Estimating "Good" Variability in Speech Production using Invertible Neural Networks

Poster No. 163

Jaekoo Kang [1,2], *Hosung Nam* [2,3], *and D. H. Whalen* [1,2,4]

[1]The Graduate Center, CUNY; [2]Haskins Laboratories; [3]Korea University; [4]Yale University

jkang@gradcenter.cuny.edu; hnam@korea.ac.kr; dwhalen@gc.cuny.edu

International Seminar on Speech Production (Dec 14 – 18, 2020)

Work supported by NIH grant DC-002717 to Haskins Laboratories.

Introduction

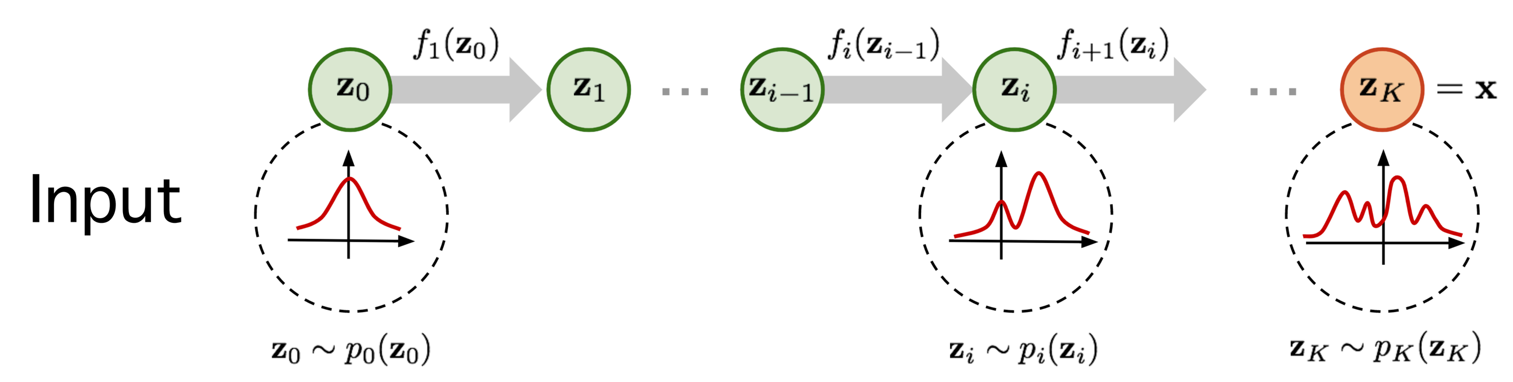
- Variability is widespread in speech, but not all of it is random. It can be decomposed into "good" and "bad" part w.r.t. speech tasks (uncontrolled manifold hypothesis [1,2], principle of motor abundance [3]).
- We examined a novel method of identifying "good" variability (flexibility) from speech data using flow-based invertible neural networks (**INN**)[6].

Data & Preprocessing

- **Data:** The Haskins IEEE electromagnetic articulography database (8 native English speakers; sentence reading; normal vs. fast speaking rate).
- **Samples:** 4 front vowels (/i, ɪ, ɛ, æ/).
- **Procedure:**
  - Normalization: Speaker-wise data normalization.
  - Feature extraction: 3 principal components (articulation) and 2 formant frequencies (F1, F2) ⇨ *3D to 2D mapping*.

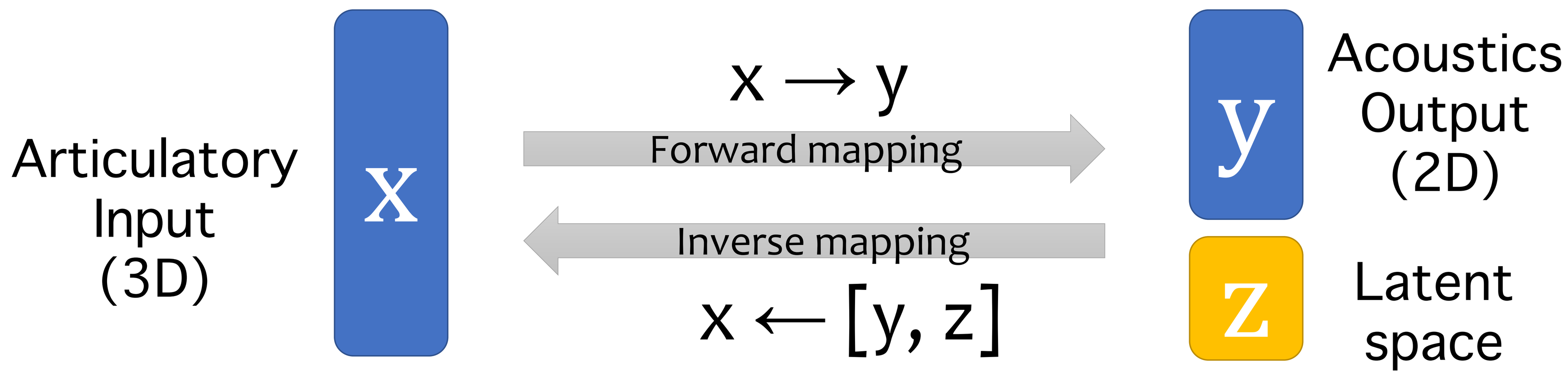
Flow-based Invertible Neural Networks

- **Normalizing-flow Technique** [4]



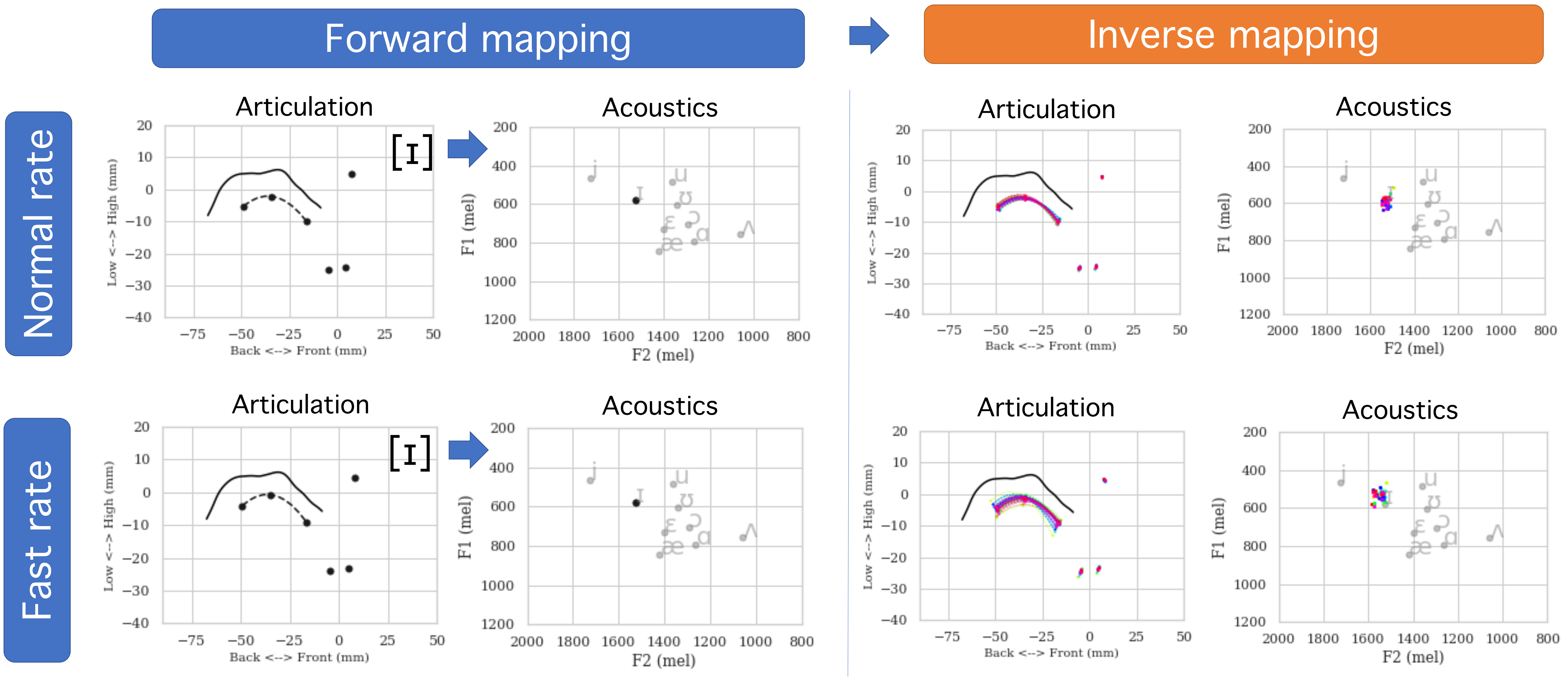
"Volume-preserving" transformations

- **Invertible Neural Networks** [6]



- **Training & Validation**
  - Model: Two INN models were trained per speaker ("normal" vs "fast" rate).
  - 3 Loss functions: forward/inverse loss with latent loss.
  - Validation: 20% of the data per speaker.

Results



Visualization from a single speaker's data (F01)

2D Gaussian latent samples were added

Why INNs?

- The learned latent space reveals "good" variability.
- There is no need for the dimensional mismatch.
- The computation of Jacobian is easy and tractable.

Conclusion

- The flow-based invertible neural networks can effectively estimate "good" variability (range of flexibility).
- More tests are required (phonetic context; neural ordinary differential equations; mixture density networks).

[1] Scholz, J., & Schöner, G. (1999). The uncontrolled manifold concept: Identifying control variables for a functional task. *Experimental Brain Research*, *126*(3), 289–306.
[2] Saltzman, E., Kubo, M., & Tsao, C.-C. (2006). Controlled variables, the Uncontrolled Manifold, and the Task-dynamic model of speech production. *Dynamics in Speech Production and Perception*, 21–31.
[3] Latash, M. (2010). Motor Synergies and the Equilibrium-Point Hypothesis. *Motor Control*, 14(3), 294–322.
[4] Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2019). Normalizing flows for probabilistic modeling and inference. *Arxiv*, 1–60.
[5] Weng, L. (2018, October 13). Flow-based Deep Generative Models. Retrieved December 03, 2020, from https://lilianweng.github.io/lil-log/2018/10/13/flow-based-deep-generative-models.html
[6] Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., Maier-Hein, L., Rother, C., & Köthe, U. (2019). Analyzing inverse problems with invertible neural networks. 7th International Conference on Learning Representations, *ICLR 2019*, 1–20.

For more, please check out
https://jaekookang.github.io/issp2020/