثر گوینده در واکهٔ /a/ و پس از آن در واکهٔ /e/ نسبت به سایر واکه‌ها قوی‌تر گزارش شده است. همچنین، میزان توزیع F در فرکانس پایهٔ واکهٔ /ɑ/ و /e/ نسبت به سایر واکه‌ها بیشتر بوده است. از آنجا که امکان مقایسهٔ مستقیم میان عدد P و F وجود ندارد، به‌منظور نشان دادن اندازهٔ اثر گوینده و قدرت تمایزدهندگی واکه‌ها از معیار مربع اتا استفاده کردیم. عدد مربع اتا بزرگی اثر گوینده یعنی نسبت واریانس توضیح داده‌شده را در آزمون مانوا[2] نشان می‌دهد. مربع اتا از تقسیم

[1]. Eta squared
[2]. MANOVA

http://dx.doi.org/10.22108/jrl.2021.128697.1577

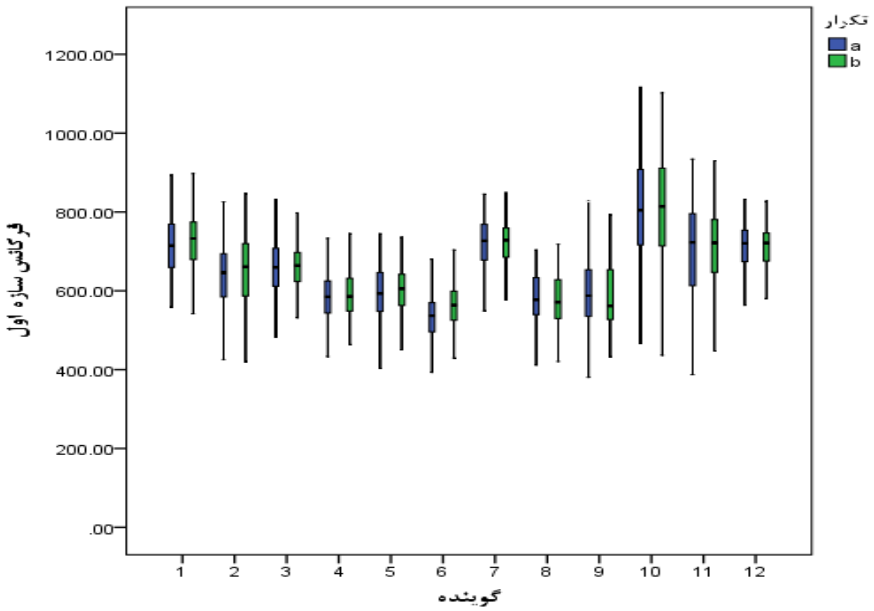
مجموع مربعات فاکتور مدنظر بر مجموع مربعات کل اثرات، تعامل‌ها و خطاها به‌دست می‌آید. همان گونه که مشاهده می‌شود واکهٔ /a/ و پس از آن واکهٔ /e/ بیشترین تغییرات میان گویندگان را نشان می‌دهند، در حالی که واکهٔ /u/ نسبت به سایر واکه‌ها کمترین میزان تغییرات را میان گویندگان نشان می‌دهد. در شکل (۲) نمودار جعبه‌ای مربوط به تغییرات بین-گوینده و درون-گوینده را در فرکانس سازهٔ اول واکهٔ /a/ در داده‌های آوایی موردبررسی مشاهده می‌کنید.

جدول ۱- تعداد واکه‌های استخراج‌شده از پیکرهٔ آوایی، فرکانس پایه و فرکانس سازه‌های اول تا چهارم و عدد مربع اتا برای هر واکه. اعدادی که در هر خانه آمده است: پررنگ=میانگین؛ معمولی= انحراف معیار بین-گوینده؛ ایرانیک=انحراف معیار درون-گوینده.

Table 1- Number of vowels analyzed from Persian speech corpus, F0, F1 to F4 and η2 values for each vowel. Figures in values cells: bold=mean; normal=between-speaker standard deviation; italics=within-speaker standard deviation.

		/a/	/e/	/o/	/ɑ/	/i/	/u/
گوینده	#	۲۹۷۵	۳۵۷۵	۱۶۶۴	۲۷۵۷	۱۵۳۷	۱۶۶۰
	F0	**۱۳۱** ۹-۳۱	**۱۳۵** ۱۱-۳۳	**۱۳٤** ۹-۳۱	**۱۳۳** ۸-۳۰	**۱۳۵** ۱۰-۳۲	**۱۳٦** ۹-۳۱
	F1	**٦۵۱** ۶۱-۱۱۶	**٤٤٦** ۴۴-۹۷	**٤۱۸** ۴۱-۸۵	**۵۱۷** ۴۶-۸۴	**۳۳۷** ۴۶-۸۴	**۳٤۲** ۶۰-۱۰۰
	F2	**۱٤۸٦** ۸۴-۱۷۰	**۱٦۱۲** ۱۶۳-۲۵۰	**۱۰۹٦** ۱۵۱-۲۱۰	**۱۱۷٤** ۱۰۸-۲۰۹	**۱۹۷٦** ۱۷۲-۳۱۳	**۱۱۵۰** ۱۷۰-۲۴۰
	F3	**۲٤۲۹** ۱۷۱-۲۹۹	**۲۵۱۱** ۱۳۰-۲۴۵	**۲٤۳۵** ۱۳۷-۲۵۲	**۲٤۳۹** ۱۴۵-۲۵۸	**۲٦۳٤** ۱۲۵-۲۶۲	**۲٤۲۳** ۱۰۷-۲۳۰
	F4	**۳۵۰۵** ۱۵۴-۲۱۲	**۳۳۲۹** ۱۶۰-۱۷۶	**۳۵۱٤** ۱۴۹-۲۱۲	**۳۳۷٦** ۱۶۰-۲۴۶	**۳٤۰۲** ۱۸۱-۲۱۳	**۳۳۲۹** ۱۱۷-۲۰۸
	مربع اتا η2	۲۰٪	۱۸٪	۱۶٪	۱۷٪	۱۵٪	۱۴٪

http://dx.doi.org/10.22108/jrl.2021.128697.1577



شکل ۲- نمودار جعبه‌ای مربوط به تغییرات بین-گوینده و درون-گوینده در فرکانس سازهٔ اول واکهٔ /a/ در داده‌های آوایی (a مربوط به جلسهٔ اول ضبط داده‌ها و b مربوط به جلسهٔ دوم ضبط داده‌هاست).

Figure 2- Boxplot of between- and within-speaker variability in F1 values of vowel /a/ in Persian speech corpus (a is related to the first session of recording while b shows the second recording session).

۵-۲. تغییرات درون-گوینده در فرکانس پایه و فرکانس سازه‌های واکه‌ها

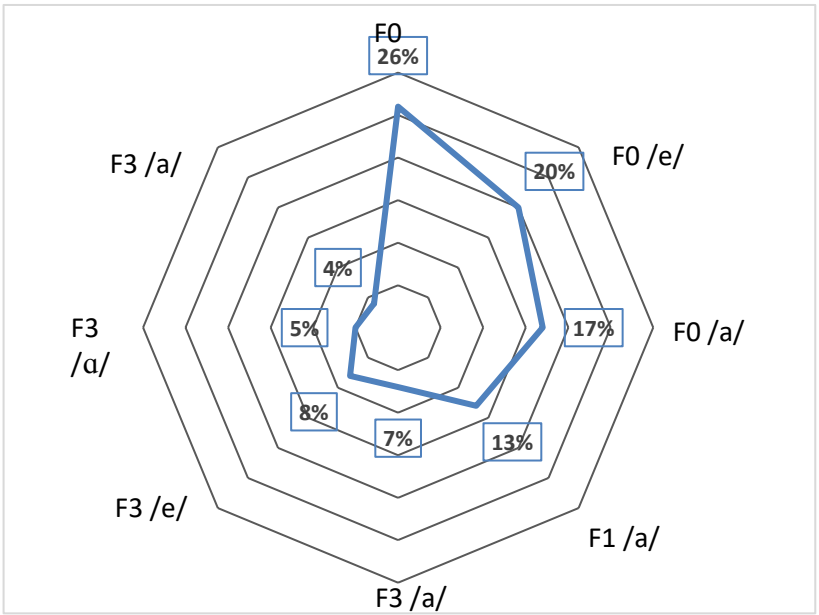
در ادامه به‌منظور بررسی عامل گذر زمان، اثر تکرار در داده‌های آوایی مدنظر اندازه‌گیری شد. همان طور که در بخش ۳-۱ اشاره شد داده‌های آوایی در دو مرحلهٔ جداگانه ضبط شده است. هدف از ضبط غیرهمزمان داده‌ها اندازه‌گیری تغییرات درون-گوینده در داده‌های آوایی موردبررسی بود. همان گونه که پیش‌تر نیز ذکر شد پارامترهایی در تشخیص هویت گوینده مناسب هستند که بتوانند نسبت به عوامل تأثیرگذار در تغییرات درون-گوینده نیز مقاوم باشند. تغییرات درون-گوینده می‌توانند از منابع مختلفی همچون شرایط جسمانی، شرایط روحی، سبک گفتار و یا گذر زمان حاصل شوند. در این پژوهش، عامل گذر زمان به‌عنوان منبع تغییر-درون‌گوینده انتخاب شده است و به همین منظور داده‌ها در دو جلسهٔ غیرهمزمان به فاصلهٔ یک تا دو هفته ضبط شد. به‌منظور بررسی اثر عامل گذر زمان، تکرار (ضبط دومرحله‌ای داده‌ها) به‌عنوان متغیر مستقل و پارامترهای آکوستیکی فرکانس پایه و فرکانس سازهٔ اول تا چهارم به‌عنوان متغیرهای وابسته وارد آزمون تحلیل واریانس شد. طبق نتایج به‌دست‌آمده، رابطهٔ معنی‌داری میان اثر تکرار و متغیرهای وابسته گزارش نشد ($p\geq0.05$). این بدان معناست که پارامترهای موردبررسی در داده‌های غیرهمزمان نسبت به عامل گذر زمان مقاوم بوده‌اند و رفتاری ثابت در دو هر مرحله از خود نشان داده‌اند.

۵-۳. سنجش میزان فردویژگیِ فرکانس پایه و فرکانس سازه‌های واکه‌ها

به‌منظور اندازه‌گیری اینکه هر کدام از پارامترهای آکوستیکی مورد بررسی می‌توانند تفاوت‌های بین-گوینده را بهتر نشان دهند و

http://dx.doi.org/10.22108/jrl.2021.128697.1577

کدام‌یک از پارامترها از این لحاظ عملکرد بهتری دارند از مدل رگرسیون لجستیک چند اسمی[1] استفاده کردیم. این آزمون را با استفاده از نرم افزار SPSS ویرایش ۲۱/۰ اجرا کردیم. گوینده به‌عنوان متغیر پاسخ اسمی[2] در نظر گرفته شد و پارامترهای آکوستیکی مورد بررسی را نیز به‌عنوان کوواریات‌های پیش‌بین[3] مدنظر قرار دادیم. به‌منظور بیان قدرت فردویژگیِ هر پارامتر براساس درصد، ابتدا نسبت درست‌نمایی مقدار خی دوی هر کدام از پارامترها را بر مجموع مقادیر خی دوها تقسیم و سپس نتیجهٔ حاصل را بر ۱۰۰ تقسیم کردیم. تحلیل فرکانس پایهٔ واکه‌های مورد بررسی نشان داد که فرکانس پایهٔ واکهٔ /ɑ/ با حدود ۲۶ درصد و پس از آن فرکانس پایهٔ واکهٔ /e/ با حدود ۲۰ درصد حاوی بیشترین اطلاعات فردویژه در مقایسه با فرکانس پایهٔ سایر واکه‌ها هستند. از میان فرکانس سازه‌های واکه‌های مختلف نیز فرکانس سازهٔ اول واکهٔ /a/ با حدود ۱۳درصد و پس از آن فرکانس سازهٔ سوم واکه /e/ با حدود ۸ درصد و فرکانس سازهٔ سوم واکهٔ /a/ با حدود ۷ درصد بیشترین میزان اطلاعات فردویژه را در گویندگان نشان می‌دهند. همان گونه که نتایج نشان می‌دهد واکهٔ /a/ در مقایسه با سایر واکه‌ها و فرکانس سازهٔ سوم در مقایسه با سایر سازه‌ها تغییرات بین-گویندهٔ بیشتری نشان می‌دهد. شکل (۳) نمایش گرافیکی توانایی نسبی هر کدام از پارامترها را در نشان دادن تغییرات بین-گوینده نشان می‌دهد. با توجه به تعداد بالای پارامترهای تنها هشت پارامتر اول که بیشترین قدرت تمایزدهندگی را داشته‌اند، انتخاب شده است.



شکل ۳- نمودار رادار نشان‌دهندهٔ توانایی نسبی پارامترهای فرکانس پایه و فرکانس سازه‌های واکه‌ها در توضیح تغییرات بین-گوینده

Figure 3- Radar chart showing relative strength of F0 and vowel formants in explaining between-speaker variability

[1]. multinominal logistic regression
[2]. nominal response variable
[3]. predicting covariates

http://dx.doi.org/10.22108/jrl.2021.128697.1577

٤-٥. همبستگی پارامترهای فرکانس پایه و فرکانس سازه‌های واکه‌ها

در گام آخر به‌منظور محاسبهٔ بهترین ترکیب پارامترها و پاسخ به اینکه چه ترکیبی از پارامترها قادرند بهترین تغییرات بین-گوینده را نشان دهند آزمون تحلیل عامل[1] (تحلیل مؤلفه‌های اصلی) بر روی داده‌های آوایی اجرا شد. پارامترهایی که کمترین همبستگی را میان همدیگر داشته باشند این قابلیت را دارند که با هم ترکیب شوند و بنابراین، می‌توان از ترکیب آن‌ها در تشخیص هویت گوینده بهره برد. نحوهٔ عملکرد آزمون تحلیل عامل به این شکل است که ابتدا ماتریسی از ضرایب همبستگی ایجاد می‌کند. پس از آن از طریق چرخش عامل‌ها حداکثر رابطه بین متغیرها و عامل‌ها اندازه‌گیری می‌شود و در نهایت، بار عاملی[2] هر پارامتر محاسبه می‌شود. بار عاملی کوچکتر از ۰/۴ نشان می‌دهد که پارامتر مدنظر تأثیر اندکی در توضیح واریانس عامل دارد و از این رو، نمی‌تواند ذیل آن عامل طبقه‌بندی شود. جدول (۲) نتایج آزمون تحلیل مؤلفه‌های اصلی را برای نشان دادن رابطه میان پارامترهای فرکانس پایه و فرکانس سازه‌های اول تا چهارم گزارش می‌دهد. خانه‌هایی که در جدول (۲) به‌صورت سایه‌دار مشخص شده است بزرگترین ارقام مربوط به هر پارامتر (بزرگتر از حد آستانه ۰/۴) را نشان می‌دهند که نشان از همبستگی بالای میان پارامترهای هر عامل را دارد. نتایج نشان می‌دهد که پارامترهای مربوط به فرکانس سازهٔ دوم تا چهارم، یعنی عامل اول، و پارامترهای مربوط به فرکانس پایه و فرکانس سازهٔ اول، یعنی عامل دوم، مقوله‌های مستقلی را تشکیل می‌دهند و این نشان می‌دهد که این دو نوع پارامتر متعامد هستند و همبستگی کمی با هم دارند. از این رو، پارامترهای فرکانس پایه و فرکانس سازهٔ اول اطلاعات مشابهی دربارهٔ گوینده رمزگذاری می‌کنند و ترکیب آن‌ها با فرکانس سازه‌های دوم تا چهارم می‌تواند در تشخیص هویت گوینده مفید باشد.

جدول ۲- نتایج آزمون تحلیل مؤلفه‌های اصلی در نشان دادن رابطه میان پارامترهای فرکانس پایه و فرکانس سازه‌ها

Table 2- Results of Principal Component Analysis for F0 and vowel formants relationship

پارامتر	بار عاملی	
	عامل ۱	عامل ۲
فرکانس پایه	-۰/۱۶۰	۰/۶۸۶
فرکانس سازهٔ اول	-۰/۱۰۷	۰/۷۵۸
فرکانس سازهٔ دوم	۰/۷۲۵	۰/۰۳۵
فرکانس سازهٔ سوم	۰/۷۹۵	۰/۰۲
فرکانس سازهٔ چهارم	۰/۶۹۷	۰/۲۳۹

٦. بحث و بررسی

در پژوهش حاضر به بررسی تأثیر پارامترهای مربوط به بخش‌های منبع و صافی، یعنی فرکانس پایه و فرکانس سازه، در تشخیص هویت گویندگان فارسی‌زبان پرداخته‌ایم. نتایج حاصل از تجزیه و تحلیل داده‌های آوایی نشان داد که اثر گوینده به‌جز در یک

[1]. factor analysis
[2]. factor loadings

http://dx.doi.org/10.22108/jrl.2021.128697.1577

مورد در پارامترهای آکوستیکی فرکانس پایه و فرکانس سازه‌های اول، دوم، سوم و چهارم همهٔ واکه‌ها معنی‌دار بوده است. اثر گوینده تنها در فرکانس سازهٔ دوم واکهٔ /u/ معنی‌دار نبوده است ($p\geq0.05$). پارامتر آکوستیکی فرکانس پایه به ترتیب در واکه‌های /ɑ/، /e/ و /a/ بهترین عملکرد را از خود نشان داده است. در رابطه با فرکانس سازه‌های واکه‌ها، بهترین عملکرد به ترتیب در فرکانس سازهٔ اول و سوم واکهٔ /a/، فرکانس سازهٔ سوم /e/ و نیز فرکانس سازهٔ سوم /ɑ/ مشاهده شده است. از میان واکه‌های سادهٔ زبان فارسی نیز واکهٔ /a/ و /e/ بیشترین اطلاعات فردویژه را نشان دادند، در حالی که واکه‌های /u/ و /i/ ضعیف‌ترین عملکرد را در تمایز گویندگان داشته است. این نتیجه همراستا با یافته‌های کان و همکاران (2011) است که نشان دادند واکهٔ افتادهٔ /a/ به همراه واکهٔ میانی/æ/ و /ɛ/ اطلاعات فردویژهٔ بیشتری نسبت به سایر واکه‌های دهانی در زبان فرانسه منتقل می‌کنند، در حالی که واکه‌های /u/ و /i/ عملکرد خوبی در تمایز گویندگان نداشتند. این نتیجه نشان می‌دهد که هرچه واکه افتاده‌تر باشد، اطلاعات فردویژهٔ آن به موازات نیز بیشتر است. در تولید واکه‌های افتاده بدنهٔ زبان افتاده است و حداکثر فاصله را با سقف دهان دارد. طبق نتایج این پژوهش، فرکانس سازهٔ اول واکهٔ /a/ یکی از پارامترهای قدرتمند در تشخیص هویت گویندگان گزارش شد. فرکانس سازهٔ اول با میزان باز بودن مجرای دهان در ارتباط است. این بدان معناست که هرچه مجرای دهان در تولید یک واکه بازتر باشد، فرکانس سازهٔ اول آن نیز بیشتر است (مدرسی قوامی، ۱۳۹۲: ۱۳۲). بنابراین، می‌توان این احتمال را در نظر گرفت که گویندگان فارسی‌زبان در میزان بازکردن مجرای دهان رفتار متفاوتی داشته‌اند که منجر به تغییرات مقادیر فرکانس سازهٔ اول در واکهٔ افتادهٔ /a/ شده است. در این پژوهش برای گویندگان فارسی‌زبان واکهٔ /i/ نسبت به سایر واکه‌ها اطلاعات فردویژهٔ اندکی نشان داد. این نتیجه برخلاف دستاورد رز (2007) است که در آن واکهٔ /i/ در گویندگان مرد استرالیایی‌زبان سودمند گزارش شده بود. با مقایسهٔ نتایج پژوهش‌های پیشین می‌توان ابراز داشت که ویژگی‌های زبان‌ویژه نیز می‌توانند در مقادیر پارامترهای آکوستیکی تأثیرگذار باشند و همان طور که کینوشیتا (2001) نیز اذعان می‌دارد هر پارامتری که در یک زبان فردویژه باشد لزوماً در زبان‌های دیگر فردویژه نیست.

طبق نتایج، اگر فرکانس سازه‌ها را برای همهٔ واکه‌ها در نظر بگیریم فرکانس سازه‌های سوم و چهارم نسبت به فرکانس سازه‌های اول و دوم قدرت تمایزدهندگی بیشتری داشته‌اند. سازه‌های اول و دوم در تشخیص زبانی واکه‌ها اندازه‌گیری می‌شوند و سرنخ‌های درکی مناسبی در تعیین کیفیت واکه هستند (Ladefoged, 2006)، در حالی که سازه‌های بالاتر نسبت به تغییرات فیزیولوژیکی شکل دستگاه گفتار و جایگاه اندام‌های گفتار حساس‌اند و این موجب می‌شود که تغییرات فردویژهٔ بیشتری نسبت به سازه‌های پایین‌تر نشان بدهند (McDougall, 2004). قدرت تمایزدهندگی فرکانس سازه‌های بالاتر به‌ویژه فرکانس سازهٔ سوم در پژوهش‌های پیشین (Asadi et.al, 2018 b ; Gold et al., 2013) نیز تأیید شده است. به نظر می‌رسد ارتباط فرکانس سازهٔ سوم با گردشدگی لب‌ها (West, 1999) نیز می‌تواند یکی از دلایل قدرت تمایزدهندگی این سازه باشد. افزون بر آن در پاره‌ای از پژوهش‌ها نیز به ارتباط میان فرکانس سازهٔ سوم با کیفیت صدا اشاره شده است (Klatt and Klatt, 1990 ; Gold et al., 2013) که این عامل نیز خود می‌تواند سبب تغییر در سیگنال‌های آکوستیکی گفتار شود.

تحلیل پارامترهای آکوستیکی منتخب در داده‌های آوایی موردبررسی نشان داد که گویندگان در دو جلسهٔ جداگانه از ضبط داده‌ها ثابت عمل کرده‌اند. این یافته بیانگر این است که پارامترهای موردبررسی نسبت به عامل گذر زمان به‌عنوان یک منبع تغییر درون-گوینده مقاوم بوده‌اند. براساس معیارهایی که برای تعیین پارامترهای مناسب برای تشخیص هویت گوینده ارائه شده است

http://dx.doi.org/10.22108/jrl.2021.128697.1577

پارامترهایی سودمند تلقی می‌شوند که تا حد امکان بین گویندگان متفاوت باشند و در عین حال درون هر گویندهٔ واحد نیز ثابت باشند (Nolan, 1983 ; Wolf, 1972). به سخن دیگر، هرچه تغییرات بین-گوینده بالاتر و همزمان تغییرات درون-گوینده در یک پارامتر کمتر باشد، آن پارامتر می‌تواند سرنخ مناسب‌تری در تشخیص صدای گویندگان باشد.

تحلیل ترکیبی پارامترهای موردبررسی با استفاده از تحلیل مؤلفه‌های اصلی بر روی داده‌های آوایی نشانگر همبستگی بالا میان پارامترهای فرکانس پایه و فرکانس سازهٔ اول بود. همبستگی بالا میان فرکانس پایه و فرکانس سازهٔ اول نشان می‌دهد که علی‌رغم فرض استقلال بخش‌های منبع و صافی دستگاه گفتار در نظریهٔ منبع-صافی، به نظر می‌رسد باید به نوعی تعامل و رابطهٔ متقابل میان این دو بخش و عوامل تأثیرگذار در این رابطه نگاهی دیگر داشت. برخی از پژوهش‌هایی که به رابطهٔ متقابل میان بخش‌های منبع و صافی اشاره کرده‌اند بر تأثیر فشار هوای درون دستگاه گفتار بر لرزش تارآوها اشاره کرده‌اند (Titze, 2008) و در پاره‌ای دیگر از پژوهش‌ها نشان داده شده است که حرکت حنجره به سمت پایین یا بالا بر مقادیر فرکانس سازهٔ اول تأثیر می‌گذارد (Gordon and Ladefoged, 2001). همبستگی میان فرکانس پایه و فرکانس سازهٔ اول نشان می‌دهد که این دو پارامتر حاوی اطلاعات مشابهی دربارهٔ گوینده هستند؛ یعنی فرکانس پایه و فرکانس سازهٔ اول نشانگر اطلاعات کمابیش مشابهی دربارهٔ گوینده هستند. همچنین، نتایج آزمون تحلیل عامل نشان داد که فرکانس سازه‌های دوم تا چهارم که همگی پارامترهای مربوط به صافی هستند نیز همبستگی بالایی دارند. بنابراین، در راستای کاهش ابعاد ویژگی‌ها می‌توان از میان پارامترهای دارای همبستگی بالا مشخصه‌های فردویژه‌تر را انتخاب کرد و از آن‌ها در شناسایی صدای گویندگان به‌ویژه در سیستم‌های خودکار تشخیص صدای گوینده استفاده کرد. طبق نتایج فرکانس پایه و فرکانس سازهٔ سوم حاوی بیشترین اطلاعات فردویژه بودند. پس می‌توان ابراز داشت که ترکیب این دو پارامتر که یکی مربوط به حنجره و دیگری مربوط به دستگاه گفتار است می‌تواند حاوی اطلاعات تکمیلی دربارهٔ گوینده باشد واین ترکیب می‌تواند سرنخ مناسبی در تشخیص هویت گوینده باشد.

۷. نتیجه‌گیری

پژوهش حاضر به بررسی نقش پارامترهای مربوط به بخش‌های منبع و صافی در تشخیص هویت گویندگان فارسی‌زبان پرداخته است. پنج پارامترآکوستیکی فرکانس پایه، فرکانس سازهٔ اول، فرکانس سازهٔ دوم، فرکانس سازهٔ سوم و فرکانس سازهٔ چهارم در شش واکهٔ سادهٔ زبان فارسی، یعنی /a, e, o, ɑ, i, u/، از یک پیکرهٔ آوایی فارسی شامل صدای دوازده گویندهٔ مرد استخراج شد و مورد تجزیه و تحلیل آکوستیکی قرار گرفت. نتایج نشان داد که پارامترهای موردبررسی توانایی نشان دادن تغییرات بین-گوینده را دارند اگرچه قدرت این پارامترها در تعیین تمایزات بین-گوینده با هم متفاوت بوده است. واکهٔ /a/ و پس از آن واکهٔ /e/ حاوی بیشترین اطلاعات فردویژه بودند. به عبارت دیگر، می‌توان ابراز داشت که واکه‌های افتاده بیشترین اطلاعات فردویژه و واکه‌های افراشته شامل کمترین اطلاعات فردویژه بودند. از میان پارامترهای آکوستیکی، فرکانس سازهٔ سوم نسبت به سایر پارامترها قدرت بیشتری در ایجاد تمایز میان گویندگان داشته است. همچنین، نتایج پژوهش حاکی از آن بود که فرکانس پایه با فرکانس سازهٔ اول همبستگی بالایی دارد که این موضوع نشان می‌دهد این دو پارامتر اطلاعات مشابهی دربارهٔ خصیصه‌های گوینده منتقل می‌کنند. این یافته نشان می‌دهد که علی‌رغم تأکید برخی از پژوهش‌ها بر استقلال میان پارامترهای منبع و صافی به نظر می‌رسد که نوعی رابطهٔ متقابل بین این دو بخش نیز وجود دارد. با این حال، فرکانس پایه با فرکانس

http://dx.doi.org/10.22108/jrl.2021.128697.1577

سازه‌های دوم، سوم و چهارم همبستگی نداشته است که این موضوع نشان می‌دهد پارامترهای یادشده حاوی اطلاعات متفاوتی دربارهٔ گوینده‌اند و ترکیبشان می‌تواند در تشخیص هویت گوینده سودمند باشد.

منابع

بی‌جن‌خان، محمود. (۱۳۹۲). *نظام آوایی زبان فارسی*. سمت.

علی‌نژاد، بتول؛ حسینی‌بالام، فهیمه. (۱۳۹۲). *مبانی آواشناسی آکوستیکی*. انتشارات دانشگاه اصفهان.

مدرسی قوامی، گلناز. (۱۳۹۲). *آواشناسی: بررسی علمی گفتار*. سمت.

نوربخش، ماندانا. (۱۳۹۲). *آواشناسی فیزیکی با استفاده از رایانه*. نشر علم.

Alinezahd, A & Hosseinibalam, F. (2013). *Fundamentals of Acoustic Phonetics*. University of Isfahan. [In Persian].

Alderman, T. (2005). *Forensic speaker identification: A likelihood ratio-based approach using vowel formants*. LINCOM Studies in Phonetics.

Asadi, H., Hosseini-Kivanani, N & Nourbakhsh, M. (2018 a). *Speaker-specificity in spectral moments of fricative /s/ in Persian*. TABU Dag, The Netherlands.

Asadi, H., Nourbakhsh, M, Sasani, F & Dellwo, V. (2018 b). Examining long-term formant frequency as a forensic cue for speaker identification: An experiment on Persian. In M. Nourbakhsh, H. Asadi, & M. Asiaee (Eds.), *Proceedings of the First International Conference on Laboratory Phonetics and Phonology* (pp. 21-28). Neveesh Parsi Publications.

Bijankhan, M. (2013). *Phonetic system of the Persian language*. Samt. [In Persian]

Boersma, P. & Weenink, D. (2013) *Praat: Doing phonetics by computer* (version 5.2.34). http://www.praat.org, Accessed 13 July 2013.

Fant, G. (1960). Acoustic theory of speech production. Mouton.

Gold, E., French, J.P. (2011). International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law*, *18*(2), 293-307.

Gold, E., French, J.P & Harrison, P. (2013). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. In *Proceedings of Meetings on Acoustics*, Montreal, Canada, (pp. 1-8).

Goldstein, U. (1976). Speaker-identifying features based on formant tracks. *The Journal of the Acoustical Society of America*, 59(3), 176-182.

Gordon, M. & P. Ladefoged. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, 29(4), 383–406

Hughes, V., Cardoso, A., Foulkes, P., French, P., Gully, A. & Harrison, P. (2019). The contribution of source and filter to speaker characterization. *The 28th Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA).* Istanbul, Turkey.

Hughes, V., Harrison, P., Foulkes, P., French, P., Kavanagh, C., Segundo, E.S. (2017). Mapping across feature spaces in forensic voice comparison: The contribution of auditory-based voice quality to (semi-) automatic system testing. *INTERSPEECH,* (3892-3896).

Jessen, M. & Becker, T. (2010). Long-term formant distribution as a forensic phonetic feature. *Conference of the Acoustical Society of America*, Cancun, Mexico.

Kahn, J., Audibert, J.F.B., & Rossato, S. (2011). Inter and intra-speaker variability in French: An analysis of oral vowels and its implication for automatic speaker verification. *International Congress of Phonetic Sciences (ICPhS)*, 17(pp. 1002-1005).

Kavanagh, C. M. (2012). *New consonantal acoustic parameters for forensic speaker comparison*

http://dx.doi.org/10.22108/jrl.2021.128697.1577

[Ph.D. dissertation]. University of York.
Kinoshita, K. (2001). *Testing realistic forensic speaker identification in Japanese: A likelihood ratio based approach using formants* [Ph.D. dissertation]. Australian National University.
Kinoshita, Y. (2002). Use of likelihood ratio and Bayesian approach in forensic speaker identification. In *Proceedings of the 9th Australian International Conference on Speech Science and Technology*. Melbourne, Australia, (pp. 297-302).
Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustic Society of America*, *87*(2), 820-857.
Ladefoged, P. (2006). *A course in phonetics*. Wadsworth Cengage Learning.
McDougall, K. (2004). Speaker-specific formant dynamics: an experiment on Australian English /aɪ/. *International Journal of Speech, Language and the Law*, *11*(1), 103-130.
McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: Toward a new approach using formant frequencies. *International Journal of Speech, Language and the Law*, *13*(1), 89-126.
Modarresi Ghavami, G. (2011). *Phonetics: The scientific study of speech*. Samt. [In Persian].
Nolan, F & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech Language and the Law*, *12*(2), 143-173.
Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge University Press.
Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). The DyViS database: Style controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech Language and the Law*, *16*(1), 31-57.
Nourbakhsh, M. (2013). *Acoustic phonetics using computer*. Nashre Elm. [In Persian].
Rose, P. (2002). *Forensic speaker identification*. Taylor & Francis.
Rose, P. (2007). Forensic speaker discrimination with Australian English vowel acoustics. In *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, (pp. 1817-1820).
Rose, P. & E. Winter (2010). Traditional forensic voice comparison with female formants: Gaussian mixture model and multivariate likelihood ratio analyses. In *Proceedings of the 13th Australian International Conference on Speech, Science and Technology*, pp. 42–45.
Rose, P., Osanai, T., & Kinoshita, Y. (2003). Strength of forensic speaker identification evidence: multispeaker formant- and cepstrum-based segmental discrimination with a Bayesian likelihood ratio as threshold. *Forensic Linguistics*, 10, 179-202.
Sambur, S. (1975). Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 23(2), 176-182.
San Segundo, E., Tsanas, A., Gómez-Vilda, P. (2017). Euclidean distances as measures of speaker similarity including identical twin pairs: A forensic investigation using source and filter voice characteristics. *Forensic Science International*, 270, 25-38.
Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85*(5), 2135- 2153.
Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *Journal of the Acoustical Society of America*, *123*(5), 2733–2749.
West, P. (1999). The extent of coarticulation of English liquids: an acoustic and articulatory study. *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, US, (1901-1904).
Wolf, J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America, 51*(6B), 2044-2056.

http://dx.doi.org/10.22108/jrl.2021.128697.1577

http://dx.doi.org/10.22108/jrl.2021.128697.1577