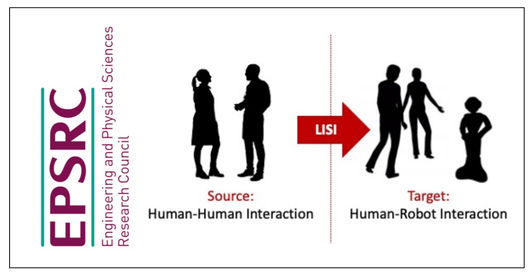
Context-Aware Body Gesture Generation for Social Robots

Nguyen Tan Viet Tuyen and Oya Celiktutan
{tan_viet_tuyen.nguyen; oya.celiktutan}@kcl.ac.uk
Social AI and Robotics Lab, Centre for Robotics Research, Department of Engineering,
King's College London, London, United Kingdom



Introduction

Social robots will progressively become widespread in many aspects of our daily lives, including education, healthcare, workplace, and home. Such practical applications require social interaction between humans and robots.

Social robots should therefore engage in interactions in a human-like manner. Along with verbal communication, successful interaction is closely coupled with the exchange of nonverbal signals.

This study proposes an approach for generating robots' non-verbal behaviours in affective human-robot interaction (HRI).

Related Works

The approach to non-verbal behaviours generation can be broadly divided into two groups: rule-based and data-driven.

***Rule-based Approach*:** it requires the design of interaction logic manually [1]. Once fixed, it will be limited, not transferrable to unseen interaction contexts, and not robust to unpredicted inputs from the robot's environment.

Data-driven Approach: the relationships between non-verbal behaviours and speech are determined through end-to-end learning manners [2]. However, a few works aim to generate communication behaviours by taking into consideration of interaction contexts encoded in the interacting partner's social signals [3].

Methodology

***Problem Definition*:** finding a mapping function F that receives the speech $A_{fo}^{0:T}$ of the target person, the interacting partner's speech $A_{ob}^{0:T}$, and body gesture $P_{ob}^{0:T}$ in order to predict the body gesture $\hat{P}_{fo}^{0:T}$ of the target person.

This study proposes a framework with context-awareness based on the conditional generative adversarial network.

The model consists of Context Encoder E, Generator G, and Discriminator D.

Encoder E encodes social signals simultaneously collected from the interacting partner in dyadic interaction into a contextual.

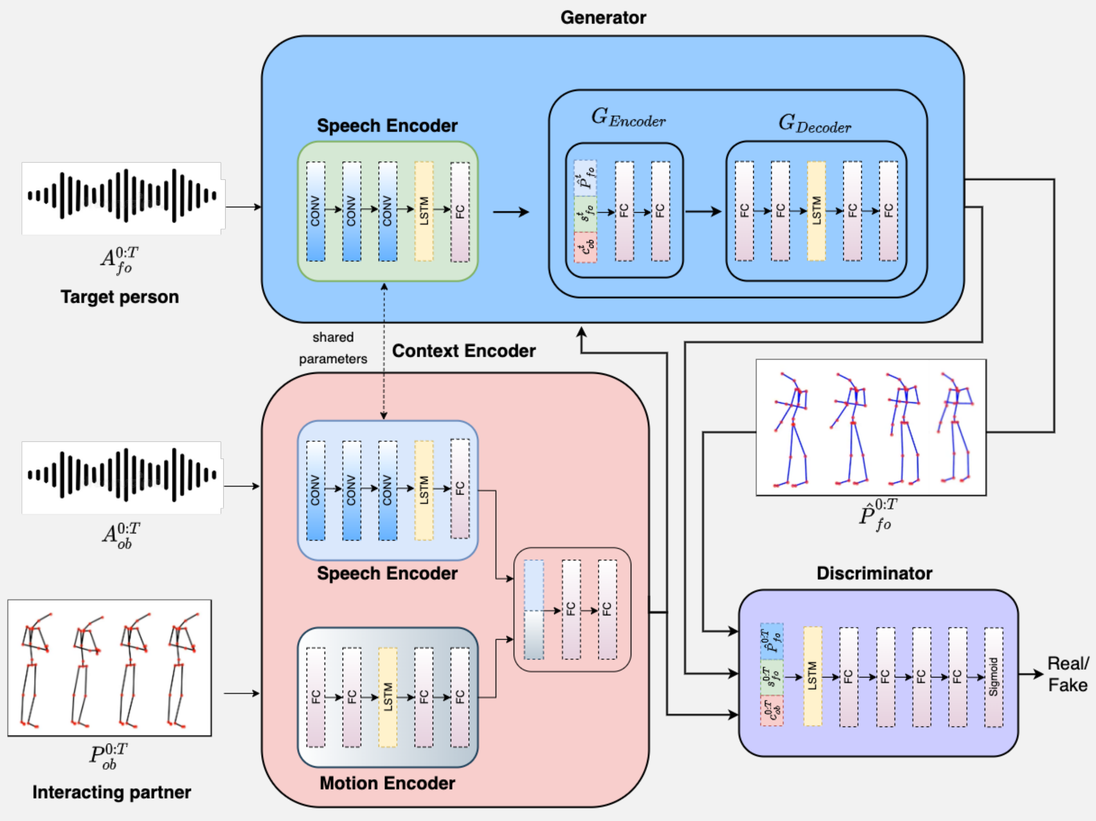


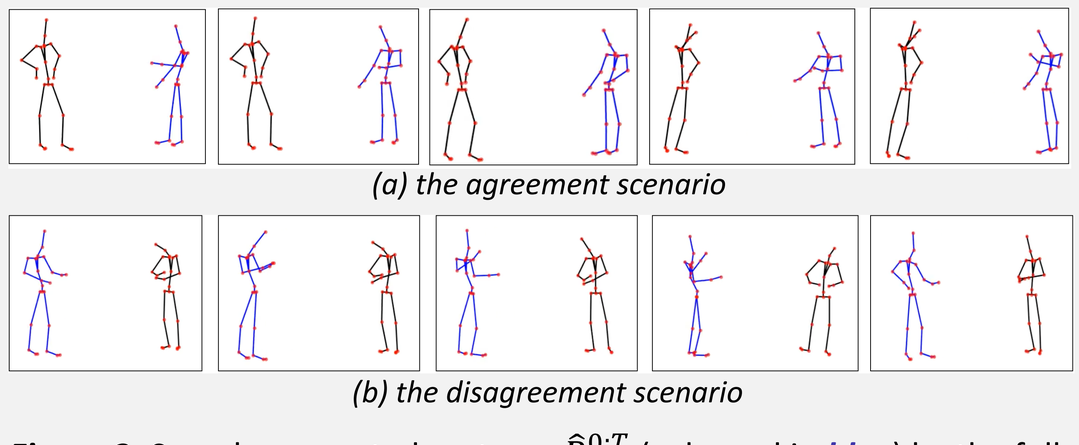
Figure 1. The proposed framework for generating body gestures of the target person from their speech (or audio) and affective contextual cues.

Experimental Results

The model was validated on the JESTKOD dataset, a time-synchronised speech and gesture dataset in affective dyadic interactions.

Scenario	Model	APE (degree)	Acceleration (degree/s^2)	Jerk (degree/s^3)
Agreement	*Full Model*	3.966 ± 1.961	5.064 ± 0.870	134.418 ± 26.040
Agreement	*without Context Encoder*	4.917 ± 1.810	145.680 ± 38.366	3999.423 ± 995.027
Disagreement	*Full Model*	3.891 ± 2.207	6.270 ± 1.448	170.298 ± 41.463
Disagreement	*without Context Encoder*	5.752 ± 2.253	166.135 ± 45.301	4518.250 ± 1197.977

Table 1. Accuracy in terms of Average Position Error (APE), Acceleration, and Jerk with respect to the two type of affective scenarios, namely, agreement and disagreement.



(a) the agreement scenario

(b) the disagreement scenario

Figure 2. Sample generated gestures $\hat{P}_{fo}^{0:T}$ (coloured in ***blue***) by the fully implemented model from the agreement and disagreement scenario.

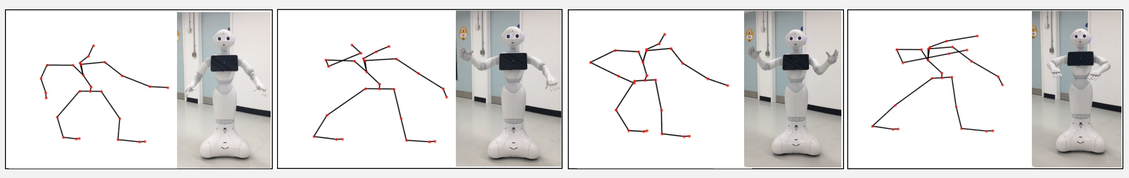


Figure 3. A generated gesture $\hat{P}_{fo}^{0:T}$ performed by the Pepper robot. Human skeleton (coloured in ***black***) represents the interacting partner motion $P_{ob}^{0:T}$.

Acknowledgement

This work has been supported by the "LISI - Learning to Imitate Nonverbal Communication Dynamics for Human-Robot Social Interaction" project, funded by the Engineering and Physical Sciences Research Council (Grant Ref.: EP/V010875/1).

References

1. Cassell, H. H. Vilhja´lmsson, and T. Bickmore, "Beat: the behavior expression animation toolkit". Life-Like Characters. Springer, 2004.
2. H. Ahn, T. Ha, Y. Choi, H. Yoo, and S. Oh, "Text2action: Generative adversarial synthesis from language to action". IEEE ICRA 2018.
3. Tuyen, Nguyen Tan Viet, and Oya Celiktutan. "Context-Aware Human Behaviour Forecasting in Dyadic Interactions". PMLR, 2022.