Journal of Positive School Psychology http://journalppw.com
2023, Vol. 7, No. 2, 1642-1648

The Conceptual Framework of Pakistani English Speech: An Experimental Study of Pakistani English for Speech Recognition and Machine Learning Modeling

Dr. Abdul Malik Abbasi[1], Dr. Imtiaz Husain[2], Dr. Sadaf Irtaza[3]

[1]Faculty of Language and Culture Studies, Department of English, SMI University, Karachi.
[2]Department of Artificial Intelligence and Mathematical Sciences, SMI University, Karachi.
[3]Department of Media Studies, University of Balochistan, Quetta.

ABSTRACT

The primary objective of this research is to provide a comprehensive overview of the potential experimental and phonetic exponents of acoustical features of Pakistani English Speech. The Pakistani English research team is determined to document the acoustic structures of Pakistani English speech, which includes young Sindhi ESL learners' Speech, Voice Onset Time (VOT), Lexical Prominence, Vowel Quality, Pitch, Intonation, and Prosodic Features for Speech Recognition, Machine Learning, and Mathematical Modeling in Pakistani English. The research study will be conducted by selecting seven pairs of disyllabic words following the methodology of Beckman (1986) and Fry (1955, 1958). These stimulus pairs will be formed from word forms such as contract, desert, object, permit, rebel, record, and subject. Each target word will be elicited in isolation and in a semantically neutral frame sentence "I said __ this time" and will be accompanied by associated context sentences created specifically for each word. The data will be collected from forty young Sindhi ESL learners, a diverse group of university undergraduate students, native speakers of Sindhi, Urdu, Punjabi, Pashto, and Balochi. The primary study will test the hypothesis that disyllabic words have either first-syllable stress in English nouns or second-syllable stress in English verbs. The initial pool of 100 students will be recruited for this study, resulting in a total of 130 participating students who speak four different L1 across the provinces of Pakistan. In the second study, 30 undergraduate Sindhi ESL learners will be recruited to record their voice samples in an anechoic chamber to examine whether Sindhi speakers transfer their L1 negative Voice Onset Time to L2 English-voiced stops. This study will result in a total of 1080 tokens for analysis. The study will generate algorithms through a Python Coding System for speech recognition and machine learning modeling based on the acoustic datasets. We are confident that our research will make significant contributions to the field of acoustic-phonetics in Pakistani English speech for speech recognition and machine learning fields and for teaching and learning Speech Science: Experimental Phonetics for BS/MS/PhD English/Linguistics/Computer Speech Scientists/researchers in the field.

Keywords: Experimental phonetics, VOT, vowel quality, duration, acoustic formant frequency, lexical stress, pitch, algorithm, Pakistani English speech, speech recognition, and machine learning modeling

INTRODUCTION

In Study 1 we aim to investigate the universal order of English grammatical morphemes among young Sindhi ESL learners through mixed research methods. In Study 2 we aim to study the prominence on lexical level patterning in Pakistani English using the Speech Processing Tool Praat Software. Keeping in view of the modern era, where everything gets digitalized, therefore, it is time to document our languages digitally. In study 2 we will analyze the average values of Voice Onset Time (VOT) in Sindhi and Pakistani English speech, which are produced by Pakistani ESL learners. The study aims to contribute to both theoretical and applied linguistics. The stimuli are based on word-initially placed plosives, such as labial [p^h], coronal [t^h], and velar [k^h] allophones in L2

English, and aspirated labial / p^h/, retroflex / $ʈ^h$/, and velar / k^h/ consonants in Sindhi. We will design stimuli with three contrastive voicing pairs from both languages. It has been hypothesized that Pakistani ESL learners do not produce English-voiced and voiceless stops native-like word-initially. Additionally, the data is expected to reveal that Pinglish and Sindhi VOT characteristics of stops are altered as a function of the place of articulation. The algorithm will be generated through a Python Coding System based on the data.

SIGNIFICANCE OF THE STUDY

Currently, there is a growing interest in the experimental study of lexicology of speech recognition through machine learning and mathematical modeling of the experimental data sets. To this end, we propose conducting a research project within the university premises. We require laboratory facilities to collect the necessary data samples for our experiment and will collaborate with researchers to formulate algorithms for speech recognition, construct mathematical models, and develop applications through machine learning modeling. The importance of lexicological studies has increased in recent years. Utilizing modern technologies like signal processing, FFT, and machine learning algorithms can greatly improve the accuracy and precision of experimental data for lexicological studies in cognitive science and speech technology. The introduction of Industry and the fourth industrial revolution has had a significant impact on the mathematical transformation of experimental data. Thus, precise and error-free contextual words and their expressions with high accuracy are essential. We propose the experimental study of lexicological speech recognition through machine learning algorithms to develop the best way to address possible issues faced by the Information Technology (IT) industry in developing new IT tools. Our goal is to support the IT industry by presenting the best model and building a state-of-the-art mathematical model for speech recognition based on the lexical datasets to aid the IT industry and security agencies.

THE GOALS OF THE STUDY

The goals of this project are to explore the English grammatical morphemes, document the acoustic characteristics of Voice Onset Time (VOT) in Pakistani English speech, investigate English grammatical morphemes among young Pakistani ESL learners, analyze lexical stress patterns, examine the relationship between pitch, stress, and intonation, and measure phonetic properties. The data collected will be modeled using state-of-the-art Fast Fourier Transformation technology and analyzed using Speech Recognition and Machine Learning Algorithms to develop computer programs. The results obtained from this research can be used for various diagnostic purposes such as speech recognition. It will further guide the English language teachers and the students to pronounce English words native like and for the teachers to teach English through technology based applications on the data collected.

RESEARCH QUESTIONS & HYPOTHESES

STUDY 1-2: RESEARCH QUESTIONS

1. Do young Sindhi learners follow the universal order of grammatical morphemes?
2. What are the average VOT values of L1 Sindhi and L2 English plosives?
3. Are there acoustic VOT variations associated with the place of articulation of L1 and L2 plosives?
4. What are the steps needed for the generation of VOT algorithm for Sindhi and English constructed on the current acoustic datasets for speech recognition and machine learning modeling?

HYPOTHESES

1. Young Sindhi ESL learners do not follow universal order of grammatical morphemes?
2. Pakistani English speakers do not produce English VOT native like.
3. Acoustic VOT variations are associated with the place of articulation for L1 and L2 plosives.

4. Pakistani English speech algorithm will be generated differently from English native speech.

STUDY 3: RESEARCH QUESTIONS

1. What are the acoustic realizations of lexical stress in Pakistani English?
2. What are the acoustic differences in ESL speakers of English speech (duration, pitch-F0 and F1-F2)?
3. Do Pakistani English speakers speak seven disyllables contrastive pairs of words in a separate phrase and in a different sentence context correctly?
4. What are the steps in generating the algorithm for Pakistani English Speech?

HYPOTHESES

1. Phonetic Correlates of Lexical stress in Pakistani English.
2. Pakistani speakers of English do not differentiate between nouns and verbs in a disyllable words while speaking.
3. Pakistani English speakers speak seven disyllables contrastive pairs of words in separate phrases and in different sentence contexts correctly.
4. Pakistani English speech algorithm will be generated differently from English native speech.

LITERATURE REVIEW

Numerous studies have examined the acoustic indicators of lexical stress in American English. These studies include research by Beckman (1986), Bolinger (1958), Campbell and Beckman (1997), Fry (1955, 1958, 1965), Lieberman (1960, 1975), Sluijter and van Heuven (1996), and Sluijter et al. (1997). Most of these studies focused on lexical stress in English disyllabic words, where stress on the first or second syllable determined whether the word was a noun or a verb, respectively. Other studies have extensively investigated vowel quality, duration, and F0 acoustics in Pinglish and Sindhi. Abbasi (2015) suggested that Indo-Aryan languages should be analyzed for intonation contours that seem to be independent of stress, which is intriguing, as it implies that lexical stress and F0 contours are entirely separate in contrast to most stress accent languages, in which pitch accents are attached to the stressed syllables. Abbasi A. M. (2015) also noted that while voice quality has been thoroughly researched in Pinglish (Pakistani English), no study based on VOT in Pinglish (Pakistani English) for Sindhi ESL learners is currently available.

In 2015, Abbasi & Hussain conducted an experiment which found evidence that stressed vowels occur at higher frequencies and unstressed vowels at lower frequencies. The study also discovered statistically significant differences between the short and long vowel values in stressed and unstressed syllables. Stressed vowels had a greater duration and stop closure compared to unstressed vowels. Additionally, F1-F2 and F0 values were higher in stressed and lower in unstressed vowels, indicating phonetic correlates of lexical stress in Sindhi. These findings provide strong evidence that Sindhi is a stress accent language. Abbasi & Hussain argue that Sindhi is unique in that it is completely orthogonal to F0 contours, unlike most stress languages where pitch accents are docked on stressed syllables. In Sindhi, pitch accent rises from the first syllable in disyllabic words, regardless of syllable weight, and is followed by a fall at the end of the word. The study concludes that Sindhi behaves like a stress accent language. According to Abbasi (2018; 2020), women have smaller vocal tracts compared to men, resulting in higher formant frequencies. Abbasi (2018) also found that the average duration of stressed English vowels was 75 milliseconds, while unstressed syllables were 66 milliseconds. The mean difference between stressed and unstressed was 40 milliseconds for long vowels in English and 9 milliseconds for Sindhi vowels. Stressed long vowels and English vowels had higher F1 and F2 values, while unstressed long and English vowels had

lower F1 and F2 values, as explained by Abbasi (2017). Abbasi and Hussain (2015) discovered strong evidence of modification of all phonetic exponents of stress, indicating that Sindhi is a stress accent language. According to Ladefoged (2011), most English speakers do not use voicing during the closure of voiced stops in silence initial position or after a voiceless sound, such as in "bag" /bæg/ or "brass band" /bæs bænd/. Sindhi has a complete system of aspirated and unaspirated plosives as different phonemes, including voiceless plosives /p/, /pʰ/, /t/, /tʰ/, /k/, /kʰ/, and voiced plosives /b/, /bʰ/, /d/, /dʰ/, /g/, /gʰ/, as pointed out by Cole J. (2005). It is important to conduct an acoustic study on Sindhi Voice Onset Time and Pinglish VOT produced by adult Sindhi ESL learners coupled with Pakistani English word stress patterning.

METHODOLOGY

FOR STUDY 1-2 We will use mixed research methods that involves 40 essays written by *(N=40)* young Sindhi ESL learners to investigate the acquisition of English grammatical morphemes. We will recruit 30 undergraduate students in second study the through acoustic recording of voices samples; who are native Sindhi speakers (18 males and 12 females) and are learning English as their second language. These students will be between the ages of 17 and 25 and will have had around ten years of exposure to learning and speaking English throughout their academic careers. We will record their voice samples in an anechoic chamber at FM Radio Studio SMI University in Karachi, which is equipped with the latest instruments such as an audio mixer, condenser MICS, monitor speaker, Mic stands, computer system, mic filter, headphone sound card, and monitor wood stand. The participating students' level of English proficiency will be determined by their English language teacher-author as the same for all students.

FOR STUDY 3 we will be using a quantitative method. Our subjects will be 100 undergraduate students from SMIU in Karachi and out if required. We will look for fluent Pakistani English speakers, with 20 participants in each L1 speaker's category. We will be selecting those who have a higher level of proficiency in both speaking and understanding English, with ages ranging from 17 to 27 years and no speech problems. For the perceptual study, we will be providing a list of words to the participants and asking them to mark the stressed syllables. Meanwhile, for the acoustic study, we will be analyzing three parameters: duration, pitch-F0, and F1-F2. We will have a total of 100 participating subjects, with 20 participants from each language category (L1 Sindh, L1 Urdu, L1 Punjabi, L1 Pashto, and L1 Balochi). The statistical results will be analyzed using SPSS, and we will document Pakistani English lexical stress patterning acoustically based on the data. From this, we will generate speech recognition and machine learning algorithms.

STATISTICAL DATA ANALYSIS

We will conduct a chi-square test of independence to investigate the relationship between young Sindhi ESL learners and their ability to follow the universal order of grammatical morphemes. We will analyze the impact of vowels on preceding stops in Sindhi and Pakistani English using a repeated measures analysis of variance (RM ANOVA) to determine if the effect is significant or not ($p > .05$). Repeated measures will be used to obtain the Voice Onset Time (VOT) for both male and female Sindhi and English stops, and the RM ANOVA will be used to analyze the effects on VOTs. A non-parametric test (Wald Chi-square) will be used to measure the overall effect of place and aspiration contrast. The algorithm coded in Python will be utilized to analyze the burst onset on the voice samples datasets for both L1 and L2, following the conceptual framework based on iterations.

FOR STUDY 2 a separate t-test will be run on the means of stressed and unstressed English vowels for Durational Values. For F1 and F2 Analysis, paired sample t-tests will be used to compare the F1 and F2 means of stressed and unstressed vocalic sounds. Paired sample t-tests for F0 Analysis will be conducted to test the effects of stress on F0. Stop Closure will be analyzed by running a paired samples t-test on the means of stop closures to compare the stressed and unstressed closure duration of onset stops.

ACOUSTIC ANALYSIS

The duration of stressed and unstressed vowels will be manually measured by visually inspecting the wideband spectrographic display on the computer screen. The start and end points of the target vowels will be measured in milliseconds on the spectrographic displays. The duration of closure will be measured in milliseconds from the end of the preceding vowel to the beginning of the release burst. The speech processing tool Praat (Boersma & Weenink, 2023) will be used to measure both the beginning and ending points. Additionally, we will manually take formant measures for the lowest two formants, F1 and F2, based on formant tracks located at the midpoint of the target vowel in stressed and unstressed tokens. If we detect any differences between the tracks and the formant band in the spectrogram, we will visually inspect the formants using a wideband spectrographic display on the computer screen. To extract the pitch contours (F0), we will manually use the Praat autocorrelation method. We will measure each target vowel's midpoint visually in stressed and unstressed tokens. These acoustic parameters will be measured in Hertz.

CONCLUSION

The conceptual framework of Pakistani English Speech provides an overview of the potential research work that will present a comprehensive study on Pakistani English (Pakistani English) speech. The study will utilize datasets that support speech recognition for machine learning modeling, leading to compelling conclusions that will serve as a valuable resource for future speech scientists to advance their research. The seminal research findings hold immense significance and are essential for researchers and practitioners in the field of speech science to make progress.

ACKNOWLEDGEMENT

I extend my sincere thanks to the Sindh Higher Education Commission for sponsoring my research project on the Experimental Study of Pakistani English Speech. Professor Jennifer Cole from Northwestern University in IL, USA, provided invaluable comments on my initial draft, which helped me significantly enhance the quality and insight of this research work. I am also grateful to the Project P2 for bringing Artificial Intelligence Assistance (Speech recognition/machine learning) to the project and Research Associates for their exceptional support with the recordings and data analysis process, and to the participating speakers across Pakistan who recorded their voice samples. Their contributions have been crucial in the successful completion of the data recording process. I am also thankful to the unwavering support and assistance provided by the Office of Research, Innovation & Commercialization (ORIC) SMI University, I am confident that the findings of this research will make a significant impact in the academic and business communities.

REFERENCES

1. Abbasi, A. M, Channa, M. A. (2020). Phonetics and cognitive linguistics in Pakistani English (Pinglish). LINCOM Europa. Germany.
2. Abbasi, A. M. (2017). The stress pattern of Sindhi and English. LINCOM Europa. Germany.
3. Abbasi, A. M. (2010). The production of English consonants. Germany. Printed in UK, USA, VDM Verlag Dr. Muller.

4. Abbasi, A. M., Channa, M. A., Husain, I., & Khan, A. (2022, August). Temporal Patterns of Voice Onset Time of English-Sindhi Stops. New York , USA. Retrieved from Research Square [Prepint Version]: https://www.researchsquare.com/article/rs-1744637/v2
5. Abbasi, A, M; Memon, M, A; Channa, M, A; John, S; Irtaza, A, Kumar, K. (2018) Acoustic Characteristics of Pakistani English Vowel Sounds. International Journal of English Linguistics, 8
6. Abbasi, A, M; Memon, M, A; Channa, M, A; John, S; Anwar, R. (2018). An Acoustic Investigation of Pakistani and American English Vowels. International Journal of English Linguistics, 8
7. Abbasi, A, M; Memon, M, A; Channa, M, A; John, S. (2018). Awareness of L2 American English word stress in English for academic purposes: Implications for teaching speakers of Indo-Aryan languages. International Journal of English Linguistics, 8 (3), 101-107, http://doi.org/10.5539/ijel.v8n3p101.
8. Abbasi, A. M., Channa, M. A., Kakepoto, I., Ali, R., & Mehmood, M. (2017). A perceptual study of phonological variations in Pakistani English. Toronto. International Journal of English Linguistics, 8(2), 91-100, http://doi.org/10.5539/ijel.v8n2p92 .
9. Abbasi, A. M., & Hussain, S. (2015a). Phonetic analysis of lexical stress in Sindhi. Sindh University Research Journal-SURJ, 47(4), 749-756.
10. Abbasi, A. M., & Hussain, S. (2015b). The role of pitch between stress and intonation in Sindhi. Annual Research Journal of English Language Forum, 17, 09-24.
11. Abbasi, A. M., & Hussain, S. (2012). Syllable structure and syllabification in Sindhi-English loanwords. International Journal of Researchers, 1(4), 120-134.
12. Abbasi, A. M, & Kimball, A. (2014). Word stress in Sindhi and English: Implications for learners of English. Poster presentation in Sixth Annual Second Language Acquisition and Teacher Education (SLATE) Graduate Research Symposium. Department of linguistics and foreign languages, University of Illinois, Urbana, Champaign USA.
13. Archibald, J. (1997). The acquisition of English stress by speakers of non-accentual languages: lexical storage versus computation of stress. Linguistics,35(1),167-181.http://doi.org/10.1515/ling.1997.35.1.167
14. Beckman, M. (1986). Stress and non-stress accent. Dordrecht, Netherlands: Foris Publications. https://doi.org/10.1515/9783110874020
15. Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. Phonology Yearbook, 3, 255- 309.
16. Beckman, M. & G. Ayers Elam. (1993). Guidelines for ToBI labelling. Columbus: Ohio State University Research Foundation. Retrieved from http://docplayer.net/8594926-Guidelines-for-tobi-labelling-version-3-0-march-1997-by-mary-e-beckman-and-gayle-ayers-elam.html
17. Baloch, N. A. (2005). New comprehensive Sindhi dictionary (1st ed.). [Jamia Sindhi Lughat] (Vols. 1-111). Hyderabad, Pakistan: Sindhi Language Authority.
18. Bian, F. (2013). The Influence of Chinese Stress on English Pronunciation Teaching and Learning. English Language Teaching, 6(11). http://doi.org/10.5539/elt.v611p199
19. Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer (Version version 6.3.18) [Computer software]. Retrieved from http://www.praat.org

20. Bolinger, D. (1986). Two kinds of vowels, two kinds of rhythm. Ms., distributed by Indiana University Linguistics Club, Bloomington, IN.
21. Coxhead, A. (2000). A New Academic Word List. TESOL Quarterly, 34(2), 213-238. http://doi.org/10.2307/3587951
22. Crosswhite, K. (2001). Stress Placement in Dihovo Macedonian. Ms., U. of Rochester.
23. Derwing, T. M., & Rossitor, M. J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. Applied Language Learning.
24. Flemming, E. (2005). Auditory Representations in Phonology. Routledge, New York.
25. Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America, 27, 765–8.
26. Fry, D. (1958). Experiments in the perception of stress. Language and Speech, 1, 120–52.
27. Gordon, M. (2010). Disentangling stress and pitch accent: A typology of prominence at different prosodic levels. Paper presented at the university of Connecticut workshop on stress and accent, Storrs, CT.
28. Gordon, M. (2004). A phonological and phonetic study of word-level stress in Chickasaw. International Journal of American Linguistics, 70, 1-32.
29. Gussenhoven, C., & Jacobs, H. (2003). Understanding phonology (1sted.). London, UK: Hodder Arnold.
30. Hahn, L. D. (2004). Primary Stress and Intelligibility: Research to Motivate the Teaching of Suprasegmentals. TESOL Quarterly, 38(2), 201-223. http://doi.org/10.2307/3588378
31. Hammond, M. (1999). The phonology of English: A prosodic optimality-theoretic approach. Oxford: Oxford University Press, UK.
32. Hussain, S. (2010). Phonetic correlates of lexical stress in Urdu. (Doctoral dissertation, Northwest University USA). National Language Authority, Islamabad, Pakistan.
33. Hualde, J. (1991). Basque phonology. London, UK: Routledge. Hualde, J. (1999). Basque accentuation. In van der Hulst, H. (Ed.), Word prosodic systems in the languages of Europe (pp.947-993). Berlin, Germany: Mouton de Gruyter.
34. Jatoi, N. A. (1996). Linguistics and Sindhi language, (3rd ed.), Hyderabad: Institute of Sindhology, Hyderabad.
35. Keerio, A. (2010). Acoustic analysis of Sindhi speech pre-curser for an ASR system (Unpublished doctoral thesis). University of Sussex, Brighton, UK.
36. Kleber, F., & Klipphahn, N. (2006). An acoustic investigation of secondary stress in German.
37. Ladd, D. R. (1996). Intonational phonology. Cambridge, UK: Cambridge University Press.
38. Lindblom, B. (1963). Spectrographic study of vowel reduction. JASA 35. 1773-1781
39. Nair, R. (2001). Acoustic correlates of lexical stress in Hindi: Papers from the proceedings of SALA XVIII Round table (1sted.). Motilal Banarsidass.
40. Nihalani, P. (1995). Sindhi. Journal of the IPA, 25(2), 95-98. Ogden, R. (2009). An introduction to English phonetics. UK: Edinburgh University Press.
41. Parsons, T. W. (1987). Voice and speech processing. New York: McGraw-Hill.
42. Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. (Unpublished doctoral dissertation), MIT, Cambridge.