APRIL 17 2026

Acoustic evidence for an emerging binary sibilant system in Telugu

Charles Redmon

JASA Express Lett. 6, 045203 (2026)
https://doi.org/10.1121/10.0043575

JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Acoustic evidence for an emerging binary sibilant system in Telugu

Charles Redmon[a)]
Department of Language and Linguistics, University of Essex, Colchester, CO4 3SQ, United Kingdom

Abstract: This paper presents production data and acoustic modelling from 16 speakers of Telugu, showing that what has typically been described as a ternary distinction between alveolar, palatal, and retroflex sibilants is rather more accurately an emerging binary system, with the majority of speakers showing no distinction between the latter two postalveolar categories. The relative acoustic discriminability of the system (and ultimate support for merger patterns) was not, however, uniform and varied by position, vowel context, and lexical characteristics such as frequency, neighbourhood density, and the presence and number of sibilant competitors in the lexicon.

[Editor: Peggy Mok] https://doi.org/10.1121/10.0043575

Received: 2 January 2026 **Accepted:** 3 April 2026 **Published Online:** 17 April 2026

1. Introduction

1.1 Prior work on Telugu

Telugu has typically been described as exhibiting a ternary sibilant distinction between alveolar /s/, alveolo-palatal /ɕ/,[1] and retroflex /ʂ/ places of articulation (Bhaskararao and Ray, 2017), though even from early descriptions, linguists have noted variation in implementation according to formality, education, and dialect, among other social variables (Krishnamurti, 1978). Such a system represents a combination of the native Dravidian alveolar sibilant (historically derived internally from intervocalic plosive lenition) with two postalveolar sibilants borrowed from Indo-Aryan (primarily Sanskrit): one laminal and one apical and therefore retroflex (Krishnamurti, 2003).

This system was already attested in Old Telugu, though at this stage, palatals and retroflexes were quite peripheral to the system as recent borrowings from Sanskrit (Ramanarasimham, 2020), and among Dravidian languages, it remains comparatively dense, with only Toda exhibiting a similar system in the modern language. Outside of Dravidian, contrasts between multiple postalveolar sibilants have disappeared in modern Indo-Aryan languages. To take just a few key examples, Hindi has merged the palatal/postalveolar and retroflex into a single postalveolar category, while Eastern Indo-Aryan languages like Bengali have merged all three sibilants into a postalveolar. More broadly, systems of three or more sibilants are cross-linguistically rare, appearing in approximately 3% of languages in the UCLA Phonological Segment Inventory Database (UPSID; Maddieson and Precoda, 1990).

While no prior study has sought to directly address the sibilant system in Telugu, numerous linguists have commented on the subject as part of wider descriptions and surveys of historical, regional, and social variation in the phonology of Telugu. While the standard description of the system remains ternary in place of articulation (alveolar, palatal, retroflex), such as in the recent IPA Illustration of Bhaskararao and Ray (2017), a common observation in most of this work has been on the constrained and variable nature of the system when it comes to the borrowed sibilants, /ɕ/ and /ʂ/.

Krishnamurti (1957, 1978, 2003), for instance, has claimed throughout his work that while all three sibilants are present in the speech of "educated" Telugu speakers, among "uneducated" speakers, all three are commonly merged into /s/. Sjoberg (1962) observes as well a stratification in the realisation of the sibilant system; however, unlike Krishnamurti, Sjoberg argues that all three sibilants are maintained at an informal level, but that /ɕ, ʂ/ are "more constrained" in their production and often arise as /s/ in the Telangana region, and /ɕ, ʂ/ in the Coastal and Southern regions.

1.2 Typology of sibilant place contrasts

In general, phonological distinctions in sibilant place of articulation are not uncommon, though many languages only exhibit a single voiceless sibilant (45% of languages in UPSID, or 205), while a similar number exhibit two or more voiceless sibilants and thus do make a place distinction (40% in UPSID, or 182). However, among those distinguishing sibilant places, the vast majority (162 in UPSID) show a single binary distinction between a dental/alveolar and apostalveolar sibilant, where the postalveolar is largely laminal (only four languages in the UPSID set of binary distinctions involve a

[a)]Corresponding autor: c.redmon@essex.ac.uk

retroflex). There are, however, a number of languages with three sibilant places, with the most common set being that presumed for Telugu: alveolar, palatal, retroflex (nine languages). Thus, while the general density of this system is cross-linguistically rare, among languages that do exhibit ternary sibilant place contrasts, Telugu's is the expected system. Finally, it must be acknowledged that while three-way sibilant contrasts show historical trends toward simplification (to binary and unary systems) across the Subcontinent, such systems have also been shown to be relatively widespread and more complex historically in other contexts, such as in the Chinese languages (Handel, 2017).

1.3 Acoustic correlates of sibilant place of articulation

Sibilant place contrasts are among the most salient phonological distinctions, not just among fricatives, but more broadly. The alveolar-postalveolar /s–ʃ/ distinction which is present in English and many other languages is particularly well studied, and is marked by many acoustic features but primarily a difference in the concentration of energy in the spectrum (both as measured by spectral peak frequency and spectral mean/centre of gravity), where postalveolar sibilants show mid-frequency peaks around 3–4 kHz, while alveolars show concentrations much higher in the range: around 6–7 kHz (Jongman *et al.*, 2000). Other characteristics, however, such as the spectral tilt (Shadle and Mair, 1996), spectral shape (Evers *et al.*, 1998), and amplitude, have been proposed as additional features of the contrast (McMurray and Jongman, 2011).

Retroflex sibilants, on the other hand, are less well studied, though one may find detailed acoustic studies on Polish (Bukmaier and Harrington, 2016; Żygis and Padgett, 2010) and Mandarin (Lee-Kim, 2014; Li, 2017), highlighting the general lowering of spectral peak frequencies as well as the importance of formant cues, particularly F3, in the preceding vowel (in cases like Polish with fewer phonotactic constraints). Therefore, we might expect a system such as Telugu's, with two postalveolar sibilants, to depend more on contextual cues such as F3 transitions than systems which can largely be distinguished on the basis of spectral characteristics of the noise.

1.4 Present study

The present study aims to provide, for the first time, detailed acoustic data on Telugu sibilant production, as well as quantitative modelling of the classification and discrimination of sibilant place of articulation in multivariate acoustic space. Further, interactions with indexical (gender, region) and lexical (word frequency, phonological neighbourhood structure) factors are considered in accounting for production patterns and the acoustic structure of the system. To this end, native speakers of Telugu were recorded in a laboratory environment reading isolated words exhibiting each sibilant type in varying prosodic positions and vowel contexts. Further details on the materials and procedure are provided in Sec. 2.

2. Methods

2.1 Participants

Sixteen native speakers of Telugu (eight female, eight male) were recruited from the English and Foreign Languages University in Hyderabad. All were students at the university (mean age, 25.5; range, 18–36), and were primarily from the Telangana region (14/16); further, the majority of participants from Telangana grew up in and around Hyderabad (9/14). Thus, the sample is not regionally diverse but should provide a good estimate of the state of the sibilant system in Telangana and in Hyderabad in particular.

Regarding the multilingual background of participants, all but one participant was raised monolingually (Speaker M07 is a Telugu-Lambadi bilingual), though all speak English and Hindi as additional languages. Further, all participants had some experience in English-medium education (all at university, 14/16 at college, 8/16 at secondary school, and 8/16 at primary school), though there was a gender imbalance here as the vast majority of those with English-medium education before college (7/8) were women. Finally, no participant had lived outside of India for any period of their life at the time of recording. All participants were compensated Rs. 300 for their participation.

2.2 Materials and procedure

A total of 120 words containing sibilant fricatives were recorded from each participant: 60 with the target sibilant in consonant-vowel (CV) position (20 each of palatal, alveolar, and retroflex places), and 60 in vowel-consonant-vowel (VCV) position. Consonant-vowel sequences were then distributed according to vowel context (the following vowel for CV position, the preceding vowel for VCV position) in the following manner: ten items each of /aː/ context and two items each of /iː/, /eː/, /oː/, and /uː/ contexts. This distribution is broadly consistent with that found in the lexicon more broadly, though with some levelling out of non-/aː/ vowels. Each word was repeated twice (in a separate, randomised block), alongside a further 240 filler words containing no sibilant fricatives and otherwise matched for length (syllable count) and word frequency. Each word was then presented to participants on a separate slide, with the total trial sequence (including repetition) divided into four separate blocks with block order counterbalanced across participants. All recordings were made with a Shure SM-35 (Shure, Niles, IL) head-worn cardioid condenser microphone in the

Phonetics Laboratory at the English and Foreign Languages University in Hyderabad, Telangana, India, and took less than 30 min each to complete.

2.3 Analysis

All consonant and adjacent vowel sequences (CV, VCV) were annotated and segmented in Praat (Boersma and Weenink, 2015). Then, two types of acoustic measurements were extracted: consonantal [root-mean-square (RMS) amplitude (AMP); spectral peak frequency (SPF); low-frequency spectral tilt (LF TILT) and high-frequency spectral tilt (HF TILT); and spectral moments μ_1-μ_4, all taken from 40 ms Hamming windows over the consonant midpoint],[2] and vocalic (second and third formant transitions into/out of the neighbouring vowel, measured at 10% increments over the 0.1–0.5 relative-duration interval, i.e., between boundary and midpoint, in each vowel to avoid consonantal boundary effects). These measures were subsequently used both in independent statistical models of the relative overlap between the three place of articulation (POA) categories along each acoustic dimension (as well as as a function of speaker and contextual characteristics), and in machine learning models (supervised and unsupervised) of category recognition in multivariate acoustic space.

3. Results

3.1 Independent parameter discriminability

Figure 1 shows the posterior distributions of each consonantal parameter by Place of Articulation (alveolar, palatal, retroflex) and Position (VCV, CV). All estimates are obtained from Bayesian multilevel linear models of the same structure (in R formula syntax),

$$y \sim POA * Position + Vowel + (1 + POA + Position|Speaker),$$

where y is one of the eight measured consonantal parameters, *POA* is the place of articulation (alveolar, palatal, retroflex), *Vowel* is one of three vowel subclasses (front = {i, e}, back = {u, o}, low = {a}), and per-*Speaker* random intercepts and slopes for *POA* and *Position* are introduced in the final random effects term.[3] All models were fit using the `brms` package in R (Bürkner, 2017).

Among the consonantal parameters tested, nearly all distinguished alveolars from postalveolars robustly. Only the spectral shape parameters: dispersion, μ_2, and kurtosis, μ_4, showed no such distinction (palatals and retroflexes were similarly not well distinguished by these measures). Otherwise, alveolars could generally be distinguished acoustically along a variety of spectral dimensions, and even by amplitude (alveolars being relatively quieter).

Considering then the system as a whole, the most robust at distinguishing all three sibilants were the spectral mean, μ_1, and spectral skewness, μ_3, where alveolars were notably higher-frequency than the postalveolars, both intervocalically and word-initially, and palatals had a marginally higher spectral mean than retroflexes (see right-margin panels of Fig. 1 for posterior contrasts; numerical details are provided in the online repository[4]). Spectral skewness, which correlates strongly with spectral mean ($\rho = -0.872$), exhibits a similar set of relations: $A \ll P$, $A \ll R$, $P < R$. Further, much of the spectral mean effects are reflective of comparable differences in spectral peak frequency, both word-initially and intervocalically. However, in all such cases where there was an aggregate distinction between palatals and retroflexes, this only emerged in half or fewer speakers (5–8/16 showed the distinction for μ_1, μ_3, and SPF).

As for the palatal-retroflex distinction along other acoustic dimensions, the two postalveolars were generally acoustically indistinguishable, with the only parameters (outside of μ_1, μ_3, and SPF) showing a significant contrast on average being low-frequency spectral tilt and spectral dispersion in VCV position, and in both cases, the majority of speakers did not show a distinction (15 and 12, respectively). Thus in terms of consonantal acoustic measures, while the alveolar-postalveolar distinction is quite robust, the palatal-retroflex contrast is much less reliable.

Turning then to vocalic information from formant transitions, Fig. 2 shows the model-estimated F2 and F3 transitions in low-vowel /a/ contexts in CV and VC positions (this is the most common vocalic environment and one of the more robust, alongside the other back vowels /o/ and /u/, in terms of cues to consonantal place of articulation). All models of vowel formant transitions used a Generalised Additive model (GAM), with fixed effects of place of articulation (POA), cubic-spline smooths for Time and POA with 3 basis functions, ridge-penalised smooths for random effects of Speaker, and factor smooths to model separate effects of Time by Speaker/POA.[5] Separate models were run for each position (VC, CV) and vowel context /i, e, a, o, u/). All models were estimated in R with the *mgcv* package (Wood, 2011), with uncertainty in model-estimated transitions modeled with Laplace sampling of the posterior distribution.

What we see from the formant transition data is that broadly the three sibilants can be distinguished from some combination of F2 and F3 transitions, either into or out of the consonant constriction. In VC position, alveolar transitions are significantly shallower than the postalveolars in F2 (i.e., had a lower offset/locus frequency, though transitions significantly diverge between alveolars and palatals by 32% before vowel offset, and 26% pre-offset between alveolars and retroflexes),[6] but with no distinction in F3, whereas in CV position retroflexes are further distinct and intermediate in F2 locus frequency between alveolars and palatals (/ʂ/ < /ɕ/ up to 42% of the vowel; /s/ < /ʂ/ up to 19%; /ʂ/ < /ɕ/

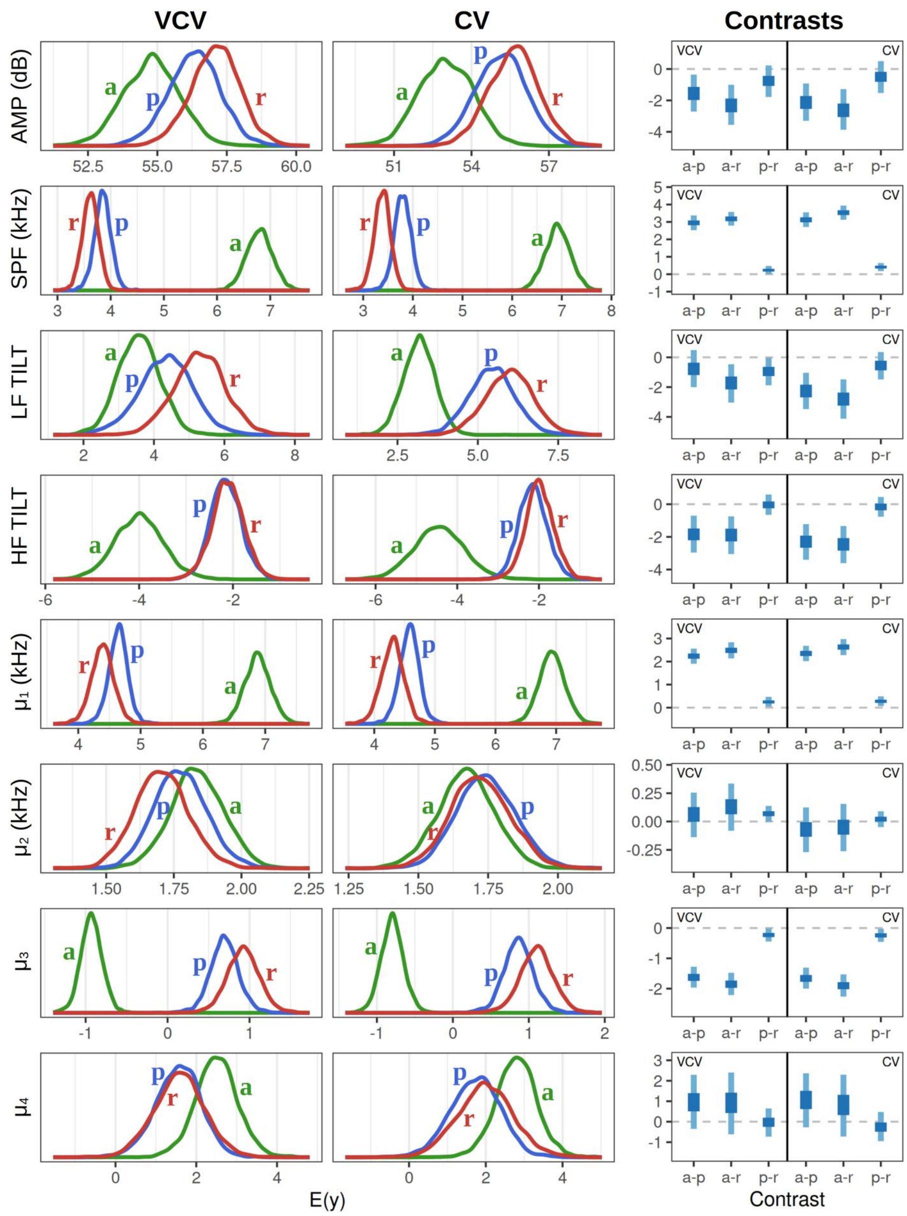

Fig. 1. Posterior marginal means for each parameter in VCV and CV positions for each place of articulation (alveolar, green; palatal, blue; retroflex, red). Corresponding contrast distributions are shown in the panels on the right.

into at least vowel midpoint). Finally, F3 transitions in CV position are not distinctive for any of the three contrasts (at least based on 95% posterior quantile intervals).

Turning then to the generalisability of these results to other vowel contexts,[7] there is a notable distinction between front and back vowels, with front vowels showing no significant distinction between palatals and retroflexes, and only marginal distinctions between alveolars and postalveolars, with differences emerging only in the final/initial 20%–30% of the vowel. These results are consistent with the observations of Li (2017), among others, that sibilants are acoustically less distinctive—acoustically, perceptually, and typologically—in front vowel contexts, particularly high-front.

Formant transitions into/out of back vowels, on the other hand, are generally robust. In the /o/-context palatals show a relatively flatter F2 trajectory than retroflexes (in VC position, higher between midpoint and 61%, lower after 87%), with no distinction in F3. However, in the /u/ context, in addition to the lower and more dynamic F2 transition in VC position, retroflexes have a significantly lower F3 transition than palatals (in the middle of the preceding vowel, 50%–79%, and within the first 23% of the following vowel). Alveolars are generally distinct in their formant transitions from postalveolars in both /o/ and /u/ contexts, with lower F2 transitions in VC and CV positions, particularly relative to

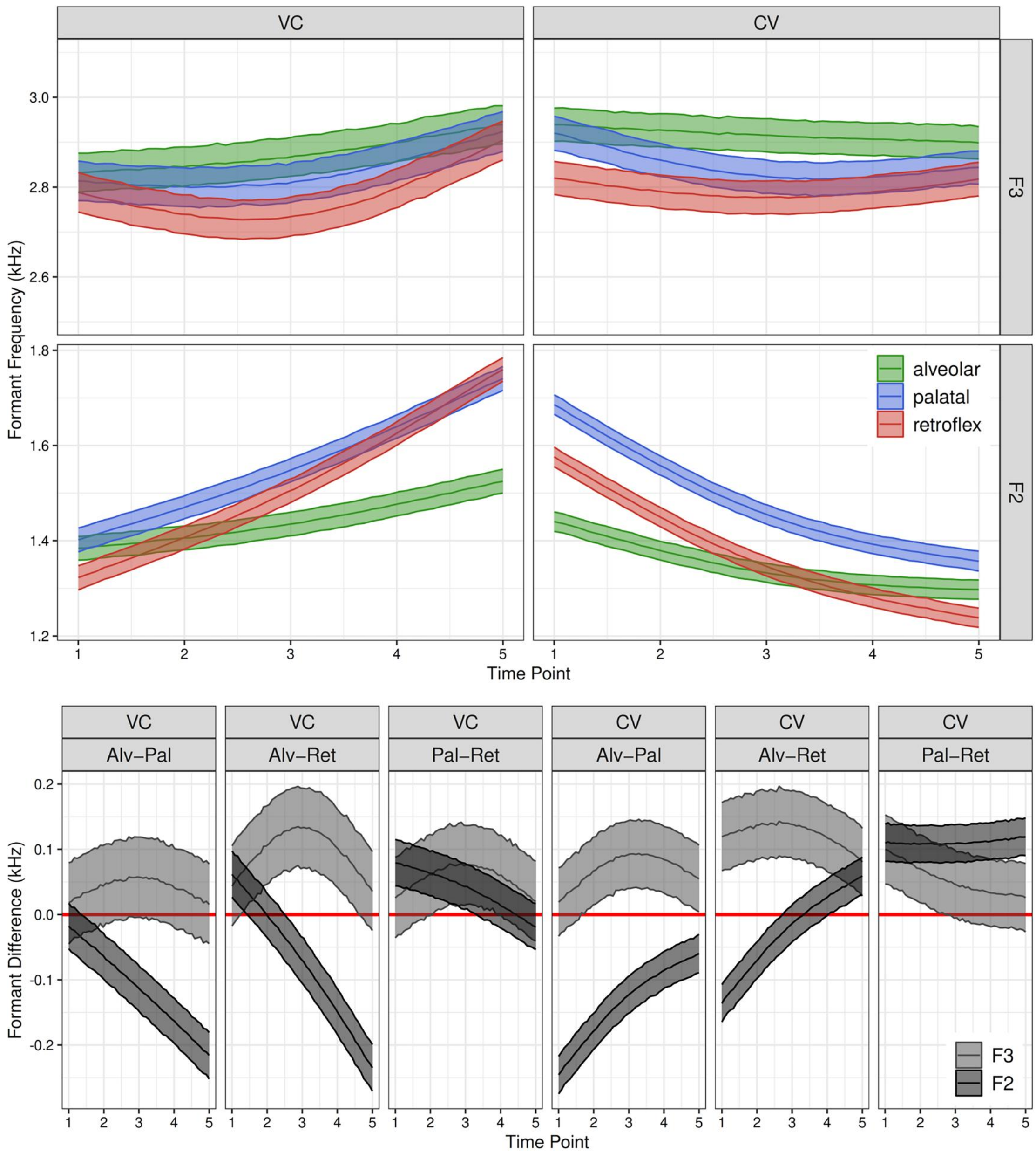

Fig. 2. Top, model-estimated formant frequencies (F2 and F3) and posterior confidence intervals (50%) in VC and CV transitions in the low-vowel /a/ context. Time points are normalised to range from vowel midpoint ($t=1$ in VC, $t=5$ in CV) to offset/onset ($t=5$ in VC, $t=1$ in CV). Bottom, model-estimated difference curves for each pairwise POA contrast.

palatals, and conversely higher-frequency F3 transitions relative to both palatals and retroflexes (see online repository for details).

3.2 Supervised category learning

In order to assess the degree to which the three sibilant categories can be distinguished on the basis of the full acoustic parameter set, a linear support vector machine (SVM) was trained with sibilant POA as the outcome (y) and the speaker- and vowel-residualised acoustic parameters (full consonantal set and F2/F3 values at onset/offset and midpoint) as predictors (i.e., the parameter values are assumed to be represented as offsets from speaker and vowel expectations, following the C-CuRE approach in McMurray and Jongman, 2011). All models are trained first on acoustic data, with the model results then analysed as a function of various lexical characteristics to understand how this system may operate in the lexicon and at a wider communicative level.

Model accuracy (from tenfold cross-validation averaged over five repetitions) in CV position was 72.0%, and was moderately lower in VCV position at 67.4%. In terms of pairwise contrasts within the sibilant system, however, results were highly imbalanced (consistent with the parameter-wise analyses above). Performance on alveolars was notably more accurate than the remainder in both CV (/s/ = 89.3%; /s-ç/ = 92.7%, /s-ʂ/ = 96.8%) and VCV (/s/ = 73.5%; /s-ç/ = 81.5%, /s-ʂ/ = 93.0%) positions. Palatal-retroflex discrimination, on the other hand, was at chance (52.5% in CV, 50.2% in VCV). Errors between the two postalveolars were relatively balanced: /ç/ → /ʂ/ = 42% in CV, 37% in VCV; /ʂ/ → /ç/ = 43% in CV, 38% in VCV. These results are further robust across speakers, with model predictions on all 16 speakers' data at or below 60% accuracy on palatals and retroflexes (/ç, ʂ/ error rates ranging from 21 to 50%).

Finally, I consider the extent to which these results are predictable from other lexical characteristics, such as the number and presence/absence of sibilant competitors in the lexical neighbourhood (i.e., are there any minimal or near-minimal sibilant-contrastive pairs and if so how many), the phonological neighbourhood density (how many words in the lexicon differ from the target by a single phonological substitution, addition, or deletion), and word frequency, all drawn from a word list extracted from the EMILLE corpus (Baker *et al.*, 2002). In the CV position, all four variables significantly predicted model accuracy, where both the presence of a sibilant competitor ($\beta = -0.781$, $z = -5.884$, $p < 0.001$) and the number of competitors ($\beta = -0.159$, $z = -8.375$, $p < 0.001$) negatively predicted model accuracy in a logistic regression (while controlling for the other lexical variables). General neighbourhood density, on the other hand, was positively associated with model accuracy ($\beta = 0.007$, $z = 7.840$, $p < 0.001$), as was log-transformed lexical frequency ($\beta = 0.601$, $z = 6.367$, $p < 0.001$), suggesting that there are potential mediating factors to increase acoustic distinctiveness in cases where there is greater communicative pressure. In VC position, only neighbourhood density was predictive, with words from more dense neighbourhoods more accurately categorised from the acoustic data ($\beta = 0.025$, $z = 4.918$, $p < 0.001$).

3.3 Unsupervised category learning

This section tests whether sibilant categories can be learned from the data without prior knowledge of category assignment (i.e., modeling the problem using unsupervised learning). For this purpose, projection-based clustering (PBC) was used due to the high dimensionality of the data, testing both 2- and 3-cluster solutions for their alignment with the true categories. Table 1 shows the classification results by each model size and position.

In 2-cluster models, alveolars showed a clear separation comprising 82% of one cluster in CV position, and 62% in VCV position; whereas, palatals and retroflexes equally comprised the other category in CV (at 41% and 42%, respectively) and VCV (41% and 49%). For 3-cluster models, in both CV and VCV, the alveolars again primarily comprised one category (at 81% and 84%, respectively), with the palatals and retroflexes splitting another category (46%/52% and 41%/49%) and the third category evenly distributed across all three sibilants. Thus, in the absence of prior categorisations, the alveolar-postalveolar distinction appears to be much more learnable than that between palatals and retroflexes.

4. Discussion and conclusions

The present data, combined with the general sparsity of minimal and near-minimal pairs in the Telugu lexicon, point toward a sibilant system which is more reliably comprised of two categories than three, though some acoustic cues to the palatal-retroflex distinction remain, particularly in the vowel transitions.[8] This result is consistent with earlier observations by Krishnamurti (1978, 2003) and Sjoberg (1962) (though note that both Krishnamurti and Sjoberg report mergers toward /s/, whereas the present study only found evidence for a postalveolar merger), among others, but crucially provides numerical evidence in support of such claims, while also providing greater specificity about differential merger patterns as a function of position, vowel context, and lexical characteristics.

Notably, while both earlier work and more recent work in Bhaskararao and Ray (2017) have described the three-way contrast as restricted to a higher register, "educated" speech, or some related social dimension, in the present study, participants were all university students and primarily Master's or Ph.D. students, who nevertheless showed little to no distinction among the postalveolars. Of course, this could be a consequence of bilingualism or limited formal

Table 1. Percentages of items classified in clusters 1, 2, and 3, by POA, Position, and model (2-category vs 3-category).

	2-category				3-category					
	CV		VCV		CV			VCV		
	1	2	1	2	1	2	3	1	2	3
Alveolar	31	2	12	21	10	21	2	7	21	5
Palatal	5	29	27	7	3	2	29	16	7	11
Retroflex	2	31	31	3	1	1	31	15	3	16

education in Telugu, though in the latter case, the male speakers were both more likely to have been educated in Telugu-medium schools and less likely to make a clear acoustic distinction between palatals and retroflexes. Further, at an informal level, these numerical results are consistent with speakers' intuitions. While these intuitions were not probed systematically, in a debriefing following the recording, many speakers, once informed of the purpose of the study, indicated that while they were taught three distinct pronunciations in school, they were only ever able to perceive or produce two.

Therefore, the acoustic data presented here indicate Telugu may more accurately be described as exhibiting a binary sibilant system, comprising an alveolar and a postalveolar (as in Hindi), with no further distinction among postalveolars. However, future research is needed to expand the regional and social distributions as well as to test for sibilant place categorisation and discrimination in perception.

Acknowledgment

This research was supported by funding from the Chancellor's Doctoral Fellowship at the University of Kansas.

Author Declarations

Conflict of Interest

The author has no conflicts of interest to disclose.

Ethics Approval

All data collection proceeded under ethical approval from the University of Kansas Institutional Review Board.

Data Availability

The data that support the findings of this study are openly available at https://github.com/chredmon/25-R-ACS_telugu-sibilants.

References

[1]Descriptions vary for this part of the system, with some describing the sound instead as *palatal*, others as *alveolo-palatal*, and others *postalveolar*, though all agree on this sound being produced somewhere between the palate and the alveolar ridge with a wide, laminal constriction; for simplicity I will refer to this sound in the remainder as *palatal*, though the alveolo-palatal symbol /ɕ/ will be used as it is more appropriate for the place where this sound is produced.

[2]Taking a single time window in the centre of the noise, rather than at multiple points, of course limits our ability to capture more dynamic information, but this choice was taken to avoid overparameterising the analysis and later modelling. This is, however, a relevant point of consideration in future research.

[3]Vowels were pooled into these larger subsets and interactions of sibilant place contrasts with vowel context were excluded because of insufficient data to model these more granular distinctions. Nevertheless we do expect given prior literature on the topic that the relative discriminability of sibilant contrasts may vary as a function of vowel context (e.g., being reduced in front and particularly high-front contexts).

[4]All acoustic parameter data, code, and detailed results and additional figures may be found on the project's Github page at https://github.com/chredmon/25-R-ACS_telugu-sibilants.

[5]In R syntax the models were formulated as follows: $F \sim POA + s(\mathrm{Time}, \mathrm{by} = \mathrm{POA}, \mathrm{bs} = \text{“cr”}, \mathrm{k} = 3) + s(\mathrm{Speaker}, \mathrm{bs} = \text{“re”}) + s(\mathrm{Time}, \mathrm{by} = \mathrm{interaction}(\mathrm{Speaker}, \mathrm{POA}), \mathrm{bs} = \text{“fs”}, \mathrm{k} = 3)$.

[6]All reports of *significant* divergences between formant transitions are based on intervals where the 95% median posterior interval excludes zero.

[7]See the online repository for formant transition figures for vowel contexts /i, e, o, u/.

[8]Of course, from the evidence presented above it is still possible to hold to the analysis of Telugu as exhibiting a three-way sibilant system, but in the main the data appear to point more strongly to a two-way system.

Baker, P., Hardie, A., McEnery, T., Cunningham, H., and Gaizauskas, R. J. (**2002**). "Emille, a 67-million word corpus of Indic languages: Data collection, mark-up and harmonisation," in *Proceedings of LREC.*

Bhaskararao, P., and Ray, A. (**2017**). "Telugu," J. IPA **47**, 231–241.

Boersma, P., and Weenink, D. (**2015**). "Praat: Doing phonetics by computer [computer program], version 6.0," https://praat.org (Last viewed August 20, 2015).

Bukmaier, V., and Harrington, J. (**2016**). "The articulatory and acoustic characteristics of Polish sibilants and their consequences for diachronic change," J. IPA **46**, 311–329.

Bürkner, P.-C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," J. Stat. Softw. **80**(1), 1–28.

Evers, V., Reetz, H., and Lahiri, A. (**1998**). "Crosslinguistic acoustic categorization of sibilants independent of phonological status," J. Phon. **26**, 345–370.

Handel, Z. (**2017**). "The Sinitic languages: Phonology," in *The Sino-Tibetan Languages*, edited by G. Thurgood and R. J. LaPolla (Routledge, London), pp. 85–113.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Krishnamurti, B. (**1957**). "Sandhi in modern colloquial Telugu," Indian Ling. **17**, 178–188.

Krishnamurti, B. (**1978**). "Language planning and development: The case of Telugu," Contributions Asian Stud. **11**, 37–56.

Krishnamurti, B. (**2003**). *The Dravidian Languages* (Cambridge University Press, Cambridge).
Lee-Kim, S. I. (**2014**). "Contrast neutralization and enhancement in phoneme inventories: Evidence from sibilant place contrasts and typology," Ph.D. thesis, New York University, New York.
Li, M. (**2017**). "Sibilant contrast: Perception, production, and sound change," Ph.D. thesis, University of Kansas, Lawrence, KS.
Maddieson, I., and Precoda, K. (**1990**). "Updating UPSID," UCLA Work. Papers Phon. **74**, 104–111.
McMurray, B., and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psychol. Rev. **118**, 219–246.
Ramanarasimham, P. (**2020**). "Old Telugu," in *The Dravidian Languages*, 2nd ed., edited by S. Steever (Routledge, Abingdon), pp. 239–259.
Shadle, C. H., and Mair, S. J. (**1996**). "Quantifying spectral characteristics of fricatives," in *Proceedings of ICSLP'96*, pp. 1521–1524.
Sjoberg, A. (**1962**). "Coexistent phonemic systems in Telugu: A socio-cultural perspective," Word **18**, 269–279.
Wood, S. N. (**2011**). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," J. R. Stat. Soc. (B) **73**, 3–36.
Żygis, M., and Padgett, J. (**2010**). "A perceptual study of Polish fricatives, and its implications for historical sound change," J. Phon. **38**, 207–226.