Automatic Rank Ordering of Singing Vocals with Twin-Neural Network

Chitralekha Gupta, Lin Huang, Haizhou Li

National University of Singapore

1. Introduction

- Singing is a desirable skill to develop
- Online platforms to showcase singing talent such as Smule Sing! and We Sing
- We explore automated methods of assessing the quality of singing
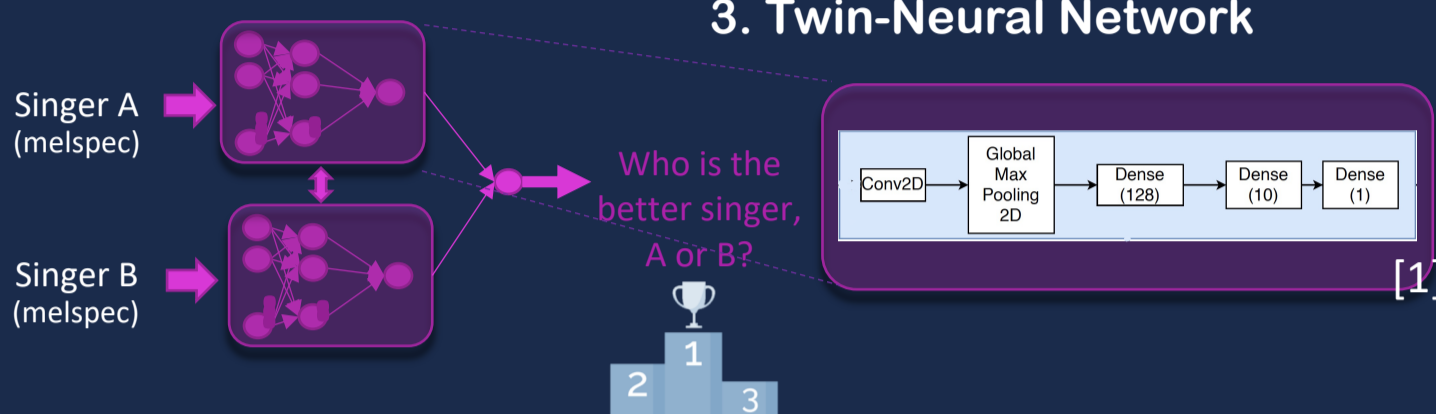 - For singing talent search
  - For feedback to amateur singer

2. Inspiration

Humans are known to be better at choosing from a small number of options, rather than absolute ranking many options

THE PARADOX OF CHOICE: Why less is more when it comes to engagement

We propose a preference-based framework to generate a rank-order of singing vocals

3. Twin-Neural Network

Singer A (melspec)

Singer B (melspec)

Who is the better singer, A or B?

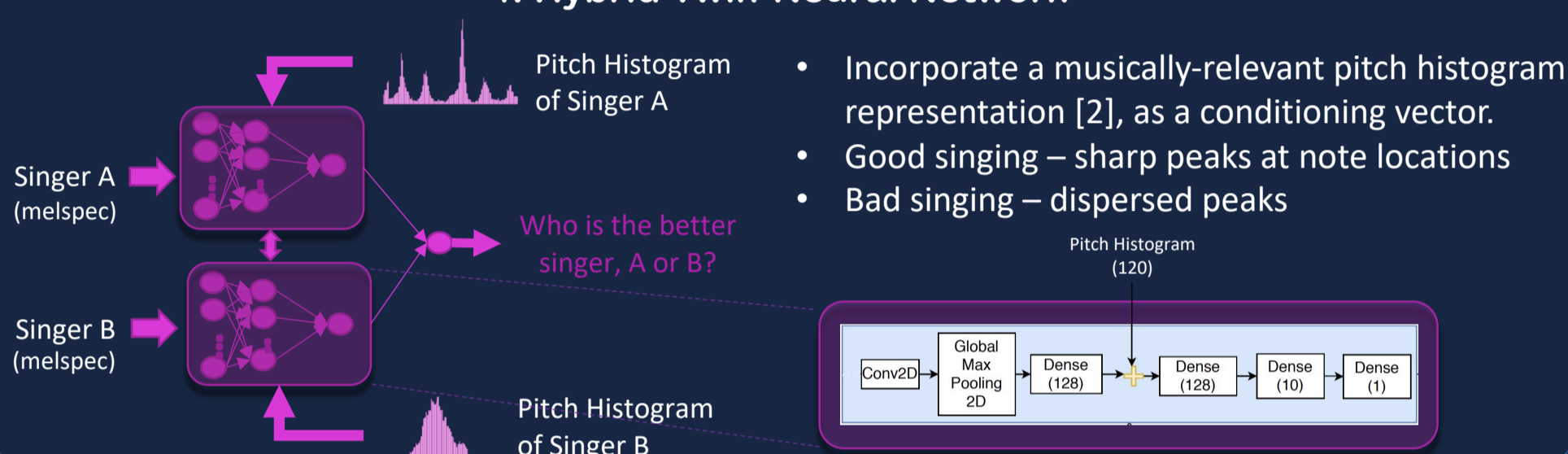Conv2D → Global Max Pooling 2D → Dense (128) → Dense (10) → Dense (1) [1]

Proposed Modifications

1. Preference Metric, instead of Distance Metric

$$D_p = |f(x_A)| - |f(x_B)|$$

2. Comparative Loss, instead of Contrastive Loss

4. Hybrid Twin-Neural Network

Pitch Histogram of Singer A

Singer A (melspec)

Singer B (melspec)

Who is the better singer, A or B?

Pitch Histogram of Singer B

- Incorporate a musically-relevant pitch histogram representation [2], as a conditioning vector.
- Good singing – sharp peaks at note locations
- Bad singing – dispersed peaks

Pitch Histogram (120)

Conv2D → Global Max Pooling 2D → Dense (128) → Dense (128) → Dense (10) → Dense (1)

5. Dataset

[3]	# singers	Pairs of vocals singing same song	Pairs of vocals singing same + different songs
Train	320	25,280	102,080
Validation	40	180	780
Test_1 (seen songs)	40	180	780
Test_2 (unseen songs)	10	90	190

6. Some setup details

- Human score for a singer is b, then we label a pair of singers (A,B), as:

$$y = \begin{cases} 1, & if\ b_A \geq b_B \\ 0, & otherwise \end{cases}$$

- Rank-ordering is done according to the aggregate score of each singer through pairwise comparisons

7. Results

Spearman Rank Correlation between rank order from machine versus human

Gupta et al. [3]		This work	
Framework	Corr	Framework	Corr
Relative Measures	0.64	Hybrid Twin-Net	**0.68**
Absolute Measures	0.48	Absolute score prediction [1]	0.62

	Test Data	Twin Network	Hybrid-Twin Network
Only same song pairs for training	Unseen singers, but seen songs	0.61	**0.68**
	Unseen singers and songs	0.41	**0.65**

Training pairs	Test Data	Hybrid-Twin Network
Only different song pairs for training	Unseen singers and songs	0.68
Same + different song pairs for training	Unseen singers and songs	**0.73**

The inter-judge correlation amongst music experts is 0.82

8. Contributions

- Modified twin network for preference selection
- Hybrid pitch histogram + twin network
- Song and singer independent singing quality evaluation

9. References

[1] K. A. Pati, S. Gururani, and A. Lerch, "Assessment of student music performances using deep neural networks," Applied Sciences, vol. 8, no. 4, p. 507, 2018.

[2] C. Gupta, H. Li, and Y. Wang, "Automatic evaluation of singing quality withouta reference," in Proceedings of APSIPA Annual Summit and Conference, 2018.

[3] C. Gupta, H. Li, and Y. Wang, "Automatic leaderboard: Evaluation of singing quality without a standard reference," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 13–26, 2020.