ICPhS 2023
August 7–11, 2023
Prague Congress Center, Czech Republic

20th International Congress of Phonetic Sciences (ICPhS)

August 7–11, 2023
Prague Congress Center, Czech Republic

Using machine learning to model the three-way laryngeal contrast in Korean

Jeremy Perkins[a], Yu Yan[b], Dahm Lee[c], Seunghun J. Lee[d]

University of Aizu[a], Ritsumeikan University[b], Seoul National University[c], International Christian University[d], IIT Guwahati[d]

Introduction

- **Random Forest** models (a kind of Machine Learning (ML) tech.) were trained to learn the laryngeal contrast in Korean via acoustic measurements.
- Training data: from 7 acoustic measurements.
  - F0, VOT, spectral tilt, closure duration, frication duration, aspiration duration, release duration (frication + aspiration duration)
- Goal #1: To identify which measures form necessary & sufficient conditions for ML models with accuracy over 95%.
- Goal #2: To identify which measures are relatively more important for the models.

Background – Korean Obstruents

- Korean obstruent inventory:
  - By place, manner & laryngeal setting: **lenis** [p], **tense** [p’], **aspirated** [pʰ] (Kim, 1965; Kim & Duanmu, 2004).

Type	Labial	Coronal	Dorsal
Stops	p p’ pʰ	t t’ tʰ	k k’ kʰ
Affricates		t͡ʃ t͡ʃ’ t͡ʃʰ	
Fricatives		s’ s	

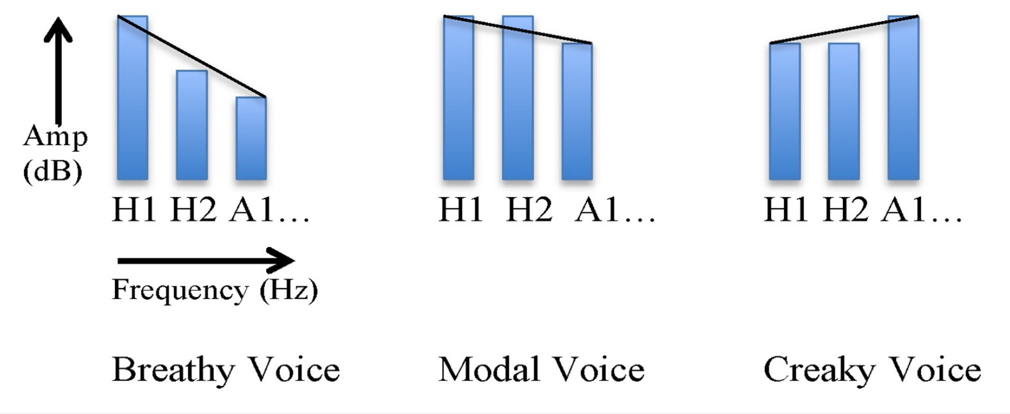
- In early work, laryngeal setting differences depended on VOT:
  - Tense: short positive VOT.
  - Lenis: intermediate VOT (Lisker & Abramson, 1964).
  - Aspirated: long positive VOT.

Background – Changing Phonetic Cues

- Aspirated & lenis **VOT** differences are no longer distinct.
  - F0 in a following vowel has replaced VOT as the primary difference between aspirated & lenis (Silva, 2006; Wright, 2007; Kang & Guion, 2008; Kang, 2010; Kang, 2014).
- Young speakers used **lowered f0**, not VOT to recognize lenis obstruents.
  - Tense & aspirated are recognized via a combination of f0, VOT and spectral tilt (Kim, Beddor & Horrocks, 2002; Kim, 2004; Schertz et al., 2015).

Background – Tense Obstruents

- Tense obstruents have laryngeal constriction - creaky phonation at the start of a following vowel (Cho et al., 2002).
- Measured via **spectral tilt**: At higher frequencies, intensity drops less in creaky voice than modal voice.
  - Tense obstruents have reduced values for H1*–H2* and H1*–A1* at following vowel onset.



Methods - Data Collection

- 24 Seoul Korean speakers (14 F, 10 M), aged 20-27.
- 66 bisyllabic **C**aCa words, most nonce (2 repetitions).
  - Target: C1, allowing all possible variations of laryngeal setting, manner & place.
  - C2: Lenis or Aspirated stop.
- Carrier sentence: 단어 X 는 무슨 뜻인가요
  - “What does the word X mean?”

Methods: Acoustic Measurements (1)

- F0 & spectral tilt in V1 extracted via VoiceSauce (Shue, 2010).
  - Formant-normalized:
    - H1*–H2*, H1*–A1*, H1*–A2*, H1*–A3*, H2*–H4*.
  - 1 measure chosen per model, to maximize accuracy.
- 1 measurement per word, closest to the vowel.
- Duration via Praat (Boersma & Weenink, 2022) segmentation of consonant closure, frication & aspiration.

Methods – RF Model Training Method

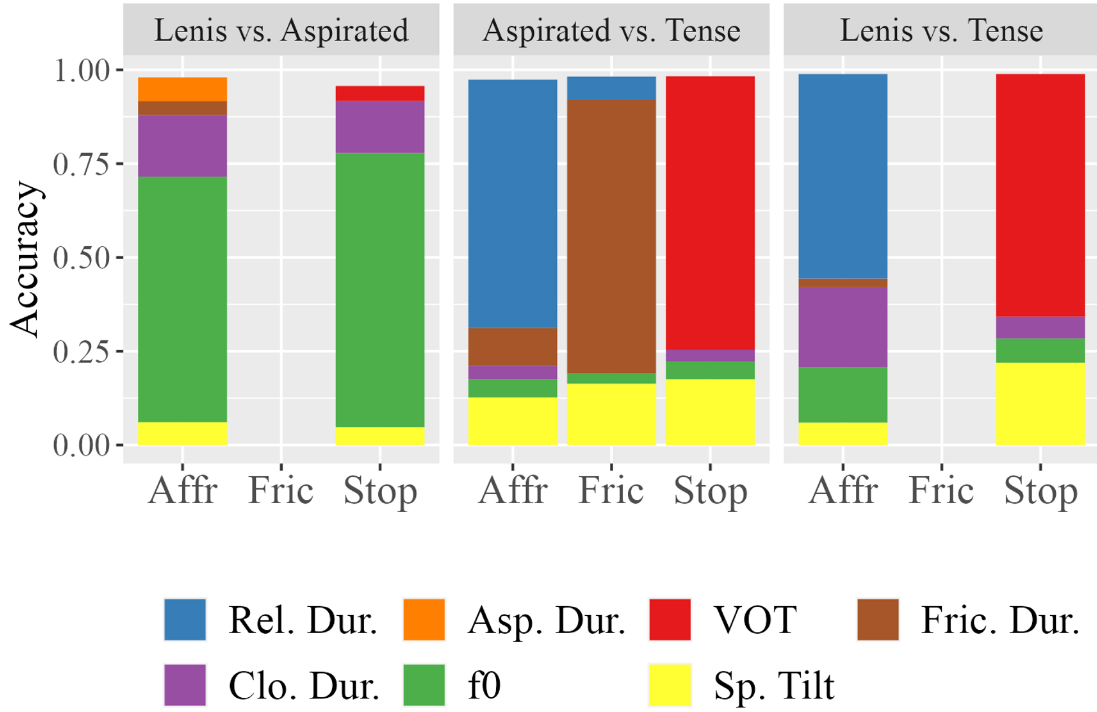
- RF models trained by separate data partitions defined by pairs of laryngeal features & by manner:
  - (1) aspirated & lenis; (2) aspirated & tense; (3) lenis & tense.
  - (A) stops; (B) affricates; (C) fricatives.
  - 7 total data partitions (3 for stops, 3 for affricates, 1 for fricatives).
- Training data: 1 of 2 repetitions randomly selected per word & speaker.
- Therefore, data partitions for testing/training were fully balanced by word & speaker.

Methods – RF Model Interpretation

- Models were built containing all possible combinations of the measures for each data partition (Xue & Zhao 2008).
  - A **successful** model: overall accuracy ≥ 95%.
- A measure or group of measures was **sufficient** if all models containing those measures were successful.
- A measure or group of measures was **necessary** if the only models that were successful contained those measures.
- Relative **importance** values (summing to 1) quantify each measurement's relative contribution to a model.

Results (1)

- Fully-loaded RFs ≥ 95% acc.
- Lenis vs. Aspirated: mostly f0
- Aspirated/Lenis vs. Tense:
  - Affricates: Rel. duration
  - Fricatives: Fric. Duration
  - Stops: VOT



Importance values are plotted.

ICPhS 2023
20th International Congress
of Phonetic Sciences (ICPhS)
August 7–11, 2023
Prague Congress Center, Czech Republic

Results (2)

- Necessary and sufficient acoustic measures for learning:

Lar. Contrast	Manner	Necessary Measures	Sufficient Measures
Lenis-Aspirated	Affr.	f0	f0
	Stops	f0	f0, closure duration
			f0, spectral tilt, VOT
Aspirated-Tense	Affricates	release duration	release duration, any other
	Fricatives	frication duration	frication duration, release duration
			frication duration, spectral tilt
	Stops	VOT	VOT
Lenis-Tense	Affricates	**none**	release duration
			F0, spectral tilt, closure duration, frication duration
	Stops	**none**	VOT
			F0, spectral tilt, closure duration

Conclusions

- ML models can get relative weights of acoustic cues.
  - Separate perception research is needed to confirm a cue.
- Previous findings were confirmed by RF modelling:
  - The aspirated-lenis contrast is primarily distinguished via f0.
  - Tense obstruents are primarily distinguished via VOT among stops.
  - No single acoustic measure is necessary for Tense vs. Lenis.
- Affricates & fricatives: release duration and frication duration substitute for VOT in distinguishing tense obstruents.

Acknowledgements

- This research was supported by JSPS KAKENHI Grant-in-Aid for Early Career Scientists #JP19K13162 and by the ILCAA joint research project “Phonetic typology from cross-linguistics perspectives (PhonTyp)” (2021-2023).

References

- P. Boersma and D. Weenink, “Praat: doing phonetics by computer,” 2022.
- T. Cho, S.-A. Jun, and P. Ladefoged, “Acoustic and aerodynamic correlates of Korean stops and fricatives,” J. of Phonetics, vol. 30, pp. 193–228, 2002.
- K.-H. Kang and S.G. Guion, “Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies,” JASA, vol. 124, no. 6, pp. 3909–3917, Dec. 2008.
- K.-H. Kang, “Generational differences in the perception of Korean stops,” Phonetics & Speech Sciences, vol. 2, no. 3, pp. 3–10, 2010.
- Y. Kang, “Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study,” J. of Phonetics, vol. 45, pp. 76–90, 2014.
- C.W. Kim, “On the autonomy of the tensity feature in stop classification,” Word, vol. 21, pp. 339–359, 1965.
- M.R. Kim, P.S. Beddor, and J. Horrocks, “The contribution of consonantal and vocalic information to the perception of Korean initial stops,” J. of Phonetics, vol. 30, pp. 77–100, 2002.
- M. R. Kim and S. Duanmu, “‘Tense’ and ‘lax’ stops in Korean,” J. of East Asian Linguistics, vol. 13, no. 1, pp. 59–104, 2004.
- M. Kim, “Correlation between VOT and f0 in the perception of Korean stops and affricates, in 8th Int’l. Conf. on Spoken Language Processing, 2004.
- L. Lisker and A.S. Abramson, “Cross-language study of voicing in initial stops: acoustical measurements,” Word, vol. 20, pp. 384–422, 1964.
- J. Perkins, “Acoustic measurement of laryngeal constriction in Thai consonants,” in 35th General Meeting of the Phonetics Soc. Of Japan, Sept. 25-26, 2021
- J. Schertz, T. Cho, A. Lotto, and N. Warner, “Individual differences in phonetic cue use in production and perception of a non-native sound contrast,” J. of Phonetics, vol. 52, pp. 183–204, 2015.
- Y.-L. Shue, “The voice source in speech production: Data, analysis and models,” Ph.D Thesis, UCLA, 2010.
- D.J. Silva, “Acoustic evidence for the emergence of tonal contrast in comtemporary Korean,” Phonology, vol. 23, pp. 287–308, 2006.
- J. Villegas, K. Markov, J. Perkins, and S.J. Lee, “Prediction of creaky speech by recurrent neural networks using psychoacoustic roughness,” IEEE J. of Selected Topics in Signal Processing, vol. 14, no. 2, pp. 355–366, 2020.
- J. Wright, “Laryngeal contrast in Seoul Korean,” Ph.D. dissertation, U. of Penn., 2007.
- J. Xue & Y. Zhao, “Random Forests of Phonetic Decision Trees for Acoustic Modeling in Conversational Speech Recognition,” IEEE Trans. On Audio Speech & Language Processing, vol. 16, no. 3, pp. 519–528, 2008.

Random Forest (RF)

- RF is an ML procedure used to develop prediction models.
  - In the RF settings, many classification and regression trees are constructed using randomly selected training datasets and random subsets of predictor variables for modelling outcomes.
  - Results from each tree are aggregated to give a prediction for each observation.
- In this study, the basic parameter settings of all of the RF models are: {number of estimators: 200, random state: shuffle}

Background – Psychoacoustic Roughness

- Tense obstruents have laryngeal constriction.
- Leads to creaky phonation at the start of a following vowel, traditionally measured via spectral tilt (Cho et al., 2002).
- Psychoacoustic roughness: Used to measure creaky phonation in Burmese (Villegas et al. 2020).
    - Roughness extracted via a Matlab routine in Villegas et al., 2020.
    - Research question 2: How do roughness & spectral tilt compare in identification of Tense obstruents?

Measurements of Laryngealization (1)

- Research Question 2: Which of the various spectral tilt & roughness measures are optimal for learning the laryngeal contrast?
- Minimally sufficient models with one measure of laryngealization are compared for model accuracy:

Measure	Lenis-Aspirated		Aspirated-Tense			Lenis-Tense	
	Affricates	Stops	Affricates	Fricatives	Stops	Affricates	Stops
Roughness	.967	**.960**	**.965**	.942	**.982**	.944	.907
H1*–H2*	.967	.959	.944	.942	.976	.946	.923
H2*–H4*	.967	.952	.958	.942	.964	.953	.915
H1*–A1*	.951	.952	.963	**.971**	.973	.946	**.958**
H1*–A2*	**.972**	.950	.963	.942	.970	.944	.925
H1*–A3*	.967	.949	.963	.947	.972	**.958**	.931

Measurements of Laryngealization (2)

- Minimally sufficient models with one measure of laryngealization are compared for relative importance of that measure:

Measure	Lenis-Aspirated		Aspirated-Tense			Lenis-Tense	
	Affricates	Stops	Affricates	Fricatives	Stops	Affricates	Stops
Roughness	**.135**	**.116**	.160	.178	.152	.060	.142
H1*–H2*	.135	.085	.141	.211	.159	.073	.182
H2*–H4*	.133	.103	.177	.187	.114	.117	.193
H1*–A1*	.101	.085	.218	**.257**	**.278**	**.184**	**.398**
H1*–A2*	.096	.068	.218	.202	.250	.131	.328
H1*–A3*	.096	.074	**.235**	.218	.189	.182	.277

Measurements of Laryngealization (3)

- Tense contrasts: spectral tilt measures had higher importance values than roughness.
  - H1*–A1* was most important in 5/6 models.
- Lenis-aspirated contrast, roughness had higher importance values than spectral tilt.
  - Roughness inversely correlates with f0 (in Thai; Perkins, 2021), so that roughness is raised following lenis obstruents (low tone).

Measurements of Laryngealization (4)

- Spectral tilt correlates with creakiness & breathiness.
  - Lenis are breathy, (raised spectral tilt) (Cho et al, 2002).
  - Spectral tilt diverges for lenis (raised) and tense (lowered).
  - Roughness converges for lenis and tense (both raised).
- Therefore, aspirated-tense contrast is most informative.
  - Accuracy: Higher roughness for affr. & stops (not fricatives).
  - Importance: Higher spectral tilt in all three models.
    - Reason: Correlation between roughness and f0.