

Speech Recognition Challenges in a coffee-making game

- 1. Project:** a voice-controlled coffee-making simulation game. Players will give voice commands to interact with the game.
- 2. Experiment:** several coffee-related words were tested, including: (1) types of coffee: Americano, coconut water Americano, jasmine flower milk coffee, Irish coffee, cappuccino, mocha (2) Coffee making method: drip coffee, French press, espresso machine (3) common coffee ingredients: oat milk, hazelnut
- 3. Results:** Common terms like "Irish coffee" and "French press" were recognized with high confidence, while words like "Americano" and "hazelnut" had much lower confidence scores. The word "Americano" was often transcribed as "American" (confidence: 0.357), and "hazelnut" was misrecognized several times (confidence: 0.471).
- 4. Challenges:** (1) Some specific or innovative drink names, such as "Americano" and "coconut water Americano" are less common in the training data, leading to difficulty in accurate recognition. These terms are not frequently found in typical speech recognition models, resulting in errors during recognition. For example, "Americano" was recognized as "American," and "hazelnut" was repeatedly misrecognized. (2) Some words may have pronunciations that are unfamiliar to the speech model, making them harder to recognize accurately. (3) Background noise may also interfere with the recognition process and reduce accuracy.
- 5. Solutions:** (1) Optimize the recognition system by adding specific drink names, ingredient names, and other specialized terminology to improve recognition accuracy for these terms. (2) Implement a confirmation step, allowing players to repeat their command if the recognition confidence is low.