



The retroflex-dental contrast in Punjabi stops and nasals: A principal component analysis of ultrasound images

Alexei Kochetov¹, Matthew Faytak², Kiranpreet Nara¹

¹University of Toronto

²University of California, Berkeley

al.kochetov@utoronto.ca, mf@berkeley.edu, kiranpreet.nara@mail.utoronto.ca

Abstract

Many languages of South Asia show a phonemic contrast between retroflexes and dentals across different manners of articulation. This contrast, however, tends to be less phonetically distinct and more variable in nasals.

The goal of this paper is to examine the overall similarity of the retroflex-dental contrasts in Punjabi stops and nasals. Ultrasound tongue imaging recordings were obtained from 14 Punjabi speakers producing /t, n, t, n/ in the /ba_ab/ nonsense word context. Selected video frames were fed to a principal component analysis (PCA); the output was used for (1) training a linear discriminant model on one manner that discriminates place and (2) testing it on the other manner.

The results showed 100% correct classification of the contrast (retroflex or dental) in stops and 92% correct classification in nasals in the training data. The classification was much poorer across different manners: on average 67% of stops and 57% of nasals were classified correctly based on training sets with nasals or stops, respectively. In both cases, retroflex responses were more common.

These results suggest that the tongue configurations for Punjabi retroflex and dental consonants differ by manner of articulation. The contrast is also overall less robust in nasals than in stops, confirming previous reports.

Index Terms: speech production, articulation, retroflex, manner, ultrasound, Punjabi

1. Introduction

The phonemic contrast between retroflex and dental consonants is common in languages of South Asia, both belonging to the Indo-Aryan and Dravidian groups/families [1, 2]. The contrast can involve stops (/t/ vs. /t/) and nasals (/n/ vs. /n/), among other manners of articulation. However, retroflexion is considerably more common in stops than nasals, being reported in 75% and 35% of South Asian languages, respectively [3]. In Indo-Aryan languages that have retroflexes of both manners, the contrast tends to be phonetically more robust and less variable in stops than in nasals. For example, in Hindi and Nepali retroflex stops are quite frequent and occur in a variety of phonetic contexts, while retroflex nasals are limited to a small number of lexical items and certain contexts [4, 5]. Many speakers of these languages tend to neutralize the nasal contrast to /n/, and a related sound change has occurred historically in many Indo-Aryan languages, such as Urdu [6] and Bengali [7]. The more limited occurrence of the contrast in nasals and its

susceptibility to neutralization can be attributed to two facts. First the 'dental' /n/ in many South Asian languages is produced as more retracted, alveolar, compared to its clearly dental stop counterpart (e.g. [8]). Second, the retroflex nasal is typically shorter in duration and is often flapped, produced with an incomplete closure ([ĩ] [9, 10, 11]). Both factors thus reduce the phonetic realization of the place contrast in nasals compared to stops.

In this paper, we examine the general similarity and difference in the production of retroflex and dental stops and nasals in Standard Punjabi. All four consonants /t, n, t, n/ are phonemically contrastive in the standard variety [12, 13, 14, 15]. Retroflex stops occur word-initially, medially, and finally, while retroflex nasals occur medially and finally, but not initially (with dental stops and nasals occurring in all three contexts). Neutralization of the retroflex nasal to dental has also been noted in some urban varieties of Punjabi [16], but the extent of this variation is unclear. The only articulatory (palatography and x-rays) work on Punjabi we are aware of, Sandhu (1986) [16], observed that retroflex and dental nasals were produced somewhat differently from their stop counterparts – either as more retracted or with less contact (cf. [12, 13]). Being limited to a handful of speakers, these findings require a more thorough investigation.

To study the Punjabi retroflex-dental contrast, we used ultrasound tongue imaging [17], the method increasingly employed to investigate lingual posture contrasts. Specifically, ultrasound was used to study retroflex-dental contrasts in the Dravidian languages Kannada [18] and Malayalam [19] and in the Australian aboriginal language Arrernte [20]. These ultrasound studies were based on tongue tracings of individual tokens. The approach taken in this paper is different: we used the principal component analysis (PCA) method applied in Faytak (2018) [21], variants of which have previously been used to analyze ultrasound images by [22], [23], and [24].

Based on the previous work on Punjabi and other South Asian languages, it was predicted that the place contrast in stops would be more easily discriminated than in nasals. This is because the stops were expected to be characterized by more extreme, dispersed tongue shape configurations compared to the nasals. Given the expected difference in the realization of stops and nasals of the same place, it was also expected that any classification based on stops may not be fully applicable to nasals and vice versa. In addition, the discrimination of nasals could be affected by possible within-speaker variation in the realization of the place, as noted in the literature [16].

2. Method

2.1. Participants

Ultrasound and simultaneous audio data were collected from 16 native speakers of Punjabi (P1-P16). Due to technical issues with the recordings, data from the first 2 speakers were discarded, and the analysis was based on 14 participants. The speakers were originally from a variety of locations in Punjab, India: Abohar, Barnala, Ludhiana, Malerkotla, Moga, Nabha, Phagwara, and Tarn-Taran. Seven of them were female and 7 were male, with a mean age of 33 (range 20–47); their mean age of arrival in Canada was 28 (range 18–44, except 11 for P3), giving an average time lived outside of India of 5 years. All reported speaking English and (all but 3) Hindi as their other languages. At the time of the experiment, all the participants resided in the Greater Toronto Area. Speaker P3 was the third author of this paper.

2.2. Materials

The materials included nonsense words of the type bVCVb with various vowels and consonants (cf. [10, 11]). The analysis presented here is based on 4 items – the words with the stops /t, t/ and nasals /ŋ, n/ in the low vowel environment /a_a/, as shown in Table 1. The speakers were instructed to read the words, presented in the Gurmukhi script, at a comfortable speaking rate. On average, 9 repetitions per item per speaker were collected with a total of 517 tokens. Table 1 shows the breakdown of tokens by item.

Table 1: *This is an example of a table.*

Consonant	Word (IPA)	Word (Gurmukhi)	Tokens
retroflex stop	baʈab	ਬਾਟਾਬ	130
dental stop	batab	ਬਾਤਾਬ	131
retroflex nasal	baŋab	ਬਾਣਾਬ	129
dental nasal	banab	ਬਾਨਾਬ	127

2.3. Instrumentation and the procedure

Ultrasound data were collected using the *Teleded Echo Blaster 128 CEXT-1Z* system with an *Articulate Instruments* pulse-stretch unit [25]. The frame rate was 35.39 fps, the probe field of view and depth were 93.27 degrees and 150 mm. The probe was stabilized using an *Articulate Instruments* stabilization headset [26]. Audio recordings were done using an *AT831b* lavalier microphone and a *Sound Devices USBPre2* pre-amp. The audio, collected at a sampling rate of 22,050 Hz, was synchronized with the ultrasound video using the *Articulate Assistant Advanced* software [27]. Half of the speakers' (P3-P8, P16) data were collected in the Linguistics Department Phonetics Lab and for the other half (P9-15) at two public library locations in Brampton, a suburb with a sizeable Punjabi population. The equipment and the procedure were the same. All the experiments were conducted in Punjabi by the third author, except for her own experiment.

2.4. Data preparation and analysis

For each token, the frame of maximum tongue displacement during the acoustic consonant constriction was identified and extracted for analysis. On average, there were 9 such frames per item per speaker. Sample frames for the four consonants are shown in Figure 1.

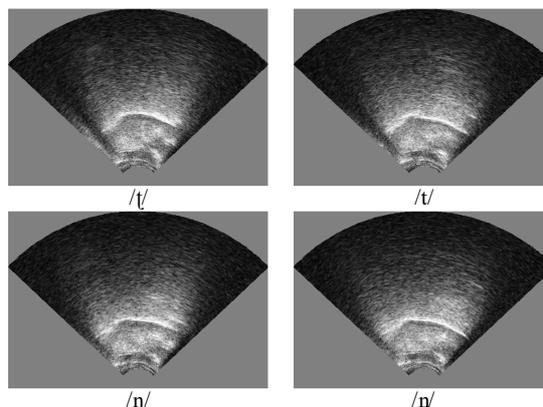


Figure 1: *Sample extracted frames for four consonants (Speaker P14, the 7th repetition).*

Raw extracted frames were processed with a median filter to reduce speckle noise. Following Faytak (2018) [21], a PCA was carried out on each speaker's set of collected and filtered frames (numbering at most 13 and at least 6 per item) to reduce the dimensionality of the data. Rather than carry out a single PCA including all productions by all speakers, a separate PCA was carried out for each speaker to avoid incorporating non-linguistic variation across speakers (due to probe positioning and morphological variation) into the speaker-specific realization of the linguistic contrasts being examined. Loadings of PC1-4 for one speaker's model, or "eigentongues" (after [22]), are shown in Figure 2.

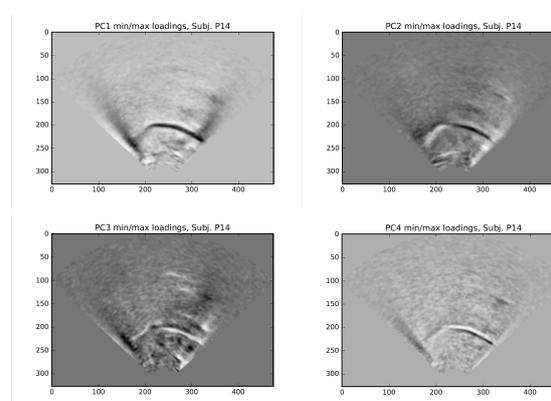


Figure 2: *PC1-4 "eigentongues" (explaining over 50% of variation) for Speaker P14.*

Taking each speaker's PC scores as a representation of tongue shape, two linear discriminant analyses (LDAs) were performed for each speaker as follows. A first LDA was trained using PC scores for the dental and retroflex stops /t, t/; the resulting linear discriminant was used to classify the dental and retroflex nasals /n, ŋ/ as either /t/ or /t/ in terms of tongue shape. A second LDA reversed the roles of each manner, training the model on the nasals /n, ŋ/ and classifying the stops /t, t/ as either /n/ or /ŋ/ in terms of tongue shape. The two linear discriminants served as indices of the speaker-specific

dental-retroflex contrast for the manner each was trained on. Classifying one manner according to the linear discriminant trained on the other manner in turn indicated how comparable the dental-retroflex contrast was across manners for a given speaker. Since the range of linear discriminant scores differed among subject-specific LDAs, any pooled data presented below were range-normalized (0.00 for retroflex and 1.00 for dental) for each token using mean LD values means for each training data category (calculated separately for each speaker).

Principal component analyses and linear discriminant analyses were carried out using the *scikit-learn* toolkit in *Python* [28]. Eigentongue images were produced using *Matplotlib* in *Python* [29]. Other data visualizations below were produced using *ggplot2* [30] for *R* [31].

In addition to the automatic classification, all tokens were examined auditorily by the third author, to determine whether the contrast was consistently distinguished by each speaker.

3. Results

The results are presented below separately based on which manner of articulation was used in the LDA training set: stops or nasals.

Before turning to these, however, it is important to note that the nasal data were inherently variable. Based on our auditory examination of all tokens, the speakers clearly distinguished the retroflex-dental contrast in stops, while only half of them did so for nasals (P3, P6-8, P11-13). Among the other speakers, some appeared to produce all nasals as dental (P5, P9, P16) or as retroflex (P15), or showed within-speaker variation (in / η / by P14, in / n / by P10, and in both by P4). Some of the nasal realizations could be of intermediate quality, and thus were not easily classified as retroflex or dental. These observations, however, should be taken with caution, as apparently neutralized consonants may still exhibit somewhat different tongue shape configurations.

3.1. Analysis of nasals based on stop training

The results for the stop-training set showed 100% correct discrimination of place in the training data: / t / was consistently classified as retroflex and / t / was classified as dental. The nasals were on average very poorly discriminated, although slightly above the chance level (59% for / η / and 54% for / n /). This can be seen in the aggregated density plot in Figure 3: the peaks for the stops were clearly separated, being aligned with 0.0 (retroflex) and 1.0 (dental), respectively. The nasals, on the other hand, largely fell in between and overlapped considerably.

Individual LD classification scores are shown in Table 2. It is clear that the training set classification was perfect – 0.00 for / t / and 1.00 for / t / – for all 14 speakers. Individual classifications of nasals, however, varied considerably. Only P3, P14, and, to a lesser degree, P11 showed the contrast between retroflex and dental nasals, that is, both their / η / and / n / were classified correctly more than 50% of the time. For 6 speakers (P4, P5, P7, P9, P12, P15), both consonants were classified as retroflex; for the other 5 speakers both were classified as dental. Recall that some of the speakers (indicated with an asterisk) were observed to produce an auditorily indistinct or variable contrast in nasals.

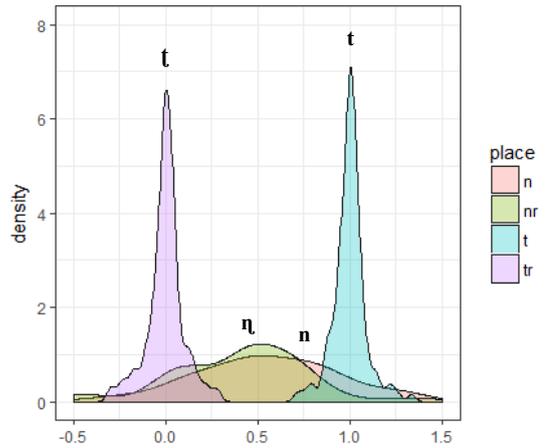


Figure 3: Density plot for dental-score in the stop-training set (aggregated data), nr, tr = η , t

Table 2: LD classification scores by speaker for the stop training set; 0.00 = all retroflex, 1.00 = all dental; mostly retroflex values (<0.50) are shaded

Speaker	Training set		Testing set	
	t	t	η	n
P3	0.00	1.00	0.00	1.00
P4*	0.00	1.00	0.10	0.36
P5*	0.00	1.00	0.00	0.00
P6	0.00	1.00	1.00	1.00
P7	0.00	1.00	0.00	0.44
P8	0.00	1.00	0.60	1.00
P9*	0.00	1.00	0.44	0.44
P10*	0.00	1.00	0.89	1.00
P11	0.00	1.00	0.44	0.56
P12	0.00	1.00	0.33	0.00
P13	0.00	1.00	0.89	0.22
P14*	0.00	1.00	0.00	0.89
P15*	0.00	1.00	0.00	0.00
P16*	0.00	1.00	1.00	0.67
Mean	0.00	1.00	0.41	0.54

Overall, these results showed that the place distinction in stops was clearly manifested in ultrasound images. Yet, the differences in stops were not well suited for discriminating place in nasals. This suggests two possible reasons, or a combination of them: (i) the place contrast in nasals was neutralized by most speakers or (ii) the place contrast in nasals was produced differently from the stops, and thus not well analyzed based on the stops criteria.

3.2. Analysis of stops based on nasal training

Turning to the nasal training set, the results showed 92% correct discrimination of place in nasals: 91% for / η / and 94% for / n / . This clearly shows that the place contrast in nasals was not fully neutralized in production, albeit reduced compared to stops (which showed 100% correct discrimination). Notably, this is despite the auditory impressions for some of the speakers' data (see above). Among the stops, / t / was correctly classified retroflex 78% of the time, while / t / was correctly classified as dental 56% of the time. These differences among the consonants can be observed in Figure 4. Note that the density peaks for the training consonants (nasals) were much

lower here than in the previous set. While the density peaks for the stops were further apart here, there was an overall greater spread for the two categories, with many data points found beyond the plot scale cut-offs (-0.5 to 1.5).

In sum, the results for this run were partly similar: while place in the training set consonants was well discriminated, the performance was much poorer once we switched to the different manner. It should be noted that the discrimination of target consonants in this run was lower, while the discrimination of novel consonants was better. The latter, however, could be due to the greater bias towards retroflex responses.

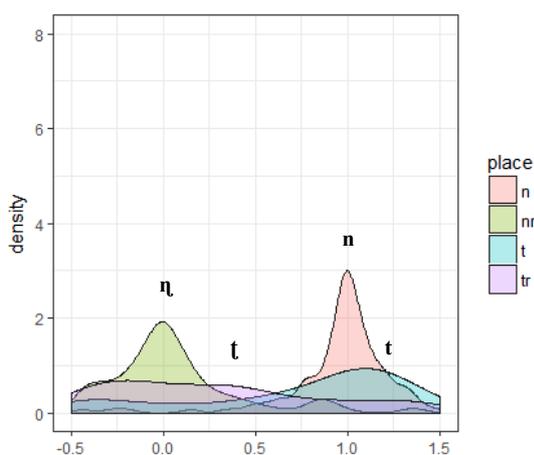


Figure 4: Density plot for dental-score in the nasal-training set (aggregated data), nr, tr = η , t .

Table 3: LD classification scores by speaker for the nasal training set; 0.00 = all retroflex, 1.00 = all dental; mostly retroflex values (<0.50) are shaded

Speaker	Training set		Testing set	
	η	n	t	t
P3	0.00	1.00	0.00	1.00
P4*	0.30	1.00	0.00	0.36
P5*	0.18	0.73	0.18	0.00
P6	0.00	1.00	0.00	0.43
P7	0.00	1.00	0.56	1.00
P8	0.00	1.00	0.00	0.00
P9*	0.33	0.89	0.00	0.00
P10*	0.22	0.89	0.11	1.00
P11	0.00	1.00	0.11	1.00
P12	0.11	1.00	0.00	0.00
P13	0.00	1.00	1.00	1.00
P14*	0.00	1.00	0.00	1.00
P15*	0.13	0.78	0.75	1.00
P16*	0.00	0.89	0.36	0.00
Mean	0.09	0.94	0.22	0.56

Table 3 presents individual LD classification scores (with asterisks indicating speakers with auditorily variable nasals). Overall, all speakers' data showed good-to-perfect discrimination of the contrast, with / η / receiving an average dental score of 0.09 (i.e. mostly retroflex) and /n/ receiving 0.94 (overwhelmingly dental). Note that the contrast was discriminated 100% correctly for half of the speakers (P3, P6-

8, P11, P13, P14); the other speakers' data were prone to errors in the range 0.11-0.33, either for one of the categories or for both. The variation in stops, however, was considerably greater. Here, only 4 speakers distinguished the contrast at least 50% of the time (P3, P10, P11, P14). Both nasals were classified as retroflex for seven other speakers (P4-6, P8-9, P12, P16), and for the remaining three speakers, both were classified as dental. There was thus a clear bias for the analysis to classify consonants as retroflex.

4. Discussion and conclusion

Given the known manner-specific phonological restrictions and variability in the retroflex-dental contrast in Punjabi [12, 13, 14] and Indo-Aryan languages in general [3, 4, 5, 6, 7], we predicted that the linear discriminant analysis of ultrasound tongue images would be more successful for stops than nasals. Moreover, the previously reported articulatory manner differences [16] suggested that the classification across manners may not be successful, even if the contrast was consistently produced in both.

The results provide support for these predictions, as stops in our data were overall better discriminated than nasals, and the discrimination criteria used for one manner were not particularly useful for the other. The latter indicated that stops and nasals of the same place were produced with somewhat different tongue configurations (cf. [16]). The interpretation of the results, however, was complicated by the fact that place in nasals was auditorily indistinct or variable for many speakers. Interestingly, place in nasals was still successfully classified in the nasal training run, suggesting an incomplete neutralization. Thorough articulatory (including tongue tracing), acoustic, and perceptual analyses of the data are needed to determine the nature of this variation. It is also important to determine the source of this variation – whether it is related to dialect differences, sociolinguistic factors, or second language (English or Hindi) influences.

It should be noted that the poor LD classification of some of the data can be due to relatively small numbers of tokens, which may have resulted in PCs that did not capture the major dimensions of variation in the data. Thus, the results should be further confirmed using larger-scale individual datasets, containing higher numbers of tokens for each consonant. Further, given the inherently dynamic nature of retroflex articulations [32, 18], it would be useful to examine tongue configurations over multiple points in time – towards, during, and after the constriction. This information is likely to improve the LD classification of the retroflex-dental contrast. It would be also useful to compare the Punjabi contrasts to those in Dravidian languages, where retroflex nasals are known to be more stable and overall more retracted [8].

To conclude, the results provide new evidence for systematic articulatory differences between consonants of the same place but different manner of articulation (cf. [33, 24]), raising questions about their sources in general aerodynamic and physiological constraints inherent in speech production.

5. Acknowledgements

We would like to thank Maida Percival for assistance with the data collection and the Punjabi speakers for their participation in the experiment. This work was partly supported by a *Social Sciences and Humanities Research Council* (#435-2015-2013) grant to the first author.

6. References

- [1] D. N. S. Bhat. "Retroflexion: An areal feature," *Working Papers on Language Universals*, vol. 13, pp. 27–67, 1973.
- [2] A. K. Ramanujan and C. Masica. "Toward a phonological typology of the Indian linguistic area," In T. A. Sebeok (Ed.), *Current Trends in Linguistics, vol. 5: Linguistics in South Asia* (pp. 543–577), Paris: Mouton, 1969.
- [3] P. Arsenault. "Retroflexion in South Asia: Typological, genetic and areal patterns," *Journal of South Asian Languages and Linguistics*, vol. 4, pp. 1–53, 2017.
- [4] M. Ohala. "Hindi," *Journal of the International Phonetic Association*, vol. 24, pp. 35–38, 1994.
- [5] R. Khatiwada. "Nepali," *Journal of the International Phonetic Association*, vol. 39, pp. 373–380, 2009.
- [6] S. G. M. Qadri. *Hindustani Phonetics*. Villeneuve-Saint-Georges: Imprimerie l'Union Typographique, 1930.
- [7] E. Šrámek. "Les consonnes rétroflexes du Bengali," *Revue de Phonétique*, Tome Ve, Paris, pp. 206–259, 1928.
- [8] O. Švarný and K. Zvelebil. "Some remarks on the articulation of the "cerebral" consonants in Indian languages, especially in Tamil," *Archiv Orientální*, vol. 23, pp. 374–407, 1955.
- [9] R. Dave. "Retroflex and dental consonants in Gujarati: A palatographic and acoustic study," *Annual Report of the Institute of Phonetics*, University of Copenhagen (ARIPUC), vol. 11, pp. 27–156, 1977.
- [10] R. P. Dixit. "Linguotectal contact patterns in the dental and retroflex stops of Hindi," *Journal of Phonetics*, vol. 18, 189–201, 1990.
- [11] R. P. Dixit. "Tongue-palate contact during retroflex consonants /D/, /R/, /N/ in Hindi," In Braun, A. (Ed.), *Advances in Phonetics: Proceedings of the International Phonetic Sciences Conference (IPS)*, Bellingham, WA, June 27-30, 1998, vol. 106, Franz Steiner Verlag, 1999.
- [12] H. S. Gill and H. A. Gleason. *A Reference Grammar of Punjabi*. Department of Linguistics, Punjabi University, 1969.
- [13] A. N. Malik. *The Phonology and Morphology of Punjabi*. New Delhi: Munshiram Manoharlal Publishers, 1995.
- [14] C. Shackleton. "Punjabi" In Cardona, G. and Jain, D. (Eds.), *The Indo-Aryan Languages* (pp. 581–621), New York: Routledge, 2003.
- [15] Q. Hussain, M. Harvey, M. Proctor, and K. Demuth. "Punjabi (Lyallpuri variety)," *Journal of the International Phonetic Association*, to appear.
- [16] B. S. Sandhu. *The Articulatory and Acoustic Structure of the Punjabi Consonants*. Patiala: Publication Bureau, Punjabi University, 1986.
- [17] M. Stone. "A guide to analysing tongue motion from ultrasound images," *Clinical Linguistics & Phonetics*, vol. 19, pp. 455–501, 2005.
- [18] A. Kochetov, N. Sreedevi, M. Kasim, and R. Manjula. "Spatial and dynamic aspects of retroflex production: An ultrasound and EMA study of Kannada geminate stops," *Journal of Phonetics*, vol. 46, pp. 168–184, 2014.
- [19] J. M. Scobbie, R. Punnoose, and G. Khatib. "Articulating five liquids: A single speaker ultrasound study of Malayalam," In L. Spreafico and A. Vietti (Eds.) *Rhotics: New Data and Perspectives* (pp. 99–124), Bozen-Bolzano: BU Press, 2013.
- [20] M. Tabain, and R. Beare. "An ultrasound study of coronal places of articulation in Central Arrernte: Apicals, laminals and rhotics," *Journal of Phonetics*, vol. 66, pp. 63–81, 2018.
- [21] M. Faytak. "Motor attractors mediate articulation of the Suzhou Chinese fricative vowel," Poster presented at the 92nd Annual Meeting of the LSA, Salt Lake City, 2018.
- [22] T. Hueber, G. Aversano, G. Cholle, B. Denby, G. Dreyfus, P. Oussar, P. Roussel, and M. Stone. "Eigentongue feature extraction for an ultrasound-based silent speech interface," *ICASSP '07*, Honolulu, HI, pp. I-1245–I-1248, 2007.
- [23] J. Mielke, A. Baker, and D. Archangeli. "Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /s/," *Language*, vol. 92, pp. 101–140, 2016.
- [24] P. Hoole and M. Pouplier. "Öhman returns: New horizons in the collection and analysis of imaging data in speech production research," *Computer Speech and Language*, vol. 45, pp. 253–277, 2017.
- [25] A. A. Wrench and J. M. Scobbie. "Very high frame rate ultrasound tongue imaging," *Proceedings of the 9th International Seminar on Speech Production (ISSP)*, pp. 155–162, 2011.
- [26] J. M. Scobbie, A. Wrench, and M. van der Linden. "Head-probe stabilization in ultrasound tongue imaging using a headset to permit natural head movement," *Proceedings of the 8th International Seminar on Speech Production*, pp. 373–376, 2008.
- [27] Articulate Instruments Ltd., *Articulate Assistant Advanced User Guide: Version 2.16*. Edinburgh, UK: Articulate Instruments Ltd., 2012.
- [28] J. Hunter. "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, pp. 90–95, 2007.
- [29] 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.
- [30] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.
- [31] R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <http://www.R-project.org/>. 2014
- [32] S. Narayanan, D. Byrd, and A. Kaun. "Geometry, kinematics, and acoustics of Tamil liquid consonants," *Journal of the Acoustical Society of America*, vol. 106, pp. 1993–2007, 1999.
- [33] F. E. Gibbon, I. Yuen, A. Lee, & L. Adams. Normal adult speakers' tongue palate contact patterns for alveolar oral and nasal stops. *Advances in Speech-Language Pathology*, vol. 9, pp. 82–89, 2007.
- [34] M. Tabain, A. Kochetov, R. Beare, and N. Sreedevi. "A comparative ultrasound study of manner contrasts in Arrernte and Kannada: manner contrasts," *Proceedings of the 16th Australasian International Conference on Speech Science and Technology (SST2016)*. December 2016, 4 pp, 2016.