AUTOMATIC CLASSIFICATION OF VOWELS IN SEMI-SPONTANEOUS PAKISTANI PUNJABI

Farhat Jabeen, Moritz Wackerbarth, Petra Wagner

Bielefeld University, Germany
firstname.lastname@uni-bielefeld.de

ABSTRACT

This is the first machine learning based analysis of oral vowels produced in semi-spontaneous Pakistani Punjabi. Our investigation is based on lower frequency characteristics (f0, F1, F2, F3) and vowel duration. We use random forest classification to investigate the prediction of Punjabi vowels based on these acoustic features. For allophonic variation, we use a confusion matrix to identify the false positive and false negative vowel predictions. Our data indicate that frequency characteristics do not reliably classify different vowels of the same class (e.g., front or high). This results in varying rates of successful predictions in the classification of vowels. The analysis of feature importance shows that F1 and F2 play the most important role in the classification of Punjabi vowels, followed by duration. The influence of F3 and f0 is negligible in this regard. The misclassification of vowels in the same group highlights the phonemic vs. allophonic variation in Pakistani Punjabi.

Keywords: vowels, formants, semi-spontaneous speech, Punjabi, automatic classification

1. BACKGROUND

Punjabi is an Indo-Aryan language spoken mainly in India and Pakistan. Indian Punjabi is written in Gurumukhi script, whereas Pakistani Punjabi uses the Shahmukhi script. Notwithstanding, speakers of Pakistani and Indian Punjabi appear to understand each other fairly well. The literature review in the following subsections specifies if an analysis is based on an Indian or a Pakistani dialect of Punjabi.

1.1. Vowel inventory

Being an Abugida language, every consonant in Punjabi is associated with an inherent vowel whose quality changes depending on the associated diacritic [1]. The ten oral vowels in Punjabi are distinguished on the basis of their quality and quantity. [2] reported an inventory of three short /ɪ ʊ ə/ and seven long vowels /i e ɛ u o ɔ a/. [3] described that the tense/lax distinction in Punjabi is based on vowel quantity. However, [4] stated that vowel quality plays a more important role in Punjabi compared to vowel duration. They made a distinction between central lax vowels /ɪ ʊ ə/ and peripheral tense vowels /i e ɛ u o ɔ a/. Examples of words listed by [5] show that this difference is reflected in their distribution as peripheral vowels occur at all positions in a word, while central vowels are placed word finally in close syllables only.

For their analysis of oral vowels produced by speakers of Majhi dialect in India, [6] used word lists consisting of monosyllabic words. Their proposed inventory of vowels differed from the ones reported in existing literature. For example, [6] included the open vowel /æ/ in their inventory instead of /ɛ/ reported by [7] in their analyses of Punjabi spoken in Delhi. Moreover, [6] characterized schwa as a near-open vowel compared to the schwa reported as a central vowel by [3]. The source of these differences in vowel quality remains unclear as [6] did not provide much information about their participants.

1.2. Allophonic variation

The issue of allophonic distribution contributes to the disagreement regarding the vowel inventory of Punjabi. [2] reported that the vowel pairs /ɪ e/ and /ʊ o/ are allophones. But she offered no explanation regarding their distribution. It is interesting to note that the allophonic alteration of these vowels is mentioned for Pakistani Punjabi only and has not been reported in the analyses of Indian Punjabi.

Most of the existing analyses of Punjabi vowels are based on Indian dialects. [5] is the only acoustic analysis of Punjabi spoken in Pakistan (Lyallpuri dialect). Their data showed a closely shared acoustic space by the following set of vowels: /i ɪ e/, /ʊ o/, and /ə a/. Despite this, [5] do not refer to the allophonic variation reported by [2]. Moreover, as [5]'s analysis is based on one speaker's data, further investigation is warranted.

The acoustic analyses of Punjabi mentioned above are based on data produced by a few speakers reading out word lists. To date, there is no analysis of Punjabi vowels produced in fluent speech. Moreover, no detailed acoustic investigation of vowels in Pakistani Punjabi has been carried out so far. This study aims to fill these gaps by analysing vowels produced in fluent speech by speakers of Pakistani Punjabi (Lahori variety). We also investigate the importance of different acoustic features in the automatic classification of Punjabi vowels. The details of our analysis are as follows.

2. EXPERIMENT

2.1. Participants

Ten male[1] speakers of Punjabi were recorded for this experiment. The participants could not reliably state their dialect of Punjabi. Hence, the geographical area where they spent their early years was used as a proxy for dialect. All the participants had grown up in Greater Lahore and surrounding regions in Pakistan. One speaker's data was excluded as he audibly spoke a different dialect (Multani).

2.2. Data collection

Participants were shown an animation video (4:28 minutes) on YouTube[2]. At the end of the video, they were asked to retell the story in their own words providing as much detail as they could remember. The data was recorded using the university license of the Zoom software[3] (sampling rate: 32KHz, 16 Bit). The data consisted of 26:18 minutes of speech.

2.3. Data annotation

Every speaker's recording was divided into inter-pausal units (IPU) separated by a pause of at least 150ms. Given the lack of consensus on the inventory of diphthongs in Punjabi [4, 5], we excluded all instances of consecutive vowels as well as vowels preceded/followed by approximants. When vowels were preceded by a plosive, boundaries were placed after the burst or the end of aspiration for voiced and voiceless stops respectively. For vowels preceded by a sibilant, boundaries were marked at voice onset for voiceless and at regions of decline in high frequency energy for voiced sibilants. Vowel quality was annotated using the vowel inventory provided by [5].

Vowel boundaries were marked manually by the first (native speaker) and second (no knowledge of Punjabi) authors. The annotators reported uncertainty in the annotation of high front vowels /i ɪ e/, high back vowels /ʊ u o/, and the central low vowels /ə a/. All the annotations by the second author were verified by the first author. In case of inter-annotator disagreement, the formant values of other vowels produced by the same speaker were used to annotate a given vowel.

2.4. Data analysis

A Praat [8] script was used to extract vowel duration as well as the first three formants and f0 measured in the middle of vowels. The formant values were Bark normalized using the "PhonR" package in R [9]. As the data consists of fluent speech, we used z-transformed (by vowel) vowel duration.

Automatic classification was carried out using the random forest algorithm implemented in the Scikit-learn toolkit [10] in Python. The data was randomly divided into training (70%) and test (30%) sets. The model included z-transformed vowel duration (durZ) along with f0 and the first three Bark-normalized formant frequencies. Moreover, permutation based feature importance[4] was calculated on the held-out test set. This helped investigate the contribution of each acoustic feature to the power of the model. The results are discussed in the following section.

3. RESULTS & DISCUSSION

3.1. Frequency of vowels

In total, 3,182 oral vowels were found in our corpus of semi-spontaneous speech. Table 1 presents the frequency and percentage of each vowel. It shows that front vowels occurred more frequently (44%) in the data compared with the back vowels (18%). The high frequency of central vowels is largely due to /a/ that covers a quarter of the data. Moreover, the open-mid back vowel /ɔ/ was found very rarely in this corpus. This raises questions about the status of this vowel in Pakistani Punjabi as it might be an allophone of another back vowel. However, our corpus is based on the narration of a specific story which is not phonetically balanced. Future studies based on other texts and speech genres can shed further light on the frequency of this vowel.

3.2. Formant analysis

Fig. 1 illustrates the Bark-normalized F1 and F2 values for Punjabi oral vowels. Each vowel in the figure presents an individual data point for a given vowel type. The figure shows that Bark-normalized F1 and F2 capture the distinction between front and back vowels on one hand and between high and low vowels on the other. However, there is a high degree of overlap between front vowels /i ɪ e/ and between

i	ɪ	e	ɛ	ʊ	u	o	ɔ	ə	a
242	378	657	141	137	129	280	33	388	797
8%	12%	21%	4%	4%	4%	9%	1%	12%	25%

Table 1: Vowel frequency & percentage.

back vowels /u ʊ o ɔ/. In fact, the ellipse of /u/ is placed entirely within the ellipse for /ʊ/. The overlap is reminiscent of the allophonic realization of /ɪ e/ and /ʊ o/ reported by [2]. Furthermore, there is considerable variation in vowel quality as we find instances of schwa with low F1 and some fronted productions of /u/ and /a/. As our data consists of fluent speech, there is bound to be a high rate of coarticulation in the realization of vowels.

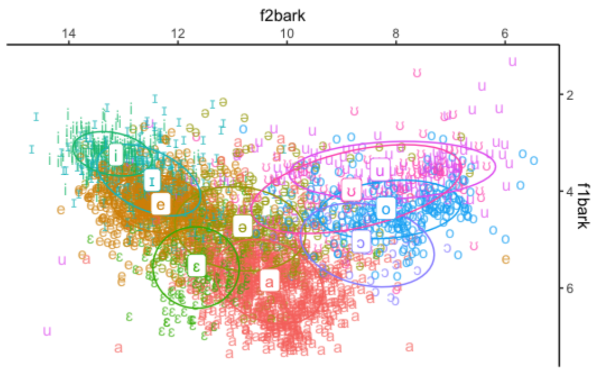

Figure 1: Bark normalized F1 and F2. The text boxes show mean formant frequencies.

Figure 1 differs only slightly from the vowel inventory offered by [5]. In our data, /e/ is realized with a higher F1 (close-mid) in comparison with [5]'s data. The most prominent difference may be observed in the quality of schwa produced as a central vowel in our data, whereas [5] had reported it to have a higher F1, in close approximation with /a/. This difference in vowel quality may be attributed to speech styles. As we have analysed fluent speech, the vowels are produced in different syllable structures (simplex and different types of complex syllables) placed at different positions in a word. In comparison, the data reported by [5] was based on vowels produced in CVC syllables in a word list. Moreover, the vowels in our data are produced in varying prosodic contexts that influence their quality. Finally, the difference between our and [5]'s results may also be attributed to dialectal variation. However, [5] claimed that the Layallpuri dialect analysed in their study is very similar to the Lahori variety spoken by our participants. Hence, the difference may be better explained by style and prosody based differences instead of dialect.

3.3. Classification report

The report for the random forest classification is given in Table 2[5]. It was computed using the true vowel labels in the held-out test set compared with the labels predicted by the model. The overall classification metrics show that the model performed fairly well. Table 2 further illustrates that the model's performance varied for different vowels. The metrics for the classification of /a/ are quite high, whereas our model made no predictions for /ɔ/ because only 12 data points were included in the model for this vowel. It could be argued that the size of the data set for each vowel included in the model is the main predictor of the classifier's performance. While its importance – unsurprisingly – cannot be denied, we argue that the performance of the classifier does not solely result from the number of data points available for each vowel in the model. Consider the classification of /ɛ/, /ʊ/, and /u/. Although the data sets for these vowels are similar in size (cf. Table 2), their precision, recall, and F1-scores differ greatly. Furthermore, the data support for schwa is high, yet the model makes better predictions for /ɛ/ than it does for schwa. In fact, the metrics are generally low for the lax vowels /ɪ ʊ ə/ compared with the peripheral ones. In future research, we aim to include other acoustic features to analyse the tense/lax distinction in Punjabi vowels.

Vowel	Precision[6]	Recall[7]	F1-score[8]	Support[9]
i	0.69	0.64	0.67	81
ɪ	0.59	0.46	0.52	126
e	0.65	0.85	0.73	178
ɛ	0.65	0.47	0.54	47
ʊ	0.76	0.33	0.46	40
u	0.71	0.71	0.71	38
o	0.67	0.76	0.71	87
ɔ	0.00	0.00	0.00	12
ə	0.45	0.42	0.43	97
a	0.83	0.90	0.86	230
Overall	0.66	0.68	0.66	

Table 2: Classification report.

The metrics given in Table 2 only provide a partial picture, as we do not know the false positives predicted by the classifier. This is illustrated in the confusion matrix shown in Table 3. The x-axis here presents the true vowel labels from the test set, while the vowel labels on the y-axis are predicted by the classifier. The confusion matrix is a good reflection of the overlap observed in F1 and F2 values plotted in Fig. 1. The front vowels are frequently confused with each other, whereas the distinction between front-back and high-low vowels is maintained more successfully. Schwa is an exception to that as it was incorrectly predicted as /a/ as well as /e/.

	i	ɪ	e	ɛ	ʊ	u	o	ɔ	ə	a
i	52	19	10							
ɪ	16	58	43	1			1		7	
e	6	8	151	2		1	1		9	
ɛ		1	5	22					7	12
ʊ		3			13	3	13		6	2
u	1		1		2	27	7			
o			1		2	6	66		6	6
ɔ						1	5	0	1	5
ə		9	20	4		1	5		41	18
a		1	2	5					15	207

Table 3: True labels are presented on x-axis and the labels predicted by the classifier are given in the first column. True positives are shown in dark cells and frequent false negatives in lighter cells.

Taken together, Table 2 and 3 illustrate the need for investigating the phoneme/allophone distinction for Punjabi vowels. As mentioned earlier, [2] claimed that the vowel pairs /ɪ e/ and /ʊ o/ are allophonic. The output of the classification algorithm indicates this further, as the recall and F1-score for /ɪ/ are low compared with these metrics for /e/. Similarly, the recall and F1-score for /ʊ/ are lower than the comparative metrics for /o/. Furthermore, the confusion matrix shows that /e/ is the most frequent false negative prediction for /ɪ/, and /ʊ/ was frequently predicted as /o/. This indicates that there is considerable overlap in the formant frequencies of these vowel pairs. Another overlap observed in our data is between /ə e/ and /ə a/. It's intriguing that schwa is misclassified with either a close-mid or an open vowel. This aspect of schwa in Punjabi has never been reported before. The differing quality of schwa is not surprising as it is considered the "default" vowel in Punjabi. It is not written orthographically [11] and is claimed to be inherently associated with every consonant. [1] reported that the quality of this inherent vowel may vary. Our data provides first acoustic evidence of this change in the quality of schwa in Pakistani Punjabi. Further research is needed to investigate this variation.

Fig. 2 illustrates the importance of different acoustic features when it comes to the classification of vowels in our data. It shows that Bark-normalized F1 and F2 play the most important role followed by z-normalized vowel duration. The lower importance of duration is not surprising as it is influenced by factors such as syllable type and structure, position in a word, and lexical stress [12]. As our data comprises fluent speech, we can not control for these variables. For future analyses, we aim to include syllable and stress level information to investigate the vowel quantity in Punjabi. As the decrease in mean accuracy is also affected by speakers, it is indicative of speaker-based variation in the classification of these vowels.

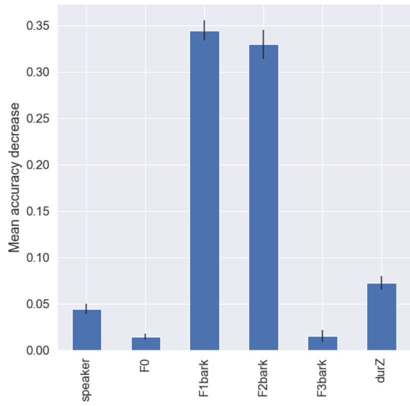

Figure 2: Importance of acoustic features. Bigger decrease in accuracy indicates higher importance.

4. CONCLUSION & FUTURE WORK

Overall, our results confirm the vowel inventory reported by [2, 5]. However, the output of automatic classification indicates that some vowels, hitherto considered as phonemes, may be allophonic in Punjabi. This highlights the importance of using machine learning algorithms to investigate acoustic phenomena. In future, we plan to conduct a perception experiment regarding the identification and discrimination of vowels in Punjabi. Furthermore, we have shown that F1 and F2 play the most important role in the classification of Punjabi vowels. However, these acoustic features led to poor predictions for the automatic classification of lax vowels. Future research should include other variables to analyse the tense/lax distinction in Punjabi vowels.

This is the first detailed analysis of vowels in Pakistani Punjabi. Despite being spoken by millions, research on this variety is lacking. Our methods and findings can be used to inform acoustic analyses of other South Asian languages.

ACKNOWLEDGEMENT

Many thanks go to Fritz Seebauer for the informed discussions on the selection and interpretation of the classification algorithm.

5. REFERENCES

[1] V. Goyal and G. S. Lehal, "Comparative study of Hindi and Punjabi language scripts," *Nepalese Linguistics*, vol. 23, pp. 67–82, 2008.

[2] N. Karamat, "Phonemic inventory of Punjabi," 2012, retrieved from Center for Language Engineering and Technology: http://www.cle.org.pk/information/people/nayyarakaramat.html.

[3] S. Lata, P. Verma, and S. Kaur, "Acoustic characteristics of schwa vowel in Punjabi," in *Proceedings of 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages*, Gurugram, India, 2018.

[4] O. Maxwell and J. Fletcher, "Acoustic and durational properties of Indian English vowels," *World Englishes*, vol. 28, no. 1, pp. 52–69, 2009.

[5] Q. Hussain, M. Proctor, M. Harvey, and K. Demuth, "Punjabi (Lyallpuri variety)," *Journal of the International Phonetic Association*, vol. 50, no. 2, pp. 282–297, 2020.

[6] P. Singh and K. Dutta, "Formant analysis of Punjabi non-nasalized vowel phonemes," in *Proceedings of International Conference on Computational Intelligence and Communication Systems*, 2011.

[7] V. Narang and D. Misra, "Redefining acoustic space in language contact situation: Case of Hindi and Punjabi in Delhi," *Interdisciplinary Journal of Linguistics*, vol. 5, pp. 13–24, 2012.

[8] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program, [v. 6.0.56]," 2013, available at http://www.praat.org/ [retrieved 20.11.2017].

[9] R Core Team, *R: A language and environment for statistical computing [v. 4.2.0]*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] P. Singh and G. S. Lehal, "A rule based schwa deletion algorithm for Punjabi TTS system," in *Proceddings of International Conference on Information Systems for Indian Languages*. Springer, 2011, pp. 98–103.

[12] S. G. Nooteboom and G. J. N. Doodeman, "Production and perception of vowel length in spoken sentences," *Journal of the Acoustical Society of America*, vol. 67, no. 1, pp. 276–287, 1980.

[13] L. Yu and N. Zhou, "Survey of imbalanced data methodologies," 2021, preprint at arXiv.

---

[1] For our preliminary study, we decided to analyse only male speakers' data. This allowed us to compare our findings with those of [5] who reported data from one male speaker. We plan to include female speakers in future works.

[2] https://www.youtube.com/watch?v=J_t2Z6_fuvU

[3] https://www.zoom.us/

[4] For details on permutation based feature importance, see https://scikit-learn.org/stable/modules/permutation_importance.html##permutation-importance

[5] As the vowel frequencies in our data are highly skewed, this is reflected in the number of data points for each vowel (support) included in the model. We decided against using the under/oversampling of vowels in our data as [13] has shown that the F-score for random forest classification is not affected significantly by imbalanced categories.

[6] Ratio of true positives & false positives

[7] Ratio of true positives & false negatives

[8] Weighted harmonic mean of precision & recall

[9] Number of observations for each vowel type