IOWA STATE UNIVERSITY
Dependable Data Driven Discovery

Spoken Fluency in Interaction with Conversational Voice AI

Christopher Thompson Litten, Young-Joo Lee, Ahmad Zubaidi Amrullah, Benjamin Godard, & Dr. Mahdi Duris
{Litten, yjlee1, ubedm2g, godardb@iastate.edu} mduris@ut.edu.sa

Summary

Does AI talk like people?

Measures of speech
- Filled pauses – uh, um, er, erm, etc.
- Speech rate – actual syllables produced per minute
- Silent pauses – frequency & length
- Silent pause location

Listener judgments
- Perceived fluency
- 9-point scale
- 15-20 second audio clips

Pause type	Location	Example
MID	Interrupts a clause	"I think the best option is … the oxygen tanks."
INTER	Between connected clauses	"We need the oxygen … because we can't survive without it."
END	After a complete spoken unit	"I would put the oxygen first. … The water should probably be second."
PRE	After a discourse marker or hesitation before the clause begins	"Well … I would say the oxygen." / "Hmm … that's a good question."

Problem

- More practice ≠ the same conversational experience
- Conversation is co-created by all participants (Voss & Waring, 2025)
- Appropriate language use depends on the social context
- Fewer pauses ≠ better conversation (Clark & Fox Tree, 2002)
- Measured fluency may not match perceived fluency (Bosker et al., 2013)

Listen to audio samples

Method

- **Participants:** 11 English speakers & 11 listeners and Sesame.
- **Tasks:** Narrative, debate, and problem-solving conversations.
- **Ratings:** Listeners rated 15-20 second speech clips for fluency on a 9-point scale. (adapted from Bosker et al., 2013)

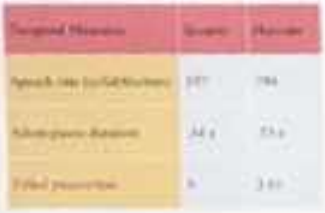


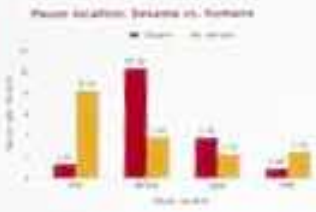
Task	Name	Type
1a	Narrative	Monologue
1b	Narrative follow-up	Dialogue
2	Open-ended debate	Dialogue
3	Collaborative problem-solving	Dialogue

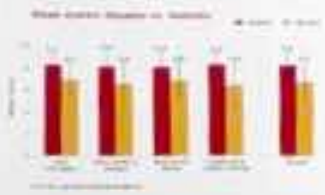


Results

- **Similar speech rates:** Sesame averaged 197 syllables/minute; humans averaged 194.
- **Different pause patterns:** Sesame produced no filled pauses and fewer mid-clause pauses than humans.
- **Higher average fluency ratings for Sesame:** 8.25 versus 6.73 out of 9.
- **Filled pauses and listener judgments:** More filled pauses were associated with lower perceived-fluency ratings.









Solution / contribution

- Documents L1 speakers' fluency across three AI speaking tasks.
- Examines which speech features relate to listeners' fluency judgments.
- Provides a reference point for future research on AI speaking practice.

Similar speech rates and average pause lengths can conceal differences in how speech is organized and perceived.

Limitations and Future Work

Small sample: 11 speakers and 11 listeners	Replicate with a larger sample, including second language speakers
Brief (1–20 second) samples	Use longer conversation samples and examine full conversations
Single AI system and voice	Compare multiple systems, AI systems and voices
No human-human comparison	Add matched human–human interaction comparisons
No analysis of the words produced	Conduct analyses of the conversational content

Related Work

- **Fluency perception:** Speech rate and pausing contribute to listeners' fluency judgments (Bosker et al., 2013).
- **Pause location:** Where pauses occur influences perceived fluency (Kahng, 2018).
- **Filled pauses:** *Uh* and *um* can signal upcoming delays and help manage conversational turns (Clark & Fox Tree, 2002).
- **AI naturalness:** Evaluating natural AI conversation requires attention to context and adaptation to the conversational partner (Voss & Waring, 2025).

This work is partially supported by the National Science Foundation under Grant No. 2152117. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.