himalayan linguistics

A free refereed web journal and archive devoted to the study of the languages of the Himalayas

Himalayan Linguistics

Spoken and sung vowels produced by bilingual Nepali speakers: A brief comparison

Arnav Darnal

Independent Scholar

ABSTRACT

Speech and singing both make use of the same vocal apparatus, but studies analyzing the formant frequencies of spoken and sung vowels produced by the same subject generally show a difference in vowel quality as a result of articulatory modifications. Though such modifications may be codified and systematized in traditional musical styles, which place special emphasis on pedagogy, they appear more arbitrary in contemporary genres, which are usually passed down from mentors to students as aural traditions. While multiple studies have been conducted on the effects of singing on vowel space in various languages, this study is the first of its kind to take a look at such effects with reference to Nepalese pop rock. Since this study deals with bilingual speakers, the spoken vowels here have been compared with their sung counterparts only after establishing some deviations from those produced by monolingual speakers as referenced in previous phonetic studies of the language. This study elucidates the variation in Nepali vowels while speaking versus singing and attempts to derive an orderly, albeit preliminary, pattern of articulatory modifications that must have led to such variations.

KEYWORDS

singing, speech, music, vowels, formants, articulatory modifications, bilingualism, Nepalese pop rock

This is a contribution from *Himalayan Linguistics, Vol. 23(1): 5-11.*
ISSN 1544-7502

Himalayan Linguistics, Vol. 23(1). © Himalayan Linguistics 2024
ISSN 1544-7502

Spoken and sung vowels produced by bilingual Nepali speakers: A brief comparison

Arnav Darnal
Independent Scholar

1 Introduction

Though speech and singing both utilize the same vocal apparatus, studies analyzing the formant frequencies of spoken and sung vowels produced by the same subject generally show a difference in vowel quality as a result of articulatory modifications (Bradley 2010; Sundberg 1970). In various classical musical styles throughout Eurasia, theoretical learning is considered to be as important as learning through practice; as a result, such articulatory modifications may be codified and systematized. In operatic singing, for instance, *aggiustamento*—the adjustment of vowel resonances at different pitches—as employed by sopranos, involves the lowering of front vowels and the raising of back vowels such that /i/ is realized as [ɪ] while /ʊ/ is realized as [u] (Yu Chan & Do 2021). In contrast, contemporary musical genres are often passed down as aural traditions, in which younger singers try to replicate their older and more experienced mentors, with limited emphasis on pedagogy. Indeed, in Nepal, contemporary rock bands have been found to foster suitable environments for peer learning such that band members learn from each other and experiment with different techniques to develop their respective personal styles (Shrestha 2018). As a result, articulatory modifications in such genres tend to be more arbitrary and less systematic, varying even within the same subject on different occasions. Nevertheless, this study aims to compare the acoustic components of spoken and sung vowels produced by bilingual speakers of Nepali and Nepal Bhasa—the group that has produced a sizeable number of pop rock musicians since the inception of the genre in Nepal—and postulate the articulatory modifications that might have led to such variations in an attempt to derive an orderly, albeit preliminary, pattern thereof.

2 Methodology

For the purpose of this study, the lead vocalists of five Nepali-language pop rock bands were selected primarily on the basis of their backgrounds in bilingual Nepali and Nepal Bhasa-speaking Newar families inside Kathmandu Valley, and partially on the basis of their streaming numbers in various audio streaming platforms online. For each singer, three of their songs and two of their interviews were chosen at random and downloaded from YouTube as WAV files with a sample rate of 44.1 kHz. Backing tracks and vocals were isolated and removed such that only the phones

produced by the lead singers remained. The names of the singers, the pop rock bands they form part of, and the songs analyzed herein are listed in Table 1.

Singers	**Bands**	**Songs**
Divesh Mulmi	Cobweb	"Baadal Paari"
		"Mercedes Benz"
		"Zindagi Yo"
Mukti Shakya	Mukti & Revival	"Dal Bhat Ko Khacho"
		"Dalli Resham"
		"Ghar Jam"
Niran Shahi	Anuprastha	"Din"
		"Guff"
		"Sukha Ra Dukha"
Sarun Tamrakar	The Uglyz	"Dherai Aasu Thorai Haaso"
		"Khate"
		"Saathi (Saath Dinchu)"
Shirish Dali	Albatross	"Adhar"
		"Awaz"
		"Khaseka Tara"

Table 1. Names of the singers, the bands they form part of, and the songs analyzed for the purpose of this study

Morphemes were then filtered out such that only syllables of the structure CV, CVC or VC were chosen for analysis. Here, V refers to any oral monophthong and C to any obstruent, bar the glottal fricative, which tends to be elided in intervocalic positions but still superimposes a trace of breathiness on the resulting vowel (Hari 1971). For each monophthong that could be considered a phoneme, 15 tokens each of their spoken and sung instances were taken from each subject, and the mean of their first three formant frequencies in variable intervals according to the duration of the downloaded audio were extracted using a non-time-normalized Praat script with the formant ceiling set to the standard value of 5 kHz for male speakers (DiCanio 2008; Boersma 2001).

Finally, these raw measurements were normalized with NORM, a web-based interface to the **vowels** R package, using the speaker-extrinsic Labov ANAE method to remove variation caused by physiological differences among speakers (Thomas & Kendall 2010). The formant frequencies of spoken and sung vowels, as normalized using the suite, were then plotted on the same graph for ease of comparison. Because this study deals with bilingual subjects, the spoken vowels herein have first been compared with those referenced in the work of Khatiwada (2009) to establish deviations from those produced by monolinguals and to set the stage for comparing them with their sung counterparts as produced by the same subject.

Due to the independent nature of this study and a consequent lack of funding, the audio clips used herein were taken off the internet with limited information about the original format in which

they were recorded or about the appliances utilized in the process. With this limitation in mind, it should be noted that the results of this study may be subject to change in case a more comprehensive data set obtained after taking a more systematic and thorough approach is used for analysis. Still, this study proves successful in achieving the target outlined in Section 1.

3 Results

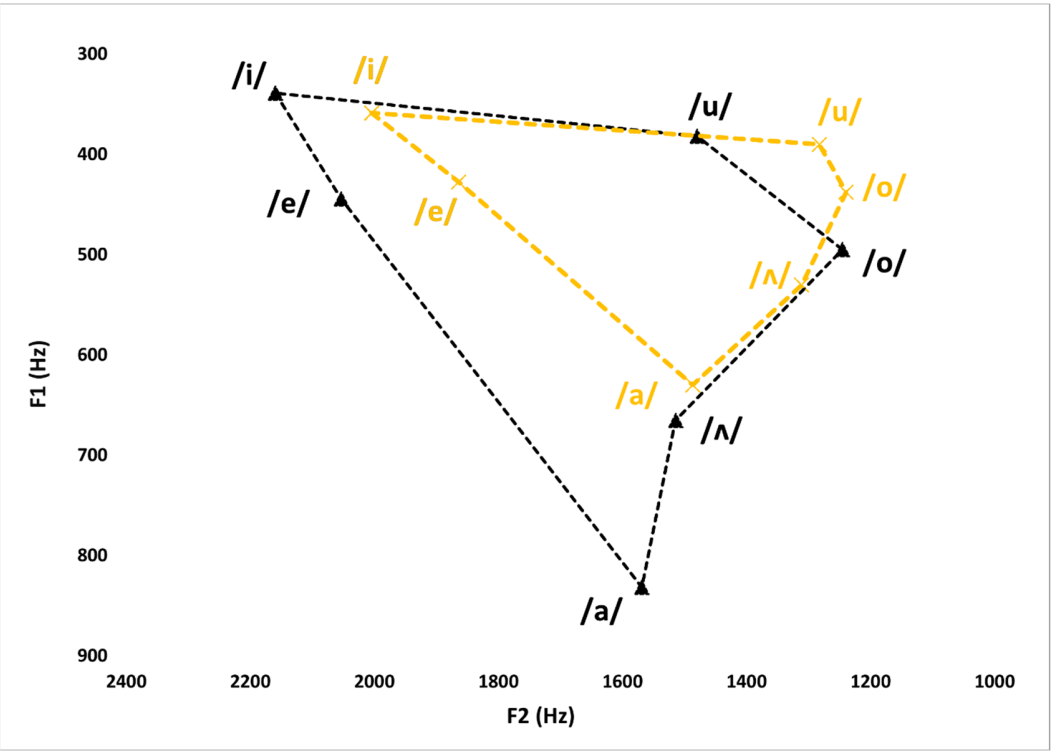

Figure 1. Vowel polygons with vowels labeled with standard phonemic symbols. Crosses: spoken; triangles: sung.

In comparison with that referenced in Khatiwada (2009: 6), the vowel polygon obtained in this study for spoken instances of vowels generally shows an increase in F1 for close vowels and a decrease for open ones. As for F2, the polygon shows a decrease for front vowels and an increase for back ones. Given that F1 is most responsive to changes in mouth openings and F2 is most responsive to changes in the size of the oral cavity, it can be concluded that the bilingual subjects tend to centralize their vowels and produce close and open vowel phonemes with more open and more closed mouths, respectively, relative to the monolingual subjects referenced in previous studies (Raphael et al. 2011).

Among the vowel phonemes analyzed herein, the open-mid back unrounded /ʌ/, which is considered the schwa phoneme in Nepali, and the open central unrounded /a/ appear to have the greatest number of allophones. The schwa phoneme, which has averaged out to a mid back unrounded [ɤ̟] or a fronted realization thereof, was variably realized by the subjects as a close-mid central unrounded [ɵ], a close-mid back unrounded [ɤ], a mid central [ə], a mid back unrounded [ɤ̟], an open-mid central unrounded [ɜ], an open-mid back unrounded [ʌ], or a near-open central [ɐ]. At

times, a decrease in F3 was observed when /ʌ/ was produced in the vicinity of labial consonants or in isolation, suggesting a rounded or compressed articulation, consistent with the findings of Khatiwada.

Though stress is non-distinctive in Nepali, words in the language are primarily trochaic, in which the first syllable carries the greatest stress, with all following odd-numbered syllables carrying weaker stress. In declarative sentences, the last word typically carries the least stress. This is a feature Nepali shares with Bengali (Khatiwada 2009; Chatterjee 1921). In this study, the subjects were found to merge /a/ with /ʌ/ in unstressed positions. In stressed syllables, the phoneme was realized as a near-open front unrounded [æ], a near-open central [ɐ], an open front unrounded [a], an open central unrounded [ä], or an open back unrounded [ɑ]. The various instances of the phoneme, as produced by the subjects in casual speech, averaged out to a near-open central [ɐ] or a raised articulation thereof.

On the contrary, the close-mid /e/ and /o/ were frequently lowered to mid [e̞] and [o̞] or an open-mid [ɛ] or [ɔ] in unstressed syllables. Albeit rarely, the first three formants of /o/ at the end of sentences were found to increase, suggesting not only a fronted and lowered articulation but also an unrounded one. That is, in rare occasions, /o/ partially merged with and surfaced as /ʌ/. This merger appeared to be particularly advanced in the verb *bhayo* (/bʱʌjo/ 'it happened' or 'it became'), which each subject realized as [βʌ̟ ~ βo̞] in their speeches at least once.

Even in stressed syllables, however, the close /i/ and /u/ were lowered to the vicinity of the near-close [ɪ] and [ʊ] in closed syllables, with /i/ sporadically surfacing as a close-mid [e]. Taking into consideration instances of the phoneme in both closed and open syllables, the vowel averaged out to become a retracted close front unrounded [i̠], articulated with the body of the tongue pulled downward into the pharynx. The close back rounded /u/, however, averaged out to the vicinity of a near-close near-back rounded [ʊ] instead of that of a close back rounded [u] as had been the case with the monolingual speakers referenced by Khatiwada. Still, [u] surfaced as one of the most common allophones of the phoneme, in addition to the close central rounded [ʉ] and the close front rounded [y].

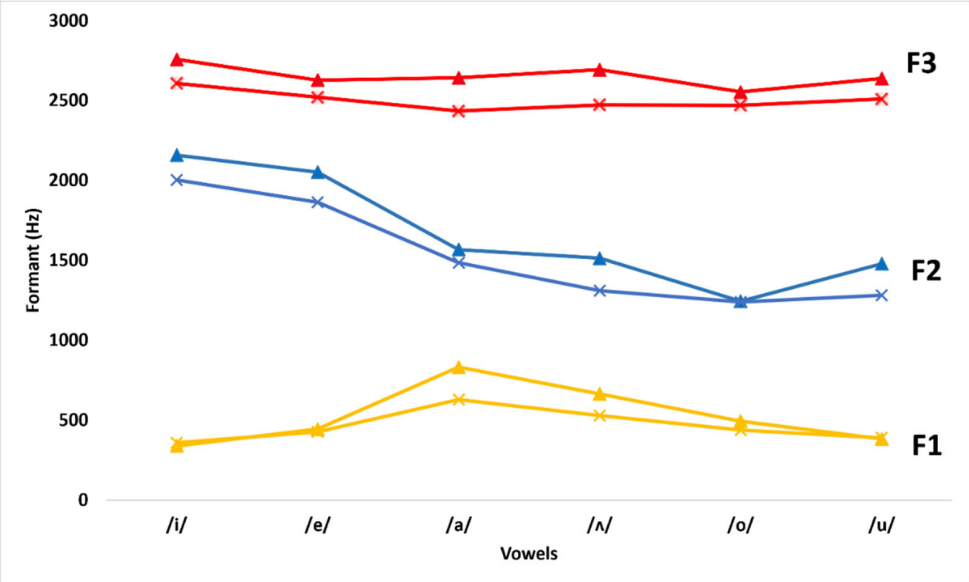

Figure 2. F-patterns of vowels. Crosses: spoken; triangles: sung

As shown in Figure 1, the vowel polygon obtained from sung vowels is asymmetrically fronted and generally covers a larger area than that obtained from spoken vowels. A closer look indicates that the subjects in this study articulated the close and the close-mid vowel phonemes with a smaller mouth opening than they did in casual speech. The close-mid back rounded /o/ appears to be an exception because it was lowered. On the contrary, the F1 values for the open-mid and the open vowel phonemes, which were already rather unstable in speech, were found to increase while singing, suggesting a wider mouth opening.

It is clear that the fronting of the sung vowel polygon is a direct result of increase in F2 values, which suggests a shortening of the vocal tract resulting from the subjects elevating their larynges and/or protruding their lips less while singing. Because the F3 values were also found to increase, as can be seen in Figure 2, a less rounded or compressed articulation seems likely. It can thus be postulated that the concomitancy of lip protrusion and larynx movement affected the articulation of vowels in this study, causing an increase in the fundamental frequency of sung vowels (Perkell 1969). The close-mid back rounded /o/ presents an exception in this regard as well because it was the only phoneme that ended up being backed on average while singing. Why this is, however, remains unclear.

The sung phonations of vowels typically tended to follow the same rules of stress and reduction as their spoken counterparts, causing /a/ and /ʌ/ to merge in unstressed positions quite often. Instances of /o/ were not found to merge with /ʌ/, though, in the tokens analyzed herein. All in all, the phonemes /i e a ʌ o u/ correspond to [i̟ e̟ ä ɐ̟ o̟ ʉ] or some variation thereof while singing.

4 Discussion

As a fusion genre, pop rock is characterized by its strong commercial appeal, with more emphasis on professional songwriting and recording craft, and less emphasis on attitude than standard rock music (James 2023). As such, pop rock typically employs different phonetic and linguistic innovations than those seen in other subgenres of music. In this study, the sung vowel polygon was found to be fronted and cover a larger area than its spoken counterpart. The sung vowels also had a greater fundamental frequency than their spoken counterparts. Analyzing and comparing the normalized formants in spoken and sung phonations, it can be surmised that the subjects yielded to the concomitancy of larynx movement and lip protrusion while singing. That is, it can be postulated that the subjects elevated their larynges while singing, causing their lips to be less protruded as a result. The close-mid back rounded /o/ raised an exception, though, because it was backed unlike all of the other vowels and lowered unlike the other close vowels. Why this is remains unclear. However, addressing the limitations presented in this study might yield more consistent and accurate results.

In comparison to other phonemes, the close back rounded /u/ doesn't stand out much in terms of the phonetic changes it goes through while singing, but its spoken instances averaging out to [ʊ] is interesting because it apparently hasn't been recorded in previous studies of Nepali. While phonetic studies of Nepal Bhasa do list a fronted close back rounded [u̟], a lowered close back rounded [ụ], a centralized close back rounded [ü] and even a close-mid back unrounded [ɤ] as allophones of /u/ in certain environments, the subjects in this study realized /u/ as [ʊ], [ʉ] or [y] without being bound by any phonetic constraints (Friedman et al. 1983). Further studies may be needed to conclude whether this realization is an influence of Nepal Bhasa or a feature of Nepali spoken in Kathmandu Valley.

The general pattern of subjects elevating their larynges while singing, causing the vowel space to become more extensive and asymmetrically fronted, is found to be the exact opposite of the pattern observed in similar studies in Swedish, Australian English, and American English (Clermont 2002; Bradley 2018). It is important to note, however, that although the genre of songs has been shown to affect the realization of vowels to a certain extent, making accurate cross-linguistic comparisons solely in a pop-rock context has been difficult due to an apparent lack of genre-specific research (Bradley 2011). This inclusion of various musical genres in an attempt to contextualize the results obtained herein on a cross-linguistic scale can be taken as one of the limitations of this study.

Some phonetic changes noted while analyzing sung phonations but not included within the scope of this study were spontaneous nasalization and denasalization of vowels, smoothing of diphthongs, gliding of monophthongs, and rhotacization of vowels. Including these features, which form the most salient parts of what is colloquially called cursive singing, and removing the addressed limitations in future works might help yield a fuller and more accurate picture of all the variations that occur in vowels produced by bilingual Nepali speakers while singing pop rock.

References

Boersma, P. 2001. "Praat, a system for doing phonetics by computer". *Glot International* 5.9/10: 341–345.

Bradley, E. D. 2010. "An investigation of the acoustic vowel space of singing". Paper presented at the 11th International Conference on Music Perception and Cognition. University of Washington, Seattle. Retrieved from https://www.researchgate.net/publication/253639193

Bradley, E. D. 2011. Song style and the acoustic vowel space of singing. Paper presented at the 2011 meeting of the Society for Music Perception and Cognition. University of Rochester, Rochester. doi:10.13140/RG.2.2.15071.46246

Bradley, E. D. 2018. "A comparison of the acoustic vowel spaces of speech and song". *Linguistic Research* 35.2: 381-395.

Chatterjee, S. K. 1921. "Bengali phonetics". *Bulletin of the School of Oriental and African Studies* 2.1: 19–20. https://doi.org/10.1017/S0041977X0010179X

Clermont, F. 2002. Systemic comparison of spoken and sung vowels in formant-frequency space. Paper presented at the 9th Australian International Conference on Speech Science and Technology (SST). Melbourne.

DiCanio, C. 2008. *Spectral Analysis / Phonation Analysis scripts.* Retrieved May 2, 2024, from Christian DiCanio: https://www.acsu.buffalo.edu/~cdicanio/scripts/Get_Formants.praat

Friedman, L. C.; Kansakar, T. R.; Tuladhar, J.; and Hale, A. 1983. "On the variants of Newari vowels: A study in phonological non-alignment". *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 27: 49-50 doi: 10.31356/silwp.vol27.05

Hari, A. M. 1971. *Conversational Nepali*. Kathmandu: Summer Institute of Linguistics and Institute of Nepal Studies, Tribhuvan University.

James, M. 2023. *What is pop rock music? With 7 top examples and history.* Retrieved May 2, 2024, from Music Industry How To: https://www.musicindustryhowto.com/what-is-pop-rock-music/

Khatiwada, R. 2009. "Nepali". *Journal of the International Phonetic Association* 39.3: 373–380. https://doi.org/10.1017/S0025100309990181

Perkell, J. S. 1969. *Physiology of speech production.* Cambridge: The MIT Press.

Raphael, L. J.; Borden, G. J.; and Harris, K. S. 2011. *Speech science primer: Physiology, acoustics, and perception of speech*. Philadephia: Lippincott Williams & Wilkins.

Shrestha, J. 2018. Peer Teaching and Learning in a Nepali Rock Band. Helsinki: Sibelius Academy, University of the Arts. Retrieved May 2, 2024, from https://taju.uniarts.fi/bitstream/handle/10024/6667/JohnShrestha.pdf

Sundberg, J. 1970. "Formant structure and articulation of spoken and sung vowels". *Folia Phoniatrica et Logopaedica* 22.1: 28–48. https://doi.org/10.1159/000263365

Thomas, E. R.; and Kendall, T. 2010. NORM: The vowel normalization and plotting suite. Retrieved May 2, 2024, from http://ncslaap.lib.ncsu.edu/tools/norm/

Yu Chan, M. P.; and Do, Y. 2021, November. "Vowel modification (Aggiustamento) in Soprano voices". *Music and Science* 4: 1-25. https://doi.org/10.1177/20592043211055168

Arnav Darnal
darnalarnav.05@gmail.com