

Virtual Speech Therapist: A Clinician-in-the-Loop AI Speech Therapy Agent for Personalized and Supervised Therapy

Shakeel SHEIKH

shakeelzmail1608@gmail.com

Novartis Institute of Biomedical Research/The Kashmir Hub for Artificial Intelligence Research

Patrick Marmaroli

Microsoft / Vocametrix, Tallinn, Estonia.

MD Sahidullah

IAI, TCG CREST, West Bengal, India.

Slim Ouni

Université de Lorraine, CNRS, Inria, LORIA, Nancy, 54000, France.

Fabrice Hirsch

Laboratoire Praxiling, UMR5267, CNRS et Université Paul-Valéry Montpellier 3, France.

Gonçalo Leal

Speechcare iStutter, Portuguese Catholic University, Portugal.

Björn W. Schuller

CHI – Chair of Health Informatics, TUM University Hospital, Germany. /GLAM – Group on Language, Audio, & Music, Imperial College London, United Kingdom

Article

Keywords: AI Agents, Virtual Speech Therapist, Pathological Speech, Digital Health, Stuttering

Posted Date: March 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9204155/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Virtual Speech Therapist: A Clinician-in-the-Loop AI Speech Therapy Agent for Personalized and Supervised Therapy

Shakeel A. Sheikh^{1*}, Patrick Marmaroli^{2*}, Md Sahidullah³,
Slim Ouni⁴, Fabrice Hirsch⁵, Gonçalo Leal⁶, Björn W. Schuller⁷

^{1*}Novartis Institute of Biomedical Research, Novartis AG, Basel, 4056,
Switzerland.

¹The Kashmir Hub for Artificial Intelligence Research.

²Microsoft / Vocametrix, Tallinn, Estonia.

³IAI, TCG CREST, West Bengal, India.

⁴Université de Lorraine, CNRS, Inria, LORIA, Nancy, 54000, France.

⁵Laboratoire Praxiling, UMR5267, CNRS et Université Paul-Valéry
Montpellier 3, France.

⁶Speechcare iStutter, Portuguese Catholic University, Portugal.

⁷CHI – Chair of Health Informatics, TUM University Hospital, Germany.

⁷GLAM – Group on Language, Audio, & Music, Imperial College
London, United Kingdom.

*Corresponding author(s). E-mail(s): shakeelzmail608@gmail.com;
patrick.marmaroli@gmail.com, schuller@ieee.org;

Contributing authors: md.sahidullah@tcgcrest.org; slim.ouni@loria.fr;
fabrice.hirsch@univ-montp3.fr; goncalo.leal@speechcare.center;
schuller@ieee.org;

Abstract

This paper develops Virtual Speech Therapist (VST)*, an intelligent agent-based platform that streamlines stuttering assessment and delivers customized therapy planning through automated and adaptive AI-driven workflows. VST integrates state-of-the-art deep learning-based stuttering classification, and multi-agent large language model (LLM) reasoning to support evidence-based clinical

*Code: <https://github.com/pmarmaroli/vocametrix-platform/tree/main/python/vstagent>

decision-making. The VST begins with the acquisition and feature extraction of patient speech samples, followed by robust classification of stuttering types. Building on these outputs, VST initiates an agentic reasoning process in which specialized LLM agents autonomously generate, critique, and iteratively refine individualized therapy plans. A dedicated critic agent evaluates all generated therapy plans to ensure clinical safety, methodological soundness, and alignment with peer-reviewed evidence and established professional guidelines. The resulting output is a comprehensive, patient-specific therapy draft intended for clinician review. Incorporating clinician feedback, the system then produces a finalized therapy plan suitable for patient delivery, thereby maintaining a clinician-in-the-loop paradigm. Experimental evaluation by expert speech therapists confirms that VST consistently generates high-quality, evidence-based therapy recommendations. These findings demonstrate the system’s potential to augment clinical workflows, reduce clinician burden, and improve therapeutic outcomes for individuals with speech impairments. An interactive user interface for the proposed system is available online at: <https://vocamatrix.com/ai/stuttering-therapy-planning-agent>, facilitating real-time stuttering assessment and personalized therapy planning.

Keywords: AI Agents, Virtual Speech Therapist, Pathological Speech, Digital Health, Stuttering

1 Introduction

Speech is a fundamental neuro-motor function central to human communication, self-expression, and social participation. Speech impairment—encompassing a spectrum of communication disorders characterized by deficits in the motor production of sounds essential for intelligible and fluent verbal output—can significantly diminish quality of life and daily functioning (Beilby 2014).

Stuttering, a prevalent neurodevelopmental disorder, is defined by involuntary disruptions in the forward progression of speech, characteristically expressed as sound or word repetitions, prolongations, and blocks (Guitar 2013). Epidemiological data indicate that stuttering affects nearly 5% of children at some stage in development and approximately 1% of the adult population worldwide (Yairi and Ambrose 2013; Reilly et al. 2013).

Given its developmental nature and potential long-term psychosocial impact, early and accurate identification of stuttering plays a critical role in determining effective clinical outcomes (Onslow 1992; Yairi et al. 2025). Beyond identification, personalized therapy strategies are crucial for effective stuttering management (SheikhBahaei et al. 2023). Accordingly, this work presents an integrated multi-agent framework for automated stuttering detection and personalized therapy generation.

Beyond identification, personalized therapy strategies are crucial for effective stuttering management (SheikhBahaei et al. 2023). Stuttering varies widely across individuals due to linguistic, cognitive, emotional, motor, and socio-demographic factors. Personalization in voice therapy may depend on age, educational background,

professional communication needs, and social interaction demands. For example, therapy for children often focuses on developmental fluency, whereas adults may require strategies tailored to academic or occupational communication. Additional considerations such as cultural and linguistic diversity, co-occurring disorders, and psychological factors further necessitate individualized intervention. Therefore, therapy planning informed by detailed speech analysis and patient-specific characteristics is essential for optimizing outcomes (Kaur 2025). Artificial Intelligence (AI) driven models enable adaptive and scalable personalized therapy recommendations. Accordingly, this work presents an integrated multi-agent framework for automated stuttering detection and personalized therapy generation.

1.1 Automated Stuttering Detection: From Handcrafted Features to Deep Learning

Traditional stuttering assessment has historically depended on manual auditory-perceptual judgments conducted by trained *speech-language pathologists* (SLP). These evaluations, while clinically valuable, are inherently time-consuming, expensive and susceptible to inter- and intra-rater variability due to the subjective nature of human perception (Guitar 2013; Yaruss et al. 2012; Lea et al. 2021). Such assessments create bottlenecks in clinical throughput and introduce inconsistencies, particularly in large-scale screening. The drive to augment and partially automate this process led to the development of computational methods for stuttering detection. Early research focused on conventional machine learning pipelines using handcrafted acoustic features (Sheikh et al. 2022c). Machine learning models such as support vector machines and hidden Markov models were trained on features like jitter, shimmer, and spectral features designed to capture the prosodic and spectral disruptions characteristic of disfluent speech (Sheikh et al. 2022c). While these systems demonstrated the feasibility of automated detection, their performance was fundamentally constrained by the representational limits of manually engineered features and their lack of generalizability across diverse speakers and recording conditions.

Recent advancements in deep learning have significantly transformed the speech domain by enabling models to learn hierarchical representations directly from raw signal or minimally processed speech data. This capability is exemplified by the effective application of convolutional neural networks to time–frequency representations like spectrograms, which capture the local spectral-temporal patterns associated with different disfluencies (Kourkounakis et al. 2020, 2021; Sheikh et al. 2022d, 2023a, 2021). To model the sequential nature of stuttered speech, recurrent neural networks, and particularly bidirectional long short-term memory networks, have been employed to capture long-range temporal dependencies and contextual information (Kourkounakis et al. 2021). Recent frameworks have successfully integrated metadata information via multi-task learning to address the challenge of speaker variability. Through adversarial learning, these models effectively unlearn (other) speaker characteristics, ensuring that the detected stuttering patterns are robust to the biological and acoustic diversity of the population (Sheikh et al. 2022d; Liu et al. 2025). More recently, transformer-based architectures such as wav2vec2 (Baevski et al. 2020), which leverage self-attention

mechanisms (Vaswani et al. 2017), have shown remarkable success in capturing long-range contextual interactions in speech signals, achieving state-of-the-art performance in stuttering detection (Sheikh et al. 2022c,a; Schuller et al. 2022). The automated stuttering detection has been further accelerated by benchmark competitions, such as the *Stuttering Sub-Challenge* within the *Computational Paralinguistics Challenge* (ComParE), which have provided standardized datasets and evaluation frameworks, fostering innovation and comparative analysis of state-of-the-art approaches (Schuller et al. 2022; Bayerl et al. 2023). Collectively, these advances have matured automated stuttering detection into a robust component of computational speech pathology and the broader field of computational paralinguistics. These developments indicate that state-of-the-art stuttering detection systems are approaching a level of reliability and generalizability that supports pilot deployment in real-world environments. With continued validation and clinician-in-the-loop integration, such technologies hold strong potential for translation into practical clinical workflows.

1.2 Towards Agentic AI in Speech Therapy: The Clinician-in-the-Loop Paradigm

While state-of-the-art detection models represent a critical technological foundation, they address only the diagnostic component of the therapeutic pipeline. The translation of these computational insights into effective, personalized intervention remains a significant challenge. Most existing digital tools function as passive practice platforms or telepractice conduits, lacking the dynamic, adaptive, and interactive capabilities required to simulate a therapeutic alliance or deliver truly responsive therapy.

This gap motivates the evolution towards *agentic AI systems*—intelligent agents that can perceive a user’s state (through speech, behavior, or engagement metrics), reason about optimal therapeutic strategies based on evidence-based guidelines, and execute personalized interventions (Sheikh et al. 2025; Karunanayake 2025). More importantly, the goal is not to replace the clinician but to augment their capabilities through a *clinician-in-the-loop* (CITL) paradigm (Tang et al. 2020). This framework delineates a collaborative workflow between an AI agent and a human clinician (Zou and Topol 2025). The framework accepts a speech sample, classifies its stuttering type, and processes an ASR-generated transcription to propose an initial, data-driven therapy plan. This plan is then presented to an SLP for clinical validation, modification, and approval. This division of labor aims to optimize the therapeutic process: the AI enables scalable, high-frequency intervention, while the SLP contributes indispensable clinical reasoning, psychosocial insight, and personalized therapeutic adaptation.

The CITL model is particularly vital in speech therapy, where the therapeutic relationship and the ability to adapt to a user’s emotional and motivational state are key determinants of success (Guitar 2013). An agentic system can maintain engagement through principles of motivational design (e.g., oral speech exercises, gamification, adaptive challenge), but it requires clinician approval and final feedback to ensure interventions remain ethically sound, clinically appropriate, and tailored to the individual’s holistic needs. Furthermore, by continuously collecting and analyzing granular

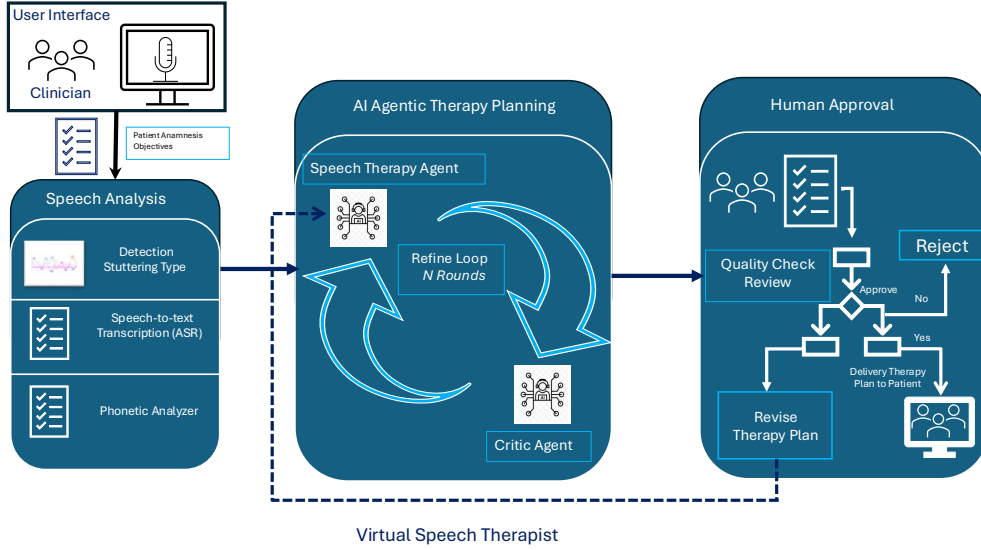


Fig. 1 End-to-end pipeline for a *virtual speech therapist* agent. The process begins with a user providing an input speech sample. This audio is first analyzed by a stuttering classification module, which identifies specific disfluency types such as *repetition*, *block*, or *prolongation*. Parallely, the speech sample is also passed to ASR and phonemizer modules. Based on this diagnostic output, phonetic and textual transcription, this *virtual speech therapist* generates a recommendation therapy for a personalized therapy session. Before being delivered to the user, this AI-generated therapy plan is sent to a CITL for review. The clinical speech therapist provides feedback to the VST output and comes up with a modified proposed therapy plan. The clinical speech therapist has the final authority to approve, reject, or request modifications to the therapy plan, ensuring clinical oversight and safety.

speech and interaction data, the AI agent can generate actionable insights for the clinician, transforming therapy from episodic assessments into a data-informed, continuous care model.

Although agentic AI has witnessed growing adoption across several medical and healthcare domains (Karunanayake 2025; Powell et al. 2026; Dietrich 2025), including diagnostic decision support, patient monitoring, and treatment recommendation systems, its application in speech and language therapy remains largely unexplored despite its significant potential. In particular, the integration of clinician-in-the-loop (CITL) paradigms within agentic AI frameworks for speech disorders has received minimal attention. Given the inherently interactive, behavioral, and psychosocial nature of speech therapy, the absence of clinically supervised agentic systems represents a critical research and translational gap. Addressing this gap is essential for developing trustworthy, clinically reliable, and ethically grounded AI-driven therapeutic solutions for speech impairments.

1.3 Proposed System: Virtual Speech Therapist

Building on the advancements in deep learning-based stuttering detection and the conceptual framework of agentic, CITL systems, we propose a novel, integrated *Virtual Speech Therapist* (VST) platform. This system is designed to bridge the gap between automated assessment and dynamic, personalized therapy generation. As outlined in Section 3, the VST’s architecture synergistically combines: (1) a **robust stuttering detection engine** built on state-of-the-art self supervised learning models for real-time fluency analysis; (2) a planning and reasoning **agentic module that formulates personalized therapy tasks** based on user performance, and evidence-based protocols; and (3) a **clinician-supervised interactive platform** where he reviews, modifies, and grants final approval for the treatment plan.

The clinician remains central to reviewing generated therapy plans, refining therapeutic goals, and validating AI-driven recommendations, thereby ensuring that interventions remain safe, effective, and human-centered. Maintaining standardized clinical protocols and enabling systematic evaluation through controlled clinical trials are essential for ensuring regulatory acceptance and real-world clinical adoption. This integrated approach aims to create a scalable, high-fidelity digital therapeutic that democratizes access to high-quality speech rehabilitation while empowering clinicians with advanced data-driven insights and decision-support tools. Consequently, such a system has the potential to extend evidence-based care to underserved regions with limited access to specialized SLPs. The following section presents an overview of the VST architecture and its underlying platform.

2 Methods

2.1 Virtual Speech Therapist

The proposed *agentic* VST transforms a standard stuttering detection pipeline into a dynamic, adaptive, and CITL-guided therapeutic experience. The workflow can be decomposed into four primary phases: (1) preprocessing, analysis, and detection, (2) therapy planning with critique feedback, (3) clinician oversight, and (4) delivery to patient.

2.2 Phase 1: Processing, analysis and detection

This phase involves processing the raw speech input to generate a structured, machine-readable diagnosis with following steps:

- **Step 1.1: Patient speech sample input:** In this phase, a speech sample is provided to the *virtual speech therapist* agent through a user interface¹, which supports both direct audio recording and audio file upload functionalities. The interface is designed to enable seamless user interaction while ensuring data integrity and standardized input formats. Once a speech sample is provided by a person who stutters

¹<https://www.vocametrix.com/ai/stuttering-therapy-planning-agent>

(PWS), the speech utterance undergoes preprocessing and is segmented into 3–5 second windows (default: 4 s) using a sliding window approach with an overlap of $k\%$ ($k \in \{0, 25, 50, 75\}$, default: 50%). This segmentation step prepares the data for subsequent feature extraction and analysis within the processing pipeline.

- **Step 1.2: Stuttering Detection:** The segmented speech chunks are processed by a pre-trained stuttering detection model (Section 3.2) to identify the type of stuttering. In parallel, an automatic speech recognition (ASR) module generates the corresponding textual transcription. The detected stuttering type, together with the transcribed text, is then forwarded to the *Virtual Speech Therapist* agent module for personalized therapy generation.

2.3 Phase 2: Initiation of Agentic Overflow

Following the detection phase described in the preceding module, the *Virtual Speech Therapist* system initiates an agentic orchestration workflow, a coordinated multi-agent process designed to generate, evaluate, and iteratively refine individualized therapy plans. This workflow comprises several specialized agents, each dedicated to a specific function within the therapy generation and evaluation pipeline. These include agents responsible for therapeutic content formulation and clinical relevance review based on LLM prompts described in *Appendix A*. The agent orchestration proceeds through \mathcal{N} iterative refinement rounds, wherein the system evaluates and adjusts the proposed therapy recommendations based on feedback from critic agent until convergence toward an optimized, patient-tailored therapy plan is achieved. This iterative, agentic feedback mechanism ensures both clinical validity and personalization of the therapeutic intervention. The structure of the multi-agentic *Virtual Speech Therapist* is depicted in Figure 1.

Prompt Selection:

The effectiveness and reliability of the *Virtual Speech Therapist* agent are strongly influenced by the design of its prompts. The prompts define the agent’s role, behavioral constraints, and response objectives, ensuring clinically appropriate, consistent, and safe interactions with PWS. The prompt was iteratively designed through a combination of domain expertise (SLPs and phoneticians) and empirical evaluation. Key design principles included: (i) adherence to evidence-based speech therapy practices, (ii) use of supportive and non-judgmental language, (iii) adaptation to the detected stuttering type and linguistic context, and (iv) avoidance of medical diagnosis beyond the system’s intended scope. These constraints were embedded to ensure alignment with ethical guidelines and to reduce the risk of hallucinated or inappropriate therapeutic advice. Multiple candidate prompts were evaluated using representative speech samples and stuttering profiles. Selection criteria focused on response relevance, therapeutic clarity, linguistic accuracy, and consistency across sessions. The final prompt demonstrated robust performance across stuttering categories while maintaining a patient-centered and empathetic communication style. For transparency and reproducibility, the complete system prompt used in the VST agent is provided in *Appendix A*.

Critic Prompt:

The preliminary therapy plan generated by the *TherapyAgent* is subsequently evaluated by the *CriticAgent*, a complementary reasoning module implemented using an LLM-based architecture. The behavior and evaluation criteria of the *CriticAgent* are governed by a dedicated critic prompt, which is described in detail in *Appendix A*. The critic prompt instructs the *CriticAgent* to review the therapy plan across six domains: (1) *clinical soundness* — alignment with stuttering type and severity, (2) *safety and psychological impact* — risks of physical strain, vocal misuse, or psychological harm, (3) *evidence alignment* — grounding in peer-reviewed research and clinical guidelines such as ASHA and IFA, (4) *structure, clarity, and practicality* — readability, feasibility, and usability for SLPs, (5) *improvements and missing elements* — specific, actionable revisions to sequencing, intensity, or coverage, and (6) *explainability and reasoning transparency* — verification that each strategy includes a complete `clinicalReasoning` chain with observation, rationale, expected outcome, and evidence base. For each domain, the *CriticAgent* produces structured observations, strengths, concerns, and concrete recommendations. When discrepancies, inconsistencies, or suboptimal therapeutic decisions are identified, the *CriticAgent* flags the affected components and generates explicit feedback to guide revision. This feedback typically includes suggested modifications, alternative therapeutic approaches, and references to established clinical protocols or empirical research that support the recommended changes. By functioning as a supervisory validation layer, the *CriticAgent* ensures that therapy plans generated within the *Virtual Speech Therapist* framework meet high standards of clinical validity, safety, and linguistic accuracy.

The internal architecture of the proposed *Virtual Speech Therapist* framework is structured around a multi-agent reasoning pipeline, comprising the following interconnected components:

1. **TherapyAgent:** Following the stuttering detection and transcription stages, the proposed agentic framework transitions into the therapy generation phase. In this phase, the system receives the diagnostic outputs produced by the detection modules, including the classified stuttering types (e.g., *prolongation*, *sound repetition*, etc.) and their corresponding phonetic and textual transcriptions. In addition, relevant acoustic characteristics are integrated to provide a comprehensive multimodal representation of the speaker’s disfluency profile. This set of diagnostic features is subsequently passed to the central reasoning module, referred to as *TherapyAgent*. The *TherapyAgent* constitutes an LLM-based component specifically configured for therapeutic reasoning and personalized recommendation generation (Section 3.3). Its configuration relies on a combination of predefined prompts, which encode clinical guidelines and behavioral therapy principles, patient-specific context as described in *Appendix A*. Exploiting this structured and context-enriched input, the *TherapyAgent* autonomously generates a detailed and adaptive therapy plan. This plan includes individualized fluency-shaping exercises, cognitive-behavioral tasks, and progression schedules aimed at addressing the identified disfluency types.
2. **CriticAgent:** The preliminary therapy plan generated by the *TherapyAgent* is subsequently evaluated by the *CriticAgent* (along with the critic prompt as described in *Appendix A*), a complementary reasoning component also implemented using an

LLM-based architecture (Section 3.3). The *CriticAgent* is designed to provide the critical feedback of the proposed therapy plan, encompassing linguistic coherence, therapeutic appropriateness, clinical relevance, and adherence to evidence-based speech-language pathology practices. It assesses the logical alignment between the identified stuttering types, the corresponding therapeutic exercises, and the progression strategies outlined by the *TherapyAgent*. When discrepancies, inconsistencies, or suboptimal therapeutic decisions are detected, the *CriticAgent* flags these components for revision and provides explicit, traceable feedback. This feedback typically includes suggested modifications, alternative therapeutic approaches, and relevant references to established clinical protocols or empirical research that support its recommendations. The *CriticAgent* thereby functions as a supervisory validation layer, ensuring that the therapy plans generated within the *Virtual Speech Therapist* framework meet rigorous standards of clinical validity and linguistic accuracy.

3. **Iterative Refinement and Agentic Loop:** The *Virtual Speech Therapist* framework employs an iterative refinement strategy governed by an agentic feedback loop that promotes progressive optimization of the therapy plan. Within this loop, the outputs produced by the *CriticAgent*—comprising evaluative comments, flagged inconsistencies, and suggested modifications—are recursively fed back to the *TherapyAgent*. This feedback mechanism enables the *TherapyAgent* to reassess, revise, and enhance the initial therapy recommendations, incorporating the critic’s guidance into updated therapeutic formulations. The refinement proceeds for a predefined number of iterations (\mathcal{N} , default $\mathcal{N} = 2$, configurable up to 5 via the user interface). Each iteration within this agentic loop represents a distinct reasoning cycle in which the *TherapyAgent* and *CriticAgent* interact cooperatively to minimize discrepancies, reinforce evidence-based decision-making, and improve personalization of the therapy plan. This LLM based agentic interplay effectively emulates a multi-expert review process, allowing the VST system to self-correct and refine therapeutic outputs before human intervention.
4. **Clinical Speech Therapist in the Loop:** Following the iterative refinement process, the optimized therapy plan advances to a critical clinical validation phase, which constitutes the final stage of the *Virtual Speech Therapist* workflow. This CITL mechanism serves as an essential safeguard, ensuring that all automated therapeutic recommendations align with clinical appropriateness, patient safety, and evidence-based standards of speech-language therapy practice. By integrating human oversight into the decision-making loop, the *Virtual Speech Therapist* maintains both therapeutic validity and ethical accountability while harnessing the efficiency of agentic automation. The clinical validation protocol proceeds through the following steps:
 - (a) **Clinical Expert Review:** A licensed SLP/*speech-language therapist* conducts a comprehensive evaluation of the therapy plan generated by the VST system. The review focuses on multiple dimensions, including (i) the therapeutic appropriateness of recommended interventions for the identified stuttering types, (ii) the feasibility and safety of proposed exercises, and (iii) the congruence of the

therapy plan with evidence-based clinical guidelines. The clinician also evaluates linguistic clarity, treatment progression, and personalization to the patient’s fluency profile.

- (b) **Decision:** Based on the expert assessment, the clinician exercises one of three possible actions:
- **Approval:** Direct endorsement of the generated therapy plan, confirming its clinical adequacy, safety, and potential for therapeutic benefit.
 - **Rejection:** Termination of the workflow if the plan contradicts patient-specific considerations or clinical judgment.
 - **Modification:** Provision of targeted, actionable recommendations for improvement, which may include modifications to therapy intensity, sequencing, or intervention type. The system resubmits the feedback to the therapy-critic agent loop for one automated refinement round.
- (c) **Last Refinement Loop:** If modifications are requested, the clinician’s feedback is reintegrated into the *TherapyAgent* for a final optimization cycle. The agentic model then revises the therapy plan by incorporating domain-specific corrections, such as adjusting phonetic targets, tailoring fluency drills, or modifying temporal pacing strategies. This step reflects a synergistic human-AI co-adaptive process, in which the clinician’s expertise directly informs the system’s generative reasoning.
- (d) **Final Authorization:** The refined therapy plan is subsequently resubmitted to the clinician for final evaluation. Two outcomes are possible:
- **Final Approval:** The clinician validates that the revised plan meets all therapeutic and safety criteria, authorizing it for patient delivery.
 - **Rejection:** The clinician determines that the plan remains suboptimal or clinically inappropriate, and simply rejects the generated therapy plan.

This closed-loop clinical validation framework ensures that automated therapy generation remains subordinate to human expertise, upholding the principles of clinical accountability, patient safety, and ethical responsibility. By embedding human oversight within the agentic workflow, the *Virtual Speech Therapist* exemplifies a translational paradigm in which computational efficiency complements, rather than replaces, professional judgment. This hybrid validation process represents a model for deploying LLM-driven systems within regulated clinical therapy contexts, maintaining transparency and clinical rigor at each stage of decision-making.

3 Experimental Setup

Dataset:

We utilize the SEP-28k-Extended (SEP-28k-E) dataset (Bayerl et al. 2022), an extension of the SEP-28k stuttering dataset² (Lea et al. 2021). The dataset comprises approximately 28k three-second audio clips sourced from podcasts, annotated by three trained non-clinicians. The annotations employ a multi-label scheme, capturing the following stuttering labels: *prolongation*, *block*, *sound repetition*, *word repetition*, and

²The largest publicly available dataset for stuttering detection.

interjection. We used this dataset to train our stuttering detection model for the *Virtual Speech Therapist*.

3.1 Models

This section describes the computational model employed for stuttering detection and the LLMs serving as the backbone for our proposed *Virtual Speech Therapist*.

3.2 Stuttering Detection

We adopted a dual-strategy approach for classification, leveraging the robust feature extraction capabilities of the pretrained wav2vec2-XLSR-53 model³. First, we extracted contextual embeddings from the frozen model to serve as input features for a downstream shallow neural network classifier. Second, to explore the potential for task-specific adaptation, we also performed full fine-tuning of the entire wav2vec2-XLSR-53 architecture. This comparative framework allows us to assess the relative merits of using the model as a fixed feature extractor versus an end-to-end trainable system for our specific classification objectives. The deployed web-based VST system currently uses the frozen-feature approach with the shallow classifier for local inference, while an Azure ML online endpoint hosting the fine-tuned XLSR-53 model is available as a configurable alternative.

Shallow Neural Network: To classify stuttering types, we train a lightweight, three-layer neural network⁴ on the SEP-28k dataset. This approach builds upon established methodologies for paralinguistic and disfluency analysis (Bayerl et al. 2023; Sebastian Peter Bayerl and Dominik Wagner and Elmar Noeth and Korbinian Riedhammer 2022; Sheikh et al. 2022b, 2023b, 2022a; Sheikh 2023). The input to the network consists of contextual embeddings extracted using state-of-the-art pretrained models, in particular, the multilingual XLSR-53 model (Conneau et al. 2020). These embeddings have been demonstrated to effectively capture a rich array of paralinguistic cues, including those directly relevant to stuttering, making them a robust learned representation for our task (Sheikh et al. 2022a; Bayerl et al. 2023,?; Sheikh 2023). As these embeddings are variable-length sequences, we first apply statistical pooling over the temporal dimension to convert them into a fixed 1024-dimensional feature vector. This vector is then propagated through the network’s two hidden layers, each containing 64 units with *exponential linear unit* (ELU) activations. Finally, a fully-connected output layer maps the processed features into a logit space corresponding to our target stuttering event classes, yielding a probability distribution for the classification of each disfluency type.

XLRS53 Finetuning: While leveraging pretrained features from the XLSR-53 wav2vec 2.0 architecture, we demonstrate that model fine-tuning provides considerable performance advantages in stuttering detection. We fine-tuned the complete XLSR-53 model on the SEP-28k-E dataset, adhering to the validated speaker-disjoint data

³*It is important to emphasize that enhancing stuttering detection accuracy is not the primary focus of this study. We assume the availability of state-of-the-art detection models that already achieve competitive performance. Rather, the principal contribution of this work is the design of an agentic framework for automated and adaptive therapy planning within the Virtual Speech Therapist system.*

⁴<https://github.com/shakeel608/stutternet/tree/main>

split (Bayerl et al. 2022) which minimizes overfitting and enables robust generalization (Bayerl et al. 2023). This fine-tuning strategy resulted in a substantial performance enhancement across most disfluency types. Notably, detection accuracy improved for sound, word repetition, and interjection categories. Crucially, even for the most challenging category, blocks, we see a modest but positive performance gain. This overall improvement came at the minimal cost of a negligible decrease in performance for the prolongation class as depicted in Table 1.

Table 1 Comparative F1 Scores with fine tuning (FT) XLR53 and without finetuning (%)

Class	Fine tuning	SOTA (No FT)
Soundrepetition	43.00	32.07
Wordrepetition	56.00	41.23
Block	32.00	31.02
Fluent	82.00	66.92
Interjection	77.00	51.63
Prolongation	44.00	46.23
Weighted Average F1	67.00	44.85

3.3 Large Language Models

With the rapid progress of LLMs across diverse downstream tasks spanning natural language processing, and speech (Mohamed et al. 2022), multiple model families have emerged, including OpenAI’s GPT, DeepSeek, Mistral, Anthropic’s Claude, Google’s Gemini amongst others. As most state-of-the-art LLMs require paid access tiers, we focus on a single family to ensure methodological consistency and cost-efficiency. Specifically, we adopt *Gemini 3 Pro (Preview)* (Comanici et al. 2025), selected for its favorable balance of long-context handling and low-latency inference, which aligns with the requirements of our proposed *Virtual Speech Therapist* framework. Our preliminary investigations identified Gemini 3 Pro (Preview) as the most suitable model for therapeutic content generation. Its performance offered the best equilibrium between advanced reasoning capability, operational efficiency, and budgetary constraints, leading to its selection for this study (Comanici et al. 2025).

3.3.1 Gemini 3 Pro Hyperparamters

Low sampling temperatures are generally preferred for tasks demanding precision and factual reliability, such as technical writing, code generation, or question answering. Conversely, higher temperatures are advantageous for creative applications, including narrative generation, poetry, or ideation. Following the findings of Renze et al. (Renze 2024), we adopt task-specific temperature settings. As detailed in Section 2.3, both agents are built upon the *Gemini 3 Pro (Preview)* LLM. The *TherapyAgent* employs a temperature of $T = 0.3$ to introduce controlled variability and promote adaptive speech therapy interactions, whereas the *CriticAgent* is configured with $T = 0$ to

maximize precision and factual consistency (Renze 2024). The temperature parameter modulates the stochasticity of the model’s token sampling, governed by the softmax formulation:

$$P_i = \frac{e^{\frac{z_i}{T}}}{\sum_j e^{\frac{z_j}{T}}},$$

where:

- P_i is the predicted probability of the i -th token.
- z_i is the raw output score (logit) for the i -th token.
- T is the temperature parameter, controlling the sharpness of the probability distribution.
- \sum_j is the sum over all token (or indices j).

The temperature parameter (T) acts as a scaling factor on the model’s logits prior to the softmax operation, directly controlling the stochasticity of the output. A lower temperature ($T < 1$) sharpens the resulting probability distribution by amplifying the relative differences between logits. This increases the probability mass assigned to the most likely tokens, leading to more deterministic and focused model generations. Conversely, a higher temperature ($T > 1$) flattens the distribution by compressing the logit values, thereby increasing the likelihood of sampling from a broader vocabulary. This introduces greater randomness and diversity in the generated text, often at the cost of coherence.

3.4 Implementation

The *Virtual Speech Therapist* is realised as a web-based client–server application. The stuttering detection pipeline (Section 3.2) is implemented in *PyTorch*, with pre-trained model weights loaded via the *Hugging Face Transformers* library. Phonetic transcriptions are derived from speech segments using a wav2vec2-based phonemizer loaded through *torchaudio*. Orthographic transcriptions are generated by *Azure Speech Services* (Microsoft Cognitive Services), a cloud-based ASR module.

The multi-agent reasoning workflow described in Section 2.3—encompassing the *TherapyAgent*, *CriticAgent*, and the iterative refinement loop—is orchestrated with *LangGraph*, a framework for building stateful, graph-based agent pipelines. *LangChain* complements the stack with prompt management utilities and a unified interface through which both agents invoke the *Gemini 3 Pro (Preview)* model (Section 3.3).

All machine-learning models and agent graphs are initialised once at service startup and retained in memory across requests, eliminating cold-start overhead during interactive therapy sessions. The resulting user interface, data-flow architecture, and clinician approval workflow are detailed in Section 4.

4 User Interface and Clinical Integration

The *Virtual Speech Therapist* system integrates a web-based frontend with a persistent *Python* backend service to deliver therapy generation with clinician oversight. This

section describes the user interface design, data flow, and approval workflow that bridges VST recommendations with clinical judgment.

4.1 Frontend Design and Workflow

The clinical interface supports two operational modes: a *classification-only* mode for rapid stuttering analysis without therapy generation, and a *full therapy generation* mode for comprehensive assessment and plan development. In classification-only mode, cached classification results can be carried forward into full therapy generation, avoiding redundant audio analysis while enriching the cached chunks with phonemic transcription and ASR text during the second pass.

The clinician begins by entering patient demographics, clinical history, therapy background, and current treatment goals. Audio input may be either a live recording or an uploaded speech sample. Table 2 summarises the configurable audio segmentation parameters. This segmentation strategy ensures continuous temporal coverage of stuttering events by preventing artificial boundaries from disrupting event detection.

Table 2 Audio segmentation parameters available in the user interface.

Parameter	Options	Default
Segment duration	3 s, 4 s, 5 s	4 s
Overlap (k)	0%, 25%, 50%, 75%	50%

4.2 Data Flow and API Architecture

The system implements asynchronous REST API endpoints to handle the computationally intensive therapy generation workflow (typically 120–240 s). Table 3 outlines the four-stage request lifecycle.

Table 3 Asynchronous API workflow stages.

Stage	Description
Submission	Patient data and audio are submitted; the system returns a unique session identifier for asynchronous tracking.
Processing	The backend analyses audio, performs chunked classification, and generates therapy recommendations. Real-time progress updates are streamed to the frontend.
Results	A comprehensive therapy plan is displayed, including stuttering classification, therapeutic rationale, weekly progression, and intervention strategies.
Export	HTML documentation is generated for clinical records and patient handouts.

4.3 Results Presentation

Once processing completes, the interface presents results through several expandable sections:

- **Analysis Summary:** Audio characteristics (duration, sample rate, number of chunks) and an aggregate stuttering-type distribution visualised as a donut chart.
- **Overall Classification:** The primary and secondary stuttering types, weighted average confidence, a classified severity level (mild, moderate, severe), stuttering percentage across all chunks, and—when phonemic transcription is available—a list of phonemes correlated with disfluent regions (*problematic phonemes*).
- **Chunk Heatmap:** A compact grid where each cell represents one audio chunk. Cell colour encodes the classified stuttering type and opacity reflects classification confidence. Hovering over a cell reveals the time span, type, confidence, phonemic transcription, and textual transcription; clicking a cell plays the corresponding audio segment directly in the browser.
- **Therapy Recommendations:** Primary goals, therapeutic rationale, week-by-week progression (typically 6–8 weeks), and specific evidence-based techniques with clinical reasoning traces linking observations to expected outcomes.
- **Generation History:** Records of all refinement iterations when modifications were requested during approval, including the prompts generated by the therapy and critic agents for full transparency.

4.4 CITL Approval Workflow

A structured three-action approval process, summarised in Table 4, ensures clinical oversight. All actions are logged with timestamps and clinician information for compliance and audit purposes.

Table 4 Clinician-in-the-loop approval actions.

Action	Description
Approve	The clinician validates the therapy plan as clinically appropriate. Final documentation is generated and the plan is marked as ready for patient delivery.
Reject	The clinician terminates the workflow if the plan contradicts patient-specific considerations or clinical judgment.
Modify	The clinician provides specific refinement feedback (e.g. “emphasise respiratory control for prolongation-type stuttering”). The system resubmits the feedback to the therapy–critic agent loop for one automated refinement round, after which the revised plan returns for re-review.

4.5 Backend Architecture

The persistent backend service maintains machine-learning models and therapy agents in memory to eliminate model-loading overhead (30–120s per cold start). Table 5 describes its principal components.

Table 5 Backend service components.

Component	Role
Model Persistence	Speech-analysis models (wav2vec2 phonemiser, stuttering classifier) and VST agents are initialised at service startup and retained in memory across requests.
Classification Caching	When a classification-only session transitions to full therapy generation, cached chunk classifications are reused. Only phonemic transcription and ASR are executed as a post-processing step, avoiding redundant ML inference.
Asynchronous Coordination	Frontend and backend communicate through inter-process mechanisms with health monitoring to prevent cascading failures.
Request Processing	Patient data and audio are received; the therapeutic reasoning workflow (classification, planning, critique, refinement) is executed and structured results are returned.

5 Discussion and Evaluation

We conducted a structured, qualitative evaluation of the proposed *Virtual Speech Therapist* framework with a *licensed*, practicing clinical SLP. The clinician performed a focused assessment using a sample of $N = 16$ clinical speech recordings, comprising four samples from each of core stuttering behaviors: blocks, prolongations, word repetitions, sound repetitions and interjections. The evaluation criteria were: (1) diagnostic alignment between the VST’s output and expert clinical judgment, (2) operational feasibility within existing clinical workflows, and (3) potential capacity to address unmet needs in speech pathology practice.

5.1 Evaluation Framework

Model evaluation focused primarily on the clinical validity of the agent’s reasoning and therapeutic sequencing, rather than on fluency outcomes or automated correctness. The aim is to assess whether the generated therapy plans are aligned with contemporary clinical practice in adult stuttering intervention, particularly in light of the speech samples provided to the system and their corresponding analysis. Specific attention was given to how the multi-agentic VST integrated information related to avoidance-driven behaviors, emotional reactivity, and the appropriate ordering of intervention strategies.

Evaluation criteria included: (a) the differentiation between core stuttering behaviors and avoidance-driven behaviors and struggle, and how this differentiation influenced therapeutic goals; (b) the appropriateness of the proposed therapeutic focus; (c) the coherence of therapeutic sequencing; and (d) the system’s capacity to adjust therapy plans following clinical feedback. All outputs were reviewed by a practicing *licensed* senior speech- language pathologist with clinical and research expertise in fluency disorders.

5.2 Overall Model Performance

Overall, the multi-agentic VST consistently generated structured and clinically interpretable therapy plans, demonstrating the ability to integrate multiple therapeutic frameworks commonly used in adult stuttering intervention, including *Stuttering Modification*, *Fluency Shaping*, *Avoidance Reduction Therapy*, and *CBT- and ACT-informed* approaches. The plans were generally coherent and clinically readable, with clear rationales linking observed speech behaviors to proposed intervention strategies.

Importantly, the multi-agentic VST did not reduce intervention planning to surface fluency metrics alone and, in several cases, demonstrated emerging sensitivity to constructs such as struggle, avoidance, and emotional reactivity. In one out of 16 speech analysis, where the automatic stuttering detection system did not identify overt moments of stuttering, the multi-agentic VST appropriately proposed a more in-depth clinical evaluation to explore the possibility of covert stuttering, a presentation in which overt stuttering behaviors are minimized due to persistent word substitution and avoidance. Covert stuttering has been shown to be closely associated with heightened avoidance, social anxiety, and emotional burden, despite apparently fluent speech (Boyle and Rosen 2025; Iverach and Rapee 2014). In this case, the VST proposed an intervention with a strong focus on the emotional and cognitive components of stuttering, rather than fluency control.

This approach is consistent with contemporary views of stuttering as a multidimensional condition, in which the speaker’s lived experience, emotional responses, and behavioral adaptations are central to intervention planning (Yaruss and Quesal 2006; Chang et al. 2025).

The tasks and practice activities suggested by the VST were generally well aligned with the therapy goals and easy to use in practice. They were clearly explained, easy to understand, and flexible enough for the speech therapist to adapt them to each person and situation. Importantly, these activities were not strict instructions, but practical examples that help therapists apply the therapy plan in real clinical settings. This balance between clear guidance and flexibility makes the system useful in everyday speech therapy practice and easy to integrate into the therapist’s workflow. Nevertheless, systematic evaluation revealed recurring limitations related to behavioral interpretation and therapeutic sequencing, detailed below

5.3 Observed Limitations in Initial Therapy Plan Generation

One observed limitation involved the premature introduction of *Fluency Shaping* strategies, before sufficient work had been conducted to address fear, urgency, and

struggle. This contrasts with integrated clinical models that emphasize reducing emotional reactivity and experiential avoidance prior to targeting speech motor control (Sønsterud et al. 2020). Additionally, some early-phase plans included an excessive number of exercises, increasing the risk of cognitive and therapeutic overload.

Interjections and pre-speech fillers were sometimes classified as prolongations; clinically they functioned as avoidance behaviors associated with anticipation, fear, and urgency. This misclassification resulted in therapy plans that emphasized motor-based fluency strategies rather than prioritizing desensitization and avoidance reduction. Such sequencing is inconsistent with current clinical understanding that avoidance and emotional reactivity play a central role in the maintenance of stuttering in adults (Chang et al. 2025).

These limitations highlight the challenges inherent in automated interpretation of complex speech behaviors and reinforce the necessity of clinician oversight. One additional limitation is that the intervention proposed for an 8-week period may be relatively intensive, with a potential risk of increased cognitive load for some patients. However, the system explicitly incorporates continuous self-monitoring mechanisms, prompting patients to regularly report perceived overload and anxiety associated with task execution. This ongoing feedback enables dynamic adjustment of task intensity, pacing, and prioritization by the clinician, helping to mitigate cognitive overload and maintain therapeutic engagement.

5.4 Impact of CITL Fine-Tuning and Iterative Feedback

Targeted fine-tuning and the explicit introduction of clinical constraints resulted in meaningful improvements in the VST’s therapeutic reasoning. Following refinement, the VST system demonstrated improved differentiation between core stuttering behaviors, avoidance-driven responses, and struggle, as well as more appropriate therapeutic sequencing.

Specifically, the VST agent increasingly prioritized reduction of fear, urgency, and avoidance, introducing desensitization strategies such as voluntary stuttering before motor-based techniques. *Stuttering Modification* strategies were more consistently positioned prior to *Fluency Shaping*, reflecting clinical evidence that reducing emotional reactivity facilitates later implementation of fluency-enhancing strategies (Sønsterud et al. 2020; Chang et al. 2025). Primary therapeutic goals also shifted away from percentage-based success metrics toward reductions in struggle, tension, avoidance, and increased communicative ease, aligning with contemporary outcome frameworks that move beyond fluency alone (Yaruss and Quesal 2006).

Importantly, once an initial therapy plan was generated and explicit clinical feedback was provided, subsequent outputs showed a high degree of responsiveness to correction. The VST agent consistently integrated feedback related to behavioral interpretation, therapeutic focus, and sequencing, producing revised plans that were clinically more coherent and better aligned with established principles of adult stuttering intervention. This iterative responsiveness supports the suitability of the VST agent for a CITL framework, in which expert input directly shapes and refines therapeutic reasoning rather than merely adjusting surface-level recommendations.

5.5 Representative Case Examples

Representative cases were examined to illustrate both initial limitations and subsequent improvements following clinical feedback. In one case, speech initially interpreted as containing prolongations was reclassified as avoidance-related interjections, leading to a revised therapy plan centered on fear reduction, approach behaviors, and voluntary stuttering rather than fluency control. In another case characterized by syllable repetitions without overt signs of anxiety, the VST agent appropriately emphasized motor timing support without prioritizing avoidance reduction. These cases illustrate the importance of accurate behavioral interpretation and demonstrate how CITL-guided refinement substantially improves the clinical plausibility and therapeutic coherence of the generated plans.

5.6 Final Reflections on Clinical Readiness

Taken together, these findings indicate that the system is ready for robust clinical trials. The next essential step is to test its efficacy and integration in diverse practice environments under clinician supervision. By assisting clinicians in structuring, sequencing, and refining therapy plans grounded in contemporary evidence-based frameworks, the system appears to have the potential to meaningfully enhance the quality and consistency of intervention for people who stutter. Rather than replacing clinical expertise, the model functions as a knowledgeable ally, supporting clinicians in delivering individualized, evidence-informed treatment that reflects current understanding of stuttering as a multidimensional condition.

6 Conclusion

In this paper, we proposed a novel *Virtual Speech Therapist*, a multi-agentic AI-powered platform designed to address the persistent and multifaceted barriers to accessing high-quality speech therapy. By integrating state-of-the-art deep learning for real-time stuttering detection with an adaptive, agentic LLM-based intervention engine, the VST offers a scalable solution for delivering intensive, personalized stuttering therapy within a CITL framework—an essential paradigm for delivering effective speech rehabilitation in motor disorders such as developmental stuttering. In evaluations conducted by licensed SLP, the VST agent was found to generate therapy plans that are clinically sound, safe, and adhere to established therapeutic guidelines.

Future validation through randomized controlled trials will be essential to establish the clinical efficacy, real-world usability, and sustained therapeutic impact of this intervention. By conceptualizing AI as a force multiplier rather than a replacement for the expert SLPs, the VST framework embodies a novel paradigm in digital therapeutics. It offers a scalable pathway to democratize access to high-fidelity speech rehabilitation, address global disparities in care, and improve long-term functional outcomes and quality of life for individuals with communication disorders, while also serving as a dynamic knowledge ally for clinicians.

Moreover, the underlying human-AI collaborative model is not specific to stuttering; it represents a generalizable framework for therapeutic intervention that could

be adapted to other speech and language disorders, broadening the potential scope of scalable, precision digital therapeutics.

A key technical direction involves extending the VST with an integrated audio modality for audio generation.

This advancement would not only generate text-based therapeutic protocols but also synthesize real-time, personalized audio exemplars for targeted practice, directly addressing specific stuttering typologies and enhancing motor learning.

Acknowledgements. Declaration of generative AI and AI-assisted technologies in the manuscript preparation process. During the preparation of this work the author(s) used ChatGPT in order to finetune. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Data availability

The SEP-28k-E dataset used in this study is available in a public repository at <https://github.com/th-nuernberg/ml-stuttering-events-dataset-extended>.

Code availability

The full Virtual Speech Therapist multi-agent codebase is available at <https://github.com/pmarmaroli/vocametrix-platform/tree/main/python/vstagent>. An interactive user interface for the proposed system is available at <https://vocametrix.com/ai/stuttering-therapy-planning-agent>, facilitating real-time stuttering assessment and personalized therapy planning.

Appendix A Prompts

TherapyAgent Prompt

You are a *LICENSED*, expert speech-language pathologist (SLP) specializing in stuttering disorders for: ({language_desc}).

You apply contemporary approaches:

1. Stuttering Modification (voluntary stuttering, pull-outs, cancellations).
2. Fluency Shaping (easy onsets, light contacts, continuous phonation).
3. Avoidance Reduction (ARTS) (desensitization, fear reduction).
4. CBT/ACT/REBT (cognitive restructuring, acceptance).

ROLE: Clinical Decision Support. You do NOT provide medical diagnosis or replace human judgment.

SAFETY & ETHICS:

- *STRICTLY AVOID*: Voice strain, breath holding, excessive practice, or shame-inducing language.
- *PRIORITIZE*: Empowering, non-judgmental, person-centered language.

- *ADAPT*: Tailor vocabulary and examples to patient’s age and cultural context.

RED FLAGS: If patient indicates severe distress/trauma, include: ”*URGENT CLINICAL NOTE*: Patient profile indicates serious concerns requiring immediate human assessment. Recommend urgent consultation.”

OUTPUT REQUIREMENT:

- Output ONLY valid JSON following the schema below.
- NO markdown, NO explanatory text.

THERAPY PLAN JSON SCHEMA: {schema_str}

The schema includes a `clinicalReasoning` object inside each strategy, with the fields: `observation`, `clinicalRationale`, `expectedOutcome`, and `evidenceBase`.

INSTRUCTIONS:

Generate an evidence-based, personalized plan following ASHA guidelines.

- Target specific stuttering types with practical exercises.
- Progress from awareness to real-world application.
- Include measurable goals, home practice, and psychoeducation.

EXPLAINABILITY REQUIREMENT (CRITICAL):

For EVERY strategy in the plan, you MUST provide a `clinicalReasoning` object that makes your decision-making transparent. This follows a Chain-of-Thought (CoT) Input–Thought–Output cycle:

1. **observation** (Input): What specific pattern in the patient’s stuttering analysis motivates this strategy?
Example: “Patient exhibits 60% block-type disfluencies concentrated on plosive consonants /p/, /b/, /k/ with visible jaw tension.”
2. **clinicalRationale** (Thought): Why is this strategy the clinically appropriate response? Reference the therapeutic approach.
Example: “Blocks on plosives indicate laryngeal tension. Light articulatory contacts (Fluency Shaping) reduce subglottic pressure and facilitate airflow, directly addressing the tension source.”
3. **expectedOutcome** (Output): What measurable improvement should result?
Example: “Reduction in block duration from >2s to <0.5s on target phonemes within 3 weeks, measured via timed readings.”
4. **evidenceBase**: Reference the clinical guideline or literature.
Example: “Guitar & McCauley (2010), ASHA Practice Portal: Fluency Disorders”

This reasoning chain will be presented to the reviewing clinician for informed validation.

Create a personalized therapy plan in JSON format for

PATIENT INFO:

- Type: `classification.get('stutteringType', 'unknown')`
- Transcription: `classification.get('transcription', 'N/A')`

- Phonemes: classification.get('phonemes', 'N/A')
- Characteristics: classification.get('characteristics', ['Not specified'])
- Locale: locale, #defines language of a speech sample
- (Infer age/context from characteristics if not explicit)

Note: At runtime, the prompt is dynamically enriched with ACOUSTIC PROFILE (chunk-level type distribution and stuttering percentage), PER-CHUNK ANALYSIS (temporal breakdown with time ranges, types, confidence, transcriptions, and phonemes), and PHONEME-DISFLUENCY CORRELATION (over-represented phonemes in disfluent segments) when these data are available from the classification pipeline.

STRATEGY:

- Analyze patterns (repetitions, blocks, tension) & severity.
- Select focus:
 - Repetitions -¿ Fluency Shaping.
 - Blocks/Tension -¿ Stuttering Modification.
 - Mixed -¿ Blended.
- Formulate Goals: Functional, linked to real contexts.
- Design trajectory: Awareness -¿ Isolated -¿ Words/Phrases -¿ Sentences -¿ Real contexts.
- Create Materials: Interest-based, phonetically aligned (e.g., vowels for easy onset).
- Integrate avoidance reduction & CBT (reframing beliefs).
- Monitoring: Define tracking metrics (e.g., 0-9 ease scales).

REQUIRED CONTENT MAPPING:

- 'explanation.stutteringTypeDefinition': Assessment of profile.
- 'explanation.therapeuticRationale': Approach & WHY. MUST end with combined warning: " IMPORTANT LIMITATION: This AI-generated plan is a decision support tool requiring review by a qualified SLP. It does not substitute for professional judgment." (Translate if locale is not en-US/en-GB, but ALWAYS include both cautions in one block).
- 'primaryGoal.rationale': Psychoeducation (normalize stuttering, effectiveness ¿ fluency).
- 'strategies.clinicalReasoning': For each strategy, provide a complete CoT reasoning chain:
 - observation: Link directly to patient's stuttering data (type, phonemes, confidence scores)
 - clinicalRationale: Explain the clinical logic connecting observation to this specific exercise
 - expectedOutcome: State a measurable, time-bound expected outcome
 - evidenceBase: Cite the relevant clinical guideline or research
- 'strategies.instructions': For each strategy:

- Approach name (1-2 sentences)
- Purpose (2-3 sentences explaining clinical rationale)
- 5-8 Practice Items with detailed instructions:
 - * Each item: 3-4 sentences covering context, execution, frequency, progression
 - * Include specific examples with sample words/phrases
 - * Specify duration, repetitions, and success criteria
- Home Practice (detailed daily schedule with 0-9 rating scales)
- Troubleshooting tips (2-3 common issues and solutions)
- Example: "STRATEGY: Easy Onset... PURPOSE: Reduce tension... PRACTICE: 1) ... 2) ... 3) ... HOME: Practice 5min, rate ease 0-9."

OUTPUT ONLY THE JSON OBJECT.

CriticAgent Prompt

The *CriticAgent* receives a two-part input: a *system message* defining its role and a *human message* containing the therapy plan and evaluation instructions.

System Message.

You are a clinical quality assurance specialist for speech therapy ({language_desc}). Your role is to ensure therapy plans are:

- Clinically sound and evidence-based
- Ethically appropriate and safe
- Properly tailored to patient needs
- Following professional association guidelines

Provide structured, constructive, and critical feedback focusing on safety, effectiveness, and evidence strength.

Human Message.

THERAPY PLAN TO REVIEW: {therapy_plan_json}

TARGET CONDITION:

- Primary Stuttering Type: {primary_type}
- Chunk Distribution: {type_distribution}
- Stuttering Percentage: {stuttering_percentage}%
- Phoneme-Disfluency Correlation: {phoneme_correlation_summary}

CRITICAL REVIEW INSTRUCTIONS:

Evaluate the therapy plan in detail, covering the following domains:

1. **Clinical Soundness** — Assess whether the plan matches the patient's stuttering type and severity. Determine whether techniques are appropriate, realistic,

and developmentally suitable. Highlight any misalignments or inappropriate strategies.

2. **Safety Concerns** — Identify potential risks or contraindications (physical, vocal, or psychological). Evaluate whether the plan avoids strain, fatigue, or anxiety triggers.
3. **Evidence Strength** — Check whether each technique is supported by peer-reviewed research or recognized guidelines (e.g., ASHA, IFA). Note the quality of evidence (strong, moderate, limited) for each intervention.
4. **Improvements Needed** — Suggest specific, actionable changes to improve the plan. Include guidance on frequency, duration, intensity, or sequencing of exercises. Highlight areas for progress monitoring, home practice, and carryover strategies.
5. **Structure and Clarity** — Evaluate whether the plan is well-organized, readable, and easy for clinicians or caregivers to follow. Suggest improvements for formatting, stepwise instructions, and clarity.
6. **Explainability and Reasoning Transparency** — Verify that each strategy includes a complete `clinicalReasoning` object with `observation`, `clinicalRationale`, `expectedOutcome`, and `evidenceBase`. Assess whether the reasoning chain is logically sound: does the observation correctly link to the rationale? Are cited evidence bases legitimate and relevant? Flag any strategies where reasoning is missing, circular, or insufficiently grounded.

OUTPUT FORMAT:

Provide your assessment in a structured, bullet-point format for each domain:

- Domain Name (e.g., Clinical Soundness):
 - Observation: ...
 - Strengths: ...
 - Concerns: ...
 - Recommendations: ...

Repeat for all domains: Clinical Soundness, Safety Concerns, Evidence Strength, Improvements Needed, Structure and Clarity, Explainability and Reasoning Transparency.

Refinement Prompt

When the *CriticAgent* identifies areas for improvement, the therapy plan is refined using a dedicated refinement prompt. The *TherapyAgent* receives the original plan together with the critic feedback and is instructed to address all flagged issues while preserving what is working well.

System Message.

You are an expert clinical speech-language pathologist and therapist specializing in stuttering therapy (`{language_desc}`).

Your task is to REFINE an existing therapy plan based on critic feedback.
REFINEMENT REQUIREMENTS:

1. Address ALL points raised in the critic feedback.
2. Maintain the therapeutic goals and overall approach.
3. Strengthen weak areas identified by the critic.
4. Add missing components or details.
5. Ensure all evidence-based recommendations are incorporated.
6. Keep what is working well, improve what needs enhancement.
7. Every strategy MUST include a `clinicalReasoning` object with: `observation`, `clinicalRationale`, `expectedOutcome`, and `evidenceBase`.

Human Message.

Refine the existing therapy plan to address the critic feedback below.

- PATIENT INFORMATION: {patient_info}
- ACOUSTIC PROFILE: {acoustic_profile} (*when available*)
- PHONEME-DISFLUENCY CORRELATION: {phoneme_correlation} (*when available*)
- EXISTING THERAPY PLAN: {previous_plan_json}
- CRITIC FEEDBACK: {critic_feedback}

Revise the therapy plan to directly address all points raised in the critic feedback. Preserve what is working well, and strengthen areas flagged for improvement. Maintain clinical soundness and safety throughout.

Human Revision Prompt

When the clinician selects **Modification** during the approval phase (Section 2.3), a dedicated human revision prompt integrates the clinician's feedback into the plan.

Prompt.

You are refining an existing therapy plan based on clinician feedback.

IMPORTANT: Do NOT create a new plan from scratch. Instead, take the existing plan below and modify it according to the clinician's feedback.

- CLINICIAN FEEDBACK: {human_feedback}
- EXISTING THERAPY PLAN TO MODIFY: {current_plan_json}
- TARGET CONDITION:
 - Stuttering Type: {primary_type}
 - Phonetic textual content: {phonemes}

Modify the existing plan to address the clinician's concerns while preserving the overall structure and elements that are working well.

Appendix B Clinician-Supervised Speech Therapy Protocols Generated by Virtual Speech Therapist Agents

Final Therapy Plans - Ready for Patient Delivery

Stuttering Type: Prolongation 1

B.1 Explanation:

- **Stuttering Type:** Prolongation involves the involuntary stretching of a sound (audible sound flow) while articulatory movement is temporarily suspended. It often indicates excessive supraglottic tension (lips/tongue) or laryngeal tension.
- **Patient Characteristics:** The patient exhibits prolongations specifically on fricatives (e.g., /ʃ/ in 'she') and complex clusters. This indicates a need to prioritize Light Articulatory Contacts (LAC) over general Easy Onsets, as the tension is likely articulatory rather than purely laryngeal.
- **Therapeutic Rationale:** The plan utilizes an Integrated Approach modified to prioritize desensitization and tension reduction. Stuttering Modification (Van Riper) is introduced first to address emotional reactivity and the fear-tension-struggle cycle. Once the patient can manage tension and reactivity, Fluency Shaping strategies (Webster/Guitar) are introduced to refine articulatory kinematics.

B.2 Primary Goal

- **Goal:** The patient will demonstrate increased ease of speech during stuttering moments by reducing physical tension and emotional reactivity, as measured by self-rating scales and clinician observation of struggle behaviors.
- **Target:** Consistently low self-reported tension (e.g., ≤ 3 on a 10-point scale) during disfluencies, rather than a specific fluency percentage.
- **Baseline:** Patient currently exhibits uncontrolled prolongations on approximately 10-15% of syllables with associated physical tension and no modification attempts.

B.3 Step by Step Plan

1. Step 1: Foundation: Identification & Proprioception

- **Week Range:** Weeks 1-2
- **Objective:** Patient will identify physical tension sites (specifically articulatory vs. laryngeal) and differentiate between 'tense' and 'loose' speech production.
 - **Proprioceptive Awareness & Voluntary Prolongation:** Developing high-resolution awareness of the articulators to locate the specific point of tension during a prolongation.

- *Instructions:* RATIONALE: You cannot relax a muscle if you don't know it's tense. We must map the mouth. DRILL ITEMS: 1) 'The Tension Scan': Prolong /s/ in 'Sun'. Scan your body. Is the tongue pressing the roof of the mouth hard? 2) 'Tense vs. Loose Contrast': Say 'She' with maximum pressure on the /sh/ (100% tension). Then say it with lips barely touching (10% tension). Repeat 10 times. 3) 'Voluntary Stretching': Intentionally stretch the first sound of 'She could buy' for 2 seconds without struggle. Keep the airflow moving. HOME PRACTICE: Read aloud for 5 mins/day using Voluntary Stretching. Rate anxiety (0-9). TROUBLESHOOTING: If you feel panic, shorten the stretch. Use a mirror to ensure lips aren't pressing hard.

2. Step 2: Stuttering Modification & Desensitization

- **Week Range:** Weeks 3-5
- **Objective:** Patient will reduce emotional reactivity and physical tension during blocks using Pull-outs and Preparatory Sets prior to attempting fluency shaping.
 - **Pull-outs & Preparatory Sets (Van Riper)** Pull-out: Modifying tension **during** a block. Preparatory Set: Pre-modifying articulators **before** an anticipated block.
 - *Instructions:* RATIONALE: We must address the struggle behavior before trying to shape fluency. These tools reduce the 'fight' response. DRILL ITEMS: 1) 'Pull-out': Fake a hard block on 'She'. Catch it. Reduce tension by 50%. Slide out. 2) 'Preparatory Set': Anticipate a block on 'Minimum'. Pause. Relax the lips **before** starting. Say it with a 'Soft Start'. 3) 'Resilience': If you push through a block, stop. Say 'I pushed.' Re-do it with a Pull-out. HOME PRACTICE: Practice Pull-outs on the phrase 'She could buy a minimum'. TROUBLESHOOTING: Focus on the feeling of release, not just getting the word out.
 - **Desensitization & Self-Disclosure:** Reducing the emotional weight of stuttering to lower baseline tension.
 - *Instructions:* RATIONALE: Lowering emotional reactivity makes physical modification easier. DRILL ITEMS: 1) 'Self-Disclosure': Practice saying 'I stutter, so I might take a minute' to a friend. 2) 'Voluntary Stuttering': Enter a low-stakes situation (e.g., asking a family member a question) and intentionally stutter on one word without tension. HOME PRACTICE: Log one instance of self-disclosure or voluntary stuttering daily.

3. Step 3: Fluency Shaping & Generalization

- **Week Range:** Weeks 6-8
- **Objective:** With tension reduced, patient will apply Light Articulatory Contacts (LAC) and Continuous Phonation to facilitate smoother forward flow.

- **Light Articulatory Contact (LAC):** Reducing the physical pressure of the articulators (lips/tongue) during consonant production to prevent the 'locking' of the prolongation.
- *Instructions:* RATIONALE: Now that the fear-tension cycle is managed, we refine the movement. Prolongations on /sh/ occur when the tongue presses too hard. DRILL ITEMS: 1) 'Feather /sh/': Say 'She'. Bring lips forward but imagine they are made of cotton. Barely make the sound. 2) 'Fricative Drills': Practice /f/, /s/, /sh/, /th/. Touch the articulators gently. 3) 'Negative Practice': Press hard on 'Shoe' (Old Way), then touch lightly on 'Shoe' (New Way). Feel the difference. HOME PRACTICE: 10 mins daily reading. Highlight all fricatives. Apply LAC to them.
- **Continuous Phonation & Real World Transfer:** Connecting words smoothly to keep the voice box vibrating, applied in hierarchical situations.
- *Instructions:* RATIONALE: Keeping the voice 'on' prevents the system from resetting. We now apply this to real life. DRILL ITEMS: 1) 'Chaining': Say 'She-could-buy' as if it is one long word. 2) 'Hierarchy Level 1': Call a business recording. Use LAC and Chaining. 3) 'Hierarchy Level 2': Ask a stranger for the time using the target phrase 'She could buy...'. HOME PRACTICE: Complete one hierarchy task daily. MAINTENANCE: Create a plan to revert to Step 2 (Pull-outs) if tension spikes.

Stuttering Type: Prolongation 2

B.4 Explanation:

- **Stuttering Type:** Prolongation involves the involuntary stretching of a sound (typically the initial phoneme) where airflow continues, but the articulators remain fixed in position for an extended period.
- **Patient Characteristics:** Analysis of 'It's fine, but sometimes' indicates a pattern of tension on fricatives (/s/, /f/). The patient maintains airflow but delays the vowel transition. This plan is framed as 'Phase 1: Acquisition & Transfer' based on Barry Guitar's Integrated Approach.
- **Therapeutic Rationale:** Integrating Van Riper's Stuttering Modification (for tension) and Webster's Fluency Shaping (for airflow) addresses both the mechanics and the 'Iceberg' of stuttering (Sheehan). We utilize the OASES to track quality of life. IMPORTANT: This plan focuses on acquisition; long-term stabilization may require follow-up.

B.5 Primary Goal

- **Goal:** The client will modify tension during prolongations using 'Light Contacts' or 'Pull-outs' in sentence-level speech, reducing the average duration of stuttering moments to ≤ 1 second.

- **Target:** 80% accuracy in self-correction AND Clinician-rated reduction of prolongation duration to ≤ 1 second.
- **Baseline:** Client exhibits prolongations on 15% of syllables (%SS) in conversation, with an average duration of 2.0 seconds per moment of stuttering.

B.6 Step by Step Plan

1. Step 1: Identification & Desensitization (Safety First)

- **Week Range:** Weeks 1-2
- **Objective:** Increase proprioceptive awareness and reduce emotional reactivity to prolongations using a safety hierarchy.
 - **Voluntary Prolongation with Video Feedback:** Intentionally stretching sounds to neutralize fear (Desensitization) and calibrating self-perception using video.
 - *Instructions:* APPROACH: Negative Practice / Desensitization (Byrd et al.). SAFETY PROTOCOL: Establish a 'Safe Word' (e.g., 'Pause'). If anxiety spikes $\geq 7/10$, use the word to stop immediately and breathe. PRACTICE ITEMS: 1. Video Calibration: Record a 1-minute monologue. Watch it together. Identify 'real' tension vs. 'felt' tension. Often the stutter looks less severe than it feels. 2. The 'Slide': Intentionally stretch /s/ or /f/ for 3 seconds. Keep eye contact. Do this ONLY in the therapy room initially. 3. Contrast Drill: Produce 'Sun' with 100% tension (hard), then 50%, then 10% (light). Feel the difference. HOME PRACTICE: Practice 'Sliding' in front of a mirror for 5 mins. Do NOT do this in public yet. Rate anxiety 0-9.

2. Step 2: Fluency Shaping: Light Contacts

- **Week Range:** Weeks 3-5
- **Objective:** Reduce articulatory pressure on fricatives to prevent the 'locking' mechanism (Webster/Perkins).
 - **Light Articulatory Contacts:** Touching articulators with minimum pressure to facilitate airflow.
 - *Instructions:* APPROACH: Fluency Shaping. PURPOSE: Prevent the physical block before it starts. PRACTICE ITEMS: 1. The 'Feather' Concept: Touch the roof of the mouth for /s/ as lightly as a feather. 2. Word Drill: 'Five, Fine, Fish'. If the lip turns white, pressure is too high. Aim for a 'whisper touch'. 3. Carrier Phrases: 'I see a...' naming objects. Focus ONLY on the light contact sensation. HOME PRACTICE: Read a paragraph. Highlight /s/ and /f/ words. Apply light contacts. If speech sounds 'slurred', increase precision but NOT pressure.

3. Step 3: Modification & Bridging the Gap

- **Week Range:** Weeks 6-7

- **Objective:** Apply 'Pull-outs' in structured interactions before attempting real-world challenges.
 - **The Pull-Out & Structured Roleplay:** Modifying tension during a block (Van Riper) within a controlled conversational script.
 - *Instructions:* APPROACH: Stuttering Modification. PURPOSE: Bridge the gap between drills and spontaneous speech. PRACTICE ITEMS: 1. The Pull-Out: Fake a prolongation on 'Sssseven'. Catch it, reduce tension, and slide to the vowel. Do not stop and restart (Cancellation). 2. Scripted Roleplay: Clinician plays a 'Store Clerk'. Client asks specific questions from a script. Goal: Use 3 pull-outs intentionally. 3. Hierarchy Worksheet: Create a list of speaking situations ranked 1-10 (e.g., 1=Talking to pet, 5=Ordering coffee, 10=Public speech). HOME PRACTICE: Choose a level 2 or 3 situation from your Hierarchy (e.g., talking to a close friend). Aim for one intentional pull-out.

4. Step 4: Generalization & Maintenance

- Week Range: Week 8
- **Objective:** Transfer skills to high-load situations and establish a relapse management plan.
 - **Real-World Transfer & Relapse Planning:** Testing skills in high-pressure contexts and planning for future fluency fluctuations.
 - *Instructions:* APPROACH: Generalization. PURPOSE: To ensure skills survive outside the clinic and prepare for the chronic nature of stuttering. ACTIVITIES: 1. The Phone Call: Make a call to a recorded line (e.g., movie times) then a live person. Use a Light Contact on the first word. 2. Relapse Plan: Write down 'What do I do if I have a bad speech day?' (Answer: Go back to Step 1 - Voluntary Stuttering to reduce tension). 3. Support: Discuss joining the National Stuttering Association (NSA) for community support. HOME PRACTICE: Complete one 'Challenge Call'. Log the outcome not by fluency, but by successful management of tension.

Stuttering Type: Silent Block 1

B.7 Explanation:

- **Stuttering Type:** Re-evaluation of the audio sample ('We do not own a freshness') identifies the pauses not as linguistic rests, but as 'Silent Blocks.' This indicates a laryngeal or articulatory posture where airflow is completely stopped despite the attempt to speak.
- **Patient Characteristics:** The patient exhibits 'Silent Blocking,' characterized by a stoppage of airflow and sound. This suggests high laryngeal tension (glottal closure) and likely significant anticipation anxiety. The silence represents the moment of highest physical struggle.

- **Therapeutic Rationale:** Therapy must pivot from general 'fluency maintenance' to specific 'Airflow Management' and 'Tension Reduction.' The priority is to re-establish airflow during the silent moments to prevent the vocal folds from locking. We utilize the ICF Model to address the frustration associated with these silent struggles.

B.8 Primary Goal

- **Goal:** The patient will identify the onset of silent blocks and utilize airflow strategies to release laryngeal tension, converting 'struggled silence' into 'easy sound.'
- **Target:** Reduction in duration of silent blocks; Self-rated physical tension during blocks \leq 3/10.
- **Baseline:** Presence of silent blocks disguised as pauses; high laryngeal tension presumed.

B.9 Step by Step Plan

1. Step 1: Airflow Initiation & Laryngeal Release

- **Week Range:** Weeks 1-2
- **Objective:** To replace the 'holding breath' mechanism of the silent block with passive airflow and gentle phonation onset.
 - **The 'Pre-Voice' Exhale (Airflow Management):** Teaching the patient to release a small amount of air *before* engaging the vocal folds to prevent the glottis from slamming shut (the silent block).
 - *Instructions:* DOSAGE: 1x Weekly Clinical Session; 10 mins Daily Home Practice. [PATIENT INSTRUCTIONS] 1. Identification: Recognize the 'stuck' feeling in your throat during the silence. That is your vocal cords closing tight. 2. The Sigh: Practice a passive sigh. Feel the air leave your mouth without sound. 3. Application to 'We': The word 'We' (/w/) is a common blocking point. Sigh out air, then gently turn on the voice: 'hhh-We'. 4. Negative Practice: Intentionally hold your breath and try to say 'We'. Feel the block. Then, stop, release the air, and use the sigh method. [RATIONALE] Sound cannot travel without air. Silent blocks occur when you try to speak without airflow. We are prioritizing 'air over voice'. [TROUBLESHOOTING] If you feel dizzy, take a break. You are hyper-ventilating. Focus on *gentle* airflow, not forceful pushing.

2. Step 2: Proprioception & Soft Contacts for Blocks

- **Week Range:** Weeks 3-5
- **Objective:** To use light articulatory contacts to navigate through blocks without forcing, specifically targeting the transition out of silence.
 - **Light Contacts & Pull-outs:** Using reduced pressure on consonants to prevent the build-up of intra-oral pressure that sustains the block.

- *Instructions*: DOSAGE: 1x Weekly Clinical Session; 15 mins Daily Home Practice. [PATIENT INSTRUCTIONS] 1. Target /f/ in 'Freshness': If you block on the /f/, do not push. Lightly touch your lip to your teeth and let air hiss through. 'ffff-reshness'. 2. Target /d/ in 'Do': If you block silently on 'Do', visualize your tongue tip barely touching the roof of your mouth. Whisper the 'd' first. 3. The 'Pull-out': When you feel a silent block, do not push through it. Stop. Release tension by 50%. Re-attempt the word with a 'Light Contact'. [RATIONALE] Fighting a block increases tension and prolongs the silence. 'Softening' the contact breaks the neurological loop of the block. [TROUBLESHOOTING] Ensure you are not just whispering. You need to engage the voice, but only *after* the air is moving.

3. Step 3: Desensitization to Silence

- Week Range: Weeks 6-8
- **Objective**: To reduce the panic associated with the silence of a block, allowing the patient to remain calm enough to apply strategies.
 - **Voluntary Pausing & Reality Testing**: Intentionally inserting silence into speech to desensitize the fear response, proving that a pause is not a catastrophe.
 - *Instructions*: DOSAGE: 1x Weekly Clinical Session; Real-world application tasks. [PATIENT INSTRUCTIONS] 1. Intentional Silence: In a safe conversation, pause for 2 seconds before saying a word. 'We do not... [2 sec pause]... own.' 2. Observation: During the silence, keep your throat loose. Do not lock up. Watch the listener. Did they interrupt? 3. Reframing: A silent block is involuntary, but a pause is voluntary. By practicing voluntary pauses, you gain control over the silence. 4. Video Feedback: Record yourself. Notice that the silence often feels longer to you than it sounds to the listener. [RATIONALE] Panic tightens the throat, worsening the block. If you are not afraid of the silence, your throat will remain open, reducing the frequency of blocks. [TROUBLESHOOTING] If the silence triggers a real block, use your 'Airflow' strategy (Step 1) to ease out of it.

Stuttering Type: Silent Block 2

B.10 Explanation:

- **Stuttering Type**: Blocks involve a temporary stoppage of airflow and voicing due to laryngeal or articulatory tension. While the patient history notes prolongations, current observation indicates blocks are the primary feature, characterized by struggle and silence before sound initiation.
- **Patient Characteristics**: Clinician observation identifies blocks as the dominant behavior (e.g., on plosives in 'pay', 'talk', 'can'). The patient likely exhibits high physical tension/struggle during these moments. Desensitization is now prioritized to reduce the fear of 'getting stuck'.

- **Therapeutic Rationale:** Based on clinician feedback, the hierarchy is inverted to place Stuttering Modification **before** Fluency Shaping. Addressing the struggle behavior (blocks) and fear (desensitization) first is crucial to prevent the patient from forcing speech. The sequence is now: Desensitization/ID → Stuttering Modification (Unlearning struggle) → Fluency Shaping (Learning ease) → Stabilization.

B.11 Primary Goal

- **Goal:** The client will demonstrate the ability to reduce physical tension during blocks using Pull-outs and Cancellations in 80% of perceived stuttering moments, following a reduction in avoidance behaviors.
- **Target:** 80% successful tension reduction; Reduced Avoidance Scale score
- **Baseline:** Client exhibits hard blocks with struggle behaviors on approx. 15% of syllables; high reactivity to stuck moments.

B.12 Step by Step Plan

1. Step 1: Phase 1: Identification & Desensitization

- **Week Range:** Weeks 1-2
- **Objective:** Identify the physical sensation of 'blocking' and reduce the panic response through voluntary stuttering.
 - **Mapping the Block:** Locating exactly where the airflow stops (Lips for 'Pay', Tongue for 'Talk', Larynx for 'I').
 - **Instructions:** APPROACH: Identification. PURPOSE: To separate the 'block' from the 'struggle'. ACTIVITIES: 1. Freeze Technique: When you block on 'Talk', freeze instantly. Don't push. Ask: 'Is my tongue glued to the roof of my mouth?' 2. Tension Rating: Rate the block 0-10. If it's a 9, voluntarily release it to a 5 without trying to say the word yet. 3. The Mirror: Watch a block happen. Observe if eyes blink or neck tightens. Acknowledge it neutrally.
 - **Voluntary Stuttering (Desensitization):** Intentionally mimicking blocks to break the fear-tension cycle.
 - **Instructions:** APPROACH: Avoidance Reduction. PURPOSE: To prove that blocking is not catastrophic. ACTIVITY: In a safe environment, read a sentence. Intentionally insert a light block on 'Pay' or 'Can'. Hold it for 2 seconds, then let go. Notice that you are in control of the duration.

2. Step 2: Phase 2: Stuttering Modification (Cancellations & Pull-outs)

- **Week Range:** Weeks 3-6
- **Objective:** Learn to modify the tension of the block **after** or **during** the moment, prioritizing struggle reduction over perfect fluency.
 - **Cancellations (Post-Block Correction):** Pausing after a hard block to re-attempt with loose articulators.

- *Instructions*: APPROACH: Stuttering Modification. PURPOSE: To erase the motor plan of 'pushing' and replace it with 'releasing'. STEPS: 1. Block on 'Pay'. Finish the word with struggle. 2. STOP. Pause for 3 seconds. Exhale. 3. Re-say 'Pay' with the lips barely touching (light contact). RULE: You are not allowed to repeat the word until you have fully exhaled and relaxed the tension site.
- **Pull-Outs (In-Block Release)**: Catching the block in real-time and voluntarily relaxing the tension to slide into the vowel.
- *Instructions*: APPROACH: Stuttering Modification. PURPOSE: To stop the struggle immediately. STEPS: 1. Feel the block on 'Talk'. 2. Do NOT push harder. Instead, stabilize the tongue posture. 3. Voluntarily reduce tension (melt the ice). 4. Slide slowly into the vowel /ɑ/. METAPHOR: Easing off the gas pedal when the wheels spin in mud.

3. Step 3: Phase 3: Fluency Shaping (Light Contacts & Easy Onsets)

- Week Range: Weeks 7-8
- **Objective**: Introduce fluency shaping techniques now that fear and struggle have been managed.
 - **Light Articulatory Contacts (For Consonant Blocks)** Touching articulators together gently to prevent the buildup of intra-oral pressure that causes blocks.
 - *Instructions*: APPROACH: Fluency Shaping. PURPOSE: Prevention of blocks on plosives (/p/, /b/, /t/, /k/). TARGETS: 'Pay', 'Talk', 'Can', 'Maybe'. INSTRUCTION: Imagine your lips/tongue are sore. Touch them together as lightly as possible to produce the sound. Reduce the 'explosion' of air.
 - **Easy Onsets (For Vowel Blocks)** Gradual initiation of voicing for vowel-initial words.
 - *Instructions*: APPROACH: Fluency Shaping. PURPOSE: To prevent laryngeal locking on words like 'I', 'About'. INSTRUCTION: Exhale a tiny amount of air (silent /h/) before engaging the vocal folds. 'h-I', 'h-About'. Smooth out the start.

4. Step 4: Phase 4: Stabilization & Transfer

- Week Range: Weeks 9-10
- **Objective**: Generalize strategies to complex sentences and daily life.
 - **Hierarchy of Transfer**: Moving from clinical practice to real-world application.
 - *Instructions*: APPROACH: Generalization. STEPS: 1. Phone call to a recorded line (low pressure). 2. Asking a question to a stranger. 3. Presentation/Conversation. TASK: Use a Pull-out intentionally during a conversation. Report back on the feeling of control.
 - **Maintenance Plan** : Preparing for relapse and high-stress days.

- *Instructions:* APPROACH: Long-term Management. GUIDANCE: If blocks return with high tension, stop Fluency Shaping (Phase 3). Return immediately to Phase 1 (Desensitization) and Phase 2 (Cancellations). Do not fight the block; analyze it.

Stuttering Type: Repetition 1

B.13 Explanation:

- **Stuttering Type:** The patient presents with specific 'initial syllable repetitions' at the start of utterances, despite a general classification of fluency. The clinician notes an absence of anxiety, ruling out 'Covert Stuttering' or Social Anxiety. This profile suggests a specific motoric difficulty with the initiation of phonation and articulation, rather than a psychological avoidance mechanism.
- **Patient Characteristics:** Unlike the previous assumption of high internal monitoring, the patient is not anxious but struggles with the motor execution of starting sentences. The behavior is characterized by repetitive attempts to initiate the first syllable. The focus shifts from desensitization (treating fear) to fluency shaping (treating the motor lag), specifically targeting the initiation of speech using the target phonemes /ʊ/ and /ʃ/.
- **Therapeutic Rationale:** Since anxiety is absent, Cognitive Restructuring and Voluntary Stuttering are removed as they are unnecessary. The plan pivots to 'Fluency Shaping' techniques. The primary objective is to smooth the transition from silence to speech (initiation) to prevent the initial syllable repetition. Techniques like Easy Onsets and Preparatory Sets are indicated.

B.14 Primary Goal

- **Goal:** The client will eliminate initial syllable repetitions and achieve smooth utterance initiation using fluency shaping techniques.
- **Target:** Frequency of initial syllable repetitions (measured per 100 syllables) & Client self-rating of initiation effort (0-10).
- **Baseline:** Frequent repetition of the initial syllable at the beginning of utterances; low anxiety.

B.15 Step by Step Plan

1. Step 1: Motor Initiation Strategies

- **Week Range:** Weeks 1-2
- **Objective:** Establish smooth initiation of phonation to replace syllable repetitions at the start of utterances.
 - **Easy Onsets (Gentle Initiation):** Gradually increasing vocal fold tension to start a word, preventing the 'hard' attack or repetition loop often seen at the start of sentences.

- *Instructions:* PURPOSE: To smooth the transition from silence to speech, preventing the 'stuttering block' or repetition loop at the start. MECHANISM: Exhale slightly before speaking, then gently turn on the voice like a dimmer switch, not a light switch. PRACTICE ITEMS (Targeting /ʃ/ and /v/): 1. Inhale, exhale slightly, then gently slide into: 'Shhh...ould' (/ʃʊd/). 2. 'Shhh...ugar' (/ʃʊgər/). 3. 'L...ook' (/lʊk/). 4. 'P...ut' (/pʊt/). ACTION: Practice 20 sentence starters. Focus ONLY on the first word. If a repetition occurs, stop, exhale, and try the Easy Onset again.
- **Rhythmic Initiation:** Using a rhythmic cadence to assist in the timing of utterance onset.
- *Instructions:* PURPOSE: To provide an external timing cue that overrides the internal motor lag causing the repetition. PRACTICE: Use a finger tap or a metronome beat to trigger the start of the sentence. EXERCISE: Tap your leg once, and coordinate the onset of your voice exactly with the tap. Do not speak *after* the tap; speak *on* the tap. TARGETS: 'Should we go?', 'Look at that.', 'Push it over.' (Focus on the /v/ and /ʃ/ sounds in the initial position).

2. Step 2: Stabilization & Generalization

- Week Range: Weeks 3-4
- **Objective:** Integrate initiation techniques into natural speech and complex sentences.
 - **Preparatory Sets:** Mentally rehearsing the motor plan for the first word before attempting to speak, ensuring articulators are in place.
 - *Instructions:* APPROACH: Before speaking, pause for 1 second. 1. SCAN: Identify the first sound of the sentence (e.g., /ʃ/ in 'Should'). 2. PLAN: Physically place your tongue/lips in the position for that sound without turning on your voice. 3. ACT: Initiate airflow gently (Easy Onset) through that pre-formed shape. GOAL: To replace the 'repetition' habit with a 'planning' habit.
 - **Continuous Phonation (Linking):** Keeping the voice 'on' between words to maintain momentum after the successful initiation.
 - *Instructions:* PURPOSE: Once the initial word is successfully started without repetition, maintain the flow to prevent downstream disfluencies. TECHNIQUE: Imagine your voice is a violin bow that doesn't leave the string. Connect the end of the first word to the start of the second. EXAMPLE: 'Shhhould_we_go.' (Instead of 'Should. We. Go.'). APPLICATION: Practice reading short paragraphs, focusing heavily on the Easy Onset of the first word and the Linking of the subsequent words.

Stuttering Type: Repetition 2

B.16 Explanation:

- **Stuttering Type:** Sound repetition involves the involuntary repetition of a specific phoneme (e.g., 's-s-s-supermarket'). In this profile, it manifests as multiple iterations of the initial sound, indicating a temporary disruption in the transition to the next sound.
- **Patient Characteristics:** The patient exhibits repetitions on vowels ('I') and voiceless fricatives ('s'). This indicates laryngeal tension (hard starts) and articulatory tension (tongue pressure).
- **Therapeutic Rationale:** This plan follows an Integrated Approach (ASHA Practice Portal). Per clinician feedback, the sequence has been adjusted to prioritize Stuttering Modification (Pull-outs) to manage active tension first, followed by Fluency Shaping (Light Contacts, Easy Onsets) to prevent stutters. It aligns with Yaruss & Quesal's ICF Model by addressing both the impairment (fluency) and the reaction (cognitive/emotional). **IMPORTANT:** This is a decision support tool. Clinical judgment is required.

B.17 Primary Goal

- **Goal:** The client will manage sound repetitions using 'Pull-outs', 'Light Articulatory Contacts', or 'Easy Onsets' in variable contexts with a self-rated Severity Rating (SR) averaging $\leq 3/9$, and demonstrate reduced negative reaction to stuttering.
- **Target:** Spontaneous speech accuracy $\geq 70\%$; Daily Severity Rating (SR) $\leq 3/9$; Reduced avoidance behaviors.
- **Baseline:** Uncontrolled repetitions (3-4 iterations) on initial sounds; Presumed moderate tension; No current modification strategies.

B.18 Step by Step Plan

1. Step 1: Foundation: Awareness & Desensitization

- **Week Range:** Weeks 1-2
- **Objective:** Client will identify physical tension sites and reduce fear of stuttering through desensitization.
 - **Proprioceptive Awareness & Voluntary Stuttering:** Learning to feel articulatory tension and reducing the emotional charge of stuttering.
 - **Instructions:** APPROACH: Sensory Awareness / Desensitization. PURPOSE: You cannot change what you do not feel. We must identify 'tight' vs. 'loose' muscles. PRACTICE ITEMS: 1. TENSION SCANS: Tense tongue against roof of mouth (Level 10), then release (Level 0). Notice the contrast. 2. THE 'BOUNCE' STUDY: Intentionally repeat 's-s-s' with high tension, then with loose muscles. 3. VOLUNTARY STUTTERING: Purposefully stutter on 'Supermarket' without struggle. SAFETY

WARNING: Ensure you feel safe before attempting voluntary stuttering. If anxiety spikes $\geq 7/10$, stop and revert to private practice. Do not force this in high-stress situations yet. 4. SEVERITY RATING (SR): Introduce a 1-9 scale (1=no stuttering, 9=severe). Rate your speech daily to track patterns, not just 'good' or 'bad' days.

2. Step 2: Stuttering Modification: Pull-outs

- Week Range: Weeks 3-5
- **Objective:** Client will learn to modify tension *during* a moment of stuttering (Pull-out).
 - **Pull-outs (In-block Modification):** Catching the stutter, stabilizing the tension, and sliding out of it voluntarily.
 - *Instructions:* APPROACH: Stuttering Modification. PURPOSE: Before focusing on prevention, we must learn to handle the moment of stuttering. When you get stuck, don't push harder. Pull out. STEPS: 1. CATCH: Feel the 's' or 't' repetition starting. 2. FREEZE: Hold the tension momentarily. Don't retreat, don't push. 3. RELEASE: Voluntarily reduce tension in the tongue or throat while keeping airflow moving. 4. SLIDE: Transition slowly to the next sound. PRACTICE: Fake a tense stutter on 'Supermarket', freeze on the 'S', relax the tongue, and say the word.

3. Step 3: Fluency Shaping: Light Articulatory Contacts (LAC)

- Week Range: Weeks 6-8
- **Objective:** Client will apply LAC to voiceless fricatives (/s/) to prevent tension buildup.
 - **Light Articulatory Contacts:** Reducing contact pressure of the tongue to prevent 'bouncing'.
 - *Instructions:* APPROACH: Fluency Shaping. VISUAL ANALOGY: Imagine your tongue is a butterfly landing on a flower (the roof of your mouth). It should not stomp; it should land gently. PRACTICE ITEMS: 1. THE FEATHER TOUCH: Say /s/ with barely any pressure. Sustain 'sssss'. 2. SYLLABLE SLIDE: 'Saaa', 'See'. Smooth transitions. 3. WORD DRILL: 'Supermarket'. If you feel pressure $\geq 2/10$, stop and restart lighter. 4. CARRIER PHRASES: 'I see a...' focusing on the /s/. TROUBLESHOOTING: If speech sounds 'mushy', you are too light. Aim for distinct but gentle contact.

4. Step 4: Fluency Shaping: Easy Onsets

- Week Range: Weeks 9-10
- **Objective:** Client will use Easy Onsets to manage vowel-initial words ('I') without hard glottal attacks.
 - **Easy Onset:** Gradual initiation of phonation to prevent laryngeal locking.

- *Instructions*: APPROACH: Fluency Shaping. VISUAL ANALOGY: A ramp vs. stairs. Slide your voice on (ramp) rather than jumping on (stairs/hard start). PRACTICE ITEMS: 1. THE 'H' TRICK: Place a tiny, silent 'h' before 'I'. 'h-I went'. 2. THE SLIDE: Inhale, and start speaking exactly as exhalation begins. SAFETY WARNING: Do not use a 'whisper voice'. The goal is full resonance. If you feel lightheaded, you are exhaling too much air before speaking (hyperventilation risk). Shorten the 'h' significantly.

5. Step 4: Generalization & Maintenance

- Week Range: Weeks 11-12
- **Objective**: Transfer strategies to variable contexts and establish a long-term maintenance plan.
 - **Hierarchy of Transfer & Cognitive Resilience**: Moving from clinical tasks to real-world resilience.
 - *Instructions*: APPROACH: Generalization / CBT / Maintenance. HIERARCHY TASKS: 1. VARIABLE CONTEXTS: Practice strategies with background noise (TV on), while walking, or on a phone call. 2. COGNITIVE REFRAMING: Identify thoughts like 'I must be perfect.' Replace with 'I can communicate effectively even if I stutter.' 3. LISTENER STRATEGIES: Teach family members to maintain eye contact and wait (do not finish sentences). MAINTENANCE: - RELAPSE PLAN: Stuttering is cyclical. If tension returns, go back to Week 1 (Proprioception) for 2 days. - SUPPORT: Consider joining a peer group (e.g., National Stuttering Association) for long-term support.

Stuttering Type: Interjection 1

B.19 Explanation:

- **Stuttering Type**: A 'block' is a moment where airflow and voicing are temporarily stopped due to excessive muscle tension. In this context, interjections are identified as secondary avoidance behaviors used to postpone the feared block.
- **Patient Characteristics**: The patient exhibits blocks on initiation ('Sometimes I...') and utilizes interjections as avoidance strategies. This indicates that the core issue is anticipatory fear and threat perception rather than solely a motor breakdown.
- **Therapeutic Rationale**: Based on clinician feedback, the plan is restructured to prioritize cognitive-emotional processing and desensitization over immediate fluency shaping. Interjections are treated as avoidance markers. Motor techniques are introduced only after fear is reduced, with Stuttering Modification preceding Fluency Shaping.

B.20 Primary Goal

- Goal: The patient will reduce avoidance-driven interjections and manage moments of blocking with decreased fear, tension, and cognitive load, demonstrating increased tolerance and adaptive responses during speech initiation across structured and spontaneous communication contexts.
- Target: 1. Reduction of avoidance-driven interjections by 50% in structured tasks. 2. Ability to identify and differentiate tension levels in 80% of blocks. 3. Successful use of Pull-outs (adaptive release) in 70% of identified blocks.
- Baseline: Patient currently uses interjections to avoid blocks on vowel initiations and experiences high fear/tension. Baseline severity via SSI-4; Cognitive impact via OASES.

B.21 Step by Step Plan

1. Step 1: Cognitive-Emotional Foundation & Desensitization

- Week Range: Weeks 1-4
- **Objective:** Increase awareness of interjections as avoidance behaviors and build tolerance for speech variability through voluntary stuttering.
 - **Psychoeducation & Avoidance Recognition** Understanding the role of threat perception and identifying interjections as safety behaviors that maintain anxiety.
 - *Instructions:* APPROACH: Cognitive Awareness. CONCEPT: Interjections (like 'um', 'uh' before 'Sometimes') are 'safety behaviors' that briefly lower anxiety but keep the fear alive. PRACTICE ITEMS: 1. The Avoidance Log: Track moments where you used a filler word to delay saying a feared word. Note the feeling immediately before (e.g., 'Panic'). 2. Neutral Observation: When you hear yourself say 'um... Sometimes', pause. Acknowledge it internally: 'That was an avoidance strategy.' Do not judge it, just label it. 3. The Pause: Instead of filling the silence with an interjection, practice simply pausing. Allow the silence to exist.
 - **Voluntary Stuttering (Desensitization):** Intentionally stuttering to reduce the fear of the block and the urge to avoid. *Instructions:* APPROACH: Desensitization/Approach Behavior. PURPOSE: To prove that stuttering is not a catastrophic event and to reduce the need for interjections. PRACTICE ITEMS: 1. Voluntary Repetition: On 'Sometimes', intentionally bounce on the 'S' (S-s-s-sometimes) with eye contact. Do not hide it. 2. The 'Clean' Block: Intentionally block on 'I' without using an interjection first. Hold the silence for 2 seconds, then say the word. 3. Reflection: After a voluntary stutter, ask: 'Did anything bad happen?' (Answer is usually no).
-

2. Step 2: Stuttering Identification & Differentiation

- Weeks 5-8
- **Objective:** Develop in-the-moment awareness of the speech mechanism and differentiate tension levels without urgency to fix them.
 - **Exploration of the Speech Mechanism:** Systematic identification of laryngeal involvement and airflow during blocks. *Instructions:* APPROACH: Proprioception. FOCUS: Where is the tension? (Larynx/Throat vs. Lips). PRACTICE ITEMS: 1. Tension Mapping: Replicate the block on 'I'. Is the throat tight or loose? Rate the tension 1-10. 2. Differentiation: Produce a 'Hard Block' (10/10 tension) on 'Sometimes'. Then produce a 'Soft Block' (3/10 tension). Feel the difference in the larynx. 3. Duration Awareness: Identify if a block is 'Long' (i2s) or 'Short' (i1s). Just observe.
 - **Stuttering Modification: The Pull-out:** Introducing agency within the block to reduce struggle, framed as an adaptive response rather than a correction.
 - *Instructions:* APPROACH: Stuttering Modification. GOAL: To increase flexibility, not to force fluency. PRACTICE ITEMS: 1. Catch the Block: When stuck on /aɪ/ (I), catch yourself pushing. Stop pushing, but don't back up. 2. Release Tension: While holding the articulatory posture, consciously lower laryngeal tension by 50%. 3. Slide Out: Slowly transition into the vowel. 'S... (release tension)... ometimes'. NOTE: If you use an interjection, stop. Reset. Try the Pull-out again.

3. Step 3: Supportive Fluency Shaping & Generalization

- Week Range: Weeks 9-12
- **Objective:** Introduce Easy Onsets as supportive tools (not primary targets) and generalize adaptive responses to spontaneous speech.
 - **Easy Onsets (Supportive Tool):** Using gentle airflow to facilitate initiation, introduced only after fear/avoidance has decreased.
 - *Instructions:* APPROACH: Fluency Shaping (Secondary). USAGE: Use this as a preventative tool when you anticipate high laryngeal tension, but do not use it to 'hide' stuttering. PRACTICE ITEMS: 1. Gentle Airflow: On vowel initiations like 'I' or 'Apple', start with a microscopic exhale before engaging the vocal folds. 2. Phrase Practice: 'h(silent)... I think so'. 3. Integration: If you feel a block coming, you have a choice: Voluntary Stutter OR Easy Onset. Both are acceptable.
 - **Resilience & Real-World Application:** Applying desensitization and modification in spontaneous contexts.
 - *Instructions:* APPROACH: Generalization. PRACTICE ITEMS: 1. The 'No Avoidance' Challenge: Enter a conversation with the specific goal of NOT using interjections, even if it means blocking silently. 2. Disclosure: Tell a listener, 'I'm working on my speech, so I might pause or repeat sounds.' 3. Spontaneous Speech: Answer open-ended questions

(e.g., 'What did you do today?') focusing on 'Agency'—handling blocks with Pull-outs rather than avoiding them.

Stuttering Type: Interjection 2

B.22 Explanation:

- **Stuttering Type:** Interjections and pre-speech fillers are avoidance behaviors used to postpone the initiation of a feared word. Unlike motor prolongations, these are behavioral strategies to delay speech due to anticipatory anxiety.
- **Patient Characteristics:** Patient exhibits interjections and fillers (e.g., 'uh', 'um', or schwa insertions like '/thɛnə/') primarily before initiating speech. These function as safety behaviors to avoid the feared initial sounds (/th/, /n/).
- **Therapeutic Rationale:** The therapy focus shifts from motor execution (managing tension) to anxiety reduction and avoidance reduction. The goal is to eliminate the 'starter' sounds and encourage direct initiation of the target word, even if disfluent. We must decouple the fear of the word from the need to use a filler.

B.23 Primary Goal

- **Goal:** The patient will initiate target phrases directly without the use of antecedent interjections or fillers in 8 out of 10 trials during structured tasks, tolerating potential disfluency on the target word.
- **Target:** 8/10 direct initiations (zero fillers) in structured reading
- **Baseline:** Patient currently uses frequent interjections/fillers before initial fricatives and nasals to delay speech onset.

B.24 Step by Step Plan

1. Step 1: Identification & Desensitization of Avoidance

- **Week Range:** Weeks 1-2
- **Objective:** Identify the urge to use fillers/interjections and decouple the fear response from the target word.
- **The 'Clean Start' & Voluntary Stuttering:** Identifying the 'urge' to use a filler and choosing to stutter openly on the target word instead.
- **Instructions:** APPROACH: Identification & Negative Practice. SAFETY PROTOCOL: Monitor anxiety. If the urge to say 'um' or 'uh' arises, pause. Do not speak until the urge subsides. PRACTICE ITEMS: 1. Identification: Read a sentence. Raise a finger every time you feel the urge to use a filler before a word. 2. Voluntary Stuttering on TARGET: Instead of saying 'Uh... then', go directly to 'Then' and intentionally stutter on the 'Th' (e.g., 'Th-th-th-then'). 3. Rationale: Show the brain that the word 'Then' is safe and does not require a shield (the filler). SUCCESS CRITERIA: Patient can

identify the urge to use a filler and choose a voluntary stutter on the target word instead (10 trials).

2. Step 2: Modification Phase I: Cancellation of Avoidance

- Week Range: Weeks 3-4
- **Objective:** Learn to stop immediately after an avoidance behavior (filler) is used and restart without it.
 - **The Avoidance Cancellation:** Stopping immediately if a filler is used, pausing to reset, and attempting the target word directly.
 - *Instructions:* APPROACH: Behavioral Modification. PURPOSE: To break the habit of using fillers as a springboard into speech. INSTRUCTIONS: 1. If you say 'Uh... Then', STOP immediately. 2. Exhale completely. Wait 3 seconds. 3. Attempt the word 'Then' again directly. Do not use a run-up. 4. If you stutter on 'Then', that is acceptable. The goal is NO FILLER. PRACTICE ITEMS: 1. Word List: 'Then', 'This', 'No'. Deliberately use a filler ('Um-No'), stop, pause, say 'No' directly. SUCCESS CRITERIA: Patient successfully cancels a filler and re-attempts with a clean start in 8/10 trials.

3. Step 3: Modification Phase II: Preparatory Sets

- Week Range: Weeks 5-7
- **Objective:** Pre-speech planning to reduce anticipatory anxiety and initiate airflow directly.
 - **Preparatory Sets (The Silent Pause):** Replacing the audible filler with a silent, deliberate pause to prepare articulators for direct onset.
 - *Instructions:* APPROACH: Stuttering Modification (Preparatory Sets). PURPOSE: To replace the 'panic' filler with a 'planning' pause. INSTRUCTIONS: 1. Before the feared word (e.g., 'Then'), PAUSE intentionally. 2. Do not say 'um'. Instead, visualize the articulators moving for the /th/ sound. 3. Start the airflow GENTLY before the voice. /h/ -i /th/. 4. Initiate the word directly. PRACTICE ITEMS: 1. Phrase practice: 'And... (silent prep)... then'. 2. Sentence completion: 'I will go to the... (silent prep)... theater.' SUCCESS CRITERIA: Patient utilizes a silent preparatory set instead of an audible filler in 8/10 sentence completions.

4. Step 4: Direct Initiation & Transfer

- Week Range: Weeks 8-9
- **Objective:** Maintain direct speech initiation (no fillers) in connected speech and real-world scenarios.
 - **Direct Entry & Tolerating Silence** Practicing the tolerance of silence (rather than filling it) during conversation.
 - *Instructions:* APPROACH: Generalization. PURPOSE: To become comfortable with the silence that occurs while formulating thoughts, reducing

the need for fillers. PRACTICE ITEMS: 1. Timed Response: Partner asks a question. Patient must wait 2 seconds (in silence) before answering directly. NO 'Well...', 'Um...', or 'You know...'. 2. Transfer Task: Order a coffee. Focus entirely on starting the order directly (e.g., 'Can I have...' vs 'Um, can I have...'). 3. Anxiety Check: If you block on the word, let it happen. Do not retreat to a filler. SUCCESS CRITERIA: Patient completes 3 transfer tasks with zero reported avoidance fillers.

5. Step 5: Maintenance & Avoidance Monitoring

- Week Range: Week 10
- **Objective:** Long-term monitoring of avoidance behaviors to prevent the return of fillers.
 - **Avoidance Audit & Support:** Self-monitoring tools to track if fillers are creeping back into speech patterns.
 - *Instructions:* APPROACH: Relapse Prevention. INSTRUCTIONS: 1. The 'Filler Audit': Record a 1-minute conversation once a week. Listen back ONLY for interjections/fillers. If you hear more than 3, return to Week 3 (Cancellations). 2. Cognitive Reframing: Remind yourself that 'Silence is better than avoidance.' 3. Support: Discuss the fear of silence with a support group or mentor. SUCCESS CRITERIA: Patient demonstrates ability to self-analyze recordings and identify avoidance behaviors.

References

- Baevski A, Henry Z, Mohamed A, et al (2020) wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems* 33:12449–12460
- Bayerl SP, Wagner D, Nöth E, et al (2022) The influence of dataset-partitioning on dysfluency Detection systems. In: *Proc. of Text, Speech, and Dialogue*
- Bayerl SP, Gerczuk M, Batliner A, et al (2023) Classification of stuttering—the ComParE challenge and beyond. *Computer Speech & Language* 81:101519
- Beilby J (2014) Psychosocial impact of living with a stuttering disorder: Knowing is not enough. In: *Seminars in speech and language*, Thieme Medical Publishers, pp 132–143
- Boyle MP, Rosen AL (2025) More than meets the eye: Self-rated covert stuttering is linked to reduced psychosocial and communicative outcomes. *Journal of Fluency Disorders* p 106162
- Chang SE, Below JE, Chow HM, et al (2025) Stuttering: Our current knowledge, research opportunities, and ways to address critical gaps. *Neurobiology of Language*

- Comanici G, Bieber E, Schaekermann M, et al (2025) Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. arXiv preprint arXiv:250706261
- Conneau A, Baevski A, Collobert R, et al (2020) Unsupervised cross-lingual representation learning for speech recognition. arXiv preprint arXiv:200613979
- Dietrich N (2025) Agentic AI in radiology: emerging potential and unresolved challenges. *British Journal of Radiology* 98(1174):1582–1584
- Guitar B (2013) *Stuttering: An Integrated Approach to Its Nature and Treatment*, 5th edn. Lippincott Williams & Wilkins
- Iverach L, Rapee RM (2014) Social anxiety disorder and stuttering: Current status and future directions. *Journal of Fluency disorders* 40:69–82
- Karunanayake N (2025) Next-generation agentic AI for transforming healthcare. *Informatics and Health* 2(2):73–83
- Kaur R (2025) *Navigating Childhood Stuttering: A Guide to Management of Stuttering at Home and School*. Springer Nature
- Kourkounakis T, Hajavi A, Etemad A (2020) Detecting multiple speech disfluencies using a deep residual network with bidirectional long short-term memory. In: *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp 6089–6093
- Kourkounakis T, Hajavi A, Etemad A (2021) Fluentnet: End-to-end detection of stuttered speech disfluencies with deep learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29:2986–2999
- Lea C, Mitra V, Joshi A, et al (2021) Sep-28k: A dataset for stuttering event detection from podcasts with people who stutter. In: *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp 6798–6802
- Liu X, Li X, Yang Y, et al (2025) Addressing task conflicts in stuttering detection via MMoE-based multi-task learning. In: *Proc. Interspeech 2025*, pp 798–802
- Mohamed A, Lee Hy, Borgholt L, et al (2022) Self-supervised speech representation learning: A review. *IEEE Journal of Selected Topics in Signal Processing* 16(6):1179–1210
- Onslow M (1992) Identification of early stuttering: Issues and suggested strategies. *American Journal of Speech-Language Pathology* 1(4):21–27

- Powell K, Fishman EK, Chu LC, et al (2026) Agentic artificial intelligence: the power to change medicine and our world. *Journal of the American College of Radiology*
- Reilly S, Onslow M, Packman A, et al (2013) Natural history of stuttering to 4 years of age: A prospective community-based study. *Pediatrics* 132(3):460–467
- Renze M (2024) The effect of sampling temperature on problem solving in large language models. In: *Proc. of Findings of the Association for Computational Linguistics: EMNLP 2024*, pp 7346–7356
- Schuller B, Batliner A, Amiriparian S, et al (2022) The ACM multimedia 2022 computational paralinguistics challenge: Vocalisations, stuttering, activity, & mosquitoes. In: *Proc. of the 30th ACM International Conference on Multimedia*, pp 7120–7124
- Sebastian Peter Bayerl and Dominik Wagner and Elmar Noeth and Korbinian Riedhammer (2022) Detecting Dysfluencies in Stuttering Therapy Using wav2vec 2.0. In: *Proc. of Interspeech 2022*, pp 2868–2872
- Sheikh SA (2023) Deep learning for stuttering detection. PhD thesis, Université de Lorraine
- Sheikh SA, Sahidullah M, Hirsch F, et al (2021) StutterNet: Stuttering detection using time delay neural network. In: *Proc. of 29th European Signal Processing Conference (EUSIPCO)*, IEEE, pp 426–430
- Sheikh SA, Sahidullah M, Hirsch F, et al (2022a) End-to-end and self-supervised learning for the ComParE 2022 stuttering sub-challenge. In: *Proc. of the ACM Multimedia 2022*
- Sheikh SA, Sahidullah M, Hirsch F, et al (2022b) Introducing ECAPA-TDNN and Wav2Vec2. 0 embeddings to stuttering detection. *arXiv preprint arXiv:220401564*
- Sheikh SA, Sahidullah M, Hirsch F, et al (2022c) Machine learning for stuttering identification: Review, challenges and future directions. *Neurocomputing* 514:385–402
- Sheikh SA, Sahidullah M, Hirsch F, et al (2022d) Robust stuttering detection via multi-task and adversarial learning. In: *Proc. of 30th European Signal Processing Conference (EUSIPCO)*, IEEE, pp 190–194
- Sheikh SA, Sahidullah M, Hirsch F, et al (2023a) Advancing stuttering detection via data augmentation, class-balanced loss and multi-contextual deep learning. *IEEE Journal of Biomedical and Health Informatics*
- Sheikh SA, Sahidullah M, Hirsch F, et al (2023b) Stuttering detection using speaker representations and self-supervised contextual embeddings. *International Journal of Speech Technology* 26(2):521–530

- Sheikh SA, Sahidullah M, Kodrasi I (2025) Overview of automatic speech analysis and technologies for neurodegenerative disorders: Diagnosis and assistive applications. *IEEE Journal of Selected Topics in Signal Processing*
- SheikhBahaei S, Millwater M, Maguire GA (2023) Stuttering as a spectrum disorder: A hypothesis. *Current Research in Neurobiology* 5:100116
- Sønsterud H, Halvorsen MS, Feragen KB, et al (2020) What works for whom? multi-dimensional individualized stuttering therapy (MIST). *Journal of Communication Disorders* 88:106052
- Tang S, Modi A, Sjoding M, et al (2020) Clinician-in-the-loop decision making: Reinforcement learning with near-optimal set-valued policies. In: *Proc. of International Conference on Machine Learning*, PMLR, pp 9387–9396
- Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. In: *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, pp 5998–6008
- Yairi E, Ambrose N (2013) Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders* 38(2):66–87
- Yairi E, Seery CH, Singer CM (2025) *Stuttering: Foundations and clinical applications*. Plural publishing
- Yaruss JS, Quesal RW (2006) Overall assessment of the speaker’s experience of stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency disorders* 31(2):90–115
- Yaruss JS, Coleman CE, Quesal RW (2012) Stuttering in school-age children: A comprehensive approach to treatment. *Language, speech, and hearing services in schools* 43(4):536–548
- Zou J, Topol EJ (2025) The rise of agentic ai teammates in medicine. *The Lancet* 405(10477):457

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [vst.zip](#)