



Acoustic analysis of parameters affecting the between-speaker variability in Persian-English bilinguals¹

Homa Asadi², Maral Asiaee³, Batool Alinezhad⁴

Received: 28/04/2022

Accepted: 26/08/2022

1. INTRODUCTION

Human voices are unique, and for this reason, speakers can be identified by their voices. This shows that speech sounds contain speaker-specific information that can be reflected in the acoustic properties of speech signals. There are many individuals around the world who speak two or more languages, adding a fascinating dimension of variability to language perception and production. However, it remains unclear whether bilinguals alter their voice when switching between languages. A holistic view of bilingualism suggests that bilinguals are an integrated whole that cannot be separated into distinct parts; instead, they possess their own specific linguistic configuration (Grosjean, 1989). Moreover, languages differ in their segment inventories, rules of segmental combinations, as well as spectral and rhythmic characteristics of speech. Speaking styles can also contribute to within-speaker variability in acoustic parameters. Despite these factors, little is known about the influence of language and speaking style on within- and between-speaker vocal variability. This study aims to investigate how acoustic features, specifically long-term F0 and long-term formant frequencies (F1-F4), contribute to speaker individuality in Persian-English bilingual speakers and to what extent these features can discriminate between bilingual speakers.

2. MATERIALS AND METHODS

To examine between-speaker variability in Persian-English bilingual speakers, speech samples were extracted from The MirasVoice Speech Corpus (MVSC) (Vaheb et al., 2018b), consisting of 40 bilingual speakers (20 males, 20 females). The MVSC is a Persian-English voice dataset comprising both read and spontaneous speech materials. After preprocessing the dataset, 26 hours of speech data were

¹ The authors would like to express their appreciation for the support of Iran National Science Foundation (INSF) [Project Number =99029580]

² Assistant Professor in General Linguistics, University of Isfahan, Isfahan, Iran (corresponding author); h.asadi@fgn.ui.ac.ir

³ Ph.D. in General Linguistics, Alzahra University, Tehran, Iran; m.asiaee@alzahra.ac.ir

⁴ Associate Professor in General Linguistics, University of Isfahan, Isfahan, Iran; b.alinezhad@fgn.ui.ac.ir

selected for analysis. Long-term F0 and formant frequencies (F1-F4) were measured using a long-term method. Vowels were extracted and concatenated from the speech samples using the Vocal Toolkit in Praat (Boersma & Weenink, 2022). A script written by Raked Skarnitzl was then employed to automatically measure the long-term F0 and formant values at 5-millisecond intervals. These measurements were averaged for each speaker, resulting in one long-term value per formant and F0. Statistical analysis was conducted using R version 4.1.0 and the R package lme4 (R Core Team, 2022). A linear mixed-effect model was employed to assess the significance of between-speaker variability on the investigated acoustic parameters. Additionally, a random forest analysis was performed to classify speakers based on language, style, and selected acoustic features using the R package randomForest. Furthermore, Pearson's correlation test was conducted to examine the relationship between F0 and formant frequencies.

3. RESULTS AND DISCUSSION

Statistical analysis of the speech data revealed significant differences in the investigated acoustic parameters between Persian and English, as well as between the two speaking styles (read and spontaneous). Firstly, regarding language differences, the analysis showed distinct patterns in F0 and formant frequencies for Persian and English. Persian exhibited higher mean F0 values compared to English, indicating a higher vocal pitch in Persian speech. This finding aligns with previous research highlighting the tonal nature of Persian language, which often features pitch variations for conveying meaning and grammatical distinctions. On the other hand, English demonstrated lower mean F0 values, suggesting a relatively lower pitch range compared to Persian. Additionally, formant frequencies (F1-F4) differed significantly between the two languages. The formant frequencies, particularly F1 and F2, were found to be higher in Persian than in English. These differences can be attributed to variations in vowel articulation and phonetic structures between the two languages.

Secondly, an analysis of speaking style variations revealed notable distinctions in F0 and formant frequencies between read and spontaneous speech. In general, spontaneous speech exhibited greater F0 variability and wider formant dispersion compared to read speech. This finding suggests that speakers may modulate their voice differently in spontaneous contexts, potentially due to factors such as conversational dynamics, emotional expressiveness, or cognitive processing demands. The differences in formant frequencies between speaking styles indicate variations in vowel quality and pronunciation precision. Read speech displayed more stable formant patterns, indicating a more controlled and standardized articulation compared to spontaneous speech. These findings underscore the influence of speaking style on the acoustic properties of speech and highlight the importance of considering style variations in speaker classification and identification tasks.

Furthermore, the discriminatory power of acoustic features in distinguishing Persian-English bilingual speakers was explored using a random forest analysis. The results indicated that F0, F1, and F3 were particularly effective in differentiating between bilingual speakers in both male and female participants. The random forest analysis, a machine learning technique, demonstrated the ability of these acoustic features to classify speakers based on their language proficiency and individual vocal characteristics. This suggests that these specific parameters carry unique vocal

traits that can be utilized for accurate speaker classification. The observed weak correlation between F0 and formant frequencies further supports the notion that these parameters provide complementary information about speaker voices. Their combination in speaker identification tasks may enhance the accuracy and robustness of the classification models. Overall, the results highlight the potential of long-term F0 and formant frequencies as reliable indicators of speaker individuality in bilingual contexts, contributing to the understanding of vocal variability and its application in various domains such as forensic phonetics, voice biometrics, and sociolinguistic studies.

4. CONCLUSION

In conclusion, this study demonstrates that long-term F0 and formant frequencies are valuable parameters for capturing speaker individuality among Persian-English bilingual speakers. The analysis revealed significant differences in these acoustic features between languages and speaking styles. Language was found to have a greater impact on speaker classification compared to speaking style. The random forest analysis highlighted the effectiveness of F0, F1, and F3 in distinguishing between Persian-English bilingual speakers. The weak correlation between F0 and formant frequencies suggests that these parameters provide complementary information about speaker identity. These findings contribute to our understanding of vocal variability in bilingual speakers and have implications for forensic speaker identification and speech technology applications.

Keywords: Acoustic Phonetics, Bilingual Speaker Identification, Formant Frequency, Fundamental Frequency