Production and perception of tonal coarticulation: Evidence from computational simulation of communication

Huang, Po-Hsuan

Institute of Information Science, Academia Sinica
Graduate Institute of Linguistics, National Taiwan University
Taipei, Taiwan

Background

Lexical tones have been found to coarticulate with the preceding and following tones in tone languages (tonal coarticulation, or TC; e.g., [1, 2]):

For example,

- H offset + target tone → H offset + ↑target tone
- L offset + target tone → L offset + ↓target tone

In addition, such variations can influence listeners' perceptions (e.g., [1, 3])

The same surface tone is:

- perceived as **lower** when preceded by a **higher** tone offset.
- perceived as **higher** when preceded by a **lower** tone offset.

Question: How do tone language users maintain faithful tone perception under TC?

Previous findings [1, 4] Beijing Mandarin (BM) TC → target tone mistaken as another tone under TC-induced variations → retrieve the target through normalization

New findings [2, 3] Linguistic differences exist btw Taiwan Mandarin (TM) & Taiwan Southern Min (TSM):
- Magnitude of TC: TM ≈ TSM
- Magnitude of normalization: TM > TSM
- Tone acceptance ranges: TM > TSM

Hypothesis: Tonal distribution affects how languages deal w/ TC.

Small tone inventory a coarticulated target tone easily accepted as another lexical tone → retrieve the target tone through normalization

Large tone inventory stricter tone acceptance ranges → the coarticulated target faithfully perceived

A view from computational simulation

Past studies [5, 6] have proved the ability of communication simulation to capture important linguistic features through the interaction of the "speaker" and "listener" agents.

In this study, a `speaker` neural agent and a `listener` neural agent to simulate real-world tone communication under TC.

Research question

Is the hypothesis that the sizes of tone inventories affect tone perceptions under TC supported by computational simulation?

When faced w/ different numbers of tones, will the listener agents demonstrate:

- different degrees of normalization?
- different ranges of tone acceptance?

Experiment

Dataset

Two datasets are constructed to simulate the tone inventories of TM & TSM.

- TM: 55/35/21/51
- TSM: 55/35/21/51/33

→ tone contours presented as tuples of two floats: `(onset, offset)`

In this experiment, the coarticulatory effect of the preceding tone offset on the target tone was investigated.

Tone contour pairs For each possible tone pair, 2,048 tokens were generated:

`(onset, offset, onset, offset)*2048`

Singleton tone contours To train the listener agent to recognize the canonical tone contours, an additional 2,048 tokens were generated for each tone.

`(onset, offset)*2048`

Model training

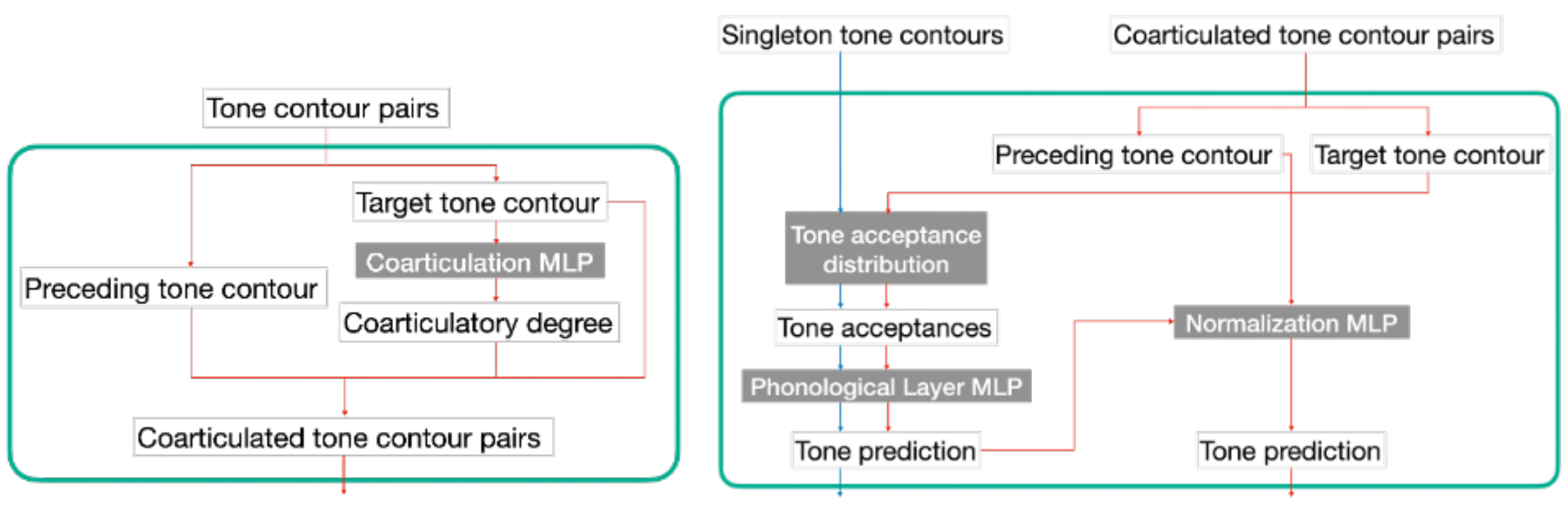

Figure 1. Structures of the speaker agent (left) and the listener agent (right). The blue lines indicate the flow of Phase A; the red lines indicate the flow of Phase B.

For each of the 256 epochs,
4×(Phase A)
sigleton tones → `listener`
→ get tone acceptance based on tone acceptance distribution
→ get tone prediction through Phonological multilayer perceptron (MLP)
→ backpropagation ← **train `listner` to recognize canonical tone contours**
4×(Phase B)
tone pairs → `speaker`
→ get coarticulated pairs based on the degree of coarticulation
coarticulated pairs → `listener`
→ get tone acceptance based on tone acceptance distribution
→ get initial tone prediction through Phonological MLP
→ get final tone prediction through Normalization MLP
→ backpropagation ← **simulate communication b/w `listner` & `speaker`**

Analyses

The three aspects proposed in [2, 3] are analyzed as follows:

- Magnitude of coarticulation: mean degree of coarticulation of `speaker` during validation
- Magnitude of normalization: a series of target tones simulating the continuum from the low tone (21) to the falling tone (51) following the different preceding tones were predicted by the TM and TSM models (cf. [4])
- Tone acceptance ranges: standard deviations of the tone acceptance ranges during validation

Results & discussion

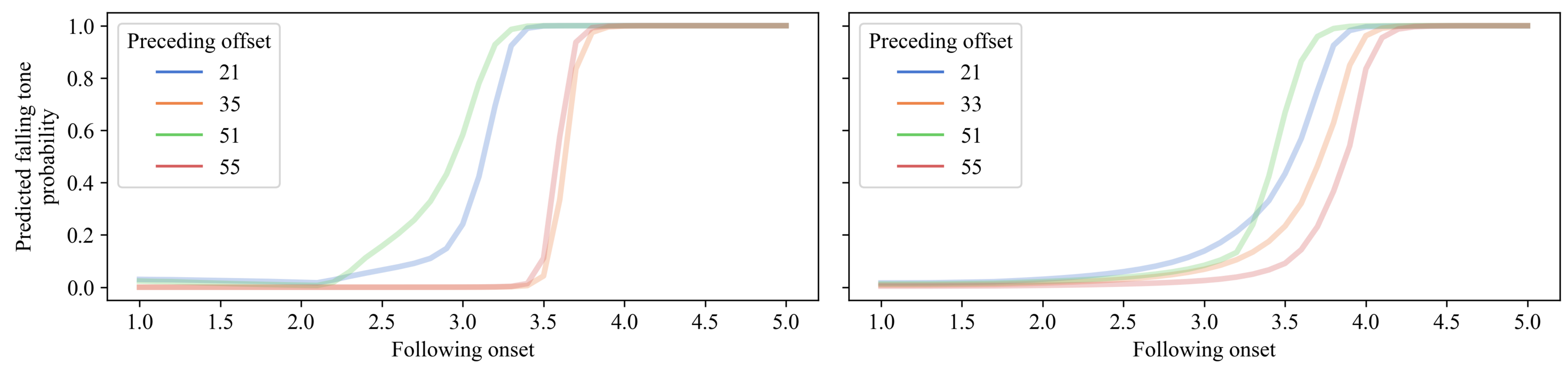

Figure 2. Normalization of the listener neural agent for different preceding tones on a low-to-falling tone continuum (left: TM; right: TSM)

- Magnitude of coarticulation: mean degrees of coarticulation in the TM & TSM models were **0.54** and **0.50**. → **similar degrees of TC**
- Magnitude of normalization: the TM model was more subject to the offset height of the preceding tones (max interval of **0.6** vs. **0.4** in the TSM model). → **larger normalizing effects in the TM model**
- Tone acceptance ranges: the mean standard deviations of the TM and TSM models were **3.35** and **2.39**. → **stricter acceptance ranges in the TSM model**

→ Different perceptual mechanisms were found when the agents delt w/ different tonal distributions.

Coarticulatory resistance's influence on accuracy

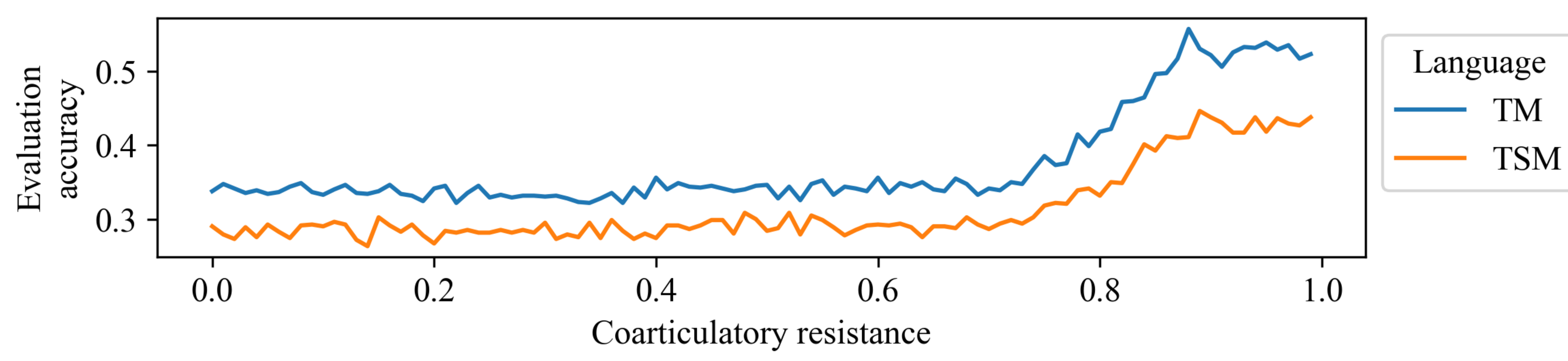

Figure 3. Evaluation accuracy over different coarticulatory resistance

→ Coarticulatory resistance does not have much impact on accuracy unless it is relatively large (>0.7).

→ Ease of articulation vs. ease of perception will need to be taken into account.

Conclusion

This study demonstrates the possibility of simulating real-world communication and human cognitive mechanisms as well as its ability to allow for more direct explanations of production and perception behaviors through computational simulation based on the interaction of the speaker and listener agents.

References

[1] Y. Xu, ``Production and perception of coarticulated tones,'' *Journal of the Acoustical Society of America*, vol. 95, no. 4, pp. 2240--2253, 1994.

[2] P.-H. Huang and C. Chiu, ``Tonal coarticulation in Taiwan Mandarin and Taiwan Southern Min,'' 2023, Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2023.

[3] ------, ``Perception of coarticulated tones in Taiwan Mandarin and Taiwan Southern Min,'' in *Proceedings of the 20th International Congress of Phonetic Sciences*, R. Skarnitzl and J. Volín, Eds., International Congress of Phonetic Sciences. Guarant International, 2023, p. 446–450.

[4] H. Zhang, H. Ding, and W.-S. Lee, ``The influence of preceding speech and nonspeech contexts on Mandarin tone identification,'' *Journal of Phonetics*, vol. 93, p. 101154, 2022.

[5] Y. Ren, S. Guo, M. Labeau, C. S.B., and S. Kirby, ``Compositional languages emerge in a neural iterated learning model.'' International Conference on Learning Representations, 2020.

[6] E. Carlsson, D. Dubhashi, and T. Regier, ``Iterated learning and communication jointly explain efficient color naming systems,'' 2023.

Our lab   This paper   Author's website