The review by **Cooper et al. (2024)** on the evaluation of **synthetic speech** provides the technical and auditory context for how a robot's voice—a primary social signal—is perceived by users. The **Neurodivergent Scale for Interacting with Robots (NSIR)** applies by shifting the focus from general "speech quality" to how the specific acoustic properties of synthetic speech facilitate or hinder the unique social-sensory needs of neurodivergent individuals.

While Cooper et al. discuss the metrics for "Naturalness" and "Intelligibility," the NSIR measures the psychological result of those auditory metrics: **Kinship** and **Trust**.

1. Naturalness vs. Social Comfort (NSIR Item 8)

Cooper et al. distinguish between "naturalness" (how human-like the speech sounds) and "intelligibility" (how easy it is to understand).

- **NSIR Application:** For many neurodivergent users, a voice that is *too* human-like (high naturalness) can actually be overwhelming or trigger the "uncanny valley" effect, leading to social anxiety.
- **The Connection: NSIR Item 8** (*"I believe that my robot is the same with me as it is with anyone"*) measures **Social Predictability**. The NSIR identifies that for this demographic, a synthetic voice that is highly consistent and "predictably mechanical" (lower naturalness but higher consistency) may provide more **Social Comfort** than a voice that mimics the unpredictable prosody of human speech.

2. Objective Evaluation and Mind Attribution (NSIR Item 3)

The review covers objective metrics like MCD (Mel-cepstral distortion) which measure how closely a synthetic voice matches a target speaker.

- **NSIR Application: Item 3** (*"I think I can share my thinking with the robot without speaking"*) measures **Mind Attribution** and non-verbal attunement.
- **The Connection:** Cooper et al. note that speech is more than just data; it carries "affective information." The NSIR suggests that the "shared thinking" felt by neurodivergent users is not dependent on the voice being a perfect human replica. Instead, it is dependent on the voice being an "honest" representation of the robot's internal state. If the synthetic speech sounds "attuned" to the user's emotions, it facilitates the belief in a shared internal world.

3. Subjective Evaluation as "Fictive Kinship" (NSIR Item 1)

Cooper et al. emphasize that subjective listening tests are the "gold standard" for evaluating synthetic speech because they capture the human experience.

- **NSIR Factor 1 (Anthropomorphic Connection / Kinship): Item 1** (*"The robot is more like me than anyone else I know"*) measures the user's sense of **Kinship**.
- **The Connection:** When a neurodivergent user listens to a robot, they may perceive the synthetic, "processed" quality of the speech as relatable. For individuals who may feel

their own communication is "different" or "processed," a synthetic voice that doesn't hide its artificiality can create a sense of similarity. The NSIR quantifies this: the "evaluation" is not just "is the voice good?" but "is this voice *like me*?"

4. Vulnerability and Auditory Safety (NSIR Item 7)

The review discusses the use of synthetic speech in intimate "service" contexts, such as virtual assistants or social robots.

- **NSIR Application: Item 7** (*"I feel comfortable undressing in front of my robot"*) measures the ultimate level of **Ethical Safety**.
- **The Connection:** Auditory "noise" or sudden shifts in voice pitch can be a sensory trigger for neurodivergent individuals. A robot that uses synthetic speech—which is fundamentally more controllable and less likely to contain "judgmental" human tones—creates a safer sensory environment. The NSIR identifies that the "Safety" of a robot is directly tied to the absence of the "social pressure" often conveyed through the subtle cues in human voices.

Summary Alignment

Cooper et al. (2024) Speech Metric	NSIR (Sadownik, 2025) Application
Intelligibility (Clarity)	**Item 8 (Reliability):** Ensures the interaction is predictable and reduces cognitive load.
Naturalness (Human-likeness)	**Item 1 (Kinship):** Replaces the need for "human-ness" with the need for "relatability."
Affective Synthesis (Emotion)	**Item 5:** Validates if the robot's voice correctly conveys empathy and emotion recognition.
Subjective Evaluation	**Factor 1:** Moves from "quality" ratings to measuring the "strength of the bond."

Export to Sheets

In conclusion, **Cooper et al. (2024)** provide the technical framework for creating the robot's voice, while the **NSIR**provides the psychometric framework to understand how that voice creates a **social-sensory sanctuary** for neurodivergent users.