Cross-modal Force & Language Embeddings

Ravi Tejwani, Karl Velazquez, John Payne, Paolo Bonato, H. Harry Asada

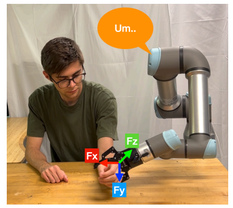
Massachusetts Institute of Technology

HARVARD MEDICAL SCHOOL

SPAULDING REHABILITATION HOSPITAL

Problem Statement

We conducted an observational study with a physical therapist from the Spaulding Rehabilitation Hospital. She demonstrated how humans naturally ***combine verbal*** instructions with precise ***physical*** forces.

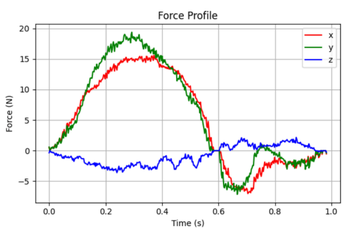


If we want robots to use both language and force, we need a method to ***translate*** between them.

Therefore, we aim to develop a framework that can learn a ***shared representation*** of natural language words to real time human reactive forces.

Force

$$\text{Force Profile} = \overbrace{\begin{bmatrix} t_0 & t_1 & \dots & t_{\mathcal{N}-1} \\ F_x(t_0) & F_x(t_1) & \dots & F_x(t_{\mathcal{N}-1}) \\ F_y(t_0) & F_y(t_1) & \dots & F_y(t_{\mathcal{N}-1}) \\ F_z(t_0) & F_z(t_1) & \dots & F_z(t_{\mathcal{N}-1}) \end{bmatrix}}^{\mathcal{N}} \Big\} 4$$



Website

(shared-language-force-embedding.github.io)

Language

Leverage Sentence-BERT to produce ***semantically meaningful*** embeddings of input phrases.

Ex: *"right"*

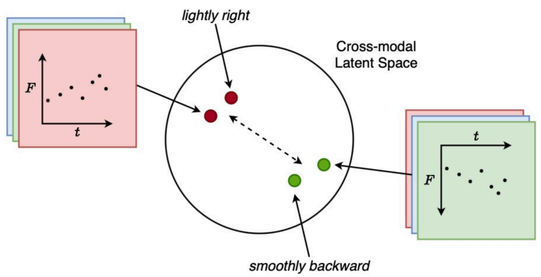
-0.3191	-0.0951	.....	0.0814

Ex: *"gently down right"*

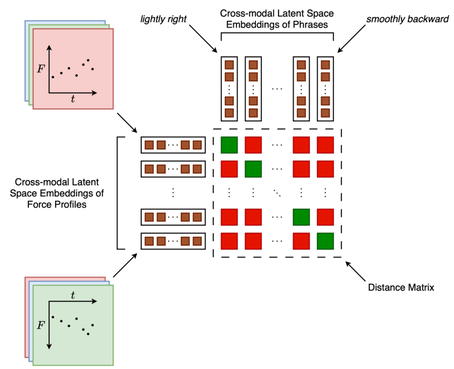
0.2173	-0.4527	.....	0.1348

Cross-modal Embedding

Corresponding force profiles and phrases and mapped ***distance-wise closer*** than non-corresponding instances.



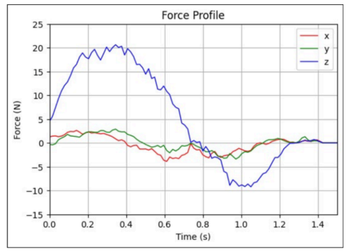
Contrastive Learning

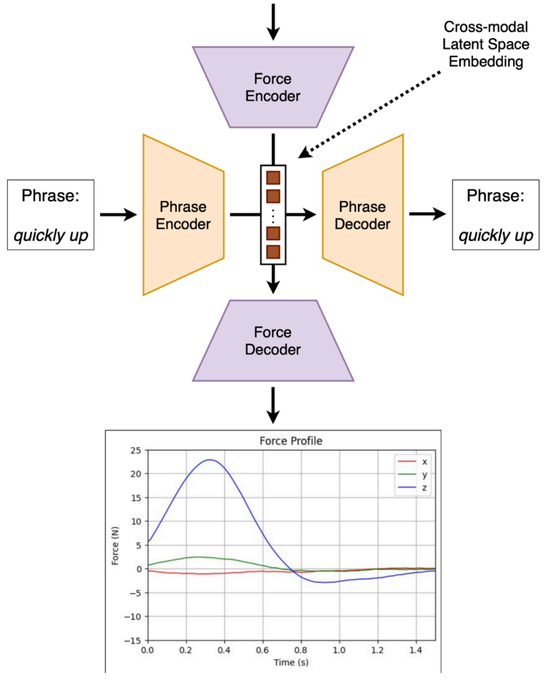


Dual Autoencoder

There is an autoencoder for each modality that is responsible for ***encoding and decoding*** to and from the shared latent space.

Contrastive learning ensures the outputs of the encoders for a given corresponding pair of force profiles and phrases are ***aligned***. This allows each decoder to generate corresponding instances even from the ***same latent space embedding***.





Data Collection

10 volunteers completed 840 trials involving ***human demonstrations*** of force and language translation.

Evaluation

Mean Model Scores for In-Distribution Samples

	SVM/KNN	$DMLP_B$	$DMLP_S$	DAE_B	DAE_S
FPAcc	11.714	4.523	4.700	4.454	4.582
FDAcc	0.902	0.975	0.973	0.977	0.972
ModSim	0.545	0.516	0.516	0.581	0.576
DirSim	0.982	0.978	0.842	0.979	0.934
PhraseSim	0.764	0.747	0.680	0.780	0.755

DAE Translates Better

Model Scores on Out-of-Distribution Modifiers

	SVM/KNN	$DMLP_B$	$DMLP_S$	DAE_B	DAE_S
FPAcc	16.912	6.762	5.861	6.815	7.239
FDAcc	0.787	0.976	0.956	0.978	0.935
ModSim	0.249	0.337	0.302	0.383	0.334
DirSim	0.973	0.974	0.846	0.975	0.923
PhraseSim	0.611	0.655	0.574	0.679	0.628

DAE Generalizes Better

Model Scores on Out-of-Distribution Directions

	SVM/KNN	$DMLP_B$	$DMLP_S$	DAE_B	DAE_S
FPAcc	21.749	25.697	11.515	31.103	9.269
FDAcc	0.449	0.044	0.789	-0.222	0.869
ModSim	0.471	0.453	0.491	0.489	0.520
DirSim	0.648	0.626	0.667	0.607	0.634
PhraseSim	0.560	0.540	0.579	0.548	0.577

SBERT Improves Generalization