# Enhancing Emotional Nuance in AI Non-Player Characters: Literature Review

Harshal Mishra
University of Adeliade
a1888203
a1888203@adelaide.edu.au

## ABSTRACT

A.I Non-Player Characters cannot effectively communicate authentic emotions. This makes it harder for the player to connect with the NPCs hence making the game less immersive. If A.I training for emotions was changed to address 4 major pitfalls the expressions of the emotions would drastically improve allowing games to have greater immersion for players.

## 1 INTRODUCTION

Modern games are currently going through a revolution from traditional Non-player Characters to AI powered NPCs which would allow for a significant increase in immersive and engaging experiences within the game. The problem with traditional NPCs is that their scripted nature means that they cannot interact to the players actions dynamically breaking the illusion of immersion of the game world (Araujo et al., 2020).

AI NPCs are also superior in several other aspects than traditional NPCs. They can understand and react to not only the player inputs but the world around them making almost ever interaction with NPCs feel unique personalized and more genius. This adaptability is crucial for more large open world games where the players except a living world to roam. AI NPCs can achieve this through the integration of natural language processing and machine learning. To process player inputs in the game to make more human esque decisions (Yannakakis & Togelius, 2018).

The emotional behavior of NPCs is vital for player immersion. As AI advancements march restlessly onwards the demand for NPCs that display not just simple basic emotions but also more complex, nuanced and thought through emotional responses that can replicate human interactions. This development of authentic AI NPCs has been riddled with challenges such as accurately identifying and replicating human emotion (Scherer, 2005), the need for extensive and large datasets (Cowie et al., 2011), finally generating nuanced facial motions to replicate complex feelings (Krumhuber et al., 2013).

## 2 Problem definition

### 2.1 Difficulty in understanding and replicating human emotions

AI programs currently struggle with understanding human emotions, which is essential for generating authentic emotional expressions. The aspects that make human emotions so complicated are that they are multilayered and lay at the cross section of context, culture and induvial experiences. AI programs are likely trained on simplified models of human emotions such as the basic emotion theory which was proposed by Ekman. This model only categories emotions into 6 buckets happiness, sadness, fear, anger, surprise and disgust (Ekman, 1992)).

These types of models are incredibly useful for identifying broad emotional states however they fall well short of mapping the subtlety and variability of real human emotions. For example, the emotion anger can have vastly different manifestations depending on the context such as ranging in silence or irrupting in fury. These AI models would fail to find the distinguishing between these two emotional states (Krumhuber et al., 2013). In addition, AI programs struggle with the time dimension of an emotion, such as how emotions can change over time to different events. An example of this is how the AI in this paper failed to change emotional states after both attempts and escalations and de-escalation (Krumhuber et al., 2013).

### 2.2 Requirement of large and diverse data sets

The creation of AI NPCs that can respond to human emotions are heavily dependent on large datasets that encompass a wide range of human expression for the AI to be trained on.( Cowie et al., 2011) However, acquiring these datasets present unique challenges.

The collection of emotional datasets is incredibly resource intensive. These datasets require annotating a vast array of emotional expressions across different contexts and cultures and individual differences. They also require not only facial expressions but also vocal tones, body language and contextual information that help interpret how emotions are understood. As noted by Paier et al. (2021)the creation of just one single actor

video corpus for training NNs is a laborious process that require planning to ensure data is usable for the training purposes.

This process of annotation creates its own set of problems. As annotation are required to understand the nuance emotions being expressed in the moment. However, due to humans biases and inconsistences especially in complex emotional situations where distinction between certain emotions is difficult to ascertain the accuracy and reliability of these datasets become diminished.

Also, the reliance on human annotations impact scalability which is a significant challenge encountered in dataset creation. An example provided by Suwajanakorn et al. (2017)which emphasized how the use of neural networks to synthesize visual speech from audio was limited by the quality and quantity of the training data which also often lacked the diversity required to cover the full range of human emotions.

Finally another challenge rests in the unique temporal dynamic of emotions. Emotions not being static and subtly ever changing to respond to internal and external stimuli (Paier et al., 2021). However most current datasets for training models only include static images or short videos that fail to demonstrate this nuance about emotions (Paier et al., 2021).

### 2.3 Difficulty in Creating Nuanced emotional expressions

Generating nuanced emotions present a monumental challenge for AI. This is due to the fact that emotions are not binary and land on a spectrum where humans express their relative position on said spectrum through very subtle ques. The ability to produce these subtle and nuanced expressions is critical to creating an NPCs that can be believable enough to resonate with the player allowing for deeper emersion (Krumhuber et al., 2013).

A technical challenge that presents itself is the AIs ability to make subtle variations in facial muscles movements that accurately correspond to different emotional states (Paier et al., 2021). This is because the current standard for facial animations set by facial action coding system (FACS) is limited in its ability to map complex niece facial expressions. FACS does provide a framework for categorizing facial expressions, however it doesn't provide enough detail to capture the interplay between small facial muscles that impact the context of an emotional response (Ekman & Friesen, 1978) for example a forced smile.

New research hopes to combat this problem by applying auto regressive models to generate more realistic facial animations. However, even these advancements struggle with the emotions timing, intensity and the surrounding context (Paier et al., 2021). This lack of surrounding context is needed for the regulation of the emotions being expressed. For example, failing to adjust the emotion shown according to the players tone of voice or body language.

Finally, the understanding of these subtle nuances requires the knowledge of social and cultural norms. These things change from place to place hence the emotions being expressed and interpreted should change accordingly. This is impactful as the lack of cultural understanding can lead to some players interpreting the interaction as either disrespectful, strange or offensive which may result in the emersion being broken.

## 3 Motivations

These challenges that are currently encountered when trying to make authentic AI NPCs need to be overcome for enhancing player immersion and engagement in games. Improving the AIs understanding of human emotions will lead to NPCs that generate more contextually appropriate responses resulting in more believable interactions. Reducing the dependency on large training datasets will increase the rate at which the development process occurs making it more scalable while still being able to produce convincing response. Finally, refining the generation of nuanced emotions will enable NPCs to convey the complexity and depth to the human emotions which will create a richer gaming experience.

## 4 Literature Review

### 4.1 "Example-Based Facial Animation of Virtual Reality Avatars Using Auto-Regressive Neural Networks"

This paper written my Paier dives into how a hybrid approach to facial animation that combing both auto regressive neural networks with example-based methods to create realistic facial animations for virtual reality avatars. The methodology tackles the third problem outlined as it creates nuanced emotion expressions by leveraging an auto regressive NN.

### 4.2 "Synthesizing Obama: Learning Lip Sync from Audio"

Suwajankorn explored a method of synthesizing lip sync from audio input with a focus on creating realistic facial animations the former U.S Presided Barack Obama. This was done by training a neural network to predict mouth shapes based on audio data. This paper explores the second problem as even with a large target dataset for an individual there were several problems with the AI limiting its expressions.

### 4.3 "Effects of Dynamic Aspects of Facial Expressions: A Review"

This review focuses on the importance of dynamic aspects of perception in recognition of facial expressions. This paper highlights how dynamic facial expressions in contrary to static images drastically increase the accuracy of the emotional recognition and increase the perceived intensive and authenticity of the emotions on display. This relates to the first problem and how understanding human emotions has a temporal dimension

### 4.4 "Facial Emotion Expression Corpora for Training Game Character Neural Network Models""

This Paper explores the creation and validation of facial emotion expressions for the sole purpose of training neural networks to drive NPC facial animations. The authors use real actors to help the model generate life like animations. This paper addresses both the first and second problem. It explores how the requirements for large datasets and the difficulty in replicating real nuanced human emotions. The authors also talk about a potential solution as they suggest training NPCs on specific NPCs requiring the size of the dataset and increasing the accuracy of the emotion replication

### 4.5 "Measuring Emotion Velocity for Resemblance in Neural Network Facial Animation Controllers and Their Emotion Corpora"

This paper explores a unique approach for evaluating the accuracy of a neural network. It suggests by measuring the velocity of emotional expression and comparing them to human actor performances this ensures the temporal dynamics of the expressions are preserved in the animations. This methodology is critical as it provides a potential solution for the third problem as it is able to create more nuanced emotional expressions that can change across time.

## 5 REFERENCES


 Araujo, P. M., Reis, L. P., & Lau, N. (2020). Affective computing in digital games: Emotions and influence on game experience. *Journal of Multimodal User Interfaces*, 14(3), 271-281.

 Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., & Schroder, M. (2011). 'FEELTRACE': An instrument for recording perceived emotion in real-time. *Proceedings of the 9th European Conference on Speech Communication and Technology*, 1655-1658.

 Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3-4), 169-200.

 Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Consulting Psychologists Press.

 Kappas, A. (2010). Smile when you read this, whether you like it or not: Conceptual challenges to affective computing. *IEEE Transactions on Affective Computing*, 1(1), 38-41.

 Krumhuber, E. G., & Scherer, K. R. (2011). Affect bursts: Dynamic patterns of facial expression. *Emotion Review*, 3(4), 442-443.

 Krumhuber, E., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, 5(1), 41-46.

 Paier, W., Hilsmann, A., & Eisert, P. (2021). Example-based facial animation of virtual reality avatars using auto-regressive neural networks. *IEEE Computer Graphics and Applications*, 41(4), 47-58.

 Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695-729.

 Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics*, 36(4), 95-108.

 Yannakakis, G. N., & Togelius, J. (2018). *Artificial Intelligence and Games*. Springer.


## 6 Project Plan

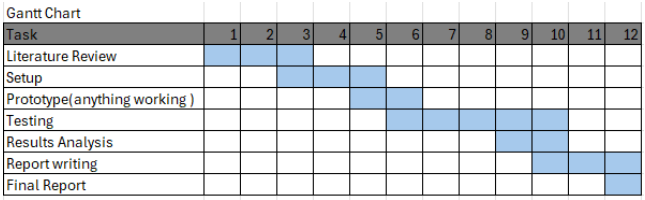


Figure 1: Shows Gantt chart of Project plan

The plan is when the literature review is completed by the end of week 3 the set-up phase can begin. I will install all necessary files and programs then initialize the project repository and make any necessary configuration. This is the most essential part of the plan and has the most chances of going wrong so I will insure to keep my supervisor and other group members update for assistance. Then after the setup process I will rapidly prototype with aiming to have the bare bones working by week 6. At which point I will begin the testing phases. Then after said test are analyzed I can get started on the report. Throughout this process I will ensure to document all steps taken to help with debugging and report writing.

Back up plan
In case any feature or functionality does not function then I will document and query my supervisor. If still unsuccessful I will research and use all other external tools such as other experts in the field. If my current improvement is not successful I will explore other ideas from highlighted in the original meeting.