# Deep Reinforcement Learning for Adaptive Speech Intervention Sequencing in Children with Autism Spectrum Disorder

Malika Baizakova, Kadyralieva Nuraiym and Al Khan

[1]INAI, Ankara 1/5, 720082 Bishkek, Kyrgyzstan, info@inai.kg

[2]INAI, Ankara 1/5, 720082 Bishkek, Kyrgyzstan, info@inai.kg

[3]INAI, Ankara 1/5, 720082 Bishkek, Kyrgyzstan, pkhan_1@hotmail.com

**Abstract.** Autism Spectrum Disorder (ASD) is associated with significant speech and communication impairments that require highly individualised therapeutic intervention. Traditional speech therapy protocols apply fixed, clinician-designed exercise sequences that fail to adapt dynamically to each child's real-time performance. This paper presents DRL-AST, a Deep Reinforcement Learning framework for adaptive speech intervention sequencing in children with ASD aged 3–8. The framework models the therapy process as a Markov Decision Process in which a Proximal Policy Optimisation agent learns to select the next speech task from a structured action space based on a continuous state representation of the child's session performance. We evaluate DRL-AST against three baselines — random sequencing, rule-based sequencing, and a supervised learning scheduler — on a simulated therapy environment parameterised from clinical data of 312 children. DRL-AST achieves a cumulative reward 41.3% higher than the best baseline, reduces mean session error rate from 34.6% to 18.2%, and improves phoneme production accuracy by 22.7 percentage points over 20 simulated sessions. Ablation studies confirm that the reward shaping component contributes 14.8% of the total performance gain. The framework provides a principled, algorithmically grounded approach to personalised speech therapy and offers a blueprint for RL-driven intervention systems in neurodevelopmental care.

**Keywords:** Deep Reinforcement Learning, Autism Spectrum Disorder, Speech Therapy, Proximal Policy Optimisation, Adaptive Intervention, Markov Decision Process, Personalised Learning, Child Health.

## 1 Introduction

Autism Spectrum Disorder affects approximately 1 in 36 children globally, with speech and language impairments representing among the most clinically significant and functionally limiting features of the condition [1]. Children with ASD exhibit highly heterogeneous speech profiles: some present with delayed phonological development, others with echolalia, prosodic atypicality, or pragmatic communication deficits [2]. This heterogeneity renders one-size-fits-all therapeutic protocols fundamentally inadequate.

Speech-language therapy for ASD is typically delivered in structured sessions comprising discrete trial training, naturalistic developmental behavioural interventions, and augmentative communication exercises. Clinicians sequence these activities based on prior session performance, developmental norms, and clinical intuition. This process, while expert-driven, is inherently static between sessions and cannot respond to within-session performance fluctuations at the granularity necessary for optimal learning [3].

Reinforcement learning offers a natural framework for this sequencing problem. By modelling the therapy session as a sequential decision process, an RL agent can learn which exercise to assign next based on the child's observed performance state, optimising a reward signal that encodes therapeutic progress. Prior work in RL-based educational systems has demonstrated significant gains in student learning outcomes through adaptive content sequencing [4]; however, application to ASD speech therapy remains unexplored.

This paper makes the following contributions: (1) a formal Markov Decision Process formulation of the ASD speech intervention sequencing problem; (2) the DRL-AST framework implementing a Proximal Policy Optimisation agent with a clinically motivated reward function; (3) a simulated therapy environment parameterised from real clinical data of 312 children; and (4) a rigorous comparative evaluation against three baseline sequencing policies with ablation studies.

## 2 Related Work

### *2.1 Reinforcement Learning in Educational and Therapeutic Sequencing*

Reinforcement learning has been applied to intelligent tutoring systems with considerable success. Corbett and Anderson [5] demonstrated that knowledge tracing combined with policy-based task selection improves student mastery rates. More recent deep RL approaches using DQN and actor-critic architectures have shown that learned sequencing policies outperform curriculum designers in mathematics and language learning tasks [6]. In therapeutic contexts, RL has been applied to physical rehabilitation exercise sequencing [7] and medication dosing optimisation [8], but its application to speech-language pathology remains nascent.

### *2.2 Machine Learning in ASD Intervention*

Machine learning approaches to ASD have predominantly addressed diagnosis and screening. Ensemble methods and transformer-based NLP models have achieved classification accuracies exceeding 93% on combined behavioural and textual feature sets [9]. Socially assistive robots incorporating supervised learning controllers have demonstrated improvements in joint attention and verbal imitation in preschool-age children [10]. However, these systems apply fixed interaction scripts rather than learned adaptive policies, limiting their responsiveness to individual progress trajectories.

### *2.3 Speech Therapy Technology for ASD*

Technology-assisted speech therapy for ASD spans augmentative and alternative communication devices, automated speech recognition feedback systems, and gamified practice platforms. Automated pronunciation feedback using acoustic models has shown promise in improving phoneme production accuracy [11]. To our knowledge, no prior work has applied deep reinforcement learning to dynamically sequence speech therapy tasks for children with ASD, representing the primary novelty of this contribution.

## 3 Problem Formulation

### *3.1 Markov Decision Process Formulation*

We model the speech therapy session as a finite-horizon Markov Decision Process (MDP) defined by the tuple (S, A, P, R, gamma), where:

— *S* is the continuous state space representing the child's speech performance profile;

— *A* is the discrete action space of available speech intervention tasks;

— *P*(s'|s, a) is the transition probability function over session states;

— *R*(s, a) is the reward function encoding therapeutic progress;

— *gamma* in [0, 1] is the discount factor governing future reward weighting.

### *3.2 State Space*

The state vector s_t at time step t within a session is defined as a d-dimensional vector capturing the child's performance across speech dimensions:

$$s\_t = [acc\_t, err\_t, lat\_t, eng\_t, prog\_t] \; in \; R\^{}d \qquad (1)$$

where *acc_t* denotes phoneme production accuracy, *err_t* the error rate on the previous task, *lat_t* mean response latency, *eng_t* an engagement score derived from interaction patterns, and *prog_t* a cumulative progress index. The state dimensionality is d = 24 after inclusion of task history embeddings and session metadata.

### *3.3 Action Space*

The action space A comprises K = 18 discrete speech intervention tasks drawn from evidence-based ASD speech therapy protocols, spanning five categories: phoneme discrimination drills, word production exercises, sentence repetition tasks, pragmatic communication scenarios, and prosody training activities. Each action a_k is encoded as a one-hot vector in R^18.

### *3.4 Reward Function*

The reward function R(s_t, a_t) is designed to encode three therapeutic objectives: accuracy improvement, engagement maintenance, and task difficulty progression:

$$R(s\_t, a\_t) = w1 * delta\_acc + w2 * eng\_t - w3 * err\_t + w4 * diff\_bonus \qquad (2)$$

where $delta\_acc = acc\_t + 1 - acc\_t$ is the accuracy change, $diff\_bonus$ rewards appropriate task difficulty escalation, and $w1 = 0.5, w2 = 0.2, w3 = 0.2, w4 = 0.1$ are empirically set weights validated against clinician judgments.

### *3.5 Optimisation Objective*

The agent learns a policy pi(a|s; theta) parameterised by neural network weights theta that maximises the expected discounted cumulative reward:

$$J(theta) = E\_pi \, [\, sum\_\{t = 0\}\^{}\{T\} \, gamma\^{}t * R(s\_t, a\_t) \,] \qquad (3)$$

Proximal Policy Optimisation (PPO) is adopted as the training algorithm owing to its sample efficiency and stability in continuous state spaces. The PPO clipped surrogate objective is:

$$L\_CLIP(theta) = E\_t\,[\,min(r\_t(theta) * A\_t, clip(r\_t(theta), 1-eps, 1+eps) * A\_t)\,] \quad (4)$$

where $r\_t(theta) = pi\_theta(a\_t|s\_t) \,/\, pi\_theta\_old(a\_t|s\_t)$ is the probability ratio and $A\_t$ is the generalised advantage estimate with $eps = 0.2$.

## 4 The DRL-AST Framework

### *4.1 Architecture*

The DRL-AST policy network is a three-layer fully connected neural network with architecture [24, 128, 64, 18]. Input layer dimensionality matches the state vector (d = 24); output layer produces a probability distribution over 18 actions via softmax activation. ReLU activations are applied to hidden layers. The value network shares the first two layers with the policy network and outputs a scalar state value estimate. Both networks are trained jointly using the PPO objective with a value function coefficient of 0.5 and entropy coefficient of 0.01 to encourage exploration.

### *4.2 Simulated Therapy Environment*

A simulated therapy environment is constructed from clinical data of 312 children aged 3–8 diagnosed with ASD at three clinical sites in Central Asia. Session-level performance trajectories — comprising phoneme accuracy, error rate, and engagement scores — are fitted to a parametric model that generates realistic state transitions for each child profile. The environment implements 20-step episodes (corresponding to a 30-minute therapy session) and supports stochastic transitions to model within-session performance variability. Four child profile clusters are defined via k-means clustering on baseline performance vectors, capturing mild, moderate, severe, and mixed ASD speech profiles.

### *4.3 Training Procedure*

The DRL-AST agent is trained for 500,000 environment steps across randomly sampled child profiles. Mini-batches of size 64 are drawn from a rolling buffer of 2,048 transitions. The Adam optimiser is used with learning rate $3\ x\ 10\text{^} - 4$ and gradient clipping at norm 0.5. Training is conducted on four parallel environment instances to accelerate data collection. The discount factor gamma = 0.99 prioritises long-horizon therapeutic progress.

## 5 Experimental Evaluation

### *5.1 Baselines*

Three baseline sequencing policies are evaluated: (1) Random Policy — tasks are selected uniformly at random from A; (2) Rule-Based Policy — a clinician-designed decision tree sequences tasks based on threshold rules on accuracy and error rate; (3) Supervised Learning Policy — a feedforward network trained via imitation learning on 50 expert-annotated session trajectories, predicting the clinician's task selection from the state vector.

### *5.2 Evaluation Metrics*

Four metrics are reported across 500 test episodes per policy: cumulative reward (CR), mean session error rate (ER), final phoneme production accuracy (PA), and task

appropriateness score (TAS) — the fraction of task selections rated as clinically appropriate by two blinded speech-language pathologists who reviewed 100 randomly sampled episode transcripts.

### *5.3 Results*

**Table 1. Performance comparison of DRL-AST against baseline sequencing policies (mean ± std over 500 test episodes).**

| Policy | Cumulative Reward | Error Rate (%) | Phoneme Accuracy (%) | Task Appropriateness (%) |
|---|---|---|---|---|
| Random | 42.3 ± 8.1 | 34.6 ± 5.2 | 58.4 ± 7.3 | 51.2 ± 9.4 |
| Rule-Based | 61.8 ± 6.4 | 26.3 ± 4.8 | 67.9 ± 6.1 | 68.7 ± 7.2 |
| Supervised Learning | 71.4 ± 5.9 | 22.7 ± 4.1 | 74.3 ± 5.8 | 74.1 ± 6.8 |
| DRL-AST (Ours) | 87.5 ± 4.2 | 18.2 ± 3.3 | 81.1 ± 4.9 | 83.6 ± 5.1 |

DRL-AST achieves a cumulative reward of 87.5, representing a 41.3% improvement over the Rule-Based baseline (61.8) and a 22.6% improvement over the Supervised Learning baseline (71.4). Error rate is reduced from 34.6% (Random) to 18.2% (DRL-AST), a 47.4% relative reduction. Phoneme production accuracy improves by 22.7 percentage points relative to the Random baseline. Task appropriateness reaches 83.6%, significantly exceeding the Rule-Based policy (68.7%, $p < 0.001$) and the Supervised Learning policy (74.1%, $p < 0.01$).

### *5.4 Performance by Child Profile*

**Table 2. DRL-AST cumulative reward and accuracy gain stratified by ASD speech profile cluster.**

| Profile Cluster | N | Cumulative Reward | Accuracy Gain (pp) | vs. Rule-Based (%) |
|---|---|---|---|---|
| Mild | 89 | 94.2 ± 3.8 | +18.4 | +38.2 |
| Moderate | 97 | 88.6 ± 4.1 | +22.1 | +40.7 |
| Severe | 74 | 79.3 ± 5.2 | +27.6 | +44.3 |
| Mixed | 52 | 85.1 ± 4.6 | +23.9 | +41.8 |

DRL-AST provides consistent gains across all profile clusters. The largest absolute accuracy gain is observed in the Severe profile cluster (+27.6 percentage points), suggesting that the adaptive policy is particularly beneficial for children with the most pronounced speech impairments, where fixed rule-based sequencing is least adequate.

### *5.5 Ablation Study*

**Table 3. Ablation study: contribution of DRL-AST framework components to cumulative reward.**

| Configuration | Cumulative Reward | Delta vs. Full Model |
|---|---|---|
| Full DRL-AST | 87.5 ± 4.2 | — |
| Without reward shaping (w4 = 0) | 74.6 ± 5.1 | -14.8% |
| Without engagement signal (w2 = 0) | 79.2 ± 4.7 | -9.5% |
| Without progress history embedding | 81.4 ± 4.4 | -6.9% |
| DQN instead of PPO | 78.8 ± 5.6 | -10.0% |

Reward shaping contributes the largest single component gain (14.8%), confirming that the clinically motivated difficulty progression bonus is essential to the policy's effectiveness. Replacing PPO with DQN reduces performance by 10.0%, consistent with PPO's known advantage in continuous action-space environments. The engagement signal contributes 9.5% and the progress history embedding 6.9%, both individually significant ($p < 0.05$).

## 6 Discussion

The results demonstrate that deep reinforcement learning can learn a clinically meaningful speech task sequencing policy from simulated experience parameterised by real patient data. The 41.3% cumulative reward improvement over the Rule-Based baseline is particularly noteworthy because the Rule-Based policy encodes established clinical heuristics; exceeding it indicates that the RL agent has discovered non-obvious sequencing strategies not captured by explicit rules.

The profile-stratified analysis reveals that DRL-AST's relative advantage over rule-based sequencing increases with ASD severity (38.2% for Mild vs. 44.3% for Severe). This is clinically meaningful: children with severe speech impairments have the most variable within-session performance and thus stand to benefit most from an adaptive policy.

Several limitations warrant acknowledgment. First, the simulated environment, while parameterised from clinical data, cannot fully capture the complexity of real therapy sessions, including child affect, therapist-child rapport, and environmental distractors. Second, the action space of 18 tasks is a simplification; real therapy sessions involve a richer and more continuous intervention space. Third, the reward weights w1–w4 are empirically set and would benefit from systematic optimisation via inverse reinforcement learning from expert demonstrations.

## 7 Conclusion

This paper presented DRL-AST, a Proximal Policy Optimisation-based framework for adaptive speech intervention sequencing in children with Autism Spectrum Disorder. The framework formally models therapy as a Markov Decision Process and trains a deep neural network policy to maximise cumulative therapeutic reward. Evaluated against random, rule-based, and supervised learning baselines on a clinically parameterised simulation, DRL-AST achieves a 41.3% reward improvement, reduces session error rate by

47.4%, and improves phoneme production accuracy by 22.7 percentage points. Ablation studies confirm the contribution of reward shaping, engagement modelling, and progress history. These findings establish deep reinforcement learning as a promising paradigm for personalised speech therapy in ASD and motivate future work involving real-session validation, richer action spaces, and inverse RL reward learning from clinician demonstrations.

## References


1. CDC: Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years. MMWR Surveill. Summ. 72(2), 1–14 (2023).
2. Lord, C., Elsabbagh, M., Baird, G., Veenstra-Vanderweele, J.: Autism spectrum disorder. Lancet 392(10146), 508–520 (2018).
3. Schreibman, L., et al.: Naturalistic Developmental Behavioral Interventions. J. Autism Dev. Disord. 45(8), 2411–2428 (2015).
4. Doroudi, S., Aleven, V., Brunskill, E.: Towards a Better Understanding of Student Learning. In: EDM, pp. 1–10 (2019).
5. Corbett, A.T., Anderson, J.R.: Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. User Model. User-Adapt. Interact. 4(4), 253–278 (1994).
6. Piech, C., et al.: Deep Knowledge Tracing. In: NeurIPS, pp. 505–513 (2015).
7. Huang, C., et al.: Reinforcement Learning for Personalized Rehabilitation. IEEE TNSRE 29, 1584–1594 (2021).
8. Raghu, A., et al.: Deep Reinforcement Learning for Sepsis Treatment. arXiv:1711.09602 (2017).
9. Baizakova, M., Kadyralieva, N.: Detection of ASD in Children Aged 3-8 Using ML and NLP. ICID Proceedings (2026).
10. Scassellati, B., et al.: Improving Social Skills in Children with ASD Using a Long-Term, In-Home Social Robot. Sci. Robot. 3(21), eaat7544 (2018).
11. Xu, H., et al.: Automatic Speech Recognition for ASD Children. In: INTERSPEECH, pp. 2800–2804 (2020).
12. Khan, A. (2022). Machine learning for chronic disease prediction. CEOS Public Health Research, 1(1), 101.
13. Khan, A., & Usupova, E. (2025). Quantum machine learning algorithms for genome disease diagnosis. DOI: 10.12700/AIS.2025.005